COMPUTATIONAL AND EXPERIMENTAL APPROACHES TO ENHANCE EXTRACELLULAR SECRETION OF RECOMBINANT PROTEINS IN *ESCHERICHIA COLI*

by

Anup Agarwal

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Chemical Engineering

Fall 2010

Copyright 2010 Anup Agarwal
All Rights Reserved
COMPUTATIONAL AND EXPERIMENTAL APPROACHES TO ENHANCE EXTRACELLULAR SECRETION OF RECOMBINANT PROTEINS IN

ESCHERICHIA COLI

by

Anup Agarwal

Approved:  __________________________________________________________
Kelvin H. Lee, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved:  __________________________________________________________
Norman J. Wagner, Ph.D.
Chair of the Department of Chemical Engineering

Approved:  __________________________________________________________
Michael J. Chajes, Ph.D.
Dean of the College of Engineering

Approved:  __________________________________________________________
Charles G. Riordan, Ph.D.
Vice Provost for Graduate and Professional Education
ACKNOWLEDGMENTS

I would like to thank my advisor, Prof. Kelvin Lee, for giving me an opportunity to pursue my graduate research in his lab. He has been a great source of inspiration over the course of my studies. His wisdom and guidance has helped me to become a better researcher. I hope that his influence on my thinking and his teachings would continue to shape my career in the future.

I would like to extend my gratitude to all the past and present Lee group members for all the fun times in the lab. Prateek has been a great source of encouragement and support over the course of my studies. Leila has been a wonderful conversationalist over the years. Pei has been a great help in the lab. Stephanie has been a great help with the drafts of my thesis. I would also like to thank Jeff Swanberg, Gilda, Diane, Yong, Shuyu, Kristin, and Jeff Foltz for their support and friendship.

I would like to thank all my friends for being there whenever I needed them. Manish, Vinit, Aditi, Sandeep, Ashwani, Harman, Rahul, Ankit, Madhur, Shamit, Avi, and Lael have been a great support system over the course of my studies. They have influenced and helped me in more ways than they can imagine.
I am very grateful and thankful to my family. My parents have been just amazing in supporting me throughout my life. My sisters, Shalu and Anshu, are simply the best and I am thankful for their love and support.

Anup Agarwal

December 2010
TABLE OF CONTENTS

LIST OF TABLES........................................................................................................... viii
LIST OF FIGURES........................................................................................................... ix
ABSTRACT...................................................................................................................... xii

Chapter
1 INTRODUCTION AND OVERVIEW .................................................................1
  1.1 Motivation.........................................................................................................1
  1.2 Recombinant Protein Production in E. coli.................................................2
     1.2.1 Cytoplasmic Expression.................................................................2
     1.2.2 Periplasmic Export.................................................................3
     1.2.3 Extracellular Secretion.............................................................4
  1.3 The Type-I Secretion System (TISS)......................................................9
     1.3.1 HlyB..............................................................................................10
     1.3.2 HlyD..............................................................................................12
     1.3.3 TolC..............................................................................................12
  1.4 Application of TISS in Extracellular Recombinant Protein Secretion ....13
  1.5 Factors Affecting Secretion Efficiency..................................................16
     1.5.1 Role of Protein Synthesis Rate in Protein Secretion.....................16
     1.5.2 Role of the CTSS in Protein Secretion via the TISS.................18
  1.6 Mathematical Modeling as a Tool to Understand Complex Biological
     System............................................................................................................19
  1.7 Thesis Goals.................................................................................................23
  1.8 Scope of the Work.........................................................................................24

2 GENETIC ENGINEERING APPROACHES TO ENHANCE
   EXTRACELLULAR SECRETION OF RECOMBINANT PROTEINS.............25
  2.1 Introduction..................................................................................................25
  2.2 Materials and Methods..............................................................................30
LIST OF TABLES

Table 2.1: Plasmids used in this study and their sources ......................................................31

Table 2.2: Strains used in this study and their origin ............................................................31

Table 2.3: Primers used in the site directed mutagenesis of the Helix1 and the Cterm1 region of the CTSS .................................................................33

Table 2.4: Primers used for the knockout of the ycdC gene and its confirmation ..........35

Table 2.5: Sequence changes (marked in red) and predicted % decrease in translation rate of different Bla mutants (Gupta and Lee 2010) .........................................................41

Table 2.6: Sequence changes (marked in red) and predicted % decrease in translation rate of different GILL-Bla mutants (Gupta and Lee 2010) .........................................................42

Table 2.7: Sequence changes (marked in red) made in the CTSS .....................................45

Table 3.1: Parameter values used in the model simulations ..............................................67

Table 3.2: Range within which the model kinetic rate constants were varied .....................79

Table 3.3: Scaling factors for all the species in the model ..................................................8
LIST OF FIGURES

Figure 1.1: Schematic overview of the *E. coli* Sec, SRP and Tat translocases ...............5

Figure 1.2: Schematic overview of the major extracellular protein secretion systems (Type I, II, III, IV and V) in Gram-negative bacteria.......................................................8

Figure 1.3: The Type-I translocator in *E. coli* ...............................................................11

Figure 2.1: Schematic of 60 amino acids of the *E. coli* HlyA CTSS showing the location of the predicted amphiphilic helix (Helix1) and Cterm1 regions where synonymous rare codon substitutions were made........................................30

Figure 2.2: The *ycdC* gene knockout from the *E. coli* W3110 genome .......................37

Figure 2.3: Effect of the *ycdC* gene knockout on the extracellular secretion of HlyA, Bla and GILL-Bla proteins via the TISS in *E. coli* ........................................................39

Figure 2.4: Design of a *de novo* gene sequence (*gill-bla*) based on *bla* gene sequence....40

Figure 2.5: Combined effect of the *ycdC* gene knockout and synonymous rare codon substitutions on the secretion of the Bla protein via the TISS in *E. coli* ..............43

Figure 2.6: Combined effect of the *ycdC* gene knockout and synonymous rare codon substitutions on the secretion of the GILL-Bla protein via the TISS in

*E. coli* ..................................................................................................................................44
Figure 2.7: Effect of synonymous rare codon changes in the CTSS on the secretion of HlyA and Bla proteins via the TISS in *E. coli* ..............46

Figure 3.1: The Type-I translocator in *E. coli* ............................................................50

Figure 3.2: Schematic of the kinetic models of secretion showing all the reactions and species involved..............................................................58

Figure 3.3: Schematic of the Model 2 of secretion showing all the reactions and species involved................................................................................59

Figure 3.4: Schematic of a bacterial growth curve showing lag phase, exponential phase and stationary phase..................................................62

Figure 3.5: Effect of SecB on the SE of HasA via the TISS for different HasA synthesis rate (*k₀*) ........................................................................70

Figure 3.6: Effect of transporter protein synthesis rate (*kₜ*) on the SE of HasA at varying HasA protein synthesis rate (*k₀*) in the wild type strain (Model 1) ..............71

Figure 3.7: Effect of transporter protein synthesis rate (*kₜ*) on the SE of HasA at varying HasA protein synthesis rate (*k₀*) in the ΔsecB strain (Model 2) ......................73

Figure 3.8: Effect of HasA synthesis rate (*k₀*) on its SE at varying transporter protein synthesis rate (*kₜ*) in the wild type strain (Model 1) ........................................74

Figure 3.9: Effect of HasA synthesis rate (*k₀*) on the rate of secretion (*rₛ*) at varying transporter protein synthesis rate (*kₜ*) in the wild type strain (Model 1) .........75

Figure 3.10: Effect of HasA synthesis rate (*k₀*) on its extracellular concentration at varying transporter protein synthesis rate (*kₜ*) in the wild type strain (Model 1) ........................................................................77
Figure 3.11: SE results of 5,196 simulation runs using different input parameter vectors

Figure 3.12: Extracellular concentration results of 5,196 simulation runs using different input parameter vectors

Figure 3.13: Normalized sensitivity index (NSI) describing the influence of all 20 kinetic rate constants on the SE and the extracellular concentration of HasA

Figure 3.14: Variation of the optimal extracellular concentration with culture time

Figure 3.15: Variation of the optimal extracellular concentration divided by culture time (ratio) with culture time

Figure 3.16: Combined effect of HasA synthesis rate (scaled: $k_0$) and transporter synthesis rate (scaled: $k_t$) on the extracellular concentration of HasA (scaled) when it is secreted from the wild type strain (Model 1)

Figure 3.17: Variation of the optimal extracellular concentration (scaled) with culture time (scaled)

Figure 3.18: Variation of the optimal $k_0$ (scaled) and the optimal $k_t$ (scaled) corresponding to the optimal extracellular concentration (scaled) with culture time (scaled)
ABSTRACT

The annual global market for biopharmaceuticals is estimated at about $99 billion. Approximately 40% of all biotechnology products made in the US and EU are derived from *Escherichia coli* expression systems. Thus, efforts to study and improve *E. coli* as the host microorganism for the production of recombinant proteins can be of considerable importance. Secretion of recombinant proteins into the extracellular medium via the Type-I secretion system (TISS) in *E. coli* offers several advantages in comparison to the cytoplasmic accumulation and the periplasmic secretion such as simplifying detection and purification of the target protein and also providing a better folding environment.

Studies have shown that the premature termination of the RutR protein, a transcription factor encoded by the *ycdC* gene, can enhance the secretion levels of recombinant proteins via the TISS. Here, the construction of an *E. coli ΔycdC* strain is described that enhances the extracellular levels of α-hemolysin (HlyA) and GILL-β-lactamase (GILL-Bla) (3.03-fold) proteins via the TISS. Previous experiments have shown that synonymous rare codon substitutions in the parent gene sequence can enhance the extracellular levels of multiple proteins via the TISS. Here, the combined effect of synonymous rare codon substitutions in the parent gene sequence and the *ycdC* gene knockout on the secretion levels of recombinant proteins via the TISS is investigated.
The results suggest that the optimal levels of expression for an enhanced secretion of Bla and GILL-Bla proteins via the TISS may vary for the *E. coli* parent and the ΔycdC strain. In addition, synonymous rare rare codon substitutions in the predicted amphiphilic helix region of the C-terminal signal sequence can enhance extracellular levels of HlyA and Bla (2.07-fold) proteins via the TISS. These results describe the experimental strategies that can be used to enhance the secretion of recombinant proteins via the TISS.

Next, two kinetic models of the TISS are illustrated that describe the secretion of the HasA protein from an *E. coli* parent and a ΔsecB strain. The SecB protein is a molecular chaperone required for efficient secretion of HasA via the TISS. The models incorporate multiple cellular events including HasA synthesis, protein folding, aggregation, chaperone interactions, degradation, transporter synthesis, reporter-transporter interactions, catalytic cycle of the transporter, protein secretion into the extracellular medium and cell growth. The model simulation results are consistent with the qualitative trends in the experimental data available from the literature. A combination of techniques including parameter variation, sensitivity analysis, and nondimensionalization shows that the optimization of the extracellular concentration of HasA requires a balance between multiple factors including the HasA synthesis rate, the transporter synthesis rate and culture time. These results demonstrate the utility of a kinetic modeling approach towards enhancing extracellular recombinant protein secretion in *E. coli*. 
Chapter 1
INTRODUCTION AND OVERVIEW

1.1 Motivation

The advent of recombinant DNA (rDNA) technology in 1973 paved the way for an industry with the potential to affect a lot of patient’s lives. The annual global market for biopharmaceuticals has been estimated at about $99 billion and nearly 40% of all marketed therapeutics in the US and EU are derived from *Escherichia coli* expression systems including vaccines, interleukins, and hormones (Langer 2009, Walsh 2010). The combination of strong knowledge base, ease of genetic manipulation, relatively rapid evaluation, and analysis of effects makes *E. coli* the host organism of choice for production of many recombinant proteins (Blight and Holland 1994). Compared to other microorganisms *E. coli* has relatively simple nutritional requirements, high growth rate, and an ability to accumulate recombinant proteins to at least 20% of total cellular protein and export them from the cytoplasm to the periplasm or into the culture medium (Blight and Holland 1994). So far, *E. coli* biotechnology has established a good record of accomplishments and efforts to improve and optimize recombinant protein production in *E. coli* are of considerable economic importance.
1.2 Recombinant Protein Production in *E. coli*

Recombinant protein production in *E. coli* can be classified into three categories depending upon the final destination of the target product. The recombinant protein can accumulate in the cytoplasm, may be exported to the periplasm, or be secreted into the culture medium.

1.2.1 Cytoplasmic Expression

The first rDNA products were accumulated in the cytoplasmic space and this continues to be a feasible method to produce recombinant proteins. While the recombinant product can accumulate up to 50% of the total cell protein, it is often insoluble or not properly folded (Schein 1989). Over-expression of a recombinant protein inside a bacterial cell can result in formation of non-functional inclusion bodies that are difficult to process (Clark 2001, Villaverde and Carrio 2003). The commonly used strategy to recover active proteins from inclusion bodies usually involves inclusion body isolation and washing, solubilization of the aggregated protein, and refolding of the solubilized protein (Clark 2001). While the efficiency of the first two steps can be relatively high, renaturation yields may be limited due to the accumulation of inactive misfolded and aggregated species (Clark 2001). The final renaturation yield can range from 0-45% (Estrada et al. 2007) and often the process optimization is protein specific (Clark 2001, Estrada et al. 2007). In addition, proteins accumulated in the cytoplasm often have an extra N-terminal methionine and might be more prone to proteolysis (Ben-Bassat et al. 1987).
Several strategies have been employed to overcome these limitations. Reduced culture temperatures (23-30 °C instead of 37 °C) often favor product solubility (Schein and Noteborn 1988). The use of slower expression rates achieved by a 500-fold decrease in the inducer concentration in the culture medium in combination with reduced growth rate also decreased the likelihood of cytoplasmic aggregation by approximately 98% (Kopetzki et al. 1989). The over-expression of \textit{E. coli} methionine amino peptidase can enhance the removal of N-terminal methionine (Ben-Bassat et al. 1987). The efficiency of the methionine removal depends upon the side chain length of the second amino acid (Hirel et al. 1989), and shorter chains allows for a more efficient methionine removal.

### 1.2.2 Periplasmic Export

Secretory proteins are translocated from the cytoplasmic space into the periplasm via the general secretion pathway (Sec-pathway) or the twin arginine translocation (Tat-pathway). The Sec-pathway translocates proteins in an unfolded confirmation whereas the Tat-pathway usually translocates proteins in a folded confirmation (Natale et al. 2008). The proteins can be targeted to the Sec translocase by a co-translational or a post-translational mechanism (Natale et al. 2008). In a post-translocation mechanism, the proteins are usually completely synthesized and released from ribosomes and targeted to the Sec translocase (Natale et al. 2008). The SecB chaperone helps to maintain the protein in an unfolded confirmation and it’s targeting to the Sec translocase (Driessen 2001). In a co-translational mechanism, the signal recognition particle (SRP) binds to the signal sequence of the nascent protein, which is attached to the ribosomes (Luirink et al. 2001).
2005). The complex of SRP, nascent protein chain, and ribosome is targeted to the Sec translocase (Luirink et al. 2005). The SRP receptor (FtsY) facilitates the transfer of protein chain to the Sec tranlocase (Luirink et al. 2005). Figure 1.1 illustrates the mechanism of protein translocation via the Sec- and Tat-pathways. Periplasmic secretion of recombinant proteins offers several advantages compared to cytoplasmic production. The signal sequence is precisely removed and usually a proper N-terminal without any methionine is obtained (Gray et al. 1985). The periplasmic environment is highly oxidized, which enhances proper protein folding (Hsiung et al. 1986, Klein et al. 1991). The periplasm contains fewer protein species and proteases than the cytoplasm, reducing the likelihood of degradation and aiding in downstream processing. However, high concentrations of recombinant proteins in the periplasmic space (e.g. 22% of total cellular protein for Human apolipoprotein E) can also form inclusion bodies (Georgiou et al. 1986, Shibui et al. 1991). Strategies to reduce inclusion body formation include the use of slower expression rates and nonmetabolizable sugars (Bowden and Georgiou 1990).

### 1.2.3 Extracellular Secretion

From a biotechnology perspective, extracellular secretion of recombinant proteins is desirable as the concentration of the target product is relatively low, decreasing the likelihood of aggregation. The extracellular medium contains even fewer protein species compared to the periplasm, simplifying detection and purification of the target protein (Ni and Chen 2009). In addition, the extracellular medium can be controlled to maintain optimum pH and osmolarity for protein folding and stability. Despite these advantages
**Figure 1.1:** Schematic overview of the *E. coli* Sec, SRP and Tat translocases (a) co-translational translocation by the SRP pathway, (b) post-translational targeting routes and translocation of unfolded proteins by the Sec-translocase, and (c) translocation of folded precursor proteins by the Tat translocase. Figure is reproduced from Biochimica et Biophysica Acta (BBA) – Biomembranes, 1778(9), Natale P, Brüser T, Driessen AJ, Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane – distinct translocases and mechanisms, 1735-1756, Copyright (2008), with permission from Elsevier.

There have been only few successful attempts to secrete recombinant proteins into the culture medium (Blight and Holland 1994, Gupta and Lee 2008, Gupta and Lee 2010, Gupta et al. 2010, Lee and Lee 2005, Li et al. 2002, Sugamata and Shiba 2005, Zhang et al. 2006). The mechanism of secretion pathways is not fully understood, making targeted and rational metabolic engineering of these pathways difficult. The extracellular secretion pathways in *E. coli* are often protein specific and the secretion efficiency is low. In addition, laboratory strains of *E. coli* do not typically secrete proteins past the outer
membrane (Pugsley and Francetic 1998). Understanding and controlling translocation mechanisms enables better control over the secretion systems, aiding in engineering of the secretion pathways for enhanced extracellular production of recombinant proteins.

In Gram-negative bacteria six distinct secretion systems (Types I, II, III, IV, V and VI) have been shown to facilitate protein transport through the inner and the outer membranes (Fronzes et al. 2009, Jeremy 2009). These systems differ widely in complexity in terms of the number of components involved (Ni and Chen 2009). Recombinant protein production using Types I, II, III and V have been reported (Blight and Holland 1994, Majander et al. 2005, Zhou et al. 1999, Zhu et al. 2006). Figure 1.2 illustrates a schematic showing major components of all the extracellular secretion systems except Type VI.

The Type-II secretion systems use a two-step mechanism for translocation of proteins into the extracellular space. First, the protein is translocated through the inner membrane by the Sec-pathway (Gold et al. 2007) or the Tat-pathway (Voulhoux et al. 2001). Next, the protein is translocated through the outer membrane and into the extracellular space via the Type-II secretion system. The Type-II secretion system is composed of 12-16 protein components (Filloux 2004) that are found in the cytoplasm, periplasm, and inner and outer membranes. The Type-II secretion system has been used for extracellular production of a recombinant protein (Zhou et al. 1999). The authors successfully expressed the entire Out system (a Type-II system from *Erwinia*...
chrysanthemi) from a plasmid in an *E. coli* B strain and demonstrated that it can be used to secrete an *Erwinia chrysanthemi* endoglucanase (Zhou et al. 1999).

