NTNU – Trondheim
Norwegian University of
Science and Technology

Optimzing the 5'-end of Coding Sequences in Recombinant mRNA to achieve high-level Expression in the Bacterium *Escherichia coli*

Adrian Ertsås Naas

Biotechnology (5 year)

Supervisor: Svein Valla, IBT
Co-supervisor: Veronika Kucharova, IBT

Norwegian University of Science and Technology
Department of Biotechnology

Preface and acknowledgements

This thesis is submitted as the conclusion of the five year Master of Science program in Biotechnology at the Norwegian University of Science and Technology. The work was performed at the Institute of Biotechnology, under the supervision of Prof. Svein Valla and Ph.D. candidate Veronika Kucharova. All presented experimental work was performed by the author, unless stated otherwise.

I would like to thank my supervisors for their support and guidance throughout the work on my Master's thesis, and especially Veronika for all her help and input, and for always pushing me to work harder, and then to work some more. I am also grateful for the pleasant working environment created by the friendly people in the MOLGEN group, which has made working in the lab a pleasure.

Finally, Hanne deserves my biggest thanks for her relentless support and encouraging words.

Trondheim, May 15th 2012

Adrian Ertsås Naas

"One, two, three, biotechnology"
-Transformation feat. Notorious GFP

Abstract

Recombinant protein production in *Escherichia coli* provides a cheap and efficient way of producing medically and industrially relevant proteins. Sequence features of individual genes and especially their 5' terminal coding sequences act on the efficiency of gene expression by complex regulatory mechanisms which are still not fully understood. This study aimed to investigate the features of the 5' coding region of recombinant mRNAs, and to optimize them for increased expression in *E. coli*.

A previous study had found that a synonymous change of the *bla* reporter gene 2nd codon leads to an increased expression, and accordingly a synonymous library in the 5' *bla* coding sequence was created by a directed evolution approach building on this feature. Variants conferring up to three-fold increases in active enzyme amounts were identified, and the increased expression was shown to stem from increased transcriptional efficiency. The effect of changing the 2nd codon synonymously was further investigated by synonymous substitutions of the 2nd codons of the *bla* and two other reporter genes, *phoA* and *celB*. These experiments showed that the effect of 2nd codon changes on the gene expression is determined by the sequence context, as changes in expression levels appeared to be gene specific. All the coding sequences of the study were also analysed *in silico*, and an application for calculating the tRNA adaptation index was programmed in Python and made freely available online.

As the synonymous codon changes did not lead to a great improvement in protein amounts and any sequence features affecting the expression were hard to pinpoint, an alternative strategy involving 5' terminal gene fusions was investigated. Combinatorial mutagenesis coupled to an effective screening technique was applied to further optimize a 5' terminal fusion partner, previously shown to improve expression of several eukaryotic genes. The application of the best identified fusion partner candidate yielded a 3.8-fold improvement in IFN-α2b protein amounts over the original fusion, and showed twice as high protein amounts than a pelB-IFN-α2b fusion previously proven to give industrial expression amounts. The developed peptide fusion is thus an eligible candidate for further development for use in heterologous protein production.

Sammendrag

Rekombinant proteinproduksjon i *Escherichia coli* er en billig og effektiv produksjonsmetode for medisinske- og industrielt relevante proteiner. Effektiviteten av genekspresjon styres av trekk i gensekvensene til de enkelte genene, og særlig deres 5' terminale kodende sekvens av kompliserte regulatoriske mekanismer som enda ikke er fullt ut forstått. Målet med denne oppgaven var å undersøke trekk i den 5' terminale kodende regionen av rekombinant mRNA, og å optimalisere regionen for økt genuttrykk i E. coli.

En tidligere studie hadde vist at en synonym endring av 2. kodon i bla reportergenet førte til økt genuttrykk, og et synonymt genbibliotek i den 5' terminale kodende sekvensen av bla ble følgelig konstruert ved en målrettet evolusjonstilnærming, der den positive endringen ble videreført. Varianter fra genbiblioteket ga opp til tre ganger økning i mengde aktivt enzym, og det økte genuttrykket ble vist til å stamme fra økt transkripsjonseffektivitet. Effekten av å endre 2. kodonet ble ytterligere undersøkt ved synonyme substitusjoner av 2. kodon i bla og to andre reportergen, phoA og celB. Disse eksperimentene viste at effekten av å endre 2. kodon synonymt avhenger av den omliggende gensekvensen, da endringene i genuttrykk var genspesifikke. Alle kodende sekvenser i denne oppgaven ble undersøkt in silico, og et script for å regne ut tRNA adapsjons indexen til en sekvens ble programmert i Python, og gjort fritt tilgjengelig på internett.

Ettersom de synonyme endringene ikke førte til store forbedringer i mengden protein, og det ikke ble identifisert bestemte trekk i sekvensene som påvirket genuttrykket, ble en annen strategi fulgt der en 5' terminal genfusjonpartner ble undersøkt. Kombinatorisk mutagenese sammen med en effektiv seleksjonsmetode ble brukt til å optimalisere en 5' fusjonspartner som tidligere hadde vist seg effektiv til å forbedre genuttrykk av flere eukaryote gener. Fusjon av den beste optimaliserte fusjonspartneren til ifn-α2b genet ga 3,8 ganger høyere IFN-α2b proteinproduksjon enn vill-type fusjonen, og dobbelt så høy proteinproduksjon enn en pelB-IFN-α2b fusjon som tidligere hadde vist industrielle produksjonsverdier. Dermed er den utviklede fusjonspartneren en godt kvalifisert kandidat for videre utvikling til bruk i heterolog protein produksjon.

Table of Contents

Abbreviations

A:	Adenine, nucleobase in RNA and DNA. Normally bonds with T or U
bla:	Gene encoding β-lactamase
C:	Cytosine, nucleobase in RNA and DNA. Normally bonds with G
celB:	Phosphoglucomutase
$celB_{23}$:	Peptide fusion of the first 23 amino-acids of celB
celB:	Gene encoding celB
$celB_{69}$:	Gene fusion of the 69 first nucleotides of *celB*
DNA:	Deoxyribonucleic acid
G:	Guanine, nucleobase in RNA and DNA. Normally bonds with C
GM-CSF:	Granulocyte-macrophage colony-stimulating factor
Gm-csf:	Gene encoding GM-CSF
IB:	Inclusion body
IFN-α2b:	Human interferon alpha 2b
ifn-α2b:	Gene IFN-α2b
mRNA:	Messenger RNA
RBS:	Ribosome binding site
RNA:	Ribonucleic acid
S:	Svedberg sedimentation coefficient (Sharma et al., 2009).)
scFv-phOx:	Single-chain antibody variable fragment
SD:	Shine-Dalgarno site
T:	Thymine, nucleobase in DNA. Normally bonds with A
TIR:	Translation initiation region of mRNA
T.I.R.:	Translation initiation rate
U:	Uracil, nucleobase in RNA. Normally bonds with A
UTR:	Un-translated region of mRNA

1 Introduction

Heterologous protein production, where an organism is reprogrammed to produce a foreign protein of interest by genetic engineering, is of major importance in the production of therapeutics. Without for example the human insulin produced in *Escherichia coli*, diabetes may have still been treated using insulin extracted from animal pancreatic tissue (Junod, 2007). Other examples of heterologously produced therapeutic proteins include Interferon α2b, used to treat viral infections and some forms of cancer (Srivastava et al., 2005), and human growth hormone used to treat growth failure (Sonoda and Sugimura, 2008). A variety of production hosts are used, such as yeasts, mammalian cells, plants, and even the mammary glands of goats, producing the proteins in their milk (Ferrer-Miralles et al., 2009). However, the gram-negative bacteria *E. coli* continues to remain the most popular choice, as it confers many advantages in the production process (Huang et al., 2012).

It has been a favorite model organism of prokaryotic research since its discovery by Theodor Escherich (Escherich, 1885; Hacker and Blum-Oehler, 2007) and further characterization in 1913 (Castellani and Chalmers, 1913), due to its high growth rate and ease of cultivation. Because of its qualities as a model organism, it was chosen for research on bacterial transformation (Lederberg and Tatum, 1946), leading to more research and increased knowledge of prokaryotic genetics. The successful recombination of genetic elements from different organisms, and their expression in *E. coli* (Cohen et al., 1973) spurred a biotechnological revolution in the 1970s and 1980s where therapeutic proteins and metabolic products could be produced cheaply and efficiently. The first recombinant therapeutic protein produced was somatostatin (Itakura et al., 1977), and in 1982 Humulin, human insulin produced in *E. coli*, became the first genetically engineered drug approved by the U.S Food and Drug Administration (FDA) (Junod, 2007).

E. coli remains a valuable production host due to its ease of cultivation on cheap substrates, well studied genetics and proven versatility in terms of the variety of proteins it can produce at high quality (Huang et al., 2012). In 2009, 29.8 % of the 151 therapeutic proteins approved by the FDA and the European Medicines Agency was produced in *E. coli* (Ferrer-Miralles et al., 2009), demonstrating its dominance over other hosts in the industry.

However, in spite of its successes in production and the vast amount of research on the field, high-level protein production is not always a straight-forward matter in the bacterium (Jana and Deb, 2005). Due to the variety of intricate mechanisms that govern the regulation of its protein production and the different genetic features of individual heterologous genes, some proteins are produced in low amounts or none at all (Makrides, 1996). Thus more knowledge is needed on these regulatory mechanisms in order to increase, and efficiently control protein production. This holds for the field of protein production, but also very much so for metabolic engineering and the emerging field of synthetic biology, where genetic circuits and even entire genomes may be synthesized from scratch, and individual proteins must be expressed at fine-tuned levels (Lale et al., 2011).