The Type-III secretion system, also known as the injectisome, facilitates a one-step translocation mechanism and is characteristic of several plant and animal pathogenic Gram-negative bacteria (Fronzes et al. 2009). The system translocates the effector protein in a Sec-independent manner (Cornelis and Van Gijsegem 2000). The secretion system is genetically, structurally, and functionally related to bacterial flagella (Cornelis and Van Gijsegem 2000). The system is composed of approximately 25 different proteins (Mota and Cornelis 2005) that form a large molecular structure, which crosses the bacterial cell envelope (Cornelis and Van Gijsegem 2000). Majander et al. developed a modified flagellar Type-III secretion system that was able to secrete recombinant proteins of varying sizes (100-400 amino acids) with a secretion level of 1-15 mg/L (Majander et al. 2005).

The Type-IV secretion systems are very adaptable systems that are found in both Gram-negative and Gram-positive bacteria (Fronzes et al. 2009). They secrete a wide range of substrates, from single proteins to protein–protein and protein–DNA complexes (Fronzes et al. 2009). They usually consist of 12 components that are organized into double-membrane-spanning complexes (Fronzes et al. 2009). The energy required for secretion is derived from ATP hydrolysis (Fronzes et al. 2009). There has been no reported study of use of Type-IV secretion systems for recombinant protein secretion.
Figure 1.2: Schematic overview of the major extracellular protein secretion systems (Type I, II, III, IV and V) in Gram-negative bacteria. OM: Outer Membrane; IM: Inner Membrane. Figure is reproduced by permission from Macmillan Publishers Ltd: [Nature Reviews Microbiology] (Fronzes R, Christie P. J., Waksman, G. The structural biology of type IV secretion systems. 7, 703-714). Copyright (2009).

The Type-V secretion systems include the autotransporter family, the two-partner secretion system and the Oca family (Ni and Chen 2009). Similar to the Type-II systems, secretion is a two-step process that involves translocation through the inner membrane and subsequently through the outer membrane (Desvaux et al. 2004). Over 700 proteins with functions including auto-aggregation, adherence, invasion, cytotoxicity, serum resistance, and proteolysis use the autotransporter and two-partner secretion systems to cross both inner and outer membranes (Henderson and Nataro 2001, Mazar and Cotter 2007). The Type-V secretion system has been used for recombinant protein secretion,
although the secretion efficiency is relatively low (6%) in comparison to an engineered Type-I system (88%) (Zhu et al. 2006).

The Type-VI secretion systems are recently discovered multi-component systems that are found in several pathogens such as *Pseudomonas aeruginosa*, enteroaggregative *E. coli*, *S. Typhimurium*, *Vibrio cholera*, and *Yersinia pestis* (Cascales 2008, Pukatzki et al. 2009). So far, very little is known about the architecture and function of these systems (Cascales 2008, Pukatzki et al. 2009). Some of known functions include pathogenesis, biofilm formation, and stress sensing (Bernard et al. 2010). It is hypothesized that the Type-VI systems could be composed of 12-25 subunits (Cascales 2008, Pukatzki et al. 2009).

1.3 The Type-I Secretion System (TISS)

The Type-I secretion system (TISS) is the most frequently used extracellular secretion system for recombinant protein production because of its simplicity and versatility compared to the other known secretion systems (Ni and Chen 2009). The *E. coli* TISS is the prototype and the most extensively studied Type-I translocation pathway (Ni and Chen 2009). A 107kDa α-hemolysin protein (HlyA) is the natural substrate of the *E. coli* TISS (Blight and Holland 1994). HlyA is a member of the Repeats in Toxin (RTX) family that is characterized by glycine rich repeats (Delepelaire 2004). It is a pore forming toxin that becomes inserted into the plasma membrane of the target cell (Delepelaire 2004). The TISS transports HlyA directly from the cytoplasm into the
extracellular space without a periplasmic intermediate (Binet et al 1997). HlyA secretion requires a 60 amino acid C-terminal signal sequence (CTSS) (Binet et al 1997). Interaction with chaperones or other factors is not necessary for translocation, but the secretion machinery may require the substrate to remain in an unfolded state prior to transport (Binet et al 1997). The secretion machinery is composed of three protein components, the ATP-binding cassette protein HlyB, the membrane fusion protein HlyD, and the outer membrane pore protein TolC (Holland et al. 2005). Figure 1.3 illustrates a schematic of the *E. coli* TISS showing all the components of the secretion machinery and a HlyA molecule in transit.

### 1.3.1 HlyB

ATP-binding cassette (ABC) proteins transport wide variety of substrates across membranes, coupled to the consumption of ATP (Higgins 1992). Despite the diversity of substrates, ABC-transporters share a common domain architecture of two transmembrane domains (TMDs), comprising 6-11 α-helices spanning the lipid bilayer, and two soluble nucleotide binding domains (NBDs) (Davidson and Chen 2004, Jones and George 2004, Moody and Thomas 2005, Schmitt and Tampé 2002). While the α-helical TMDs vary in their primary sequences, the NBDs are conserved with respect to primary sequence and three-dimensional structure. The high resolution crystal structure of the HlyB NBD without a bound nucleotide has been reported (Schmitt et al. 2003). It consists of a highly conserved catalytic domain and an ABC-transporter specific signaling domain. The catalytic domain is comprised of two β-sheets and seven α-helices and hydrolyzes ATP.
The signaling domain consists of five α-helices and a highly variable C-loop. HlyB has a single NDB and thus belongs to group of “half-size” ABC-transporters and association of two HlyB molecules is essential to form a functional ABC complex (Zaitseva et al. 2005b). The timing of the formation of the dimers during the transport cycle and cooperation between the two NBDs is still unclear (Janas et al. 2003). ATP hydrolysis is

**Figure 1.3**: The Type-I translocator in *E. coli*. Cartoon representation of TolC trimers in the outer membrane, surrounded by several HlyD monomers which also surround HlyB in the inner membrane, forming a continuous sealed channel from the cytoplasm to the exterior, with a molecule of HlyA in transit. The exact stoichiometry of HlyD with respect to other components remains unclear. Figure is reproduced with kind permission from Springer Science + Business Media: <Biotechnology Letters, Extracellular recombinant protein production from *Escherichia coli*, 31, 2009, 1661-1670, Ni Y, Chen R, figure number:1>
required for translocation but not for assembly of the translocation complex; however, it
is unknown which steps require ATP hydrolysis or how the energy is transferred to
translocate the substrate (Holland et al. 2005, Thanabalu et al. 1998).

1.3.2 HlyD

HlyD is a member of the membrane fusion protein (MFP) family, often a part of
membrane transporters and antiporters. MFP genes are often adjacent to the genes
encoding the ABC proteins (Wandersman 1996). HlyD is anchored into the cytoplasmic
membrane via a transmembrane domain (Holland et al. 2005). No structural studies of
HlyD or its homologues have been reported (Holland et al. 2005). The topological model
of HlyD indicates a cytoplasmic N-terminal region that consists of approximately 60
residues, a predicted α-helical region of 21 residues spanning the inner membrane, and a
periplasmic region consisting of approximately 160 residues, which is predicted to
contain four extended α-helices and a C-terminal, largely β-strand domain (Schulein et al.
1992). HlyD is an integral and essential part of the transport machinery and mutations in
HlyD can decrease the activity and efficiency of secreted HlyA (Pimenta et al. 2005).

1.3.3 TolC

TolC is an outer membrane protein (OMP) that forms trimers and is essential for
secretion via the TISS (Holland et al. 2005). Mutations in TolC can result in the secretion
of enzymatically inactive HlyA (Vakharia et al. 2001). There is evidence to suggest that
TolC is not substrate specific (Wandersman 1996) and is used for secretion of both
cocilin V-1 and HlyA in *E. coli* (Holland et al. 2005, Zhong et al. 1996). The crystal structure of TolC shows three subunits forming a continuous tunnel of over 140 Å in length, formed from a β-barrel domain in the outer membrane adjoining a 100 Å long cylinder in periplasm, formed mainly by 12 extended antiparallel helices (Koronakis et al. 1991, Holland et al. 2005). The β-barrel domain forms a pore that opens to the exterior with an internal diameter of 19.8 Å (Andersen et al. 2002). Using mutagenesis, cross linking studies, and electrophysiological measurements, a model for the movement of the helices to control the opening/closing of periplasmic entrance has been proposed (Andersen et al. 2002, Eswaran et al. 2003, Koronakis et al. 2000). The entrance is closed when not associated with other proteins of the secretion machinery, thus preventing the free diffusion of hydrophilic molecules into or out of the cell (Andersen et al. 2000).

1.4 Application of TISS in Extracellular Recombinant Protein Secretion

The TISS provides a reliable and flexible method for secreting recombinant proteins into the extracellular medium (Blight and Holland 1994). In an earlier study, the TISS in *E. coli* was used to secrete the cytoplasmic protein dihydrofolate reductase (DHFR) (Nakano et al. 1992). The secretion of DHFR was inversely correlated with the intracellular activity; however, a single amino acid change in a β-sheet of DHFR resulted in the extracellular secretion of enzymatically active DHFR (Nakano et al. 1992). Also, the *E. coli* TISS was used to successfully secrete a metalloprotease from *Serratia marcescens* (Suh and Benedik 1992). The metalloprotease shares a repeating motif, LXGGXGND, with HlyA. Further studies have shown that the TISS in *E. coli* is
extremely versatile and can be used for extracellular secretion of proteins unrelated to HlyA and from different species including *Salmonella typhimurium*, *Listeria monocytogenes*, and human (Gentschev et al. 1996). However, the requirements for an unfolded protein in the cytoplasm and unassisted folding in the extracellular medium restricts the transport of few proteins such as full length *E. coli* chloramphenicol acetyltransferase and β-galactosidase via the TISS (Blight and Holland 1994). The secretion efficiency via the TISS is up to 3-5% of total cellular protein, comparable to the secretion efficiency via the Sec- and Tat-pathways (Blight and Holland 1994). Reported secretion efficiency measured as a percentage of expressed proteins that is secreted into the medium ranges from 1-90% (Gentschev et al. 1996).

The TISS has also been used successfully for the extracellular secretion of proteins containing disulfide bonds. The CTSS fused to the β-lactamase (Bla) protein results in a secreted product that is folded and active (Jumpertz et al. 2010). Bla contains two disulfide bonds and it is typically secreted into the periplasm, which aids in folding and disulfide bond formation (Bowden and Georgiou 1990). Also, single-chain Fv (scFv) and single domain antibodies were successfully secreted via the TISS into the culture medium at the concentrations comparable to those obtained by their periplasmic secretion (Fernandez 2004). The exact mechanism of folding and oxidation via the TISS is still unclear but is believed to be independent of the conventional periplasmic chaperones (Fernandez 2004).
The TISS from other microbes can be successfully expressed in *E. coli* and used for extracellular recombinant protein secretion. Alkaline phosphatase (AP), normally secreted into the periplasmic space, when fused to a C-terminal region of *Pseudomonas* sp. MIS38 lipase (PML) is secreted into the extracellular medium from *E. coli* cells carrying the *Serratia marcescens* Lip system (Angkawidjaja et al. 2006). PML is composed of 617 amino acid residues and, like most of RTX proteins, consists of the N-terminal catalytic domain and C-terminal domain containing a secretion signal and 12 glycine-rich repetitive sequences. The repetitive sequences were not necessary for secretion of the fusion protein, as the extracellular level of the fusion protein containing all the repeats was only 2-fold higher compared to the extracellular level of the fusion protein without any repeats (Angkawidjaja et al. 2006). In a recent study, TliA (a lipase of *Pseudomonas fluorescens*) and lipase ABC transporter domains (LARDs) of different lengths were investigated for extracellular recombinant protein secretion in *E. coli* carrying TISS of either *Pseudomonas fluorescens* or *Erwinia chrysanthemi* (Chung et al. 2009). Both green fluorescent protein and epidermal growth factor fused with whole TliA or LARDs at their C-terminal were secreted into the extracellular medium (Chung et al. 2009).

Overall, the TISS is very versatile and provides a good method for extracellular secretion of recombinant proteins in *E. coli*. However, there are some disadvantages in using the TISS for extracellular protein production. The reported levels of secretion via the TISS is typically 10 mg/L or less, which may be sufficient for certain applications.
such as antigen delivery for vaccines (Gentschev et al. 1996) or high throughput screening; however, further improvements are required for the system to be used as a production method (Ni and Chen 2009). Also, the product is a fusion protein, which may require additional downstream processing steps after fermentation (Ni and Chen 2009).

1.5 Factors Affecting Secretion Efficiency

1.5.1 Role of Protein Synthesis Rate in Protein Secretion

Protein synthesis rate plays an important role in the secretion of recombinant proteins (Gupta and Lee 2008, Gupta and Lee 2010, Gupta et al. 2010, Lee and Lee 2005, Simmons and Yansura 1996). Some of the factors that affect protein synthesis rate include plasmid copy number (Mergulhao et al. 2003), translation initiation region (Simmons and Yansura 1996), and differences in codon usage (Williams et al. 1988). Plasmid copy number can have an impact on the utilization of cellular resources including the number of free ribosomes available for protein translation (Ikemura 1981a). Use of plasmids of moderate copy number (15-60) can enhance the periplasmic secretion of human proinsulin in comparison to using low copy number plasmids (Mergulhao et al. 2003).

The translation initiation region (TIR) plays a major role in determining the overall translation level of a protein (Simmons and Yansura 1996). Simmons and Yansura 1996 created a library of TIR variants by making silent mutations in codons 2-6 of the Heat-Stable Enterotoxin II (STII) signal sequence that translocates proteins via the
Sec-pathway. The activity of the PhoA protein fused to the library of the STII signal sequence was used for the quantification of relative strength of the TIR variants (Simmons and Yansura 1996). It was observed that optimization of TIR, i.e. protein translation level was required to maximize the secretion of heterologous proteins and the optimal TIR needed to be determined empirically for each protein (Simmons and Yansura 1996). Although these studies focused on the periplasmic secretion of recombinant proteins, they demonstrate that an optimum translation rate exists to achieve maximal recombinant protein secretion and that optimal rate may be protein specific.

Balancing the rates of translation and secretion can also enhance secretion efficiency via the TISS. Slowing translation rates by synonymous rare codon substitutions in the parent gene sequences significantly increased the extracellular secretion efficiency of multiple recombinant proteins via the TISS (Gupta and Lee 2008, Gupta and Lee 2010, Lee and Lee 2005). Rare codons are defined as those codons whose corresponding tRNA concentration is less than 1% of the total tRNA concentration (as tabulated in Solomovici et al. 1997). The use of rare codons in a gene sequence reduces the translation rate because the number of available aminoacyl-tRNAs is limited (Ikemura 1981a). Using genomic, transcriptomic, and proteomic analysis of a hypersecreting strain, it was shown that the premature termination of the RutR protein, encoded by the ycdC gene, can increase the secretion efficiency via the TISS (Gupta et al. 2010, Lee and Lee 2005). The authors proposed a role for RutR in translation and consequently protein secretion via the TISS (Gupta et al. 2010). To the best of our
knowledge this is the first report that implicates the \textit{ycdC} gene in protein secretion. These studies present a single gene target that can enhance recombinant protein secretion via the TISS.

1.5.2 Role of the CTSS in Protein Secretion via the TISS

Secretion of HlyA via the TISS is dependent upon a 60 amino acid CTSS (Binet et al. 1997). Analysis of point mutations and deletions created in the CTSS suggests that there are only few critical residues important for HlyA secretion and they are dispersed throughout the signal sequence (Chervaux and Holland 1996, Kenny et al. 1992, Kenny et al. 1994, Koronakis et al. 1989). Secretion levels are not affected if the signal sequence of HlyA is replaced by the signal sequence of \textit{Pasteurella hemolytica} leukotoxin (Zhang et al. 1993). Comparison of the two signal sequences revealed that they share very little primary sequence homology but there appears to be a common predicted helix-linker-helix motif (Yin et al. 1995, Zhang et al. 1995). Further, a similar motif for the 56 C-terminal residues of \textit{Erwinia chrysanthemi} protease G, which is secreted by a system analogous to TISS, has been reported (Wolff et al. 1994). This conservation of secondary structure has led to a hypothesis that higher order secondary structure and not the primary sequence might be more important for secretion via the TISS (Hui et al. 2000).

Combinatorial analysis of different variants created by changing primary sequence of the CTSS suggests that only the sequences that result in a predicted amphiphilic helical structure and appropriate hydrophobicity can sustain efficient transport (Hui et al. 2000, Hui and Ling 2001). Mutations in the CTSS affected the correct folding of both HlyA and
Bla fused with the CTSS (Jumpertz et al. 2010). The authors proposed dual functionality of the CTSS involved in both post-translocation folding and targeting of HlyA for secretion (Jumpertz et al. 2010). Overall, the CTSS presents an attractive target for enhancing recombinant protein secretion via the TISS.

1.6 Mathematical Modeling as a Tool to Understand Complex Biological Systems

In recent years, use of multiparameter mathematical models of biological regulation in prokaryotes and eukaryotes has become increasingly important (Chen et al. 2000, Janes and Lauffenburger 2006, Klipp and Liebermeister 2006, Papin et al. 2003). Biological regulation is now understood to be much more complex, multilayered, and branched than previously thought, limiting the understanding of these systems and their regulation by intuition alone (Brown et al. 2004). Appropriately formulated mathematical models can assist in the illustration of pathways in a physically and biologically realistic manner, include a large array of empirical observations, and generate important hypotheses that can play a major role towards better understanding of such pathways (Aldridge et al. 2006).

The correct selection of mathematical framework for any system depends upon the properties of the system being studied and on the reasons and goals for modeling. Mathematical models can be organized into three major groups depending on the level of detail: deterministic, stochastic (probabilistic), and statistical (Janes and Lauffenburger 2006). Both deterministic and stochastic are mechanism based models (Janes and
Lauffenburger 2006). In a deterministic model every set of variable states is uniquely determined, based on their previous states and the parameters in the model. Hence, for a given set of initial conditions, deterministic models will always perform the same way. However, in a stochastic model, randomness is present, and variable states are not described by unique values, but rather by probability distributions. Hence, for a given set of initial conditions, stochastic models usually have an average behavior and random fluctuations around that behavior. Statistical models are more data-driven and they are most commonly used when the prior knowledge and understanding of the system mechanism is fairly limited (Janes and Lauffenburger 2006).

Ordinary and partial differential equations (ODEs and PDEs) are commonly used to develop deterministic and stochastic models. PDEs are relations involving unknown function(s) of several independent variables and their partial derivatives with respect to those variables. Typical independent variables include time and space coordinates. Conversely, ODEs are relations involving function(s) of one independent variable and one or more of their derivatives with respect to that variable. Currently, the most common way of representing a biochemical network is through a set of coupled non-linear deterministic ODEs (an ODE network) (Aldridge et al. 2006, Balsa-Canto et al. 2010). The equations in an ODE network typically describe the production and the consumption rates for all the biomolecular species considered in the network. An ODE network can be further classified as generalized mass-action-based model and power-law model based upon the framework selected to construct the production or consumption rate equations.
mathematical modeling of complex biological systems often has several challenges. Any mathematical model based on equation of chemical kinetics will have large number of parameters, which are generally unknown, and therefore accurate quantitative predictions can be difficult (Bailey 2001). Rate constants can be obtained from in vitro experiments; however, use of in vitro rate constants is still an approximation for the actual in vivo rate constant. Statistical ensemble techniques have proven useful to
make important predictions from mathematical models with largely unknown parameters (Brown and Sethna 2003, Brown et al. 2004). Also, mathematical models are almost always incomplete as there always will be interactions and influences that are relevant to the cell function under consideration, but are ignored in the mathematical modeling process. Further, ongoing research reveals new interactions and players in biological networks and therefore mathematical models needs to be flexible enough to incorporate new information (Brown et al. 2004). Despite the simplifications, approximations, and challenges, mathematical modeling has proven of significant use in illuminating biological mechanisms and guiding experimental investigations. Mathematical modeling has been extensively used to model numerous biological processes including cell cycle in yeast (Chen et al. 2000), circadian rhythms (Goldbeter 1995), programmed ribosomal frameshifting (Liao et al. 2008), signaling pathways (Aldridge et al. 2006, Balsa-Canto et al. 2010, Brown et al. 2004, Vera et al. 2007), microbial growth (Zwietering et al. 1990), quorum sensing (Fagerlind et al. 2003), enzyme kinetics (Chen et al. 2010), protein folding, aggregation, chaperone interaction, and inclusion body formation (Hoffman et al. 2001, Kiefhaber et al. 1991, Krishnan et al. 2009), recombinant protein production (Bentley and Kompala 1989, Lee et al. 1985, Ramirez et al. 1994), single cell growth (Shuler et al. 1979), antibiotic biosynthesis (Mehra et al. 2008) etc.

To the best of our knowledge, there is no reported study of mathematical modeling of extracellular recombinant protein secretion in \textit{E. coli}. Extracellular secretion of recombinant proteins has several advantages and it can play an important role in
bioprocess economics. Significant progress has been made in recent years to engineer *E. coli* to secrete recombinant proteins into the extracellular medium and encouraging results have been obtained. However, there are still significant challenges to overcome. The product concentrations are low (Ni and Chen 2009), different proteins usually secrete with different efficiencies (Gentschev et al. 1996), and predictability of the process is limited (Ni and Chen 2009). The secretion of proteins via the TISS is a complex process involving highly non-linear branched pathways and interconnected modules involving several other cellular events. Application of mathematics and computation can play an important role in deciphering, controlling, and optimizing recombinant protein secretion via the TISS.

1.7 Thesis Goals

The overarching goal of this project is to understand, develop, and optimize strategies to engineer *E. coli* for enhanced extracellular secretion of active recombinant proteins. The major sub-goals of this thesis are as follows:

1) Construct a hypersecreter strain and study extracellular secretion of recombinant proteins via the TISS. This goal is based on our previous observations (Gupta et al. 2010, Lee and Lee 2005).

2) Employ targeted synonymous rare codon engineering strategy in the CTSS to enhance extracellular secretion of recombinant proteins. This goal aims to discover novel silent mutations that can enhance extracellular recombinant protein secretion via the TISS.
3) Develop kinetic models of the TISS. This goal aims to understand the complexity of the TISS using mathematical and computational tools and design strategies for the optimization of extracellular secretion of a model protein via the TISS.

1.8 Scope of the Work

This dissertation outlines the computational and experimental techniques to understand, control, and enhance extracellular secretion of recombinant proteins via the TISS in *E. coli*. Chapter 2 describes the construction of a hypersecretion strain using a homologous recombination technique. The results of secretion studies in the hypersecretion strain are presented. Further, a targeted synonymous rare codon engineering strategy is described that reveals novel silent mutations, which enhances secretion of recombinant proteins. Chapter 3 describes two kinetic models to understand and design strategies for the optimization of extracellular recombinant protein secretion. In chapter 4, results and major conclusions of this work are summarized and few recommendations of the future work are presented.
Chapter 2

GENETIC ENGINEERING APPROACHES TO ENHANCE EXTRACELLULAR SECRETION OF RECOMBINANT PROTEINS

2.1 Introduction

The Type-I secretion system (TISS) in uropathogenic *E. coli* when expressed in the laboratory *E. coli* K12 strains provides a reliable and an adaptable method to secrete recombinant proteins into the extracellular space (Blight and Holland 1994, Gentschev et al. 1996; Gupta and Lee 2010). The reported level of secretion via the TISS is typically 10 mg/L or less, which may be sufficient for certain applications such as antigen delivery for vaccines (Gentschev et al. 1996) or high throughput screening; however, further improvements are required for the system to be used as a production method (Ni and Chen 2009). The mechanism of protein secretion via TISS is not fully understood, making the rational design of hypersecreting strains difficult.

Several studies have focused on enhancing the extracellular levels of proteins secreted via the TISS. Experimental studies using random mutagenesis of the TISS components, HlyB and HlyD, led to an improved secretion efficiency of the subtilisin E protein fused to the HlyA signal sequence (Sugamata and Shiba 2005). More than 15,000 *E. coli* JM109 transformants, in which the *hlyB* and *hlyD* genes had been mutagenized
using the error-prone polymerized chain reaction (PCR), were screened for enhanced extracellular secretion of the fusion protein. Two HlyB mutants were isolated that had 27- and 15-fold higher activity of the subtilisin E protein in the extracellular space when compared to the wild type (Sugamata and Shiba 2005). In a separate study several point mutations were identified in the periplasmic domain of HlyD that decreased both the secretion efficiency and activity of secreted HlyA (Pimenta et al. 2005). This was the first report to show that HlyD not only forms an integral component of the TISS but also affects the folding of secreted HlyA (Pimenta et al. 2005).

Both the studies described above focused on the proteins directly involved in the secretion process; however, there can be associated pathways and events that may be important in engineering of improved recombinant protein secretion via the TISS. A systems wide approach using genomics, transcriptomics, and proteomics tools can be useful in identifying the metabolic bottlenecks in this process. A random mutagenesis approach was used to screen *E. coli* W3110 parent cells secreting HlyA for identification of hypersecreting strains (Lee and Lee 2005). Approximately 800 plates were screened for mutants, with up to 300 separate colonies per plate and the mutant B41 was selected that secreted 4-fold more active HlyA compared to the parent strain (Lee and Lee 2005). This mutant was further characterized using Affymetrix gene chip probe analysis and two-dimensional gel based proteomics strategy to examine mRNA and protein expression profiles (Lee and Lee 2005). The B41 mutant showed decreased levels of mRNA and protein expression of amino acyl tRNA synthetases compared to the parent strain,
suggesting that the B41 mutant may exhibit a decreased protein translation rate (Lee and Lee 2005). Subsequently, synonymous rare codon changes were made in a certain region of the *hlyA* gene that resulted in a 37% decrease in the predicted translation rate and an 8-fold increase in the extracellular levels of HlyA (Lee and Lee 2005). The increase in extracellular levels of HlyA resulted from a decrease in intracellular aggregation of the engineered protein (Gupta and Lee 2008). The strategy of using synonymous rare codon substitutions was successfully extended to enhance the secretion levels of several recombinant proteins fused with the CTSS including β-lactamase (Bla), scFv 13, interleukin-6, and the putative cancer vaccine NY-ESO-1 (Gupta and Lee 2010).

The genomes of the B41 mutant and the parent *E. coli* W3110 strain were sequenced at 52.8x and 55x coverage, respectively, using the Illumina sequencing technology (Gupta et al. 2010). A single nucleotide polymorphism of G to T at position 1,074,787 was found in the B41 genome relative to the parent strain (Gupta et al. 2010). The mutation results in the premature termination of the RutR protein, which is encoded by the *ycdC* gene. The hypersecretion phenotype was confirmed in an *E. coli* MG1655 *ycdC::Tn5* mutant and a 3.4-fold increase in the extracellular levels of HlyA was observed, which is consistent with the observations from the B41 mutant (Gupta et al. 2010, Lee and Lee 2005). Comparative mRNA expression profiling of the *E. coli* MG1655 parent and the *E. coli* MG1655 *ycdC::Tn5* strain showed decreased expression of nearly all the tRNA-synthetases and some amino transporters in the *ycdC::Tn5* strain,
which is also consistent with the observations from the B41 mutant (Gupta et al. 2010, Lee and Lee 2005).

The RutR protein is a putative transcription factor consisting of 212 amino acid residues belonging to the TetR family, which is a group of DNA binding transcription factors composed of a N-terminal DNA binding domain and a C-terminal ligand binding domain (Ramos et al. 2005, Shimada et al. 2008). RutR plays a major role in the synthesis of glutamine via the Gad system, the conversion of glutamine to carbamoyl phosphate, the synthesis of pyrimidines, the degradation of pyrimidines, the downstream pathway of purine degradation, as well as the synthesis of arginine (Shimada et al. 2008). Recent studies described are first to reveal a possible role of RutR in protein translation and secretion via the TISS. This makes RutR an attractive metabolic engineering target to enhance extracellular secretion of recombinant proteins in *E. coli* via the TISS.

Recent work demonstrates that synonymous codon substitutions can impact translational kinetics, which subsequently affects the final protein structure and function (Komar 2009). The ribosomal tunnel is a dynamic environment where nascent proteins fold co-translationally, which influences the nascent protein structure (Kramer et al. 2001). Hence, slight changes in the elongation rates due to synonymous codon substitutions alter secondary structure development in the nascent protein and modulate protein folding (Komar 2009).
HlyA secretion is dependent upon a 60 amino acid CTSS (Binet et al. 1997). Figure 2.1 illustrates a schematic of the CTSS showing the primary amino acid sequence and location of the predicted amphiphilic helix. Previous studies suggest that the secondary structure of the CTSS is more important for efficient secretion of HlyA compared to the primary sequence (Hui et al. 2000, Hui and Ling 2001). A combinatorial analysis of different variants generated by changing primary sequence of the CTSS indicates that only the sequences that result in a predicted amphiphilic helical structure and appropriate hydrophobicity can sustain efficient transport of HlyA (Hui et al. 2000, Hui and Ling 2001). In contrast, earlier studies have shown that the specific residues in the CTSS and not the specific secondary structure are essential for efficient HlyA secretion (Chervaux and Holland 1995). One model of the CTSS function hypothesizes that the essential residues of the CTSS are presented to the presumptive binding pocket of the transport machinery in an appropriate three-dimensional structure that is defined by the conserved secondary structure of the CTSS (Hui et al. 2000). We hypothesize that promoting the formation of the predicted amphiphilic helical structure in the CTSS by making synonymous rare codon changes in the Helix1 region (Figure 2.1) can enhance secretion levels via the TISS.

In this study, construction of an E. coli W3110 ΔycdC strain using a homologous recombination technique is described. Recombinant protein secretion via the TISS is compared between the ΔycdC strain the isogenic parent strain. The combined effect of synonymous rare codon changes in the gene sequence and the ycdC gene knockout on
secretion levels of recombinant fusion proteins is investigated. Next, synonymous rare
codon changes are made in the Helix1 and the Cterm1 region (Figure 2.1) of the CTSS
using a targeted site directed mutagenesis approach. The effect of these changes on the
secretion levels of HlyA and recombinant fusion proteins is examined. These results
demonstrate the utility of genetic engineering approaches in E. coli strain development
for enhanced extracellular secretion of recombinant proteins via the TISS.

Figure 2.1: Schematic of 60 amino acids of the E. coli HlyA CTSS showing the location
of the predicted amphiphilic helix (Helix1) and Cterm1 regions where synonymous rare
codon substitutions were made. The figure is adapted from Hui and Ling 2001.

2.2 Materials and Methods

2.2.1 Plasmids and Strains

All plasmids and strains used in this study are shown in Table 2.1 and 2.2.

2.2.2 Site-Directed Mutagenesis

Primers used for site directed mutagenesis of the CTSS are shown in Table 2.3.
The primers were obtained from IDT Technologies (Coralville, IA). The QuikChange® II
site-directed mutagenesis kit (Agilent Technologies, CA) was used with an icycler™
thermocycler (Bio-Rad, CA). The desired changes were confirmed by sequencing of purified plasmids at DNA services (Cornell University, NY).

Table 2.1: Plasmids used in this study and their sources.

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Genes</th>
<th>Resistance</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pWAM716</td>
<td>hylBD</td>
<td>Chloramphenicol</td>
<td>Felmlee et al. 1988</td>
</tr>
<tr>
<td>pWAM1097_ kan_Bla</td>
<td>bla</td>
<td>Kanamycin</td>
<td>Gupta and Lee 2010</td>
</tr>
<tr>
<td>pWAM1097_HisHlyA</td>
<td>6X-Histagged hlyA</td>
<td>Ampicillin</td>
<td>Gupta and Lee 2008</td>
</tr>
<tr>
<td>pWAM1097_kan_GILL-Bla</td>
<td>gill-bla</td>
<td>Kanamycin</td>
<td>Gupta and Lee 2010</td>
</tr>
<tr>
<td>pblaM1, pblaM2, pblaM3</td>
<td>bla</td>
<td>Kanamycin</td>
<td>Gupta and Lee 2010</td>
</tr>
<tr>
<td>pgillM1, pgillM2, pgillM3</td>
<td>gill-bla</td>
<td>Kanamycin</td>
<td>Gupta and Lee 2010</td>
</tr>
<tr>
<td>pWAM1097_HisHlyA_Helix1, pWAM1097_HisHlyA_Cterm1</td>
<td>6X-Histagged hlyA</td>
<td>Ampicillin</td>
<td>This study (site directed mutagenesis of pWAM1097_HisHlyA)</td>
</tr>
<tr>
<td>pWAM1097_ kan_Bla_Helix1, pWAM1097_ kan_Bla_Cterm1</td>
<td>bla</td>
<td>Kanamycin</td>
<td>This study (site directed mutagenesis of pWAM1097_ kan_Bla)</td>
</tr>
</tbody>
</table>

Table 2.2: Strains used in this study and their origin.

<table>
<thead>
<tr>
<th>Strains</th>
<th>Plasmid(s)</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>W3110</td>
<td>None</td>
<td>This study (W3110-based)</td>
</tr>
<tr>
<td>ΔycdC</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>W3110 HlyA</td>
<td>pWAM1097_HisHlyA, pWAM716</td>
<td>W3110</td>
</tr>
<tr>
<td>W3110 Bla-W</td>
<td>pWAM1097_ kan_Bla, pWAM716</td>
<td>W3110</td>
</tr>
<tr>
<td>W3110 GILL-W</td>
<td>pWAM1097_ kan_GILL-Bla, pWAM716</td>
<td>W3110</td>
</tr>
<tr>
<td>W3110 HlyA_Helix1</td>
<td>pWAM1097_HisHlyA_Helix1, pWAM716</td>
<td>W3110</td>
</tr>
<tr>
<td>W3110 HlyA_Cterm1</td>
<td>pWAM1097_HisHlyA_Cterm1, pWAM716</td>
<td>W3110</td>
</tr>
<tr>
<td>W3110 Bla_Helix1</td>
<td>pWAM1097_ kan_Bla_Helix1, pWAM716</td>
<td>W3110</td>
</tr>
<tr>
<td>W3110 Bla_Cterm1</td>
<td>pWAM1097_ kan_Bla_Cterm1, pWAM716</td>
<td>W3110</td>
</tr>
<tr>
<td>W3110 Bla-W (BD-)</td>
<td>pWAM1097_ kan_Bla</td>
<td>W3110</td>
</tr>
</tbody>
</table>
2.2.3 Construction of the *E. coli* ΔycdC Strain

The genomic knockout of the *ycdC* gene from the parent *E. coli* W3110 cells was created by adapting the technique of homologous recombination (Datsenko and Wanner 2000). The PCR products used for the knockout were generated using the forward and the reverse primer (Table 2.4) and the plasmid pKD4 (Datsenko and Wanner 2000) as a template. The PCR products include a kanamycin resistance gene flanked by the Flippase Recognition Target (FRT) sites along with a 50 nucleotide homology extension to the targeted region on the *E. coli* genome on either side. The *E. coli* W3110 parent cells transformed with a Red helper plasmid (pKD20) (Datsenko and Wanner 2000) were

<table>
<thead>
<tr>
<th>W3110 GILL-W (BD-)</th>
<th>pWAM1097_kan_GILL-Bla</th>
<th>W3110</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔycdC HlyA</td>
<td>pWAM1097_HisHlyA, pWAM716</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>ΔycdC Bla-W</td>
<td>pWAM1097_kan_Bla, pWAM716</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>ΔycdC GILL-W</td>
<td>pWAM1097_kan_GILL-Bla, pWAM716</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>ΔycdC Bla-M1</td>
<td>pblaM1, pWAM716</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>ΔycdC Bla-M2</td>
<td>pblaM2, pWAM716</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>ΔycdC Bla-M3</td>
<td>pblaM3, pWAM716</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>ΔycdC GILL-M1</td>
<td>pgillM1, pWAM716</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>ΔycdC GILL-M2</td>
<td>pgillM2, pWAM716</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>ΔycdC GILL-M3</td>
<td>pgillM3, pWAM716</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>ΔycdC Bla-W (BD-)</td>
<td>pWAM1097_kan_Bla</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>ΔycdC GILL-W (BD-)</td>
<td>pWAM1097_kan_GILL-Bla</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>ΔycdC Bla-M1 (BD-)</td>
<td>pblaM1</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>ΔycdC Bla-M2 (BD-)</td>
<td>pblaM2</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>ΔycdC Bla-M3 (BD-)</td>
<td>pblaM3</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>ΔycdC GILL-M1 (BD-)</td>
<td>pgillM1</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>ΔycdC GILL-M2 (BD-)</td>
<td>pgillM2</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>ΔycdC GILL-M3 (BD-)</td>
<td>pgillM3</td>
<td>ΔycdC</td>
</tr>
</tbody>
</table>
grown at 30 °C in Luria Bertani (LB) medium supplemented with ampicillin (100 μg/mL) and L-arabinose (0.4%). The Red helper plasmid is a low copy number and an easily curable plasmid that expresses phage λ Red recombinase under the control of a L-arabinose inducible promoter. Aliquots were harvested at OD$_{600}$ = 0.6 in a microcentrifuge tube by centrifuging at 11,000 rpm and 4 °C for 30 sec. The cells were made electrocompetent by washing (4 times) with 15% glycerol (pre-chilled on ice) at 11,000 rpm and 4 °C for 30 sec. The cell pellet was resuspended in 20-30 μL of 15% glycerol and ~400 ng of the PCR products were added. The contents of the microcentrifuge tube were gently mixed and transferred to a pre-chilled 1 mm electroporation cuvette (Bio-Rad, CA). Cells were electroporated by a 5 ms pulse at 1.8 KV using the Bio-Rad mini processor. The electroporated cells were resuspended in 1 mL of preheated (37 °C) rich LB media containing MgCl$_2$ (12.5 mM), MgSO$_4$ (12.5 mM), and glucose (20 mM). The cells were incubated at 37 °C for 3 hrs and then concentrated by centrifugation at 4,000 x g for 3 min and spread on LB-kanamycin (25 μg/mL) plates to select for cells containing the kanamycin resistance gene in the genome.