In the following chapters important aspects of protein expression in *E. coli* will be discussed. The regulation of gene expression will be discussed in chapter 1.1, followed by a description of various tools and strategies for improving heterologous protein production in chapter 1.2 and 1.3. Chapter 1.4 will present the recombinant proteins used in this study, and the aim of the experimental study will be described in chapter 1.5

1.1 Control of Gene Expression in *E. coli*

The central dogma of molecular biology states that genes are transcribed from DNA into a messenger RNA, which is further translated into proteins as shown in **Figure 1-1** (Crick, 1970). Thus the regulation of gene expression may be divided into transcriptional regulation and translational regulation, which both are regulated in various ways as indicated in the figure. The regulation of DNA transcription is described in chapter 1.1.1 and the regulation mechanisms of translation are described in chapter 1.1.2.

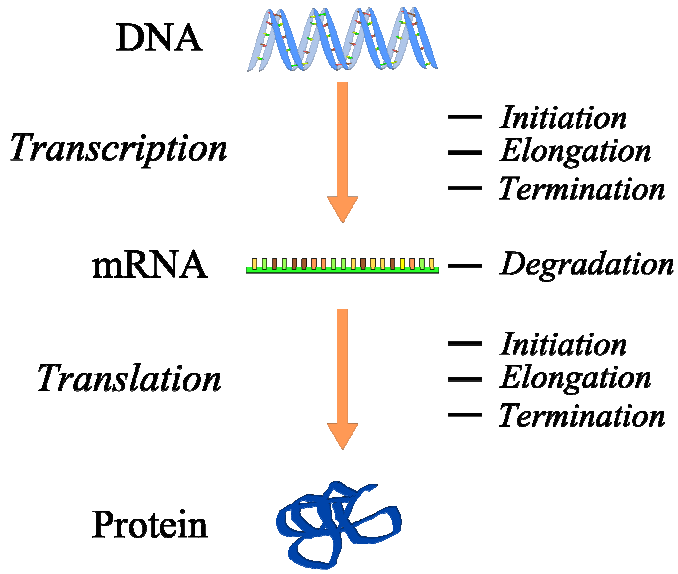


Figure 1-1: Protein expression according to the central dogma of molecular biology, important steps in expression that are under heavy regulation are shown in italics. Points of regulation for the transcription and translation processes are described in 1.1.1 and 1.1.2, respectively. Adapted from (Crick, 1970; Browning and Busby, 2004).

1.1.1 Regulation of DNA Transcription

Transcription is the primary regulatory step of protein expression; if a gene is not transcribed from DNA into mRNA, the translational machinery has no blueprint of the gene and the protein is not expressed. Transcription of genes in all organisms is under the control of up-stream DNA sequence elements called promoters that attract the transcriptional machinery with different affinities on account of sequence elements within them (Browning and Busby, 2004). The process is usually divided in three parts; transcription initiation, elongation and

termination, as seen in **Figure 1-1**, the most important for regulation being the initiation step (Browning and Busby, 2004).

Transcription initiation begins with the recognition of promoter elements by multi-domain proteins called σ-factors that attract the DNA dependent RNA-polymerase (RNAP). *E. coli* has seven σ-factors (Sharma and Chatterji, 2010), the most common one being σ^{70}. Different σ-factors recognize different sets of promoters so that global transcription can be changed, for example from the house-keeping genes controlled by σ^{70}, to heat-shock response genes under control of σ^{32} (Grossman et al., 1984).

A schematic figure of promoter features discussed in this section is shown in **Figure 1-2**. The hexameric -10 and -35 elements (relative to the transcriptional start-site) are the most important parts of the promoter, and the spacing between them and their similarities to their consensus sequences are the main determinants of promoter strength (Fournier et al., 1999). The -10 element is called the Pribnow box and has the consensus sequence 5'-TATAAT-3' (Pribnow, 1975; Lisser and Margalit, 1993). The -35 element has the consensus sequence 5'-TTGACA-'3. The -10 and -35 facilitate binding of domain 2 and 4, respectively, of the σ_{70}-subunit (Browning and Busby, 2004). The extended -10 element and the UP are additional important sequence elements of some promoters, which both act in the recruitment of σ-factors and may enhance promoter activity 300-fold (Davis et al., 2011). The relative contribution of the four elements varies from promoter to promoter (Browning and Busby, 2004). Promoter activity is also regulated by proteins known as transcription factors (TFs), which bind to regulatory regions called operators in the promoter. These may facilitate the binding of RNAP, or block its access to the promoter, acting as negative or positive regulators (Browning and Busby, 2004). TFs are produced naturally to respond to environmental changes, such as the presence of metabolites, allowing genes to be transcribed when needed.

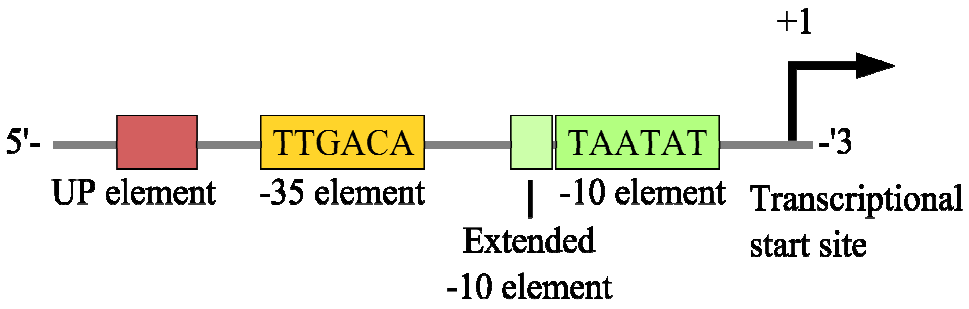


Figure 1-2: Overview of promoter elements described in the text. The gray line depicts DNA, and the transcribed DNA lies 3' of the transcriptional start site. Consensus sequences are shown for -35 and -10 elements. Adapted from (Browning and Busby, 2004).

During transcription initiation, the RNAP core-enzyme binds to the σ-factor which is attracted to the sequence elements in the promoter. This allows the now called RNAP-holoenzyme to isomerize to an "open" complex, unwinding the DNA double helix around the transcriptional start-site (Davis et al., 2011). The RNAP may undergo a repetitive process called abortive transcription (Chan and Gross, 2001), where, short RNA transcripts are produced and released before the RNAP returns to the transcriptional start-site and starts over. The duration of the process is dependent on the +1 to +20 sequence relative to the transcriptional start site, and continues in a stochastic fashion until promoter escape where the RNAP clears the promoter (Hsu et al., 2006). Thus the DNA of the 5' un-translated region of the mRNA has an affect on transcriptional efficiency (Berg et al., 2009; Berg et al., 2012), further discussed in 1.3.1.

After promoter escape, the elongation complex is formed and nucleoside triphosphates complementary to the non-coding strand in the DNA template are bound in the complex's active site and added chemically one by one to the 3'-OH group of the growing RNA chain, making a copy of the coding strand where (Roberts et al., 2008). Several mechanisms regulate the elongation. Local sequence features of the emerging RNA may cause transcriptional pausing, for example to synchronize translation with transcription in the tryptophan operon (Landick et al., 1985), and the NusA elongation factor helps prevent transcriptional termination during pauses (Roberts et al., 2008).

Termination of transcription is achieved by two mechanisms. In intrinsic termination, local sequence features allow hairpins to form in the nascent RNA which destabilizes bindings in the elongation complex, facilitating termination (Nudler and Gottesman, 2002). Rho-dependent termination depends on Rho protein monomers binding to specific sites in nascent RNA not protected by secondary structures or the translation complex and wrapping of the

RNA around the hexamer Rho complex, which may pull the RNA out of the active site of the elongation complex (Nudler and Gottesman, 2002).

The regulation of transcription is manipulated to a great extent during heterologous protein expression, especially by use- and manipulation of promoters, which will be discussed in chapter 1.2.1. This manipulation usually leads to high amounts of mRNA transcript for expression of the protein of interest, but even though transcription is considered the primary regulatory step of expression, there is a substantial variance in translational efficiencies for various mRNAs (Makrides, 1996). A wide range of translational regulation mechanisms have been identified, and the most important for heterologous protein production are discussed in the following chapter.

1.1.2 Translational Regulation

The mRNA intermediate is translated into proteins by large macromolecules composed of ribosomal RNA and proteins, called ribosomes (Kaczanowska and Ryden-Aulin, 2007). Base triplets in the transcript code for amino acids which are brought to the translation complex by aminoacyl transfer RNAs (tRNAs) and linked together (Laursen et al., 2005). As the translation efficiency of heterologous transcripts vary widely (Park et al., 2007; Kudla et al., 2009), an understanding of the regulatory mechanisms of translation is important for heterologous protein production and will be discussed below.

The translation process is usually divided into three separate events; initiation where the ribosome binds to the transcript and commences translation, elongation where the ribosome traverses the transcript and synthesizes the peptide, and termination where the peptide is released from the ribosome (Kaczanowska and Ryden-Aulin, 2007). Initiation is rate limiting step of the process (Tuller et al., 2010), and it's regulation depends heavily on the sequence and secondary structure of the translation initiation region of the transcript, as described in chapter 1.1.2.1. Elongation is also an important regulatory step that depends on the codon usage of the transcript, and is discussed in chapter 1.1.2.2. Termination is achieved when the ribosome reaches one of the three stop codons UAA, UAG, or UGA in the mRNA, recognized by Release Factors 1 or 2 which trigger hydrolysis and peptide release (Kaczanowska and Ryden-Aulin, 2007). An additional important regulatory mechanism of translation is the degradation kinetics of the mRNA, which is discussed in chapter 1.1.2.3

1.1.2.1 Translation Initiation Regulation

The translational initiation region (TIR) denotes the region of the transcript that influences the translational initiation, and includes the transcript's 5' un-translated region (UTR), its ribosome binding site (RBS), the translational start-codon, and part of the 5' coding sequence (Huang et al., 2012). Two main features of the TIR control the rate of initiation, namely the sequence of the Shine-Dalgarno (SD) site and the region's secondary structure.

In translation initiation the 70S ribosomal complex, which in *E. coli* is made up of two subunits, 50S and 30S (Julian et al., 2011), assembles on the RBS of the mRNA along with the initiating N-formylmethionine aminoacyl tRNA, which binds to the AUG start-codon. The RBS denotes the approximately 30 nucleotides of the transcript that is covered by the ribosome during binding (Laursen et al., 2005), and includes the SD sequence which attracts the ribosome by complementarity to a sequence in the 3' end of the 16S ribosomal RNA subunit of 30S. The SD was discovered in bacteriophage mRNA (Shine and Dalgarno, 1974) and interaction between the site and the 16S RNA was later confirmed (Steitz and Jakes, 1975). The SD's stimulatory effect on translational initiation depends on its sequence, its length and the spacer region between the SD and the initiation codon (Ringquist et al., 1992), and variations of these has been shown to give a 3000-fold difference in translational yields (Barrick et al., 1994).

The secondary structure of the TIR greatly affects translation initiation (deSmit and Vanduin, 1990; Kudla et al., 2009; Goltermann et al., 2011). mRNA is single-stranded, and thus bases in the transcript may form internal base-pairs giving stem-loop structures. GC bindings, which have three Hydrogen-bonds, are stronger than AU bonds with only two, and structures and their stability may be predicted based on the sequence's minimum folding energy (MFE) (Zuker and Stiegler, 1981). The stability of secondary structures forming in the UTR and around the ribosome binding site has a profound effect on translational efficiency by blocking the ribosome's access to the SD (Park et al., 2007; Huang et al., 2012) and to the translational start-codon (Zhang et al., 2006). Both the 5' UTR and the 5' coding sequence nucleotides may contribute to the secondary structures (Kudla et al., 2009; Goltermann et al., 2011), and thus their nucleotide sequence greatly affects protein expression.

As the TIR has such a profound effect on the translation process, modifications of its features are frequently utilized in improving heterologous protein expression, and such methods will be described in chapter 1.3.1.

1.1.2.2 Codon Usage Affects the Rate of Translational Elongation

The genetic code is degenerate (Nirenberg et al., 1966), in that there are 61 codons coding for only 20 amino-acids in *E. coli* (Gustafsson et al., 2004). I.e. one amino-acid may be encoded by one to six different synonymous RNA codons. Although the choice of one synonymous codon over others has no effect on the peptide sequence, the use of codons in *E. coli* and other organisms is not at all random (Tuller et al., 2010). The speed of translational elongation is affected by the availability of the codon's amino-acyl tRNA (Mitarai et al., 2008), and there is a clear codon bias where highly expressed native genes in *E. coli* are biased towards a different set of codons than other genes (Tuller et al., 2010).

The use of synonymous codons was shown to be adapted towards the abundance of the codon's respective major isoacceptor tRNAs in the organism, and the codon adaptation index (CAI) was developed to score genes on how well they are adapted to an organism (Sharp and Li, 1987). A reference set of highly expressed genes was used to score each codon against its synonymous ones by observing the frequency of which it occurs in the highly expressed genes, and a gene's score is calculated as the geometric mean of all its codon scores. In prokaryotes, several ribosomes translate each mRNA simultaneously (Goodman and Rich, 1963; Warner and Knopf, 2002), and it has been suggested that the high CAI of native highly expressed genes has evolved to increase the efficiency of the cell's global translation. In that an increased elongation rate in highly expressed proteins leads to fewer ribosomes sequestered on the transcript, and thereby increases their availability for the translation of other proteins (Kudla et al., 2009).

Another measure of the gene's adaptation to the host's tRNA pool is the tRNA Adaptation Index (tAI). It was developed on the same principle as the CAI, but is based on the tRNA genes' copy-numbers and their availability for a given codon, along with weighted contributions for wobble-interactions where one tRNA may decode several codons (dos Reis et al., 2003). Thus a codon with a low tAI is translated at a slower rate than its synonymous mutants with higher tAI. Local tAI analysis of several genomes has identified an evolutionary

conserved occurrence of slowly translated codons in the start of genes (Tuller et al., 2010). They suggested that this "ramp" may have evolved to help distribute closely spaced ribosomes from genes with high translation initiation efficiency, thus reducing chances of ribosome collisions leading to abortive protein synthesis, which is a burden to the cell.

Expressing heterologous proteins in *E. coli* often proves difficult as the source organism's genes are adapted to a different tRNA pool (Gustafsson et al., 2004), and codons rare to the host may lead to growth arrest, premature translational termination and mistranslated proteins (Huang et al., 2012). Strategies for overcoming these problems include alteration of the host's tRNA pool and codon optimization of the gene, which will be described in chapter 1.3.1.

1.1.2.3 The Stability of the mRNA Molecule Affects Translational Efficiency

The amount of time a transcript is available for translation has an effect on how much protein is produced (Belasco, 2010), and thus mRNA half-life can be used as a way to regulate gene expression. The rate of degradation of *E. coli* transcripts varies over a vast range, and half-lives last from seconds, up to 20 minutes (Sorensen and Mortensen, 2005). mRNA is degraded via two pathways in *E. coli*, both of which are dependent on the low-specificity RNA degrading enzyme (a ribonuclease) RNaseE, along with 3' exonucleases that degrade transcripts from the 3' terminus (Belasco, 2010).

Intact transcripts are protected by a stem-loop structure forming in the 3' UTR of the transcript, blocking the degradation from 3' exonucleases (Belasco, 2010). In the 5' independent pathway RNaseE repeatedly cuts the transcript in AU rich regions (McDowall et al., 1994) creating fragments with and without 3' secondary structures, allowing 3' degradation of the unprotected ones. Intermediates containing stem-loops are made accessible by 3' polyadenylation (Xu and Cohen, 1995), allowing the exonucleases to degrade them.

In addition to the stabilizing 3' structures, 5' tri-phosphate caps and 5' stem-loops also increase the stability of *E. coli* transcripts (Emory et al., 1992; Mackie, 1998). In the 5' dependent pathway, the RNA pyrophosphohydrolase H (RppH) removes a pyrophosphate from the transcript, leaving a mono-phosphorylated transcript (Celesnik et al., 2007; Deana et al., 2008). This allows its binding to a discrete pocket in RNaseE (Callaghan et al., 2005), enhancing the efficiency of RNaseE cleavage and increasing the rate of degradation by over an order of magnitude compared to tri-phosphorylated transcripts (Mackie, 1998; Belasco,

2010). Stem-loop structures in the 5' terminus of transcripts sterically block RppH, and thus lead to stabilization of the transcript (Deana et al., 2008)

Another factor influencing the degradation of mRNA is its association with ribosomes, which blocks RNaseE's access to the transcript (Deana and Belasco, 2005). A poor SD sequence lowers the rate of ribosomal binding to the transcript and thus lowers the translation initiation, and it has been shown to destabilize the mRNA (Wagner et al., 1994). An efficient translation initiation could allow closer spacing of ribosomes on the transcript thus protecting it. However a study showed that ribosome binding to the RBS of transcripts without translation also protected the mRNA, suggesting that RBS occupancy by a ribosome may work as a barrier against the 5' bound RNaseE (Deana and Belasco, 2005).

A ribosomal spacing model showed that the spacing of ribosomes does affect the transcript's stability (Pedersen et al., 2011). It predicted that inserting slowly translated codons before codon 20 of the *lacZ* transcript would slow the translating ribosomes, creating a gap in front of them as the previous ribosome would have reached the faster translated codons, speeding away from it. The gap would allow RNaseE access to the transcript, thus lowering its stability. Experiments supported the prediction, and also the stability increase predicted by inserting slowly translated codons after codon 45 to increase ribosome density in the 20-45 region, was confirmed.

The use of mRNA stabilization as a technique for increasing heterologous protein expression in *E. coli* has not been utilized as extensively as increasing transcription or the translational efficiency. However the UTR of the unusually stable *ompA* mRNA, which contains two stable hairpins in the extreme 5' terminus, has been successfully used to increase the stability of β-lactamase (Belasco et al., 1986) and β-galactosidase transcripts (Hansen et al., 1994) through UTR-gene fusions. Another study created a synthetic UTR library with stem-loop structures in the 5' terminus of the UTR, which varied *lacZ* transcript over an order of magnitude (Carrier and Keasling, 1999).

The manipulation of the translational regulation that has been described in this chapter is used extensively for optimizing heterologous protein expression, as will be discussed in chapter 1.3. However, the heterologous gene needs to be established in the host and transcribed for efficient protein production, which will be discussed in the following chapter.