Table 2.3: Primers used in the site directed mutagenesis of the Helix1 and the Cterm1 region of the CTSS.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helix1-forward</td>
<td>5′-GCC TAT GGA AGT CAG GGT AAT CTT AAT CCA CTA ATA AAT GAA ATA AGC AAA ATA ATA TCA GCT GCA GGT A-3′</td>
</tr>
<tr>
<td>Helix1-reverse</td>
<td>5′- TAC CTG CAG CTG ATA TTA TTT TGC TTA TTT CAT TTA TTA GTG GAT TAA GAT TAC CCT GAC TTC CAT AGG C-3′</td>
</tr>
<tr>
<td>Cterm1-forward</td>
<td>5′-CGG TAA TGC CAG TGA TTT TAG TTA TGG GCG GAA CTC AAT AAC TTT GAC AGC-3′</td>
</tr>
<tr>
<td>Cterm1-reverse</td>
<td>5′-GCT GTC AAA GTT ATT GAG TTC CGC CCA TAA CTA AAA TCA CTG GCA TTA CCG-3′</td>
</tr>
</tbody>
</table>
The colony PCR technique was used to confirm the integration of the kanamycin resistance gene using the primers P1, K1, P2, and K2 (Table 2.4). After the confirmation of the integration, the kanamycin resistant mutants were transformed with the plasmid pCP20 (Datsenko and Wanner 2000) and ampicillin-resistant transformants were selected at 30 °C. The pCP20 is an ampicillin and chloramphenicol resistant plasmid that shows temperature-sensitive replication and thermal induction of FLP synthesis. Few colonies of the kanamycin resistant mutants containing the pCP20 plasmid were purified at 43 °C and then tested for loss of all antibiotic resistances. The majority of the clones lost the FRT-flanked kanamycin resistance gene and the FLP helper plasmid simultaneously as confirmed by the colony PCR and spreading of the clones on the LB-kanamycin and/or ampicillin plates.

2.2.4 Protein Fractionation

The protein fractionation protocol was adapted from Gupta and Lee 2004. Single colonies of the cells transformed with the appropriate plasmids were grown at 37 °C in the LB media supplemented with appropriate antibiotics (ampicillin, 75µg/mL; kanamycin, 30 µg/mL; chloramphenicol, 85 µg/mL). Aliquots were harvested at OD₁₅₀₀ = 1.0 for all the samples and then centrifuged for 10 min at 4 °C and 4,000 x g. An aliquot of 500 µL of the supernatant fraction was mixed with two volumes of ice-cold ethanol and incubated overnight at -20 °C. The precipitated protein was pelleted by centrifugation at 5,000 rpm for 15 min at 4 °C and resuspended in 20 µL sodium dodecyl sulfate (SDS) sample buffer and analyzed using SDS-PAGE (polyacrylamide gel electrophoresis) and
western analysis. For intracellular protein fractionation from the whole cell pellets, the pellets from 1 mL of culture were dissolved in 100 µL of SDS sample buffer, heated at 95 °C for 10 min, and analyzed using SDS-PAGE and western analysis.

**Table 2.4:** Primers used for the knockout of the *ycdC* gene and its confirmation.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primers used for the knockout</strong></td>
<td></td>
</tr>
<tr>
<td>Forward</td>
<td>5’-GCC AAT CTA CAA GAG GGG AGA GCG CAT GAC GCA AGG CGC AGT GAA AAC AAG TGT AGG CTG GAG CTG CTT C-3’</td>
</tr>
<tr>
<td>Reverse</td>
<td>5’-ATG TTA CAA CCT CCT CCG GCA TCT TTA ACG TGG TCG AAT CCC CTC AAT AAC ATA TGA ATA TCC TCC TTA G-3’</td>
</tr>
<tr>
<td><strong>Primers used for the confirmation</strong></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>5’-AGG TAA CGT TAA GGT GGC AGC AGT-3’</td>
</tr>
<tr>
<td>K1</td>
<td>5’-CAG TCA TAG CCG AAT AGC CT-3’</td>
</tr>
<tr>
<td>P2</td>
<td>5’-TCG GCC CAT GTT GGT AAC CTG TAT-3’</td>
</tr>
<tr>
<td>K2</td>
<td>5’-CGG TGC CCT GAA TGA ACT GC-3’</td>
</tr>
</tbody>
</table>

**2.2.5 Western Analysis**

Protein fractions were resolved by SDS-PAGE and immunoblotted. Rabbit anti-β-lactamase (1:3,000; Millipore) was used as primary antibody for the detection of Bla and GILL-Bla. Mouse-anti-6X-His (1:3,000; Sigma) was used as the primary antibody for the detection of HlyA. Alkaline phosphatase conjugated mouse anti-rabbit IgG antibody (1:30,000; Sigma) and goat anti-mouse (1:30,000; Sigma) were used as secondary antibodies. The enhanced chemifluorescence (ECF) substrate (GE Amersham Biosciences) was used for the detection of bound antibodies. A FLA-3000G laser fluorescence scanner (Fujifilm Medical Systems) was used for imaging. Quantitative
analysis was done using ImageMaster 2D Platinum Software v5.0 (GE Amersham Biosciences).

2.3 Results and Discussion

2.3.1 Construction of the E. coli ΔycdC Strain

In a reported study, chromosomal mutants were isolated with 13 different gene disruptions by using the homologous recombination technique that involves direct transformation of the E. coli carrying a Red helper plasmid with the PCR products having short homology extensions for the targeted locus (Datsenko and Wanner 2000). Here, we have used this technique to construct the E. coli ΔycdC strain (Figure 2.2). The ycdC gene in the E. coli genome was replaced with a kanamycin resistance cassette that contains a kanamycin resistance gene flanked by the FRT sites on either side. The cassette was generated by PCR using the primers with 50 nucleotide homology extensions to the regions adjacent to the ycdC gene (Figure 2.2a, b). The insertion of the cassette into the genome was accomplished by a Red-mediated recombination in the flanking homologies. After selection of the kanamycin resistant clones, the kanamycin resistance gene was eliminated by using a helper plasmid expressing the FLP recombinase, which acts directly on the repeated FRT sites flanking the kanamycin resistance gene (Figure 2.2c). The integration of the kanamycin resistance cassette and its subsequent elimination was confirmed by agarose (1%) gel electrophoresis (Figure 2.2d).
Figure 2.2: The *ycdC* gene knockout from the *E. coli* W3110 genome; a) schematic of a section of the parent genome showing the *ycdC* gene b) schematic of the same section of the parent genome after the *ycdC* gene is replaced by the kanamycin resistance cassette c) schematic of the same section of the parent genome after the kanamycin cassette is removed d) agarose (1%) gel image to confirm the integration and then removal of the kanamycin resistance gene. Bands in each lane represent the DNA fragment generated by the PCR amplification after the integration and the removal of the kanamycin gene with the pair of primers listed above that lane. MM: Molecular marker; numbers in parenthesis over each lane: expected band size in number of base pairs; Kb: Marker band sizes in Kilobases.

2.3.2 Secretion Studies in the ΔycdC Strain

Genomics, transcriptomics, and proteomics studies have suggested that RutR, encoded by the *ycdC* gene, plays an important role in maintaining the balance of various substrates such as amino acids and tRNA synthetases of the *E. coli* translation machinery
The absence of a functional RutR may decrease the overall protein translation rate and increase extracellular secretion of HlyA via the TISS (Gupta et al. 2010, Lee and Lee 2004). This enhanced secretion may be due to a decrease in the intracellular levels of HlyA (Gupta et al. 2010). Here, the secretion of HlyA via the TISS is studied in the \( E. \ coli \Delta ycdC \) strain. The \( \Delta ycdC \) strain showed an increase in the extracellular levels of HlyA compared to the parent strain (Figure 2.3), consistent with the 3.4-fold increase from the \( ycdC::Tn5 \) mutant (Gupta et al. 2010) and the 4-fold increase from the B41 mutant, which has a point mutation in the \( ycdC \) gene (Gupta et al. 2010, Lee and Lee 2004). HlyA is the natural substrate of the TISS in \( E. \ coli \) (Blight and Holland 1994). To investigate if the hypersecretion phenotype of the \( \Delta ycdC \) strain can be extended to other recombinant proteins, secretion of \( \beta \)-lactamase (Bla) and GILL- \( \beta \)-lactamase (GILL-Bla) proteins fused to the CTSS is investigated. GILL-Bla is a modified version of Bla, which was custom synthesized by incorporating the codons \( \text{(GGT ATC CTG CTG)} \), encoding the amino acid sequence GILL, at three different positions along the parent \( bla \) gene sequence (Figure 2.4; Gupta and Lee 2010). A 3.03-fold increase in the extracellular levels of GILL-Bla is observed when it is secreted from the \( \Delta ycdC \) strain compared to the parent strain. However, similar extracellular levels of Bla are observed when it is secreted from either strain (Figure 2.3). In contrast, a 1.42-fold increase in the extracellular levels of Bla was reported when secreted from the \( ycdC::Tn5 \) mutant as compared to the parent strain (Gupta et al. 2010). These results suggest that hypersecretion phenotype observed due to absence of the RutR protein might be dependent upon the background strain as the \( ycdC::Tn5 \) mutant is MG1655 based
Figure 2.3: Effect of the \textit{ycdC} gene knockout on the extracellular secretion of HlyA, Bla and GILL-Bla proteins via the TISS in \textit{E. coli}. Number of ‘+’ indicates qualitative differences in the levels. Equivalent amount of sample loaded in all the lanes. (Gupta et al. 2010) and the \textit{ΔycdC} mutant is W3110 based. The enhanced secretion of HlyA is consistent across the strains whereas Bla secretion was only enhanced in MG1655 based strain, suggesting that phenomenon may be protein specific. Together, these results suggest that the \textit{ycdC} gene knockout differentially enhances recombinant protein secretion via the TISS depending upon the protein and background host strain. The absence of functional RutR may decrease the overall protein translation rates due to decreased level of tRNA-synthetases and some amino acid transporters (Gupta et al.)
HlyA, Bla, and GILL-Bla vary in length and amino acid composition. The differential effects of the ycdC gene knockout on secretion of these proteins may be due to varied affects on the translational rates, which have been previously shown to impact the secretion levels (Simmons and Yansura 1996).

![Figure 2.4: Design of a de novo gene sequence (gill-bla) based on bla gene sequence. A codon cluster (encoding amino acids GILL) is incorporated at three different positions along the gene sequence. The new protein (GILL-Bla) is detected with anti-Bla serum. Figure adapted from Gupta and Lee 2010.](image)

2.3.3 Combined Effect of the ycdC Gene Knockout and Synonymous Rare Codon Substitutions in the Parent Gene Sequence

Synonymous rare codon substitutions in the parent gene sequence can decrease the predicted translation rate and subsequently enhance the secretion of recombinant proteins via the TISS (Gupta and Lee 2010). Here, we have focused on Bla and GILL-Bla. Different mutants of the Bla (Bla-M1, Bla-M2 and Bla-M3; Table 2.5) and the GILL-Bla (GILL-M1, GILL-M2 and GILL-M3; Table 2.6) proteins were created by making synonymous rare codon changes in the parent gene sequence (Bla-W and GILL-W) of respective proteins (Gupta and Lee 2010). All the mutants code for the same amino acid sequence as the wild type.
Table 2.5: Sequence changes (marked in red) and predicted % decrease in translation rate of different Bla mutants (Gupta and Lee 2010).

<table>
<thead>
<tr>
<th>Mutants</th>
<th>Sequence</th>
<th>Predicted Slowdown (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pblaM1</td>
<td>195 CTC GGT CGC CGC ATA CAC</td>
<td>Initial 14.0%</td>
</tr>
<tr>
<td></td>
<td>CTA GGG AGG AGG ATA CAC</td>
<td>Mutant</td>
</tr>
<tr>
<td></td>
<td>65 L G R R I H</td>
<td>Protein</td>
</tr>
<tr>
<td>pblaM2</td>
<td>333 TTA CTT CTG ACA ACG ATC GGA</td>
<td>Initial 13.6%</td>
</tr>
<tr>
<td></td>
<td>CTA CTA CTA ACA ACG ATA GGG</td>
<td>Mutant</td>
</tr>
<tr>
<td></td>
<td>111 L L L T T I G</td>
<td>Protein</td>
</tr>
<tr>
<td>pblaM3</td>
<td>651 CGC GGT ATC ATT GCA GCA</td>
<td>Initial 21.0%</td>
</tr>
<tr>
<td></td>
<td>AGG GGG ATA ATA GCA GCA</td>
<td>Mutant</td>
</tr>
<tr>
<td></td>
<td>217 R G I I A A</td>
<td>Protein</td>
</tr>
</tbody>
</table>

To explore the combined effect of the 
ycdC gene knockout and the synonymous rare codon substitutions in the gene sequence on the secretion levels via the TISS, secretion of Bla and GILL-Bla mutants in the ΔycdC strain is investigated. Figure 2.5 and 2.6 show the extracellular and cytoplasmic levels of different mutants of Bla (Bla-M1, Bla-M2 and Bla-M3) and GILL-Bla (GILL-M1, GILL-M2 and GILL-M3) proteins in the ΔycdC strain when compared to the parent strain secreting the wild type version (Bla-W and GILL-W). To uncouple the effects of protein expression and secretion, the cytoplasmic levels were measured in secretion deficient (BD-) strains. An increase in the extracellular levels of Bla-M2 and all the three GILL-Bla mutants including the wild type (GILL-W) is observed in the ΔycdC strain when compared to the parent (W3110) strain (Figure 2.5, 2.6). In the secretion deficient cells, an increase in the cytoplasmic levels results in an increase in the extracellular levels of Bla and GILL-Bla mutants except
Table 2.6: Sequence changes (marked in red) and predicted % decrease in translation rate of different GILL-Bla mutants (Gupta and Lee 2010).

<table>
<thead>
<tr>
<th>Mutants</th>
<th>Sequence</th>
<th>Predicted Slowdown (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pgillM1</td>
<td>144 GGT   ATC  CTG  CTG  GGG  ATA  CTA  CTA  48 G   I   L   L</td>
<td>Initial Mutant Protein 5.3%</td>
</tr>
<tr>
<td>pgillM2</td>
<td>336 GGT   ATC  CTG  CTG  GGG  ATA  CTA  CTA  112 G   I   L   L</td>
<td>Initial Mutant Protein 19.0%</td>
</tr>
<tr>
<td>pgillM3</td>
<td>660 GGT   ATC  CTG  CTG  GGG  ATA  CTA  CTA  220 G   I   L   L</td>
<td>Initial Mutant Protein 10.6%</td>
</tr>
</tbody>
</table>

GILL-M2 (lanes 2-5, Figure 2.5, 2.6). These observations conflict with the previous findings that a decrease in the intracellular aggregates of HlyA, Bla, and GILL-Bla result in increased secretion levels via the TISS (Gupta and Lee 2008, Gupta and Lee 2010). In similar secretion studies of different Bla and GILL-Bla mutants in the parent (W3110) strain, it was observed that a decrease in the cytoplasmic levels, in the secretion deficient cells, results in an increase in the extracellular levels of Bla and GILL-Bla mutants (Gupta and Lee 2010).

An optimum protein translation rate is required to achieve high levels of secretion and the optimal translational rates are protein specific (Simmons and Yansura 1996). The results described here indicate that the optimal level of protein expression required for an efficient secretion of Bla and GILL-Bla may be different in the parent and the ΔycdC
Figure 2.5: Combined effect of the ycdC gene knockout and synonymous rare codon substitutions on the secretion of the Bla protein via the TISS in E. coli. Supernatant and cytoplasmic fractions from the E. coli W3110 parent and the ΔycdC strain. Bla-W: wild type Bla; Bla-M1, Bla-M2, Bla-M3: different mutants of Bla generated by synonymous codon changes in the Bla-W (Gupta and Lee 2010). BD-: secretion deficient cells. Equivalent amount of sample loaded in all the lanes.

<table>
<thead>
<tr>
<th></th>
<th>Supernatant</th>
<th>Whole cell pellet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W3110</td>
<td>ΔycdC</td>
</tr>
<tr>
<td>Bla-W</td>
<td>1X</td>
<td>0.93X</td>
</tr>
<tr>
<td>ΔycdC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bla-W</td>
<td>1X</td>
<td>0.96X</td>
</tr>
<tr>
<td>ΔycdC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bla-M1</td>
<td>(BD-)</td>
<td></td>
</tr>
<tr>
<td>(BD-)</td>
<td>1X</td>
<td>1.12X</td>
</tr>
<tr>
<td>ΔycdC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bla-M2</td>
<td>(BD-)</td>
<td></td>
</tr>
<tr>
<td>(BD-)</td>
<td>1X</td>
<td>1.47X</td>
</tr>
<tr>
<td>ΔycdC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bla-M3</td>
<td>(BD-)</td>
<td></td>
</tr>
<tr>
<td>(BD-)</td>
<td>1X</td>
<td>1.12X</td>
</tr>
</tbody>
</table>

The difference in the optimal translational levels for efficient secretion may be due to decreased level of tRNA-synthetases and some amino acid transporters in the ΔycdC strain (Gupta et al. 2010). To further investigate this effect, the bla and gill-bla gene sequences fused with the CTSS can be expressed under the control of an inducible promoter. Secretion studies via the TISS at different induction levels, and consequently different expression levels, in the parent and the ΔycdC strain may give further insights into the optimal expression levels for efficient secretion in both strains. These studies
may give further insight into role of the \textit{ycdC} gene in protein translation and secretion.

Figure 2.6: Combined effect of the \textit{ycdC} gene knockout and synonymous rare codon substitutions on the secretion of the GILL-Bla protein via the TISS in \textit{E. coli}. Supernatant and cytoplasmic fractions from the \textit{E. coli} W3110 parent and the \textit{ΔycdC} strain. GILL-W: wild type GILL-Bla; GILL-M1, GILL-M2, GILL-M3: different mutants of GILL-Bla generated by synonymous codon changes in the GILL-W (Gupta and Lee 2010). BD-: secretion deficient cells. Equivalent amount of sample loaded in all the lanes.

2.3.4 Effect of Synonymous Rare Codon Changes in the CTSS

We hypothesize that promoting the formation of the predicted amphiphilic helical structure in the CTSS by making synonymous rare codon changes in the Helix1 region (Figure 2.1) can enhance secretion via the TISS. To test this hypothesis, synonymous rare
codon substitution were made in the parent CTSS signal sequence in the Helix1 region and the Cterm1 region (Figure 2.1, Table 2.7) by changing all the abundant codons in these regions to the rare codons. Rare codons are defined as those codons whose corresponding tRNA concentration is less than 1% of the total tRNA concentration (as tabulated in Solomovici et al. 1997). The parent and the mutant sequences are shown in Table 2.7.