1.2 Heterologous Protein Expression in *E. coli*

Heterologous protein production requires the introduction and expression of heterologous genes in the host organism and is usually performed using specially designed expression vectors developed from natural plasmids (Ferrer-Miralles et al., 2009). In this chapter the features of expression vectors will be discussed, followed by a description of the vector used in this study in chapter 1.2.4.

1.2.1 General Features of Expression Vectors

Natural plasmids are circular DNA entities capable of replicating independently from the host's chromosome that usually confer an advantage to the host, like antibiotic resistance or additional metabolic pathways (Frost et al., 2005). A variety of plasmids with different origins have been developed for heterologous protein production, and vectors suited for various types of protein expression are available commercially (Promega, Invitrogen, Stratagene, Novagen). Most expression vectors have a general set of common features needed for their function (Jana and Deb, 2005), and a schematic view of these is given in **Figure 1-3**

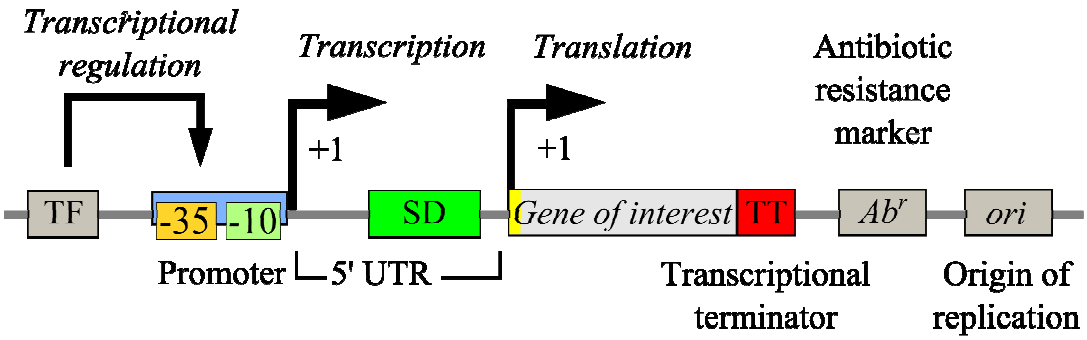


Figure 1-3: General features of protein expression vectors, not drawn to scale. TF: gene coding for transcription factor; SD: Shine-Dalgarno site. Details on features are given in the text. Adapted from (Jana and Deb, 2005).

The origin of replication (*ori*) is necessary for the plasmid to take advantage of the host cell's DNA replication machinery and proliferate along with the cell (Baker and Wickner, 1992). Their role and mechanism in expression are discussed in chapter 1.2.3. Expression cassettes containing promoters, regulatory proteins and multiple cloning sites (MCS) are used for expression of the protein, and are discussed in further detail in chapter 1.2.2. Transcriptional

terminator sequences are often positioned behind the MCS so that they hinder transcriptional read-through of the gene of interest, which may cause promoter occlusion (Makrides, 1996) and plasmid instability by transcription of the *ori* (Sorensen and Mortensen, 2005).

To keep a selective pressure for cells harboring the expression plasmid, an antibiotic resistance selection marker is normally used (Sorensen and Mortensen, 2005). The *hok/sok* system is often used in addition to increase plasmid stability, and consists of a locus specifying two RNAs (Franch et al., 1997). The *hok* mRNA encodes two reading frames, *mok* and *hok* (modulation of killing, and host killing) which are tightly coupled translationally. The sok RNA binds to the TIR of *mok,* thus inhibiting translation of *hok*, keeping the cell alive. The sok RNA degrades faster than *hok/mok* mRNA which will linger in the cytoplasm after cell division and will be translated if the cell does not contain a plasmid, killing the cell (Franch et al., 1997).

1.2.2 Expression Cassettes

To regulate and promote high expression of the protein of choice, a tightly controlled and efficient promoter system, along with its regulatory gene(s) are needed (Huang et al., 2012). Transcription factor (TF) regulated promoters are widely used in recombinant protein production, as they allow controlled induced expression of the gene of interest by adding the chemical that naturally acts on the TF to start expression. A low basal expression is desirable, meaning there should be little expression of the gene of interest before induction (Sorensen and Mortensen, 2005). This is especially important if the protein produced is host-toxic, allowing high cell density to be reached before induction, which gives a higher yield of the product (Jana and Deb, 2005). However, for industrial scale production of proteins inducer cost and its toxicity to the host may be limiting factors (Jana and Deb, 2005). Alternative induction methods have been developed, such as thermal induction (Hasan and Szybalski, 1995) or pH-dependent induction (Chou et al., 1995).

Several promoter systems have been constructed for high-level protein production (Terpe, 2006). One early example is the *tacI* promoter (de Boer et al., 1983) which was constructed from the -35 region of the *trp* promoter and the -10 region of the *lac* promoter. This synthetic promoter gave 11 and 3 times more efficient transcription than the *lac* and *trp* promoters, respectively, while maintaining the ability to be induced by the lactose analog Isopropyl β-D-1-thiogalactopyranoside (IPTG). Perhaps the most widely used expression systems today are

based on the T7 RNA polymerase (Terpe, 2006). This system utilizes the RNA polymerase from the phage T7, which only transcribes genes from T7 promoters (Studier and Moffatt, 1986). Use of the system requires a special production strain for its function where the polymerase has been cloned into the chromosome of *E. coli* strain BL21 under control of a *lac* derived promoter, allowing its strong induction from IPTG (Terpe, 2006). A third example is the pBAD/AraC promoter system, derived from the arabinose operon (Guzman et al., 1995). The AraC repressor/activator controls the pBAD promoter which is induced using L-arabinose, a relatively inexpensive inducer. *Ara* negative *E. coli* strains are maximally induced at 0.001 % L-arabinose, making the process even cheaper (Terpe, 2006).

Improving the strength and efficiency of promoter systems has received much attention in studies trying to optimize protein expression. Studies where the promoter sequence has been partially or fully randomized have proven successful in creating a wide range of promoter strengths and induction ratios (Miksch et al., 2005; Bakke et al., 2009). The TFs of promoter systems has also been targeted for improving the expression system. A randomized library followed by directed evolution of the XylS regulatory protein of the *Pm* promoter improved expression nine-fold while maintaining low basal expression (Vee Aune et al., 2010).

In addition to the promoter system, the 5' UTR of the transcript is important in protein translation as discussed in chapter 1.1.2, and downstream of the promoter there is usually a high efficiency UTR with a consensus SD sequence allowing high translational efficiency (Jana and Deb, 2005). For eased cloning of the gene of interest, a MCS with a range of restriction sites, or restriction sites that allow in-frame cloning with the start-codon is placed after the UTR. Another consideration for the vector's gene expression is the gene dosage, which is discussed below.

1.2.3 Gene Dosage

The *ori* and its regulatory system keep the plasmid replicating, but also determine how many copies of the plasmid are present in the cell (Wild and Szybalski, 2004). This is an important factor for the number of genes of interest available for transcription, the gene dosage. Expression vectors use different replicons and mechanisms for maintaining their copy number in the cells, and a high copy number generally leads to high expression of the gene of interest (Huang et al., 2012). However, the maintenance demand of keeping high numbers of plasmids

imposes a heavy metabolic burden on the cell, and may lead to low growth rates and low expression (Jana and Deb, 2005).

The detrimental effect of high copy number plasmids may be avoided by use of copy-control plasmids, where the copy number may be increased by induction (Jana and Deb, 2005). RK2-based plasmids replicate from a region called the origin of vegetative growth (*oriV*) in a manner controlled by two gene products from the *trfA gene*, and *trfA* up-mutants have been identified (Durland et al., 1990). In the pBAD/oriV expression vectors, the *trfA* up-mutant gene along with the gene of interest, are under control of the arabinose inducible promoter *P*ara (Wild and Szybalski, 2004). The increase in copy number, along with induction of the gene's promoter leads to a 50 000-fold increase in expression of the gene upon induction. In addition to lowering the maintenance burden, this gives the system a very low background expression making it useful for the production toxic proteins.

1.2.4 RK2 based Pm/XylS Expression Vector

The plasmids used in this study are based on the pJBn vectors constructed by Blatny *et. al.* to be tightly controlled broad host-range expression vectors that were easily regulated in terms of their induction (Blatny et al., 1997). The RK2 based plasmids contain the *oriV* and *TrfA* gene, which are known to be functional in many gram-negative, and some gram-positive bacteria. The XylS/*Pm* promoter system of the plasmid stems from the TOL plasmid of *Pseudomonas putida* (Ramos et al., 1997), and is a strong and active promoter in many hosts (Sletta et al., 2004), transcribed from σ^{32} and σ^{38} in *E. coli* (Marques et al., 1999). The promoter has been subjected to random mutagenesis in the vector, which strongly increased expression of three reporter genes and doubled HCDC production of GM-CSF under industry-simulated conditions (Bakke et al., 2009). The XylS protein is constitutively produced from its native promoter in low numbers, and is induced to become a transcriptional activator for the *Pm* promoter by several cheap and harmless benzoate derivatives, such as m-Toluic acid (Ramos et al., 1997; Sletta et al., 2004).