**Table 2.7: Sequence changes (marked in red) made in the CTSS.**

<table>
<thead>
<tr>
<th>Mutants</th>
<th>Sequence</th>
<th>Initial</th>
<th>Mutant</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>pWAM1097_HisHlyA_Helix1,</td>
<td>TTA ATT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pWAM1097_kan_Bla_Helix1</td>
<td>ATT AAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pWAM1097_HisHlyA_Cterm1,</td>
<td>GAA ATC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pWAM1097_kan_Bla_Cterm1</td>
<td>AGC AAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pWAM1097</td>
<td>ATC ATT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pWAM1097</td>
<td>CTA ATA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pWAM1097</td>
<td>AAT GAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pWAM1097</td>
<td>ATA AGC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pWAM1097</td>
<td>AAA ATA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pWAM1097</td>
<td>ATA ATA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial</td>
<td>L I N E</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mutant</td>
<td>I S K I I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The expectation is that decreasing the translation rate in the Helix1 region will promote the secondary structure formation and consequently increase the secretion levels via the TISS. An increase in the extracellular levels of HlyA is observed when it is fused to the CTSS carrying synonymous rare codon changes in the Helix1 region as compared to HlyA fused to the parent CTSS (Figure 2.7). To explore if this effect can be extended to other recombinant proteins, the secretion of the Bla protein was studied. A 2.07-fold increase in the extracellular levels of Bla is observed when it is fused to the CTSS.
carrying synonymous rare codon changes in the Helix1 region as compared to Bla fused to the parent CTSS (Figure 2.7). HlyA secretion is enhanced to a greater extent compared to Bla (Figure 2.7). An increase in the extracellular levels of both HlyA and Bla is observed when they are fused to the CTSS carrying the synonymous rare changes in the Cterm1 region, although to a lesser extent compared to the increase when the CTSS has mutations in the Helix1 region (Figure 2.7). This observation provides indirect evidence

**Supernatant**

<table>
<thead>
<tr>
<th></th>
<th>W3110</th>
<th>W3110</th>
<th>W3110</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HlyA</td>
<td>HlyA_Helix1</td>
<td>HlyA_Cterm1</td>
</tr>
<tr>
<td>W3110</td>
<td>+</td>
<td>++++</td>
<td>++</td>
</tr>
<tr>
<td>Bla-W</td>
<td>1X</td>
<td>2.07X</td>
<td>1.62X</td>
</tr>
</tbody>
</table>

**Figure 2.7:** Effect of synonymous rare codon changes in the CTSS on the secretion of HlyA and Bla proteins via the TISS in *E. coli*. Protein_Helix1: Protein sequence with synonymous changes in Helix1 region. Protein_Cterm1: Protein sequence with synonymous changes in the Cterm1 region. Number of ‘+’ indicates qualitative differences in the levels. Equivalent amount of sample loaded in all the lanes.
in support of the theory that secondary structure formation in the CTSS can play an important role in secretion of proteins via the TISS. To the best of our knowledge these results are the first report of synonymous codon changes in the CTSS that can enhance secretion levels via the TISS. These results help in identification of novel silent mutations in the CTSS and indicate that decreasing protein translation rate in the Helix1 region of the CTSS can result in an increase in enhanced recombinant protein secretion via the TISS. However, the increase in secretion levels may vary for different proteins. These results provide a single strategy of synonymous rare codon changes to increase recombinant protein secretion via the TISS. However, further research is required to fully understand the mechanism by which the increase occurs. In addition, the combined effect of the synonymous codon changes in both the CTSS and protein sequence can be investigated to explore the possibility of further enhancing extracellular secretion. Next, similar studies can be performed in the ΔycdC strain to further increase secretion via the TISS.

2.4 Conclusions

The absence of a functional RutR promotes the balance between translation and secretion by slowing the overall protein translation rate, consequently enhancing the secretion levels via the TISS (Gupta and Lee 2010). In this study, construction of a ΔycdC strain using the homologous recombination technique is described, which is shown to enhance secretion of HlyA and GILL-Bla via the TISS. However, the increase in secretion levels varied for different proteins. Next, the combined effect of the
synonymous rare codon changes and the \( ycdC \) gene knockout on the secretion levels via the TISS was investigated. An increase in the cytoplasmic levels, in the secretion deficient strains, led to an increase in secretion of Bla and GILL-Bla mutants except the GILL-M2 mutant. This trend is in contrast with the secretion studies in the parent strain, where a decrease in the cytoplasmic levels, in the secretion deficient strains, resulted in an increase in secretion of Bla and GILL-Bla mutants (Gupta and Lee 2010). We hypothesize that the optimal expression level to enhance the secretion levels via the TISS might be different for the parent and the \( \Delta ycdC \) strain. Further, it has been shown that making synonymous rare codon substitutions in a specific regions of the CTSS can considerably enhance the secretion levels of recombinant proteins. Overall, these results present new targets for engineering \( E. coli \) cells to enhance extracellular secretion of recombinant proteins via the TISS, but more research is required to fully understand, control, and optimize the process.
3.1 Introduction

Type-I secretion system (TISS) is the most frequently used secretion pathway for extracellular recombinant protein secretion compared to the other known secretion systems (Ni and Chen 2009). The \textit{E. coli} TISS is the prototype and the most extensively studied Type-I translocation pathway (Ni and Chen 2009). The secretion machinery is composed of three protein components, the ATP-binding cassette (ABC) protein HlyB, the membrane fusion protein HlyD, and the outer membrane pore protein TolC (Holland et al. 2005). Figure 3.1 illustrates a schematic showing the three components of the transport machinery and a HlyA molecule in transit. HlyA is the natural substrate of the \textit{E. coli} TISS and its secretion is dependent upon a 60 amino acid C-terminal signal sequence (CTSS) (Binet et al. 1997).

The mechanism of HlyA binding to the transport machinery and its subsequent secretion is still not fully understood. When HlyA is not present, HlyB and HlyD form a stable complex within the inner membrane of \textit{E. coli} (Thanabalu et al. 1998). However, the exact stoichiometry of HlyD with respect to HlyB is not known (Pimenta et al. 1999).
Deletion of 40 N-terminal residues of HlyD completely abolished HlyA transport, suggesting that HlyA specifically interacts with the N-terminal cytoplasmic domain of HlyD (Thanabalu et al. 1998). Deletion of HlyD N-terminal residues does not affect the binding to HlyB (Thanabalu et al. 1998). The N-terminal region of HlyD is also critical for the recruitment of TolC (Thanabalu et al. 1998). One hypothesis suggests that the binding of HlyA to N-terminal domain of HlyD causes a conformational change in the

**Figure 3.1:** The Type-I translocator in *E. coli*. Cartoon representation of TolC trimers in the outer membrane, surrounded by several HlyD monomers which also surround HlyB in the inner membrane, forming a continuous sealed channel from the cytoplasm to the exterior, with a molecule of HlyA in transit. The exact stoichiometry of HlyD with respect to other components remains unclear. Figure is reproduced with kind permission from Springer Science + Business Media: <Biotechnology Letters, Extracellular recombinant protein production from *Escherichia coli*, 31, 2009, 1661-1670, Ni Y, Chen R, figure number:1>
periplasmic domain of HlyD, which enables the binding with TolC and formation of a continuous channel across the periplasmic space (Thanabalu et al. 1998). In contrast, stability studies suggest that recruitment of TolC is not dependent upon HlyA-HlyD interaction and engagement of HlyD with TolC may be more stable (Pimenta et al. 1999). These studies are supported by co-purification studies of the AcrAB multidrug transport complex, which also suggest that a drug molecule is not required for recruitment of TolC (Tikhonova and Zgurskaya 2004, Touze et al. 2004, Zgurskaya and Nikaido 2000).

While, the interaction of TolC with the transport complex is highly dynamic and unstable, the interaction of HlyA may lead to a more stable complex formation (Holland et al. 2005). Surface plasma resonance studies have shown that the CTSS of HlyA specifically interacts with the nucleotide binding domain (NBD) of HlyB (Benabdelhak et al. 2003). HlyA is believed to interact with both the N-terminal domain of HlyD and the NBD of HlyB either sequentially or simultaneously and these interactions are the first step of the transport cycle (Zaitseva et al. 2005c).

These interactions activate two major events. Firstly, a conformational change occurs in the periplasmic domain of HlyD, which is either required for the engagement of TolC or stabilization of its interaction with the secretion machinery and for the opening of TolC portal in the periplasm that forms a continuous channel through the periplasm into the exterior (Holland et al. 2005). Formation of this continuous channel can explain the direct transport of HlyA into the extracellular space without a periplasmic intermediate (Holland et al. 2005). Secondly, a conformational change occurs in the
Walker A site in the NBD of HlyB, which facilitates the binding of ATP subsequently inducing the dimerization of NBDs and displacement of HlyA in the transport channel (Holland et al. 2005) constituted by HlyB and/or HlyD (Pimenta et al. 1999). The exact stoichiometry of HlyB and HlyD is unknown (Pimenta et al. 1999). The bound ATPs are sequentially hydrolyzed, which is reflected by sequential release of two inorganic phosphates, dissociating the dimer and resetting the transporter for another catalytic cycle (Zaitseva et al. 2006). Overall one catalytic cycle consumes two ATPs (Zaitseva et al. 2006). The sequential mechanism of ATP hydrolysis has also been described for the NBDs of histidine permease (Nikaido et al. 1999) and Mdl1p, an ABC transporter from yeast (Janas et al. 2003). Other models for ATP hydrolysis including alternative site model have also been described (van der Does and Tampé 2004). In the case of sequential hydrolysis of ATP, the hydrolysis of second ATP may be slower than the hydrolysis of first ATP (Janas et al. 2003). Although ATP hydrolysis is essential for secretion, the proton motive force may also be involved HlyA secretion (Koronakis et al. 1991). Using size exclusion chromatography and activity data, an active dimeric form of the HlyB NBD was detected along with an inactive dimeric form, suggesting that the ATP induced dimerization might be an essential, but not sufficient, condition for ATP hydrolysis (Zaitseva et al. 2005b). The exact mechanism of coupling ATP induced dimerization and subsequent hydrolysis to substrate transport is still unclear (Janas et al. 2003, Zaitseva et al. 2005c, Zaitseva et al. 2006). All the aspects including rotational movement of the NBD helical subdomain upon ATP binding leading to dimerization and sequential ATP hydrolysis followed by dimer dissociation provide energy and connect
HlyA is believed to be transported in an unfolded state, but how it remains in an unfolded conformation prior to transport is still unclear (Holland et al. 2005). The dimension of the translocation channel formed by TolC strongly suggests that the protein transported via the TISS may be in an unfolded conformation (Koronakis et al. 2000). Mutations in the CTSS can affect the correct folding of both HlyA and β-lactamase (Bla) fused with the CTSS, suggesting that the CTSS may be involved in post-translational folding of proteins targeted for secretion via the TISS (Jumpertz et al. 2010). Low levels of free calcium in bacteria (Jones et al. 1999) may slow folding of HlyA as it contains RTX-repeats that are rich in glycine and aspartic acid residues, which specifically binds calcium ions (Lilie et al. 2000). No conventional chaperones have yet been identified to be involved in secretion of proteins via the TISS (Holland et al. 2005) except in the case of HasA (Delepelaire and Wandersman 1998), a 20 kDa heme-binding metalloprotease in Serratia marcescens that requires antifolding activity of the SecB chaperone for an efficient secretion (Delepelaire and Wandersman 1998, Wolff et al. 2003). In the absence of SecB, HasA rapidly folds and the cytoplasmic pool of HasA interacts with the transporter and inhibits its own transport (Debarbieux and Wandersman 2001). SecB retards the folding of HasA and maintains it in a secretion competent form (Wolff et al. 2003). SecB does not interact with the transporter (Wolff et al. 2003). Single chain antibody fragments secreted via the TISS in E. coli suggest that folding and disulfide
bond formation occurs during the transport process (Fernandez and de Lorenzo 2001). The surface lipopolysaccharides may provide an environment that promotes protein folding in general and/or the CTSS promotes folding of the fusion proteins (Holland et al. 2005).

Protein secretion is tightly coupled to protein synthesis (Simmons and Yansura 1996). Periplasmic secretion of a recombinant protein in *E. coli* can be greatly enhanced by controlling the translation rate, i.e. the protein synthesis level (Simmons and Yansura 1996). Interestingly, optimization of translation rates was required to maximize the periplasmic secretion of heterologous proteins and the optimal translation rate needed to be determined empirically for individual protein (Simmons and Yansura 1996). Changing translation speed by making synonymous codon substitutions in the parent gene sequence can enhance recombinant protein secretion via the TISS (Gupta and Lee 2010). Gupta and Lee 2010 proposed a secretion model where changing the translation speed, i.e. protein synthesis rate, can change the flux between the secretion competent and secretion incompetent forms of the protein, consequently influencing its secretion.

Protein secretion via the TISS is a complex, non-linear, and a branched process as it is in kinetic competition with other parallel pathways including folding, aggregation, degradation, and inclusion body formation. Therefore, secretion of a protein of interest will depend on rate of secretion, rate of polypeptide folding and association, rate of aggregation, and rate of degradation and all these rates depend on rate of protein
synthesis (Bowden and Georgiou 1990, Hoffmann et al. 2000, Kiefhaber et al. 1991). The secretion efficiency via the TISS, which is defined as the percentage of expressed protein that is secreted into the medium, ranges from 1-90% for different proteins from different sources (Gentschev et al. 1996). The secretion levels of *Pseudomonas fluorescens* lipase (TliA) via a TISS can be enhanced by controlling the relative expression levels between the transporter and its passenger protein (Eom et al. 2006). The secretion of HasA via a TISS is also dependent on intracellular levels of HasA and the transport proteins (Debarbieuex and Wandersman 2001).

Here, we have developed kinetic models to describe the extracellular secretion of HasA via the TISS. Two kinetic models are illustrated that describe the secretion of HasA from a wild type *E. coli* strain and a ΔsecB strain, which lacks the SecB chaperone. The models include HasA synthesis, protein folding, aggregation, chaperone interactions, degradation, transporter synthesis, reporter-transporter interactions, catalytic cycle of the transporter, protein secretion into the extracellular medium, and cell growth. The models consist of coupled deterministic non-linear ordinary differential equations. This approach has been extensively used in modeling of complex biochemical reaction networks (Brown et al. 2004, Chen et al. 2000, Goldbeter 1995, Janes and Lauffenburger 2006, Klipp and Liebermeister 2006). To the best of our knowledge this is the first report of mathematical modeling of protein secretion via the TISS in *E. coli*. These models help in better understanding of the secretion process and designing of strategies to optimize extracellular secretion of recombinant proteins via the TISS.
3.2 Development of Models

Two kinetic models were developed to describe the extracellular secretion of HasA from the wild type *E. coli* strain (Model 1) and the *E. coli* Δ*secB* strain (Model 2). The schematic of both models showing all the reactions and species involved is illustrated in Figure 3.2 and 3.3, respectively. As shown in Figure 3.2, the HasA synthesis rate is assumed to follow zero order kinetics with a constant rate \((k_0)\) to produce nascent polypeptide \((P_N)\). The constant rate of nascent polypeptide production is the simplest kinetic form and has been previously adopted in kinetic modeling of competition between protein folding and aggregation (Kiefhaber et al. 1991). \(P_N\) can reversibly fold in the cytoplasm to form the folded form of HasA \((P_F)\), which can degrade in the cytoplasm. Alternatively, \(P_N\) can bind with SecB to form a complex \((\text{SecB}P_N)\). The complex \(\text{SecB}P_N\) can dissociate back to form \(P_N\), degrade in the cytoplasm, or release the folded HasA \((P_F)\). Folding of substrates when bound to SecB has been experimentally observed for HasA (Wolff et al. 2003) and barnase, a synthetic SecB substrate (Stenberg and Fersht 1997). Further, \(P_N\) can also degrade in the cytoplasm or aggregate to form \(P_N^A\). All the cytoplasmic forms of HasA such as \(P_N\), \(P_F\), \(P_N^A\), and \(\text{SecB}P_N\) can interact with the free transporter \((T_f)\). The transporter is also assumed to be synthesized at a constant rate \((k_t)\) and can also degrade in the cytoplasm. It has been proposed that the pool of cytoplasmic HasA can interact with its ABC transporter *in vivo* and inhibit the secretion of HasA (Debarbieux and Wandersman 2001). In the model, it is assumed that \(P_F\) and \(P_N^A\) can reversibly bind to \(T_f\) to form \(T_bP_F\) and \(T_bP_N^A\), respectively, both of which are inhibitory products and result in titration of the free transporter. It is assumed that \(P_N\) reversibly
binds to $T_f$ to form $T_bP_N$. The complex SecBP$_N$ interacts with $T_f$ to form $T_bP_N$. The interaction of SecBP$_N$ with $T_f$ is assumed to be irreversible because it has been proposed that SecB only has an antifolding function and does not interact with the transporter (Wolff et al. 2003). Hence, it is assumed that SecBP$_N$ delivers HasA to the transporter and if HasA dissociates from the complex then it forms $P_N$. $T_bP_N$ reversibly binds ATP to form $T_bP_N$ATP, which reversibly dimerizes to form $(T_bP_N$ATP$)^D$. The $(T_bP_N$ATP$)^D$ dimeric complex subsequently releases $P_N$ as a secreted form of HasA ($P_{SEC}$) to form $(T_b$ATP$)^D$. $(T_b$-ADP-ATP$)^D$ and $(T_b$-ADP-ADP$)^D$ are formed by sequential hydrolysis of ATP. The transporter is set free due to dissociation of the dimer following the ATP hydrolysis.

Model 1 describes the secretion of HasA from the wild type *E. coli* strain and includes the reactions and species described above (Figure 3.2). Figure 3.3 illustrates a schematic of Model 2, which describes the secretion of HasA from the *E. coli* ΔsecB strain and does not contain the reactions and species marked in yellow in the schematic illustrated in the Figure 3.2. The formation and consumption rates of all species were written based on mass action kinetics, i.e. the rate of a chemical reaction is directly proportional to the concentration of the reactants. The degradation reactions are assumed to follow typical first order kinetics (Lee and Bailey 1984). The aggregation reactions are assumed to follow second order (Kiefhaber et al. 1991). The model equations are described below:
**Figure 3.2:** Schematic of the kinetic models of secretion showing all the reactions and species involved. Model1 contains all the species and reaction. Model2 does not contain the species and reactions shown in yellow.
Figure 3.3: Schematic of the Model 2 of secretion showing all the reactions and species involved.
\[
\frac{d[P_r]}{dt} = k_0 + k_{ir}[\text{SecBP}_N] + k_{2r}[P_r] + k_{4r}[T_bP_r] - \{k_{2f} + k_d + k_{af}[P_N] + k_{1f}[\text{SecB}]
+ k_{af}[T_f] + \mu\}[P_N]
\]

(1)

\[
\frac{d[\text{SecBP}_N]}{dt} = k_{1f}[\text{SecB}][P_N] - \{k_{ir} + k_{3f} + k_d + k_{af}[T_f] + \mu\}[\text{SecBP}_N]
\]

(2)

\[
\frac{d[P_r]}{dt} = k_{2f}[P_N] + k_{4r}[T_bP_r] + k_{3f}[\text{SecBP}_N] - \{k_{2r} + k_{af}[T_f] + k_d + \mu\}[P_r]
\]

(3)

\[
\frac{d[P_N^A]}{dt} = k_{af}[P_N]^2 + k_{4r}[T_bP_N^A] - \{k_{af}[T_f] + \mu\}[P_N^A]
\]

(4)

\[
\frac{d[T_bP_N^A]}{dt} = k_{af}[T_f][P_N^A] - \{k_{4r} + \mu\}[T_bP_N^A]
\]

(5)

\[
\frac{d[T_bP_N]}{dt} = k_{af}[T_f][P_N] + k_{5r}[T_bP_NATP] - \{k_{4r} + k_{5f}[ATP] + \mu\}[T_bP_N]
\]

(6)

\[
\frac{d[T_bP_F]}{dt} = k_{af}[T_f][P_F] - \{k_{4r} + \mu\}[T_bP_F]
\]

(7)

\[
\frac{d[T_bP_NATP]}{dt} = k_{5f}[ATP][T_bP_N] + k_{6r}[(T_bP_NATP)^D] - \{k_{5r} + \mu\}[T_bP_NATP]
- k_{6f}[T_bP_NATP]^2
\]

(8)

\[
\frac{d[(T_bP_NATP)^D]}{dt} = k_{6f}[T_bP_NATP]^2 - \{k_{6r} + k_{7f} + \mu\}[(T_bP_NATP)^D]
\]

(9)

\[
\frac{d[(T_b - ATP)^D]}{dt} = k_{7f}[(T_bP_NATP)^D] - \{k_{8f} + \mu\}[(T_b - ATP)^D]
\]

(10)

\[
\frac{d[(T_b - ADP - ATP)^D]}{dt} = k_{8f}[(T_b - ATP)^D] - \{k_{9f} + \mu\}[(T_b - ADP - ATP)^D]
\]

(11)

\[
\frac{d[(T_b - ADP - ADP)^D]}{dt} = k_{9f}[(T_b - ADP - ATP)^D] - k_{10f}[(T_b - ADP - ADP)^D]
- \mu[(T_b - ADP - ADP)^D]
\]

(12)
\[
\frac{d[T_r]}{dt} = k_i + k_{10r}[(T_a - ADP - ADP)^D] + k_{4r} \{[T_9P_N] + [T_9P_F] + [T_9P_N^A]\} - k_{4f} \{[P_F] + [P_N] + [SecBP_N] + [P_N^A]\}[T_r] - \{k_d + \mu\}[T_r]
\]

(13)

\[r_s = k_{7f}[(T_bP_NATP)^D]\]

(14)

\[SE = \frac{r_s}{k_0} \times 100\%\]

(15)

where \(r_s\) is the rate of secretion, \(SE\) is the secretion efficiency, and \(\mu\) is the specific growth rate of the cells that accounts for dilution of intracellular components due to cell growth.