The expression vector system also includes the *hok/sok* system, transcriptional terminators in front of the promoter (*rrnBT1T2*) and after the MCS (*tLPP*) and *TrfA* (*t*), an antibiotic resistance marker, origin of transfer for conjugative replication (*oriT*), and an M13 origin of replication (M13 *ori*) which can yield high levels of single stranded DNA upon infection with the M13 phage (Zagursky and Berman, 1984). The system is highly versatile, and has been

used successfully in industrial scale protein production in *E. coli* (Sletta et al., 2004; Bakke et al., 2009). For research on the regulatory mechanisms governing its expression, several designs are available where features such as the promoter or UTR may be exchanged by DNA cassette exchange.

Expression vectors are excellent tools for expressing genes in *E. coli*. However, even with the optimized features of expression vectors, protein expression often encounters problems that require further strategies to be solved. The problems include features of the gene sequences and the UTR, or drawbacks *E. coli* itself as a production host, and will be discussed in the following chapter.

1.3 Strategies for Effective Heterologous Expression

Due to the intricate mechanisms regulating gene expression discussed in 1.1, many genes may require optimization before they can be efficiently produced in *E. coli*. Problems also arise when expressing eukaryotic proteins, as they often require post-translational modifications such as disulfide bridges or proteolytic cleavage, and they are often produced in insoluble inclusion bodies (Jana and Deb, 2005; Kamionka, 2011). Strategies for optimizing expression and overcoming these problems will be covered in the following chapters.

1.3.1 Modification of the TIR

The rate limiting step in translation is translation initiation, as discussed in chapter 1.1.2.1. Its efficiency is mostly determined by the secondary structure and sequence features of the TIR, which starts in the 5' UTR and extends into the 5' coding region. The sequence of the UTR may be changed for fine-tuning of translational efficiency by decreasing the stability of secondary structures and varying the sequence of the SD (Park et al., 2007). By varying the stability of a hairpin structure in the RBS, de Smit *et. al* could vary the relative expression of a target protein from <0.003 % at higher stability, to 100 % at lower stability (deSmit and Vanduin, 1990).

The 5' coding region also contributes to secondary structure formation and should be taken into consideration along with the UTR. As GC bonds are stronger than AT bonds (Zuker and Stiegler, 1981), increasing the AT content of the TIR may lower the stability of secondary structures forming and increase translational efficiency (Nishikubo et al., 2005; Krishna Rao et al., 2008). Given the degeneracy of the genetic code (Nirenberg et al., 1966), i.e. one amino-acid may be coded for by one- up to six different synonymous RNA codons (Gustafsson et al., 2004), it is possible to change the 5' coding region to alter the transcript's propensity for forming stable secondary structures in the TIR without changing the amino acid sequence of the translated protein.

A synthetic library was constructed in a study by Kudla *et. al.* with 154 synonymous variants of green fluorescent protein (GFP) that gave 250-fold variations in GFP levels (Kudla et al., 2009). Local and global RNA structures of the genes were predicted and compared to the fluorescence levels, and stable local structures forming in the -4 to +37 nucleotide window explained more than half of the variation in the dataset. In a different experiment, a

randomized synonymous mutation library was created in an artificial 10 amino acid linker in front of GFP, giving ~300-fold increase in fluorescence compared to the wild-type linker-GFP construct (Goltermann et al., 2011).

Optimization of the codon following the translational start site (+2 codon) has also been suggested as a technique for improving protein expression (Stenstrom et al., 2001; Tang et al., 2011). By exchanging the codon following the initiation codon AUG (the +2 codon), a 20 fold effect on gene expression could be obtained (Stenstrom et al., 2001), and a correlation between high adenine content of the +2 codon with high expression was observed. The effects seen were not correlated to change in secondary structures forming around the RBS, mRNA stability, or tRNA abundance. Most synonymous codons decoded by the same tRNA showed similar effects, indicating tRNA involvement in the effect on gene expression. Stenstrøm *et. al.* also found that the codons they identified as giving high expression were over-represented as +2 codons in natural genes with high expression, and vice versa.

Placing NGG (N is not G) as the +2 codon in transcripts has been shown to drastically reduce expression (Gonzalez de Valdivia and Isaksson, 2004), also in a manner not apparently affected by mRNA secondary structure, but instead a specific NGG codon effect. Replacing GUG with less G-containing +2 codons also increased expression (Tang et al., 2011). However, here it was correlated to a reduction in secondary structure stability.

Clearly modification of the UTR and the 5' coding region of transcripts may be used to enhance translational efficiency, both by altering the SD and the stability of the secondary structures. However, random mutagenesis of the UTR of the XylS/*Pm* system interestingly produced UTRs giving up to 20-fold increases in both transcript and protein levels without affecting mRNA stability (Berg et al., 2009), indicating that the DNA corresponding to the UTR also has a role in transcription as well as translation. This may be explained by changes in promoter escape efficiency by changes in the initially transcribed region (Hsu et al., 2006). The effect of the UTRs were shown to be gene specific as transcription of two other genes using the UTR were not as highly up-regulated, suggesting a dependency of the UTR sequence together with the 5' coding sequence (Berg et al., 2012).

1.3.1 Codon Optimization and tRNA Pool

Expressing mammalian proteins in *E. coli* often proves difficult, even with high amounts of transcript and an effective TIR, as the source organism's genes are adapted to a different tRNA pool (Gustafsson et al., 2004), and codons rare to the host may lead to growth arrest, premature translational termination and mistranslated proteins (Huang et al., 2012). One strategy for overcoming this problem is to modify the host's tRNA pools by overexpressing genes coding for the rare tRNAs, especially $tRNA_4^{Arg}$ reading AGG and AGA, $tRNA_2^{Ile}$ reading CUA and CUG, and $tRNA_2^{Pro}$ reading CCC and CCU as these are codons that reduce efficiency (Kane, 1995). However, this may overload the tRNA amino-acylation machinery (Wahab et al., 1993) along with other problems that lead to low expression (Gustafsson et al., 2004).

A simple, more common solution now that the cost of DNA synthesis is low, is to synthesize a codon optimized gene of interest where codons rare to the expression host are exchanged for more common ones following the CAI (Angov, 2011). This has been successful in many cases, especially with human proteins produced in *E. coli* (Gustafsson et al., 2004; Sletta et al., 2007). One study reported an increase in human phosphatidylcholine transfer protein from trace levels, using the native cDNA of the gene, to over 10 % of cytosolic protein, using an optimized version of the gene synthesized using a recursive PCR strategy (Feng et al., 2000).

Despite several successful stories involving CAI optimization of genes for production in *E. coli*, a high CAI of the gene does not universally correlate with high expression (Kudla et al., 2009; Welch et al., 2009). As in (Kudla et al., 2009), the study by Welch *et. al.* found no correlation to CAI when two genes were randomized so that common codons would appear more often in their synonymous codons, and features of the coding sequence were analyzed with respect to expression levels. The 40-fold difference in expression was instead explained by the synonymous codon choice of a subset of highly represented amino acids in the proteins, and it was suggested that the sensitivity of the respective tRNAs to amino-acid starvation influences the expression, as amino-acylation of tRNAs may become the limiting factor during protein overexpression (Welch et al., 2009).

Tuller *et. al.* describe a "ramp" of slowly translated codons with low tAI in the early coding region of genes as described in 1.1.2.2, and suggest this feature may be used in the design of genes for heterologous expression. This might reduce the burden imposed on cells by

heterologous protein production and increase the fitness of the cell, as it could increase the rate of successful translations by evenly spreading the ribosomes on the transcript (Tuller et al., 2010).

1.3.2 *E. coli* as a Production Host

Achieving high expression efficiency of a protein of interest in *E. coli* doesn't necessarily mean it is produced in high quality. Some major obstacles when expressing mammalian proteins in *E. coli* are its lack of post-translational modification systems and its inability to form disulfide bonds (Jana and Deb, 2005). Mammalian proteins may fold incorrectly, form inclusion bodies (IBs), or have lowered or no activity without modifications such as glycosylation, proteolytic maturation, or the formation of disulfide bridges in the peptide (Huang et al., 2012). The cytoplasm is a too reductive environment for the formation of disulfide bridges (Mergulhão et al., 2005), and enzymes for glycosylation and correct proteolysis aren't native to *E. Coli* (Kamionka, 2011). Many therapeutic proteins are therefore produced in mammalian cells, losing the ease-, and low cost of cultivation that comes with using *E. coli.*.

One strategy to overcome the reducing environment of the cytoplasm involves translocation of the peptide to the periplasmic space (Mergulhão et al., 2005). By fusing a translocation signal peptide, such as those from PelB (Hauser and Ryan, 2007), OmpA (Pines and Inouye, 1999), or the Consensus Signal Peptide (CSP), designed based on sequence alignments of several signal peptides (Sletta et al., 2004), to the 5' end of the protein, it is targeted to the periplasmic space between *E. coli*'s two membranes. Here the environment is less reducing and contains enzymes that help form disulfide bonds (Mergulhão et al., 2005). The signal peptide is cleaved off following translocation, which also removes the problem of the sometimes present initiation amino acid methionine (Mergulhão et al., 2005). In addition to the folding-wise more favorable environment, periplasmic expression allows for easier purification of the protein as the variety and amount of proteins in the periplasm is considerably lower than the bacterial cytoplasm (Sandkvist and Bagdasarian, 1996).

Strategies for the improvement of cytoplasmic expression have also been developed, for example for natively glycosylated proteins which may form aggregates leading to lowered activity and need for higher dosages (Kamionka, 2011).. The neutral asparagine N-glycosylation target residues in human erythropoietin were changed to lysine residues,

reducing its rate of aggregation without affecting its activity (Narhi et al., 2001). To aid in the correct folding of proteins native *E. coli* chaperone proteins may be overexpressed along with the protein of interest, a technique which has yielded higher solubility and increased production of heterologous proteins (Kolaj et al., 2009).

Human insulin is an example of a mammalian protein that needs proteolytic cleavage for activation (Son et al., 2008). It is produced natively as preproinsulin, with a signal peptide that is cleaved off, followed by proteolytic cleavage into two chains which are bonded together by two disulfide bonds (Kamionka, 2011). *E. coli* is unable to perform the proteolytic cleavage, and disulfide bonds cannot form in the reducing environment of its cytoplasm. It is produced as inclusion bodies which are resolubilized, denatured and reduced before *in vitro* enzymatic treatment for cleavage of the peptide, renaturation and disulfide bond formation (Son et al., 2008). Like insulin, the production of many commercial proteins utilizes inclusion bodies as an advantage, as it offers high yields and protection against proteases (Kamionka, 2011; Huang et al., 2012).

1.3.3 Fusion Peptides

In addition to the beneficial effects discussed above, translocation of heterologous proteins in *E. coli* may serve to increase translational efficiency (Sletta et al., 2007). The three medically important proteins granulocyte-macrophage colony-stimulating factor (GM-CSF), interferon alpha 2b (IFN-α2b), and single-chain antibody variable fragment (scFv-phOx) were expressed in the cytoplasm of *E. coli*, as they don't contain disulfide bridges, using the XylS/*Pm* vector system described above. The genes were initially poorly expressed, but when various translocation signal peptides were fused N-terminal to the GM-CSF and scFv-phOx proteins, industrial expression levels were achieved. The improvement could also be shown for a codon optimized *ifn-α2b*.

Quantitative real-time PCR analysis of the GM-CSF constructs showed only a ~three-fold increase in transcript levels for pelB and CSP constructs, suggesting that the increase in translational efficiency exceeds that of the increase in transcription (Sletta et al., 2007). A decrease in folding energy was observed for all genes when fused to signal peptides, which might explain the increased translation. However, the codon optimized IFN-α2b already had low secondary structure stability before fusion. It was also noted that the +2 codon was changed to AAA for all signal peptides, which could influence the expression as previously

discussed, but this has been shown to not be universally beneficial (Sletta et al., 2007). The authors hypothesize that it is the translocation process itself that affects the translation, as the two mechanisms are closely coupled.

However, recent studies by the Microbiology and Molecular Genetics (MOLGEN) research group at NTNU have shown similar increases in both *gm-csf* and the codon optimized *ifnα2b* gene expression when using a 5' terminal fusion of a short *celB* gene fragment (Kucharova, unpublished). As the *celB* gene is coding for cytoplasmic phosphoglucomutase enzyme, it indicates that translocation is not necessary for the increased expression effect from protein fusions. Given the high expression efficiency of *celB*, it may hold a 5' coding sequence which forms a particularly good TIR together with the *Pm* and UTR of the pJBn vectors. Another study showed that fusing the first 21 nucleotides of the *infB* gene encoding translation initiation factor II 5' to GFP increased its expression five-fold (Hansted et al., 2011). The small ubiquitin-related modifier (SUMO) tag is a popular expression enhancer for poorly expressed proteins that serves togreatly enhance expression, while also increasing the solubility of the produced protein (Satakarni and Curtis, 2011). The SUMO fusion may effectively be cleaved off by SUMO specific proteases post-production.

Another use for protein fusions are as affinity tags which may be utilized in the purification of proteins (Terpe, 2003). By fusing such a tag to the protein of interest, it may be adsorbed to columns with affinity for the tag. The tags are commonly small, so as not to interfere with the protein's tertiary structure, and some are removable by enzymatic treatment (Terpe, 2003). Tags also allow simple visualization on Western blots as described in chapter 2.2. If an antibody is not available for the protein of interest, an antibody against the affinity tag may be used.

A common affinity tag is the poly-histidine tag (his-tag) (Hochuli et al., 1987), which is a tag of normally six histidine residues that may be fused to the C- or N-terminal of the protein of interest (Terpe, 2003). The histidine residues bind to transition metal ions that are immobilized on a matrix, for example Ni^{2+}, and may be eluted by imidazole. The his-tag is believed to minimally affect proteins due to its small size but it has been shown to reinstate the dimerization of a monomeric π protein(Wu and Filutowicz, 1999), proving that structure/function may be affected.

1.3.4 Combinatorial Mutagenesis

As is evident from the subjects discussed so far, many different factors in the DNA and the transcript affect protein expression to different extents, and it is hard to predict the outcome of certain designs on a given protein in a given expression vector. Rationally incorporating all the different factors to optimize the design of transcripts would be difficult, and the testing of such designed constructs would be time-consuming (Kim et al., 2008). By utilizing randomized libraries and good screening techniques, large amounts of different sequence variants can be tested, and up-mutants that may be compromises of all the factors affecting gene expression may be successfully selected (Kim et al., 2008; Bakke et al., 2009; Goltermann et al., 2011). A directed evolution approach where positive features identified in one round are kept in a new mutagenesis could lead to even better results (Vee Aune et al., 2010). A good screening technique and quantitative measurement of mutational effects are greatly aided by the use of reporter genes, which are described in the following chapter.

1.4 Recombinant Proteins Used in the Expression Studies

Reporter genes are valuable tools when studying the molecular events associated with changes in gene expression. These are genes that, when introduced in a biological system, produce measurable phenotypes (Wood, 1995), which may illustrate the effect of the changes made to the system when compared to the wild-type. A well known example is the *lacZ* gene encoding β-galactosidase, which hydrolyzes ortho-Nitrophenyl-β-galactoside (ONPG) to a yellow product, allowing colorimetric measurement of enzyme amount and activity (James et al., 1996). Another widely popular reporter gene is *gfp*, encoding Green Fluorescent Protein, which can be measured fluorometrically without the need to lyse cells or provide additional substrates (Ghim et al., 2010). The reporter genes may be expressed in their native form, for example to measure promoter strength (Bakke et al., 2009), or mutated to study effects of for example synonymous mutations (Kudla et al., 2009).

In this chapter the reporter genes utilized in this study will be presented along with the medically important human protein IFN-α2b, which was used for demonstration of successful heterologous protein expression.

1.4.1 β-lactamase

β-lactamase encoded by the *bla* gene breaks down the β-lactam ring, and thus confers resistance to β-lactam containing antibiotics, among others ampicillin (Bakke et al., 2009). The ability to grow on increasing concentrations of ampicillin has been shown to correlate well to the amount of β-lactamase produced by *E. coli* (Winther-Larsen et al., 2000; Bakke et al., 2009; Berg et al., 2012). It is therefore a well suited reporter protein in screening for increased expression *in vivo*, as high growing mutants are likely to produce high amounts of protein (Bakke et al., 2009; Berg et al., 2009). As the β-lactam ring of penicillin G absorbs light at 240nm, the relative amounts of β-lactamase in cell extract may be measured by observing the degradation of penicillin as the decrease in absorbance at 240nm (Fukagawa et al., 1980). The *bla* gene was used in this study to investigate changes in the synonymous codons of its 5' coding region, by screening for up-mutants on increasing ampicillin concentrations. The gene was also used in this study to investigate the effects of 2nd codon synonymous mutations on gene expression.

1.4.2 Phosphoglucomutase

The phosphoglucomutase protein, encoded by *celB,* originates from *Acetobacter xylinum* (Brautaset et al., 1994) and was previously produced successfully at high levels using the XylS/Pm expression vector (Blatny et al., 1997). The intracellular protein converts glucose-1-phosphate to glucose-6-phosphate (Najjar, 1948), and has been proven useful as a reporter protein as a simple enzyme assay has been developed to measure its activity (Valla et al., 1989). The use of short 5' terminal *celB* fragments as gene fusions has been shown to dramatically increase expression of poorly expressed genes, as discussed previously (Kucharova, unpublished). The 5' *celB* gene fragment of 69 nucleotides (corresponding to 23 amino acids) was chosen to be used as a basis for constructing a genetic library of optimized 5' terminal fusion partners. The *celB* gene was also used in this study to investigate the effects of 2nd codon synonymous mutations on gene expression.

1.4.3 Alkaline Phosphatase

Alkaline phosphatase, encoded by the *phoA* gene, is expressed in a wide range of organisms and has the ability to dephosphorylate a broad range of substrates (Jiang et al., 2008), among others *p*-nitrophenyl phosphate (*p*-NPP). *p*-NPP is degraded to the colored product *p*-nitrophenol which can be colorimetrically measured at 405 nm (Akcakaya et al., 2007). The *phoA* gene was used in this study to investigate the effects of 2nd codon synonymous mutations on gene expression.

1.4.4 Aminoglycoside-(3)-acetyltransferase IV

The *aac(3)-IV* gene encodes an aminoglycoside acetyltransferase that acetylates the aminoglycoside antibiotic apramycin, thus hindering its ability to block translational elongation (Han et al., 2005). In this study it was used as a reporter gene for screening of a combinatorial genetic library of the $celB_{69}$ based 5' terminal fusion partners. The mutant library was cloned in frame with the 5' end of the *aac(3)-IV* gene, allowing screening of the library with respect to increased host apramycin tolerance.