### 3.2.1 Cell Growth Models

Bacterial growth curves often shows three phases of growth: a lag phase (\(\lambda\)), where cells get acclimated in the environment and grow very slowly; exponential growth phase, where cells grow rapidly with a specific growth rate of \(\mu\); and finally the stationary phase, where cells reach a certain number often called the carrying capacity of the environment and growth rate reduces to zero and an asymptote (\(A\)) is reached (Zwietering et al. 1990). The growth curve defined as the natural logarithm of the number of cells plotted against time can be described by a sigmoidal curve illustrated in Figure 3.4 showing all the three phases. Several sigmoidal functions have been used previously to describe such a curve including models of Gompertz 1825, Richards 1959, Stannard et al. 1985, Schnute 1981, logistic and others (Ricker 1979).
Here, we have used the modified Gompertz equation proposed by Zwietering et al. to model cell growth because it is a simple equation with three biologically relevant parameters and has been used previously to model cell growth in bacteria (Zwietering et al. 1990). The equation is described below:

$$y = A \exp \left\{ - \exp \left[ \frac{\mu \times e}{A} (\lambda - t) + 1 \right] \right\}$$

(16)

where $y = \ln \left( \frac{N_t}{N_0} \right)$; $A = \ln \left( \frac{N_\infty}{N_0} \right)$; $N_t =$ number of cells at any given time $t$; $N_\infty =$ maximum number of cells; $N_0 =$ initial number of cells, i.e. number of cells at time $t = 0$. 

**Figure 3.4:** Schematic of a bacterial growth curve showing lag phase, exponential phase and stationary phase. The figure is adapted from Zwietering et al. 1990.
3.2.2 Specific Growth Rate Models

The production of recombinant proteins adds to the metabolic burden of the cells (Bentley and Kompala 1989). The metabolic intermediates such as nucleotides and amino acids that are typically used for chromosome replication and the production of cellular proteins are being divided in the recombinant cells between the chromosome and plasmid related functions, thereby reducing the growth rate of recombinant protein producing cells (Bentley and Kompala 1989). This difference in growth rate leads to uneven partitioning of plasmids to the daughter cells (Bentley et al. 1990). In the model simulations presented here, the assumption is made that antibiotic selection pressure maintains a homogeneous plasmid bearing population and thus the effects of plasmid segregation and the growth of non-plasmid bearing cells have not been taken into account. However, the effect of HasA and transporter protein production on specific growth rate of cells still has to be considered. Several models of growth rate of recombinant protein producing cells have been proposed.

A highly detailed and structured model for a single *E. coli* cell including mechanism of plasmid replication and expression of foreign protein describing the anabolism and catabolism of 28 cellular components with 35 differential equations has been proposed by Shuler and co-workers (Shuler et al. 1979, Ataai and Shuler 1987). The model has been successfully used (Domach et al. 1984) and extended to simulate single-cell bacterial growth mechanisms under various environmental conditions (Bentley and Kompala 1989). However, integrating such a detailed and complex growth model into an
already complex system such as secretion is cumbersome and also computationally expensive to solve.

Bentley and Kompala described a less detailed mechanistic model of recombinant cells that included eight major intracellular constituent pools: cellular proteins, foreign protein, chromosomal DNA, plasmid DNA, ribosomes, nucleotides, lipids, and amino acids (Bentley and Kompala 1989). A key feature of the model is that state of the microorganism is represented in terms of the fractional mass levels of the eight intracellular constituents (Bentley and Kompala 1989). Using this technique of fractional mass units of the components, the instantaneous growth rate of the cells producing recombinant proteins was calculated (Bentley and Kompala 1989). The model was successfully used to describe hybridoma growth dynamics (Batt and Kompala 1989). The model was also used to analyze optimal induction levels of recombinant protein production in bacterial cultures (Bentley and Kompala 1990). However, in the case of an open system such as protein secretion, the technique of fractional mass/mole units is difficult to apply. Also, this model is still fairly detailed to be integrated into another complex model.

Lee et al. 1985 suggested a model equation expressing growth rate of recombinant bacteria that is a function of plasmid content, foreign protein content, and the extracellular growth-limiting substrate concentration. The proposed equation includes separate effects of plasmid replication, product expression, and limiting substrate
concentration, but the biochemical or mechanistic basis for determination of constants in the model is limited (Bentley and Kompala 1989). Here, a modified version of the Lee et al. equation is used to describe the specific growth rate of recombinant cells with some assumptions. The equation is described below:

\[
\mu = \mu_0 \left(1 - \frac{k_0}{k_{0\text{max}}} \right) \left(1 - \frac{k_t}{k_{t\text{max}}} \right)
\]

where, \(\mu_0\) = specific growth rate of the cells not expressing any recombinant product; \(k_{0\text{max}}\) and \(k_{t\text{max}}\) are the maximum rates of production of the HasA and the transporter proteins at which the specific growth rate of the recombinant cells theoretically reduces to zero. The protein synthesis rates can be estimated by using pulse-chase experiments (Bowden and Georgiou 1990). It is assumed in the model that plasmid copy number does not change and has no effect on the specific growth rate. This assumption does not mean that effects of plasmid copy number are not important in the system under consideration, but the assumption is simply made to decrease the complexity of the model. This equation for specific growth rate in combination with the cell growth model (Equation 16) is a good starting point to model the growth of bacterial cells that are producing HasA and transporter protein at different rates in absence of any quantitative data.

### 3.2.3 Equation for the Extracellular Concentration of HasA

The extracellular concentration of HasA at any given time will depend on the rate of secretion from one cell \(r_s\), number of cells \(N_t\), and degradation rate of HasA in the extracellular medium. The following equation describes the rate of change of
extracellular concentration of HasA in a fixed external volume:

\[
\frac{d[P_{\text{ext}}]}{dt} = r_s \times N_t \times \rho \frac{1}{1 - N_t \times \rho} - k_{d\text{ext}}[P_{\text{ext}}]
\]

(18)

where, \(P_{\text{ext}}\) = extracellular HasA; \(\rho = \frac{\text{cell volume}\ (v_c)}{\text{external volume}\ (v_{\text{ext}})}\); and \(k_{d\text{ext}}\) = the first order degradation rate constant for HasA in the extracellular medium. The rate of secretion from one cell \(r_s\) is calculated using Equation 14 and the value of \(N_t\) is calculated using Equations 16 and 17.

3.2.4 Parameter Values

One of the challenges in mathematical modeling is the unavailability of large number of parameter values, which makes accurate quantitative predictions very difficult (Bailey 2001). Rate constants can be estimated from in vitro experiments, but they are still an approximation for the actual in vivo rate constants. The parameter values used in this study are listed in Table 3.1 along with the references and assumptions. All the parameter values are same for both models unless otherwise indicated. It will be shown in the later sections that although a large number of parameters values are unknown, meaningful estimates and predictions can be made using the kinetic models described.

3.2.5 Simulation Software

The models described above consist of a system non-linear coupled ODE. Model 1 consists of a system of 14 equations and Model 2 consists of a system of 13 equations. Models were simulated in MATLAB 7.6 (R2008a) (Mathworks Inc., MA).
Table 3.1: Parameter values used in the model simulations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>References/Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_0$</td>
<td>-</td>
<td>Input parameter (varies across simulations)</td>
</tr>
<tr>
<td>$k_i$</td>
<td>-</td>
<td>Input parameter (varies across simulations)</td>
</tr>
<tr>
<td>$k_{1f}$</td>
<td>15.0 $\mu$M$^{-1}$min$^{-1}$</td>
<td>Assumed such that pseudo first order rate constant for chaperone binding ($k_{1f}$[SecB]) is 10-fold higher than the assumed rate constant for folding ($k_{2f}$) of HasA in presence of SecB (Model1). Consistent with findings of Krishnan et al. 2009.</td>
</tr>
<tr>
<td>$k_{1r}$</td>
<td>1.5 min$^{-1}$</td>
<td>Assumed that dissociation rate constant is 10-fold lower than the association rate constant.</td>
</tr>
<tr>
<td>$k_{2f}$</td>
<td>0.6 min$^{-1}$ (Model 1) 600.0 min$^{-1}$ (Model 2)</td>
<td>Folding ($k_{2f}$) and unfolding ($k_{2r}$) rate constant in presence of SecB (Model 1) is assumed to be 1,000-fold lower than the rate constants in absence of SecB (Model 2). Consistent with the findings of Wolff et al. 2003. The equilibrium constant is same for both models. The value of $k_{2f}$ (Model 1) assumed from Kiefhaber et al. 1991. The value is estimated for a different protein.</td>
</tr>
<tr>
<td>$k_{2r}$</td>
<td>0.006 min$^{-1}$ (Model 1) 6.0 min$^{-1}$ (Model 2)</td>
<td></td>
</tr>
<tr>
<td>$k_{3f}$</td>
<td>0.1 min$^{-1}$</td>
<td>Assumed such that the rate is slower than $k_{1f}$[SecB] and $k_{2f}$ (Model 1). Consistent with findings of Krishnan et al. 2009.</td>
</tr>
<tr>
<td>$k_d$</td>
<td>0.07 min$^{-1}$</td>
<td>Lee and Bailey 1984. The value is estimated for a different protein.</td>
</tr>
<tr>
<td>$k_{af}$</td>
<td>6.0 $\mu$M$^{-1}$min$^{-1}$</td>
<td>Kiefhaber et al. 1991 The value is estimated for a different protein.</td>
</tr>
<tr>
<td>$k_{4f}$</td>
<td>0.51 $\mu$M$^{-1}$min$^{-1}$</td>
<td>Benabelhak et al. 2003</td>
</tr>
<tr>
<td>$k_{4r}$</td>
<td>2.58 min$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$k_{5f}$</td>
<td>30.0 $\mu$M$^{-1}$min$^{-1}$</td>
<td>van der Does et al. 2006. The value is estimated for a different ABC transporter protein (Mdl1).</td>
</tr>
<tr>
<td>$k_{5r}$</td>
<td>7.8 min$^{-1}$</td>
<td>Hofacker et al. 2007. The value is estimated for a different ABC transporter protein (Mdl1).</td>
</tr>
<tr>
<td>$k_{6f}$</td>
<td>2.31 $\mu$M$^{-1}$min$^{-1}$</td>
<td>van der Does et al. 2006. Values are estimated for a different ABC transporter protein (Mdl1).</td>
</tr>
<tr>
<td>$k_{6r}$</td>
<td>0.16 min$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>$k_{7f}$</td>
<td>60.0 min$^{-1}$</td>
<td>Value assumed.</td>
</tr>
<tr>
<td>$k_{8f}$</td>
<td>156.0 min$^{-1}$</td>
<td>Hofacker et al. 2007. The value is estimated for a different ABC transporter protein (Mdl1).</td>
</tr>
<tr>
<td>$k_{9f}$</td>
<td>15.6 min$^{-1}$</td>
<td>Hydrolysis of second ATP is assumed to be 10-fold slower than the hydrolysis of first ATP ($k_{8f}$). Consistent with findings of Janas et al. 2003.</td>
</tr>
<tr>
<td>$k_{10f}$</td>
<td>60.0 min$^{-1}$</td>
<td>van der Does et al. 2006. Values are estimated for a different ABC transporter protein (Mdl1).</td>
</tr>
</tbody>
</table>
3.3 Results and Discussion

3.3.1 Effect of the SecB Chaperone on the Secretion of HasA

The SecB chaperone is required for secretion of HasA via the TISS (Delepelaire and Wandersman 1998). Here, the effect of SecB on HasA secretion is studied using theoretical simulations of the kinetic models described. Two different models are developed to describe the secretion of HasA from the wild type *E. coli* strain (Model 1) and the *E. coli ΔsecB* strain (Model 2). All the parameter values are same for both models except the folding ($k_{2f}$) and the unfolding rate constant ($k_{2r}$) for HasA (Table 3.1). In the model simulations, it is assumed that the folding and unfolding rate constant for HasA in the wild type strain, in presence of SecB (Model 1), is 1,000-fold lower compared to the respective rate constants in the ΔsecB strain (Model 2). This assumption is consistent
with *in vitro* studies, which show that SecB considerably slows the folding of HasA by \(~3,200\)-fold (Wolff et al. 2003). The simulation results are illustrated in Figure 3.5. The simulation time is fixed to 3,000 min and the steady state values are plotted. Two different simulations are illustrated for Model 1. Model 1 (i) describes secretion of HasA from the wild type strain assuming that intracellular concentration of free SecB is 0.4 \(\mu\text{M}\), which is the estimated concentration of free SecB in the wild type *E. coli* cells (Randall and Hardy 1995). Model 1 (ii) describes secretion of HasA from wild type cells with a higher concentration of free intracellular SecB of 1.2 \(\mu\text{M}\). For a given HasA synthesis rate \((k_0)\), the secretion efficiency (SE), which is defined as the percentage of expressed protein that is secreted into the extracellular medium, is considerably reduced when HasA is secreted from the \(\Delta\text{secB}\) strain (Model 2) compared to the wild type strain (Model 1 (i)) (Figure 3.5). An increase in the intracellular concentration of free SecB from 0.4 \(\mu\text{M}\) (Model 1 (i)) to 1.2 \(\mu\text{M}\) (Model 1 (ii)) has a relatively small effect on the SE of HasA (Figure 3.5). These results are consistent with previous work in which the secretion levels of HasA from a secB::Tn5 mutant was < 10 \% of the wild type strain (Delepelaire and Wandersman 1998). SecB expression from a plasmid in the secB::Tn5 mutant strain restored the wild type HasA secretion levels. The increase in intracellular levels of SecB due to plasmid based expression did not alter the HasA secretion levels in the wild type strain (Delepelaire and Wandersman 1998). The simulation results also show that qualitative trend in the data is independent of HasA synthesis rate \((k_0)\), which varies from 0.05-0.5 \(\mu\text{M}/\text{min}\). This result suggests that in absence of SecB, an increase in intracellular levels of HasA does not have a considerable effect on the SE.
Figure 3.5: Effect of SecB on the SE of HasA via the TISS for different HasA synthesis rate \((k_0)\). Model 1 (i): Secretion from the wild type strain ([SecB] = 0.4 µM); Model 1 (ii): Secretion from the wild type strain ([SecB] = 1.2 µM); Model 2: Secretion from the ΔsecB strain ([SecB] = 0.0 µM). The value of \(k_t\) was set to 0.5 µM/min for all simulations. All the other parameter values were kept same for all simulations (Table 3.1). Simulation time is set to 3,000 min and steady state values are plotted.

3.3.2 Effect of the Transporter Protein Synthesis Rate on the Secretion of HasA

Effect of the transporter protein synthesis rate \((k_t)\) on the SE of HasA in the wild type strain (Model 1) is studied. The simulation results are shown in Figure 3.6. The simulation time is fixed to 3,000 min and the steady state values are plotted. In absence of the transporter protein \((k_t = 0 \ \mu M/min)\), no HasA is secreted into the extracellular
Figure 3.6: Effect of transporter protein synthesis rate ($k_t$) on the SE of HasA at varying HasA protein synthesis rate ($k_0$) in the wild type strain (Model 1). All the other parameter values are kept same for all simulations (Table 3.1). Simulation time is set to 3,000 min and steady state values are plotted.

medium (SE = 0) (Figure 3.6). An increase in the transporter synthesis rate from 0.01 to 0.50 µM/min, results in an increase in the SE and this qualitative trend is independent of the HasA synthesis rate ($k_0$) (Figure 3.6). These results are consistent with previous work in which wild type *E. coli* cells constitutively expressing HasA were grown in the medium containing various amounts of arabinose ranging from 0-0.02 % to induce the expression of transport proteins (Debarbieux and Wandersman 2001). HasA was not
secreted in absence of the transport proteins and an addition of arabinose increased the extracellular levels of HasA (Debarbieux and Wandersman 2001). These results demonstrate that the SE of HasA can be enhanced by increasing the intracellular levels of the transport proteins.

The SE of HasA from a ΔsecB strain is considerably lower compared to the wild type strain (Figure 3.5). Effect of increase in the transport protein synthesis rates on the SE from the ΔsecB strain (Model 2) was investigated. In absence of the transporter protein ($k_t = 0 \, \mu M/min$), HasA is not secreted into the extracellular medium (SE = 0) (Figure 3.7). An increase in the transporter synthesis rate from 0.01 to 0.50 $\mu M/min$ results in an increase in the SE, but the increase is considerably less compared the wild type strain (Figure 3.6). These results indirectly support the findings of Delepelaire and Wandersman 1998 that the SecB chaperone is essential for efficient secretion of HasA. Further, these results indicate that the SE from the ΔsecB strain cannot be improved by increasing the levels of transporter proteins.

### 3.3.3 Effect of the HasA Protein Synthesis Rate on its Secretion

The effect of HasA synthesis rate ($k_0$) on its SE in the wild type strain (Model 1) is investigated. In absence of the transporter protein ($k_t = 0 \, \mu M/min$), HasA is not secreted into the extracellular medium and an increase in the HasA synthesis rate has no effect (Figure 3.8). In presence of the transporter protein ($k_t = 0.5 \, \mu M/min$), the SE initially increases and then decreases with a further increase in the HasA synthesis rate.
There are two possible explanations for such an observation. First, the initial HasA synthesis rate is low and the transporter is able to secrete large percentage of the expressed protein. As the HasA synthesis rate increases, the transporter becomes saturated with the protein and the SE decreases. Secondly, an increase in the HasA synthesis rate increases the formation of aggregated product ($P^A_N$), which binds to the transporter making it unavailable to secrete HasA. To further explore the effect of HasA synthesis rate on its secretion, the rate of secretion ($r_s$) of HasA from one cell is studied.

Figure 3.7: Effect of transporter protein synthesis rate ($k_t$) on the SE of HasA at varying HasA protein synthesis rate ($k_0$) in the ΔsecB strain (Model 2). All the other parameter values are kept same for all simulations (Table 3.1). Simulation time is set to 3,000 min and steady state values are plotted.
**Figure 3.8:** Effect of HasA synthesis rate ($k_0$) on its SE at varying transporter protein synthesis rate ($k_t$) in the wild type strain (Model 1). All the other parameter values are kept same for all simulations (Table 3.1). Simulation time is set to 3,000 min and steady state values are plotted.