1.4.5 Interferon α-2b

IFN-α2b is a human cytokine used for the treatment of viruses and certain cancers (Srivastava et al., 2005), and it is encoded by the *ifn-α2b* gene. The protein has proven difficult to express in *E. coli*, nevertheless high level production was achieved by codon optimization and N-terminal translocation peptide fusion (Sletta et al., 2007).

1.5 Aims of this Study

Recombinant protein expression in *E. coli* is of high-value, as important therapeutic proteins may be produced at low cost and high quality. However, as explained in the previous chapters, production is not always straight-forward as the expression efficiency varies from protein to protein due to intrinsic factors of its gene sequence, and thus more research is needed to be able to optimize any given gene. As the main effectors on protein expression lies in the translation initiation region, the 5' coding sequence of the transcript is an interesting target for optimization. It is part of the TIR and may affect secondary structures forming around the RBS and the translational start-site. It may also provide slowly- or rapidly translated codons, affecting translational elongation and ribosome density on the transcript. Recent findings also suggest that the UTR and the 5' coding region together play a role in transcriptional regulation.

The main focus of this study has been to investigate and optimize the 5' coding region of transcripts in an effort to obtain knowledge for enhanced heterologous protein expression, which was addressed by three different strategies. A directed evolution approach was applied to results obtained from a previous synonymous mutation library, where a 2nd codon change in the *bla* translocation signal sequence led to increased β-lactamase expression. To further investigate the cause of this effect and to achieve higher expression, a new library was created containing the identified 2nd codon mutant. Another part of the project was to further investigate the specific effect achieved by changing only the 2nd codon, and the differential expression due to 2nd codon synonymous change from the reporter genes *bla*, *phoA* and *celB* were examined.

Both the aforementioned strategies involved synonymous mutations of the 5' coding region, which is preferable in optimization as it does not lead to changes in the amino-acid sequence of the protein. However, some difficult-to-express proteins such as interferon α-2b may

require protein fusions to be expressed (Sletta et al., 2007), and recent research has shown that IFN-α2b may be expressed by fusion of the first 69 nucleotides of the *celB* gene 5' to *ifn-α2b*. In an effort to explain the reasons for this and to further enhance the positive features of the fusion, a randomized $celB_{69}$ gene fusion library was created and screened for increased expression by use of the apramycin resistance reporter gene *aac(3)-IV*.

2 Materials and Methods

This chapter first gives an overview of the various media and buffers used in this study in chapter 2.1, followed by protocols for the experimental techniques in 2.2. Chapter 2.3 contains information on the various plasmids and oligonucleotides used, and chapter 2.4 gives a short description of various bioinformatics tools that were used to analyze the results obtained.

2.1 Growth Media and Buffers

Lysogeny Broth (LB medium)

Bacto™Tryptone (Becton, Dickinson& Co.,Sparks, USA)	10 g/L
Yeast Extract (Oxoid, Basingstoke, Hampshire, England)	5 g/L
NaCl	5 g/L

Components were dissolved in MQ-water. pH was adjusted to 7.4 using NaOH, before autoclaving 20 minutes at 121 °C.

LB Agar (LA)

LB medium	
DifcoTM Agar Noble (Becton, Dickinson and Company, Sparks, USA)	15 g/L

LB components were dissolved in MQ-water before adding agar. pH was adjusted to 7.4 using NaOH before autoclaving 20 minutes at 121 °C. The agar was cooled to approximately 50 °C before addition of appropriate selective antibiotic and casting of plates

Psi medium

Bacto™Tryptone (Becton, Dickinson& Co.,Sparks, USA)	20 g/L
Yeast Extract (Oxoid, Basingstoke, Hampshire, England)	5 g/L
$MgSO_4$	5 g/L

Components were dissolved in MQ-water. pH was adjusted to 7.6 using KOH, before autoclaving 20 minutes at 121 °C.

Hi + YE reduced broth

Hi+YE broth is a rich media used for increased growth and expression, and has many nutritional additives for this purpose. It considers different nutritional needs before and after induction, and has a basic broth and an induction broth.

The basic broth is made by mixing three different solutions, A, B and C.

Solution A, concentrations denote final concentration of induced medium:

$Na_2HPO_4*2H_2O$		8.60 g/L
KH_2PO_4		3.00 g/L
NH_4Cl		0.30 g/L
NaCl		0.50 g/ L
Fe(III) citrate hydrate	(3.30 mL from 6 g/L stock solution)	19.8 mg/mL
H_3BO_3	(0.03 mL from 30 g/L stock solution)	0.9 mg/mL
$MnCl_2*2\ H_2O$	(0.50 mL from 10 g/L stock solution)	5.00 mg/mL
$EDTA*2H_2O$	(0.03 mL from 84 g/L stock solution)	2.52 mg/mL
$CuCl_2*2H_2O$	(0.03 mL from 15 g/L stock solution)	0.45 mg/mL
$Na_2Mo_4O_4*2H_2O$	(0.03 mL from 25 g/L stock solution)	0.75 mg/mL
$CoCl_2*6H_2O$	(0.03 mL from 25 g/L stock solution)	0.75 mg/mL
$Zn(CH_3COO)_2*2H_20$	(0.67 mL from 4 g/L stock solution)	2.68 mg/mL
Yeast extract (Oxoid, Basingstoke, Hampshire, England)		2.68 g/mL
Glucose*H2O		2.00 g/L
MQ-water		650 g

Solution B, concentrations denote final concentration of induced medium:

Citric acid	1 g/L

The pH was adjusted to approximately 7 with NaOH.

Solution C:

$MgSO_4*7H_2O$	24.6 g

MQ-water	100 g

The solution was autoclaved at 120 °C for 20 minutes.

Solutions A and B were mixed, and 0.9 mL solution C was added before sterile filtration. Appropriate antibiotics were added before use.

Hi+YE Induction solution

Glycerol (100 %)	12.75 g
Yeast extract (Oxoid, Basingstoke, Hampshire, England)	12 g
Tap water	to 500 mL

The solution was sterile filtered, and appropriate antibiotics were added before use.

Kanamycin stock

Kanamycin sulphate	50 mg/mL

The antibiotic salt was dissolved in MQ water, sterile filtered, aliquoted and stored in -20 °C.

Ampicillin stock solution

Ampicillin sodium salt	100 mg/mL

The antibiotic salt was dissolved in MQ water and sterile filtered before use.

Transformation buffer I (TFB I)

Potassium acetate (CH_3CO_2K)	0.03 M
RbCl	0.1 M
$CaCl_2$	0.0133 M
$MnCl_2$	0.08 M
Glycerol	15 % (v/v)

Components were dissolved in MQ-water. pH was adjusted to 5.8 using acetic acid, before sterile filtration.

Transformation buffer II (TFB II)

MOPS	10 mM
$CaCl_2$	0.1 M
RbCl	10 mM
Glycerol	15 % (v/v)

Components were dissolved in MQ-water. pH was adjusted to 6.5 using NaOH, before sterile filtration.

β-lactamase assay buffer

KH_2PO_4	80 mM
NaH_2PO_4	20 mM

Components were dissolved in MQ-water. pH was adjusted to 6.5 using NaOH, before sterile filtration.

Alkaline phosphatase assay buffer

Tris-HCl	100 mM
$CaCl_2$	5 mM
Tween 20	0.1 % (v/v)

Components were dissolved in MQ-water. pH was adjusted to 9.5 using NaOH

Phosphoglucomutase assay buffer

Immidazole	0.1 M

Components were dissolved in MQ-water. pH was adjusted to 7.4 using HCl.

SDS-PAGE 10x running buffer

Tris-HCl	250 mM
Glycine	2 M
Sodium dodecyl sulphate (SDS)	1 % (w/v)

Components were dissolved in MQ-water. pH was adjusted to 8.3 using concentrated HCl.

3x sample buffer

Tris, pH 6.8	150 mM
Glycerol	30 % (v/v)
SDS	6% (w/v)
Bromophenolblue	0.3 % (w/v)
Dithiothreitol (DTT)	300 mM

Components were dissolved in MQ-water.

Blotting buffer

Tris	25 mM
Glycine	192 mM
Methanol	20 % (v/v)
SDS	0.05 % (w/v)

Components were dissolved in MQ-water, and the buffer was stored in 4 °C.

Tris-buffered saline (TBS)

Tris-HCl, pH 7.5	20 mM
NaCl	150 mM

Components were dissolved in MQ-water.

TBST

Tween 20	0.05 % (v/v)

Tween 20 was added to TBS.

Blotto buffer

Bovine serum albumin (BSA)	3 % (w/v)

BSA was dissolved in TBS and kept at 4 °C.

2.2 Protocols

Rubidium chloride competent cells

Treatment of bacterial cells with divalent cations such as Ca^{2+} and Mn^{2+} in addition to other cations such as Rb^{+} weakens the cell membrane and induces binding of exogenous DNA to the membrane. (Weston et al., 1981; Aune and Aachmann, 2010). A subsequent heat-shock followed by cold incubation allows the DNA to permeate into the cell.

E. coli DH5α cells were inoculated ON in 10mL Psi-broth at 37 °C, 225 rpm. 1% of the ON culture was inoculated in pre-warmed Psi-broth and grown to OD_{600}=0.4 (measured with Lambda 35 UV/VIS Spectrometer) at 37 °C, 225 rpm, followed by incubation on ice for 15 minutes. The cell culture was centrifuged (4500 rpm, 5 minutes) (Sorvall RC5C) and the pellet was resuspended in 40 mL cold TFB I before incubation on ice for 15 minutes followed by centrifugation (4500 rpm, 5 minutes). The pellet was resuspended in 3 mL TFB II and the cell solution aliquoted (100µL), snap-frozen using dry-ice chilled ethanol and stored in -80 °C.