The rate of secretion ($r_s$) in absence of the transporter protein ($k_t = 0 \ \mu\text{M/min}$) is zero (Figure 3.9). In presence of the transporter protein ($k_t = 0.5 \ \mu\text{M/min}$), the $r_s$ initially increases and then decreases with a further increase in the HasA synthesis rate ($k_0$) (Figure 3.9). The decrease in the $r_s$ supports the explanation that an increase in the intracellular levels of HasA leads to the accumulation aggregated product ($P_N^A$), which binds to the transporter making it inactive for secretion. These results are consistent with
Figure 3.9: Effect of HasA synthesis rate ($k_0$) on the rate of secretion ($r_s$) at varying transporter protein synthesis rate ($k_t$) in the wild type strain (Model 1). All the other parameter values are kept same for all simulations (Table 3.1). Simulation time is set to 3,000 min and steady state values are plotted.

previous experiments in which wild type *E. coli* cells expressing HasA under the control of an isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible promoter were grown in the medium containing various amounts of IPTG ranging from 50-500 μM and arabinose, to induce the expression of transport proteins (Debarbieux and Wandersman 2001). In the absence of arabinose, addition of IPTG led to the cytoplasmic accumulation of HasA (Debarbieux and Wandersman 2001). In the presence of arabinose (0.02 %) and low level
induction of \textit{hasA} (50 \muM), a large percentage of expressed HasA was secreted into the extracellular medium. However, an increase in induction of \textit{hasA} (100-500 \muM) resulted in a decrease in the extracellular secretion of HasA and led to its cytoplasmic accumulation (Debarbieux and Wandersman 2001). Since, the western analysis was normalized, number of cells in all the samples can be assumed to be the same. Therefore, a decrease in the extracellular levels of HasA suggests a decrease in the rate of secretion from an individual cell.

The extracellular levels of HasA at any given time (t) is a function of three components: rate of secretion from an individual cell (r_s), number of cells at time t (N_t), and the degradation rate of HasA in the extracellular medium (k_{dex}). The value of r_s is indirectly dependent upon the rate of synthesis of HasA and transport protein (k_0 and k_t). The value of N_t is also indirectly dependent upon the values of k_0 and k_t as they affect the specific growth rate of the cells (\mu). Figure 3.10 illustrates the change in extracellular concentration of HasA at a fixed simulation time of 3,000 min. The extracellular concentration is zero in absence of the transport protein (k_t = 0 \muM/min) (Figure 3.10). In presence of the transport protein (k_t = 0.5 \muM/min), the extracellular concentration follows a trend similar to the rate of secretion (r_s) (Figure 3.9 and 3.10). At low HasA synthesis rate (k_0), the extracellular concentration increases and reaches a maximum after which a further increase in the value of k_0 results in a decrease in the extracellular concentration (Figure 3.10).
Figure 3.10: Effect of HasA synthesis rate ($k_0$) on its extracellular concentration at varying transporter protein synthesis rate ($k_t$) in the wild type strain (Model 1). All the other parameter values are kept same for all simulations (Table 3.1). Simulation time is set to 3,000 min and steady state values are plotted.

3.3.4 Parameter Variation and Sensitivity Analysis

One of the major challenges in mathematical modeling of biological systems is the unavailability of large number of parameter values (Balsa-Canto et al. 2010). Despite limited a priori knowledge of parameter values, variation of parameters in a feasible parameter space followed by sensitivity analysis can help in better understanding of the key regulatory domains and parameters within the model framework (Balsa-Canto et al.)
The kinetic model that describes the secretion of HasA from the wild type strain has 20 rate constants (Figure 3.2). Here, all 20 kinetic rate constants were varied in a parameter space as listed in Table 3.2. Simulation time was fixed at 3,000 min. Since, the quantitative information on the system is limited; the range is kept as large as possible. Parameter space was selected based on one-parameter and two-parameter variation studies. From the defined parameter space, 10,000 input parameter vectors were sampled. The parameter vector set was reduced to 5,196 by removing all the vectors that resulted in a SE of less than 1%. The results of the SE and the extracellular concentration of HasA for all 5,196 simulations are illustrated in Figure 3.11 and 3.12, respectively. The SE for different simulations ranges from 1-100% (Figure 3.11). Experiments have shown that SE via the TISS from E. coli varies from 1-90% for different proteins from different sources (Gentschev et al. 1996). The 5,196 model simulations can be envisioned as different models representing secretion of different proteins and each model has a unique parameter vector set. For a given SE or a range of SEs, multiple input vectors give the same output (Figure 3.11). This approach of using multiple models to explain the same output behavior of a system even when parameter values amongst different models vary by several order of magnitudes, has been applied before to study nerve growth factor signaling (Brown et al. 2004). The approach of using an ensemble of models is very useful for extracting predictive quantitative information from models that have poorly defined parameters (Brown and Sethna 2003). In this study, the constraint of SE > 1% is applied to the data set. This constraint is very minimal and therefore results in a large set of 5,196 data points. The objective of this analysis is to apply the approach of using
Table 3.2: Range within which the model kinetic rate constants were varied.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Range Varied</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_0$</td>
<td>$10^{-3} - 5 \times 10^{-2}$ µM/min</td>
</tr>
<tr>
<td>$k_f$</td>
<td>$10^{-2} - 10^{2}$ µM/min</td>
</tr>
<tr>
<td>$k_{1f}$</td>
<td>$10^{-3} - 10^{4}$ µM$^{-1}$ min$^{-1}$</td>
</tr>
<tr>
<td>$k_{1r}$</td>
<td>$10^{-3} - 10^{2}$ min$^{-1}$</td>
</tr>
<tr>
<td>$k_2f$</td>
<td>$10^{-3} - 10^{3}$ min$^{-1}$</td>
</tr>
<tr>
<td>$k_{2r}$</td>
<td>$10^{-3} - 10^{4}$ min$^{-1}$</td>
</tr>
<tr>
<td>$k_3f$</td>
<td>$10^{-3} - 10^{3}$ min$^{-1}$</td>
</tr>
<tr>
<td>$k_d$</td>
<td>$10^{-3} - 10^{3}$ min$^{-1}$</td>
</tr>
<tr>
<td>$k_{df}$</td>
<td>$10^{-3} - 10^{3}$ µM$^{-1}$ min$^{-1}$</td>
</tr>
<tr>
<td>$k_{4f}$</td>
<td>$10^{-3} - 10^{3}$ µM$^{-1}$ min$^{-1}$</td>
</tr>
<tr>
<td>$k_{4r}$</td>
<td>$10^{-3} - 10^{3}$ min$^{-1}$</td>
</tr>
<tr>
<td>$k_{5f}$</td>
<td>$10^{-3} - 10^{4}$ µM$^{-1}$ min$^{-1}$</td>
</tr>
<tr>
<td>$k_{5r}$</td>
<td>$10^{-3} - 10^{5}$ min$^{-1}$</td>
</tr>
<tr>
<td>$k_{6f}$</td>
<td>$10^{-3} - 10^{3}$ µM$^{-1}$ min$^{-1}$</td>
</tr>
<tr>
<td>$k_{6r}$</td>
<td>$10^{-3} - 10^{4}$ min$^{-1}$</td>
</tr>
<tr>
<td>$k_{7f}$</td>
<td>$10^{-3} - 10^{5}$ min$^{-1}$</td>
</tr>
<tr>
<td>$k_{8f}$</td>
<td>$10^{-3} - 10^{3}$ min$^{-1}$</td>
</tr>
<tr>
<td>$k_{df}$</td>
<td>$10^{-3} - 10^{3}$ µM$^{-1}$ min$^{-1}$</td>
</tr>
<tr>
<td>$k_{df}$</td>
<td>$10^{-3} - 10^{3}$ µM$^{-1}$ min$^{-1}$</td>
</tr>
</tbody>
</table>

ensemble of models to study protein secretion via the TISS in absence of any quantitative information. While the constraint applied here is very general, more defined constraints can be applied to study a very specific system. The extracellular concentration of HasA varies from 0-600 µM (Figure 3.12). It is interesting that this analysis gives a unique input parameter vector set that result in a very high extracellular concentration of HasA (Figure 3.12). Although, the value itself is not important, it shows the utility of such an analysis in identifying a set of parameters that result in a very high value of the output variable. However, there can be additional constraints that might define the feasibility of
such a set of parameters including constraints of thermodynamics and cellular capabilities of the host cell.

**Figure 3.11:** SE results of 5,196 simulation runs using different input parameter vectors. The simulation time is fixed at 3,000 min.

Different parameters will have varied impact on the model output variables such as SE and extracellular concentration (Balsa-Canto 2010). Parameters can be ranked in order of their relative influences on model predictions (Balsa-Canto 2010). One way to
quantify such influences is to use local parametric sensitivities (Balsa-Canto 2010). The problem with this approach is that the local sensitivities can be dependent upon the intrinsic value of the parameters which are usually unknown, making accurate ranking of parameters difficult (Balsa-Canto 2010). Alternatively, ranking for a sufficiently large number of parameter vectors in the feasible parameter space can be calculated (Balsa-Canto 2010). Parameter variation studies described above resulted in a set of 5,196 parameter vector sets.

Figure 3.12: Extracellular concentration results of 5,196 simulation runs using different input parameter vectors. The simulation time is fixed at 3,000 min.
Next, normalized sensitivity index (NSI) was calculated using the formula (Balsa-Canto 2010) described below:

\[
(\text{NSI})_j^k = \frac{1}{N} \left( \sum_{i=1}^{i=N} \left( \frac{P_{ij}}{R_k} \left( \frac{dR_k}{dP_{ij}} \right) \right)^2 \right)
\]

(19)

where \(j\) varies from 1-20 for 20 parameters; \(k\) is either 1 or 2 for two responses; \(N\) varies from 1-5,196; \(P\): parameter; \(R\): Response (SE or extracellular concentration).

The NSI value is an indication of the relative influence of a particular parameter on a model response output. The HasA synthesis rate \((k_0)\) and the transporter synthesis rate \((k_t)\) have the highest NSI values and therefore they are the most important parameters that influence both the SE and extracellular concentration (Figure 3.13). As expected the extracellular degradation rate \((k_{dext})\) is also an important parameter that influences extracellular concentration of HasA (Figure 3.13). While \(k_0\) and \(k_t\) almost equally influence the SE, \(k_t\) has a larger influence on the extracellular concentration of HasA (Figure 3.13). This result indicates that at any given time, the rates of HasA and transporter synthesis are the most important parameters that effect the SE and extracellular concentration of HasA.

3.3.5 Optimization of the Extracellular Concentration of HasA

The sensitivity analysis shows that extracellular concentration of HasA at any given time depends on the values of the HasA synthesis rate \((k_0)\) and the transporter synthesis rate \((k_t)\). Hence, at any given time there will an optimal extracellular
Figure 3.13: Normalized sensitivity index (NSI) describing the influence of all 20 kinetic rate constants on the SE and the extracellular concentration of HasA.

concentration that can be achieved and this optimal concentration will vary depending on the culture time. The extracellular concentration of HasA was optimized by varying $k_0$ and $k_1$ within the range described in Table 3.2. All other parameter values were same as described in Table 3.1. Figure 3.14 shows the change in the optimal extracellular concentration with culture time. The optimal extracellular concentration increases slowly initially and then almost linearly with culture time and finally plateaus at extended culture times (Figure 3.14). When maximization of the overall extracellular yield is the
target goal, the extracellular concentration can be enhanced by increasing the culture time. However, in some cases the reactor time can be economically important (Bentley and Kompala 1990) and increasing the culture time may not be the best possible strategy. Figure 3.15 illustrates the plot of optimal extracellular concentration divided by the culture time (ratio) with increasing culture time. The ratio increases with culture time, reaches a maximum and then decreases. This result suggests that in the cases when the reactor time can play an important role in process economics, one should operate under
conditions where the best overall yield is achieved at optimal reactor time. The exact quantitative data for the extracellular production of HasA is not available and exact operating conditions have not been established. Instead, this analysis demonstrates the utility of a secretion model and computational analysis that can shift through the factors influencing target product secretion and ultimately define an optimal strategy.

![Figure 3.15](image)

**Figure 3.15:** Variation of the optimal extracellular concentration divided by culture time (ratio) with culture time. The values of $k_0$ and $k_t$ were varied in the range described in the Table 3.2. All the other parameter values are described in Table 3.1.
3.3.6 Nondimensionalization of the Model

The simulation results described here are consistent with qualitative trends in the experimental data (see section 3.3.1, 3.3.2 and 3.3.3), but the accuracy of the quantitative predictive information from the model is limited due to unavailability of parameter values. Therefore, the model was nondimensionalized using the scaling factors described in Table 3.3 so that concentration units can be removed from the system and model analysis becomes more comparative and relative. The scaling factors were taken from the analysis described in the previous section where the values of $k_0$, $k_t$, and culture time were varied. Scaling factor for $P_N$ and $T_f$ is assumed to be the maximum amount of $P_N$ that can be achieved in the simulations. The scaling factors for all other species were assumed to be the maximum value of that species in all simulation runs described in the last section.

Table 3.3: Scaling factors for all the species in the model.

<table>
<thead>
<tr>
<th>Species/Variables</th>
<th>Scaling Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_N$</td>
<td>$1.5 \times 10^6$ µM</td>
</tr>
<tr>
<td>SecBP $N$</td>
<td>32.1 µM</td>
</tr>
<tr>
<td>$P_F$</td>
<td>92.4 µM</td>
</tr>
<tr>
<td>$P_N^A$</td>
<td>$28.2 \times 10^4$ µM</td>
</tr>
<tr>
<td>$T_bP_N^A$</td>
<td>$7.2 \times 10^2$ µM</td>
</tr>
<tr>
<td>$T_bP_N$</td>
<td>$4.4 \times 10^{-3}$ µM</td>
</tr>
<tr>
<td>$T_bP_F$</td>
<td>72.5 µM</td>
</tr>
<tr>
<td>$T_bP_NATP$</td>
<td>9.1 µM</td>
</tr>
<tr>
<td>$(T_bATP)^D$</td>
<td>3.2 µM</td>
</tr>
<tr>
<td>$(T_bP_NATP)^b$</td>
<td>1.2 µM</td>
</tr>
<tr>
<td>$(T_b-ADP-ATP)^p$</td>
<td>12.1 µM</td>
</tr>
<tr>
<td>$(T_b-ADP-ADP)^p$</td>
<td>3.2 µM</td>
</tr>
<tr>
<td>$T_f$</td>
<td>$1.5 \times 10^6$ µM</td>
</tr>
<tr>
<td>$P_{ext}$</td>
<td>$1.8 \times 10^2$ µM</td>
</tr>
<tr>
<td>time (t)</td>
<td>3,000.0 min</td>
</tr>
<tr>
<td>$k_0$ and $k_t$</td>
<td>500.0 µM/min</td>
</tr>
</tbody>
</table>
The simulation results of the scaled model are illustrated in Figure 3.16. The scaled simulation time was set at 1.0. For a given value of $k_t$, the extracellular concentration of HasA reaches a maximum and then decreases as the value of $k_0$ increases further (Figure 3.16). The maximum extracellular concentration increases with an increase in the value of $k_t$ (Figure 3.16), indicating that an increase in the levels

![Graph showing the combined effect of HasA synthesis rate and transporter synthesis rate on extracellular concentration of HasA.](image)

**Figure 3.16:** Combined effect of HasA synthesis rate (scaled: $k_0$) and transporter synthesis rate (scaled: $k_t$) on the extracellular concentration of HasA (scaled) when it is secreted from the wild type strain (Model 1). The values of $k_0$ and $k_t$ were scaled as described in the Table 3.3. All the other parameter values are described in Table 3.1. Scaled simulation time is set to 1.0 and steady state values are plotted.
of transporter enhances HasA secretion. An increase in $k_t$ also results in an increase in $k_0$ required for the maximum extracellular concentration of HasA (Figure 3.16). An increase in $k_t$ and/or $k_0$ does not result in a proportional increase in the extracellular concentration of HasA (Figure 3.16).

Next, the scaled model was used to optimize the extracellular concentration of HasA. As expected, the scaled optimal extracellular concentration of HasA (Figure 3.17)

![Figure 3.17: Variation of the optimal extracellular concentration (scaled) with culture time (scaled). The values of $k_0$ and $k_t$ were varied in the range described in the Table 3.2 and scaled as described in the Table 3.3. All the other parameter values are described in Table 3.1.](image)
follows the exact same trend as observed for non-scaled extracellular concentration (see Figure 3.14). At any given time, maximization of the extracellular concentration requires an optimal combination of the HasA and transporter protein levels, i.e. an optimal value of $k_0$ and $k_t$. Therefore, the optimal value of extracellular concentration at any given time corresponds to an optimal rate of HasA ($k_0$) and transporter ($k_t$) synthesis. Variation of these optimal rates with culture time is illustrated in Figure 3.18. The scaled simulation time was set at 1.0. At any given culture time, the optimal value of $k_t$ is less than the

![Graph](image)

**Figure 3.18:** Variation of the optimal $k_0$ (scaled) and the optimal $k_t$ (scaled) corresponding to the optimal extracellular concentration (scaled) with culture time (scaled). Parameters were scaled as described in Table 3.3.
optimal value of $k_0$ (Figure 3.18). After the hydrolysis of ATP, the dimmers dissociate and the transporter can engage another substrate molecule for transport. Therefore, low levels of the transporter protein can also allow for large percentage of expressed HasA to be secreted into the extracellular medium. The relative increase in $k_t$ with culture time is smaller than the increase in $k_0$, indicating that a relatively small increase in the transporter levels can lead to a larger increase in the secretion of intracellular HasA. In combination, these results suggest that optimization of the extracellular concentration requires a balance of HasA synthesis rate ($k_0$), transporter synthesis rate ($k_t$), and culture time. Further, the optimal level of HasA synthesis rate may be higher than the optimal level of transporter synthesis rate for an enhanced secretion of HasA.

3.4 Conclusions

Protein secretion via the TISS is a complex, non-linear, and branched process that depends on multiple cellular events including folding, aggregation, degradation, and inclusion body formation. Here, the development and analysis of two kinetic models that describe the extracellular secretion of HasA via the TISS from a wild type $E. coli$ strain and a $\Delta secB$ strain is illustrated. The models integrate HasA synthesis, protein folding, aggregation, chaperone interactions, degradation, transporter synthesis, reporter-transporter interactions, catalytic cycle of the transporter, protein secretion into the extracellular medium, and cell growth. The model results illustrate that despite the unavailability of parameter values, the secretion model can appropriately describe the qualitative experimental trends. A combination of techniques such as parameter variation,
sensitivity analysis, and nondimensionalization of the model suggests that the
optimization of extracellular concentration requires a balance between multiple factors
including HasA synthesis rate \( (k_0) \), transporter synthesis rate \( (k_t) \), and culture time. The
analysis also suggests that optimal levels of \( k_0 \) and \( k_t \) can be considerably different and
that the increase in extracellular concentration due to an increase in \( k_0 \) and/or \( k_t \) is not
proportional. Further, the optimal level of transporter synthesis rate may be lower than
the optimal level of HasA synthesis rate required for optimizing the extracellular levels of
HasA.

This chapter illustrate the utility and importance of mathematical modeling and
computational analysis that can identify factors influencing target product secretion and
ultimately define an optimal strategy. The model presented here can be extended to
incorporate new reactions, species, and pathways. This analysis describes the
optimization strategies for a model protein but the model can be modified and extended
to describe the secretion of other recombinant proteins via the TISS. To the best of our
knowledge, this study is the first report of kinetic modeling of extracellular recombinant
protein production in \textit{E. coli}. 

91
4.1 Summary and Conclusions

The first part of this thesis highlighted the use of genetic engineering approaches to enhance extracellular secretion of recombinant protein via the Type-I secretion system (TISS) in *E. coli*. Construction of an *E. coli ΔycdC* strain is described that is able to enhance the secretion levels of recombinant proteins via the TISS; however, it was observed that the increase in the secretion levels of various proteins was different and the secretion levels also varied across the background strains. In addition, the combined effect of the *ycdC* gene knockout and synonymous rare codon substitutions in the parent gene sequence on the secretion levels via the TISS was investigated. The combination of the results from this study and similar studies conducted by Gupta and Lee 2010 suggested that the optimal levels of protein expression for an enhanced secretion of recombinant proteins might be different in the parent strain and the ΔycdC strain. Further, a directed silent mutagenesis approach was employed in the C-terminal signal sequence (CTSS). This approach incorporates silent mutations in the predicted amphiphilic helix region of the CTSS keeping the amino acid sequence unchanged. It was shown that the approach can enhance the extracellular levels of recombinant proteins; however, the
increase in secretion varied across different proteins. To the best of our knowledge, this is the first report of enhanced extracellular secretion of recombinant proteins using silent mutations in the CTSS. Overall these results provide an insight into some of the experimental approaches that can be employed to increase extracellular recombinant protein secretion via the TISS in *E. coli*.

The second part of this thesis illustrated the use of mathematical modeling approaches to design strategies for optimization of extracellular secretion of a model protein (HasA). Development and analysis of two models were illustrated that describe the secretion of HasA from the wild type *E. coli* strain and an *E. coli ΔsecB* strain. The models consisted of a system of coupled non-linear ordinary differential equations that incorporated HasA synthesis, protein folding, aggregation, chaperone interactions, degradation, transporter synthesis, reporter-transporter interactions, catalytic cycle of the transporter, protein secretion into the extracellular medium, and cell growth. It was shown that the simulation results from the kinetic models were consistent with the qualitative trends in the experimental data available from the literature. Using a combination of computational and mathematical techniques including parameter variation, sensitivity analysis, and nondimensionalization, it was shown that the optimization of the extracellular concentration of HasA requires a balance between multiple factors including HasA synthesis, transporter synthesis rate, and culture time. The model analysis also suggested that the optimal level of HasA synthesis rate may be higher than the optimal level of transporter synthesis rate for an enhanced secretion of
HasA. To the best of our knowledge, this is the first report of a kinetic model of recombination protein secretion via the TISS in *E. coli*. Overall, these results demonstrated the utility of a computational and a mathematical based approach towards understanding and optimization of extracellular protein secretion via the TISS.

### 4.2 Recommendations for Future Work

#### 4.2.1 Construction of a Dual Expression System in *E. coli*

In the models described in this work, the parameter values of *k*$_{0\text{max}}$ (1,000 μM/min) and *k*$_{t\text{max}}$ (500 μM/min) are empirical values for the maximum HasA protein synthesis rate and transporter synthesis rate that the recombinant cells can tolerate before the theoretical specific growth rate (μ) reduces to zero. In the model simulations, constant values are assumed for these two parameters. However, in practice, these values will be different for each recombinant system, i.e. for each reporter protein and plasmid vector system. These parameters values can be estimated by designing a dual expression system. The reporter protein and the transporter proteins can be expressed under the control of two separate non-interacting inducers using two separate vectors that should be able to co-exist in *E. coli*, i.e. the origin of replication of the two vectors should be compatible. In addition, the two vectors should carry different resistance markers for selection after being transformed into *E. coli* cells. One combination of such vectors is pACYC-Duet1 (Novagen, NJ) and pPro18-km (Addgene, MA). The pACYC-Duet1 vector is a T7 promoter expression vector and therefore can only be expressed in an *E. coli* strain containing the chromosomal copy of the gene for T7 RNA polymerase. To solve this
problem, $P_{lac}$ promoter system from pUC18/19 (New England Biolabs, MA) can be integrated into the pACYC-Duet1 vector. The $P_{lac}$ promoter system is IPTG inducible and pPro18-km vector is induced by sodium propionate (Lee and Keasling 2005). Thus, a dual expression system can be designed in *E. coli*. Typical range of IPTG concentrations used for the induction varies from 0.005-1.0 mM (Donovan et al. 1996). Typical range of sodium propionate used for the induction of genes in the pPro18-km vector varies from 0-50 mM (Lee and Keasling 2005). Experiments can be performed where the reporter and transporter genes can be induced at different levels and growth profile of the cells under different induction conditions can be studied. The protein synthesis rates ($k_0$, $k_t$, $k_{0\text{max}}$, and $k_{t\text{max}}$) for different induction conditions can be estimated using pulse-chase experiments (Bowden and Georgiou 1990). The cell growth data can then be fitted to the equations described below to get reasonable estimates of relevant parameters; which can improve the accuracy of quantitative predictions from the model for any given recombinant system. The equations are:

$$y = A \exp\left\{- \exp\left[\frac{\mu \times e}{A} (\lambda - t) + 1\right]\right\}$$

(19)

where $y = \ln\left(\frac{N_t}{N_0}\right)$; $A = \ln\left(\frac{N_{\infty}}{N_0}\right)$; $N_t =$ number of cells at any given time “t”; $N_{\infty} =$ maximum number of cells; $N_0 =$ initial number of cells i.e. number of cells at time $t = 0$; $\lambda =$ lag time; $\mu =$ specific growth rate of the cells given by following equation:

$$\mu = \mu_0 \left(1 - \frac{k_0}{k_{0\text{max}}} \right) \left(1 - \frac{k_t}{k_{t\text{max}}} \right)$$

(20)
where, $\mu_0 =$ specific growth rate of the cells not expressing any recombinant product;

$k_{0\max}$ and $k_{t\max}$ are the maximum rates of production of HasA and transporter proteins at which the specific growth rate of the recombinant cells theoretically reduces to zero.

### 4.2.2 Extension of the Kinetic Models

The kinetic models described in this work can be extended to incorporate more details such as HlyB-HlyD interactions, HlyA-HlyB-HlyD interactions, transcription and translation kinetics (Lee and Bailey 1984), and the effect of plasmid copy number on recombinant protein production (Lee and Bailey 1984). The HasA synthesis rate ($k_0$) and transporter synthesis rate ($k_t$) are assumed to follow zero order kinetics. This assumption is a good starting point in absence of any quantitative data available from the literature. However, the mathematical equations proposed by Lee and Bailey 1984 can be incorporated in the kinetic models described in this study to understand the effect of plasmid replication, plasmid copy number, transcription, and translation kinetics on recombinant protein production. As described in the literature, these factors can play an important role in recombinant protein production (Bentley et al. 1990, Lee and Bailey 1984). Further, it is assumed in the model simulations that the recombinant protein production is induced at the start of the culture ($t = 0$). However, in practice, the cell growth phase and the recombinant protein production phase can be separated by inducing the cultures at different time points (Bentley and Kompala 1989, Lee and Ramirez 1992). To understand and simulate such conditions, the specific growth rate model can be modified to incorporate the effect of induction on specific growth rate of the cells. One
approach to incorporate such effects has been described by Lee and Ramirez 1992 where shock and recovery terms are incorporated in the equation for specific growth rate of cells. The shock and recovery terms describe the shock effect (immediate reduction in specific growth rate of the cells) after induction of recombinant proteins and subsequent recovery of the cells (Lee and Ramirez 1992). One disadvantage of the approach is the increase in number of parameters in the model. However, the experiments using a dual expression system described above can help in estimation of the parameters. Furthermore, the model simulations described in this study are restricted to batch cultures where it is assumed that for a given extracellular volume, the final cell concentration is fixed. The models can be extended to simulate the system in a fed-batch setting. This will involve incorporating a different cell growth model that can describe the concentration of cells with respect to substrate feeding rate, culture time, and induction conditions. The mathematical model developed by Lee and Ramirez 1992 can be used to simulate fed-batch production of recombinant proteins.

### 4.2.3 Construction of a Synonymous Codon Library

Previous experiments have shown that the relative contribution of specific codons, their position, and number of synonymous changes can play an important role in the secretion efficiency (Gupta and Lee 2010). To understand the effect of synonymous codon changes on the secretion efficiency it would be useful to screen the library of all the possible synonymous mutations of a particular target protein being secreted by particular secretion machinery. Recently, a 250-fold range in the expression level was
reported across the synthetic library of 154 genes that varied randomly at synonymous sites, but all encoded the same green fluorescent protein (GFP) (Kudla et al. 2009). A semi-random library of synonymous GILL-Bla mutants secreted via the TISS can be created to identify an optimal sequence for the GILL-Bla protein secretion as well as to measure the range of secretion of the library. It is not feasible to test all the possible synonymous changes (approximately $10^{34}$ mutants) and therefore library size needs to be limited to cover a feasible broad range. The 12 positions in the GILL-Bla protein sequence that incorporates three defined GILL sequences can be used as a starting point. Using the definition of rare codon as those codons whose corresponding tRNA concentration is less than 1% of the total tRNA concentration (as tabulated in Solomovici et al. 1997), G has two possible rare codons (GGA and GGA) and one most abundant codon (GGT); I has one rare codon (ATA) and two most abundant codons (ATT and ATC); and L has one rare codon (CTA) and one most abundant codon (CTG). The DNA sequence of GILL is (GGT ATC CTG CTG). For, synthesis of the G, a base mixture of G-G-A/G/T, can be used to incorporate only A, G or T at the third position. Likewise, for I, a base mixture of A-T-C/A and for L, a base mixture of C-T-G/A can be used for synthesis. The library size would be 13,824. The technique of end-to-end ligation (Isalan 2006) can be used to create the library. The library can then be fused to the CTSS and transformed into W3110 parent E. coli cells expressing the transport machinery. Screening can be performed using plate assay (Kobayashi et al. 1988), which will help in the isolation of hypersecreting strains by comparing the area of the lysis zone produced by the mutant strains relative to the parent strain (Kobayashi et al. 1988, Lee and Lee.
2005). Subsequent sequencing of the vectors secreting different amount of GILL-Bla protein will be helpful towards identifying the contribution of specific codons, their position, and number of synonymous changes that can best enhance secretion via the TISS. Further, it may help in formulating certain basic rules about synonymous codon changes to enhance secretion via the TISS that can be employed to enhance the secretion of other recombinant proteins via the TISS. Further, the hypersecreting GILL-Bla vector sequences can be tested for secretion in the ΔycdC strain as the results discussed in chapter 2 suggests that the ycdC knockout may have an additive affect on increase of the secretion levels via the TISS.
This is a License Agreement between Anup Agarwal ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Elsevier Limited</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registered Company Number</td>
<td>1982084</td>
</tr>
<tr>
<td>Customer name</td>
<td>Anup Agarwal</td>
</tr>
<tr>
<td>Customer address</td>
<td>334 East Main Street</td>
</tr>
<tr>
<td></td>
<td>Newark, DE 19711</td>
</tr>
<tr>
<td>License number</td>
<td>2545401023682</td>
</tr>
<tr>
<td>License date</td>
<td>Nov 10, 2010</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Elsevier</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Biochimica et Biophysica Acta (BBA) - Biomembranes</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—Distinct translocases and mechanisms</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Paolo Natale, Thomas Brüser, Arnold J.M. Driessen</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>September 2008</td>
</tr>
<tr>
<td>Licensed content volume number</td>
<td>1778</td>
</tr>
<tr>
<td>Licensed content issue number</td>
<td>9</td>
</tr>
<tr>
<td>Number of pages</td>
<td>22</td>
</tr>
<tr>
<td>Type of Use</td>
<td>reuse in a thesis/dissertation</td>
</tr>
<tr>
<td>Portion</td>
<td>figures/tables/illustrations</td>
</tr>
<tr>
<td>Number of figures/tables /illustrations</td>
<td>1</td>
</tr>
<tr>
<td>Format</td>
<td>both print and electronic</td>
</tr>
<tr>
<td>Are you the author of this Elsevier article?</td>
<td>No</td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>No</td>
</tr>
<tr>
<td>Order reference number</td>
<td></td>
</tr>
<tr>
<td>Title of your thesis/dissertation</td>
<td>COMPUTATIONAL AND EXPERIMENTAL APPROACHES TO ENHANCE EXTRACELLULAR SECRETION OF RECOMBINANT PROTEINS IN ESCHERICHIA COLI</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Dec 2010</td>
</tr>
</tbody>
</table>
INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

“Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER].” Also Lancet special credit - “Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier.”

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions.
conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. Translation: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.
16. **Website**: The following terms and conditions apply to electronic reserve and author websites:

**Electronic reserve**: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:

- This license was made in connection with a course,
- This permission is granted for 1 year only. You may obtain a license for future website posting,
- All content posted to the web site must maintain the copyright information line on the bottom of each image,
- A hyper-text must be included to the Homepage of the journal from which you are licensing at [http://www.sciencedirect.com/science/journal/xxxxx](http://www.sciencedirect.com/science/journal/xxxxx) or the Elsevier homepage for books at [http://www.elsevier.com](http://www.elsevier.com), and
- Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

17. **Author website** for journals with the following additional clauses:

All content posted to the web site must maintain the copyright information line on the bottom of each image, and

- The permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version,
- A hyper-text must be included to the Homepage of the journal from which you are licensing at [http://www.sciencedirect.com/science/journal/xxxxx](http://www.sciencedirect.com/science/journal/xxxxx). As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier’s online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article’s Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting,
- Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

18. **Author website** for books with the following additional clauses:

Authors are permitted to place a brief summary of their work online only.

- A hyper-text must be included to the Elsevier homepage at [http://www.elsevier.com](http://www.elsevier.com)

All content posted to the web site must maintain the copyright information line on the bottom of each image

- You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version,
- Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at [http://www.sciencedirect.com/science/journal/xxxxx](http://www.sciencedirect.com/science/journal/xxxxx), or for books to the Elsevier homepage at [http://www.elsevier.com](http://www.elsevier.com)
20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

21. **Other Conditions**:

v1.6

**Gratis licenses** (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK10881412.

Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

If you find copyrighted material related to this license will not be used and wish to cancel, please contact us referencing this license number 2545401023682 and noting the reason for cancellation.

Questions? [customercare@copyright.com](mailto:customercare@copyright.com) or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.
APPENDIX B

PERMISSION LETTER FOR FIGURE 1.2
This is a License Agreement between Anup Agarwal ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

<table>
<thead>
<tr>
<th>License Number</th>
<th>2545400225192</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Nov 10, 2010</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Nature Publishing Group</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Nature Reviews Microbiology</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>The structural biology of type IV secretion systems</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Rémi Fronzes, Peter J. Christie, Gabriel Waksman</td>
</tr>
<tr>
<td>Volume number</td>
<td></td>
</tr>
<tr>
<td>Issue number</td>
<td></td>
</tr>
<tr>
<td>Pages</td>
<td></td>
</tr>
<tr>
<td>Year of publication</td>
<td>2009</td>
</tr>
<tr>
<td>Portion used</td>
<td>Figures / tables</td>
</tr>
<tr>
<td>Number of figures / tables</td>
<td>1</td>
</tr>
<tr>
<td>Requestor type</td>
<td>Student</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis / Dissertation</td>
</tr>
<tr>
<td>High-res requested</td>
<td>No</td>
</tr>
<tr>
<td>Billing Type</td>
<td>Invoice</td>
</tr>
<tr>
<td>Company</td>
<td>Anup Agarwal</td>
</tr>
<tr>
<td>Billing Address</td>
<td>334 East Main Street</td>
</tr>
<tr>
<td></td>
<td>Apt. D10</td>
</tr>
<tr>
<td></td>
<td>Newark, DE 19711</td>
</tr>
<tr>
<td></td>
<td>United States</td>
</tr>
<tr>
<td>Customer reference info</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
</tbody>
</table>

**Terms and Conditions**

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use, subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this...
material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version. Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run). NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:

Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

For AOP papers, the credit line should read:

Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedicalcommunications.com for more information. Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:

Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication).

We are certain that all parties will benefit from this agreement and wish you the best in the
use of this material. Thank you.

v1.1

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK10881398.

Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

If you find copyrighted material related to this license will not be used and wish to cancel, please contact us referencing this license number 2545400225192 and noting the reason for cancellation.

Questions? customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.
APPENDIX C

PERMISSION LETTER FOR FIGURE 1.3 AND 3.1
This is a License Agreement between Anup Agarwal ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number 2545410248079
License date Nov 10, 2010
Licensed content publisher Springer
Licensed content publication Biotechnology Letters
Licensed content title Extracellular recombinant protein production from *Escherichia coli*
Licensed content author Ye Ni
Licensed content date Jan 1, 2009
Volume number 31
Issue number 11
Type of Use Thesis/Dissertation
Portion Figures
Author of this Springer article No
Order reference number
Title of your thesis / dissertation COMPUTATIONAL AND EXPERIMENTAL APPROACHES TO ENHANCE EXTRACELLULAR SECRETION OF RECOMBINANT PROTEINS IN *ESCHERICHIA COLI*
Expected completion date Dec 2010
Estimated size(pages) 100
Total 0.00 USD

Introduction
The publisher for this copyrighted material is Springer Science + Business Media. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at [http://myaccount.copyright.com](http://myaccount.copyright.com)).

Limited License
With reference to your request to reprint in your thesis material on which Springer Science and Business Media control the copyright, permission is granted, free of charge, for the use indicated in your enquiry. Licenses are for one-time use only with a maximum distribution
equal to the number that you identified in the licensing process.

This License includes use in an electronic form, provided it is password protected or on the university's intranet, destined to microfilming by UMI and University repository. For any other electronic use, please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com)

The material can only be used for the purpose of defending your thesis, and with a maximum of 100 extra copies in paper.

Although Springer holds copyright to the material and is entitled to negotiate on rights, this license is only valid, provided permission is also obtained from the (co) author (address is given with the article/chapter) and provided it concerns original material which does not carry references to other sources (if material in question appears with credit to another source, authorization from that source is required as well). Permission free of charge on this occasion does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

Altering/Modifying Material: Not Permitted
However figures and illustrations may be altered minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s) and/or Springer Science + Business Media. (Please contact Springer at permissions.dordrecht@springer.com or permissions.heidelberg@springer.com)

Reservation of Rights
Springer Science + Business Media reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

Copyright Notice:
Please include the following copyright citation referencing the publication in which the material was originally published. Where wording is within brackets, please include verbatim.
"With kind permission from Springer Science+Business Media: <book/journal title, chapter/article title, volume, year of publication, page, name(s) of author(s), figure number(s), and any original (first) copyright notice displayed with material>.

Warranties: Springer Science + Business Media makes no representations or warranties with respect to the licensed material.

Indemnity
You hereby indemnify and agree to hold harmless Springer Science + Business Media and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License
This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without Springer Science + Business Media's written permission.
No Amendment Except in Writing
This license may not be amended except in a writing signed by both parties (or, in the case of Springer Science + Business Media, by CCC on Springer Science + Business Media's behalf).

Objection to Contrary Terms
Springer Science + Business Media hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer Science + Business Media (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction
All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by the country's law in which the work was originally published.

Other terms and conditions:

v1.2

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK10881416.
Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

If you find copyrighted material related to this license will not be used and wish to cancel, please contact us referencing this license number 2545410248079 and noting the reason for cancellation.

Questions? customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.
REFERENCES


127


129