Transformation

Competent cells were thawed on ice, added to ligation mix / plasmid DNA, briefly vortexed and incubated on ice for 30 minutes. The cells were heat-shocked in a water-bath holding 42 °C for 37 seconds before addition of 1 mL LB and incubation on ice for 2 minutes. Cells were centrifuged (5000 rpm, 2 minutes) and the pellet resuspended in approximately 100µL of the supernatant before plating on L-agar plates with selective antibiotic followed by incubation at 37 °C.

Isolation of plasmid DNA

Plasmid DNA was isolated using the Wizard® *Plus SV* Miniprep DNA Purification System (Promega). Single colonies were picked from freshly streaked plates and incubated ON for 16-18 hours at 30 °C, 225 rpm in 15 mL LB with selective antibiotic in 125 mL Erlenmeyer flasks for high plasmid yield, or 6 mL LB with selective antibiotic in 13 mL tubes. Cells were centrifuged (10 000 g, 10 min), and the pellet was resuspended in 300 µL Resuspension

Solution and transferred to an Eppendorf tube. 400 µL Cell Lysis solution was added, along with 10 µL Alkaline Protease solution and tubes were inverted 4 times before incubation at room temperature for five minutes. 400µL Neutralization Solution was added, tubes inverted four times before centrifugation at 13 000 rpm for 10 minutes. Supernatant was decanted into spin columns, inserted into collection tubes and centrifuged at 13 000 rpm for 90 seconds. Columns were washed twice with 750 µL and 250 µL wash solution, at 13 000 rpm for 90 seconds. Plasmid DNA was eluted into Eppendorf tubes using 100 µL MQ-water, the concentration was measured using a NanoDrop 1000 spectrophotometer (Thermo Scientific), and samples were stored in -20 °C.

Restriction digests

Restriction endonucleases are part of archaeal and bacterial defense against viral DNA, and cleave DNA in specific recognition sequences. Restriction enzymes used in cloning usually recognize palindromic DNA, and cut it leaving "sticky end" overhangs of bases so that DNA digested with the same enzyme have matching ends and may be ligated together (Roberts, 2005).

All restriction digests were performed using restriction endonucleases purchased from New England Biolabs, using the recommended conditions and buffers obtained with the Double Digest Finder Tool for enzymes A and B, and the following reaction mixture.

Plasmid DNA (approximately 1500ng)	20-30 µL
NEB buffer specified by Double Digest Finder	10 µL
Restriction endonuclease A	2.5 µL
Restriction endonuclease B	2.5 µL
BSA (if recommended by Double Digest Finder)	1 µL
MQ-water to final volume 100µL	

The reaction mixtures were incubated at 37 °C over night before adding 1 µL of each enzyme in the morning followed by 1 hour incubation at 37 °C. 3 µL calf intestinal phosphatise (CIP) was added to remove 5' phosphates from the backbone to reduce religation (Wu et al., 2008), and the mixture was incubated for another 1,5 hrs. The cut DNA backbone was purified using the QIA quick PCR Purification Kit (Qiagen) and eluted with 100 µL MQ-water.

Ligation reactions

The "sticky" overhangs of DNA cut with the same restriction enzymes are attracted to each other, and may be joined together by ligases. DNA ligases catalyze the phosphodiester bond formation between the 5'-phosphoryl and 3'-hydroxyl termini of DNA, thus ligating the two ends together (Brown and Ray, 1992). The T4 DNA ligase is commercially available and often used in genetic recombination *in vitro* for cloning purposes.

All ligation reactions were performed using T4 DNA Ligase from New England Biolabs, along with its buffer, and in the following reaction mixture.

Purified cut plasmid backbone DNA	16 µL
T4 DNA ligase buffer	2 µL
T4 DNA ligase (10x diluted in Diluent A (NEB))	1 µL
DNA insert (diluted annealed oligos)	1 µL

Ligation efficiency and religation were evaluated by including negative ligation mixes without DNA inserts. Ligation reaction mixtures were incubated over night at 4 °C before transformation into *E. coli* competent cells.

Glycerol stocks

10 mL LB containing appropriate antibiotic was inoculated with a single colony from freshly transformed sequence confirmed plasmids, and incubated at 37 °°C, 225 rpm for 20 hours. 400 µL 60 % glycerol and 800 µL cell culture were mixed in an Eppendorf tube, and immediately frozen at -80 °C.

Annealing of synthetic oligonucleotides

Synthetic DNA oligonucleotides as described in **Table 2-2** were ordered from Sigma-Aldrich®, resuspended to 100 µM. The oligonucleotides were treated with T4 polynucleotide kinase (PNK) to add 5' phosphate of the oligonucleotides, to increase ligation efficiency (Rittie and Perbal, 2008). The following reaction mixture was incubated in 37 °C for 30 minutes.

Forward oligo	7µL

Reverse oligo	7 µL
T4 DNA ligase buffer	1.5 µL
T4 Polynucleotide Kinase	0.85 µL

T4 PNK was inactivated by incubation at 65 °C for 30 minutes, before the oligo mixes were annealed using an Eppendorf Mastercycler® with the "ANNEAL" program, showed in appendix Appendix A.

Construction of *bla* 5' coding sequence synonymous library, SII

A new wild-type for the library containing the TCT 2nd codon, pBS_2P1, was constructed by ligation of the annealed product of Bla.TCT.fwd and Bla.TCT.rev, shown in **Table 2-2**, into *NdeI*, *NcoI* digested pBSP1 backbone. The randomized oligonucleotide Bla-TC-5'spiked.fwd was annealed to its wild-type oligonucleotide Bla.TCT.rev, ligated into *NdeI*, *NcoI* digested $pBS_2P1$and used to transform competent cells as described above. Transformants were plated on 15 cm L-agar plates, and in dilution series to count the number of clones. The library was harvested by adding 1 mL LB to the centre of the plates before homogenizing the colonies in the LB using a sterile rod. The cell medium was extracted from the plates with a pipette, and half the volume collected of 60 % glycerol was added before mixing. Five 1 mL samples were distributed in Cryo tubes for long time storage, and the remaining library was aliquoted in Eppendorf tubes, 500 µL each before freezing samples at -80 °C. All work was performed in a sterile fume hood, and samples were kept on ice at all times.

Construction of a randomized library of 5' $celB_{69}$ based fusion partners

A randomized library was designed in the $celB_{69}$ sequence fused to the apramycin resistance reporter gene *aac(3)-IV*, encoding aminoglycoside-(3)-acetyltransferase IV resitance, to screen for mutants with higher tolerance of the antibiotic. The fp_{69}-*aac(3)-IV* library was created by ligation into *NdeI*, *NcoI* digested pAR69 (**Table 2-1**) of the annealed insert of the oligos celB69-mut.Fwd and celB69-wt.Rev shown in **Table 2-2**. The annealed library was then used to transform competent cells as described above, and plated and harvested as described for the *bla* 5' coding sequence synonymous library.

Screening libraries for increased antibiotic tolerance

To search for mutants with increased antibiotic resistance created in the libraries, the libraries were screened in the following way. L-Agar plates (15 cm) were prepared with increasing concentration of the antibiotic in question, and the appropriate amount of inducer. Control plates were prepared without inducer for the *bla* library, and with only the resistance marker antibiotic for both. 500 µL library from -80 °C stock was inoculated in 500 mL LB with kanamycin, and incubated for 2 hours at 30 °C, 225 rpm. OD600 was measured to be between 0.7 and 1.5, before 300 µL 1:1000 diluted culture was applied and distributed on the plates and incubated at 30 °C for 2 days. High growing colonies were picked and their plasmids were isolated, as described above, for further analysis.

Replica plating to determine antibiotic concentration tolerance

To determine tolerance levels of the antibiotic concentration from the screening candidates, plasmids isolated from the library screening were retransformed as described, and single colonies were inoculated in 100 µL LB with appropriate selection antibiotic in 96-well plates (Nunc) and incubated at 30 °C overnight. The cultures were diluted to approximately 1:10 000 by twice transferring culture with a 96-pin replicator to 100 µL fresh LB with selection antibiotic in 96 well plates, before using the 96-pin replicator to apply the cells to agar plates with increasing concentrations of the screening antibiotic, with and without inducer. Plates were incubated at 30 °C for two days, the highest growths of the candidates were recorded and their gene variants were sequenced.

Gene sequencing

All gene sequencing was performed by Eurofins MWG Operon, using the PmUTR.fwd sequencing primer shown in **Table 2-4**. The primer targets the Pm promoter in all used plasmids, starting 97 base-pairs upstream of the translational start-site of the genes of interest.

Cell culturing for enzyme assays and accumulated transcript evaluation

Cells harboring plasmids containing the enzyme to be studied were inoculated in 10 mL LB + appropriate antibiotic in 125 mL Erlenmeyer flasks from fresh overnight streaks from glycerol

stock and grown overnight at 30 °C and 225 rpm. Cultures were inoculated from the overnight culture to a final OD of 0.05 in 20 mL LB + appropriate antibiotic, and incubated in a water bath at 30 °C, 225 rpm. Cultures were induced to the appropriate final concentration (0.05 mM or 0.1 mM) with 200 mM m-Toluate at OD 0.1, and grown for 5 hours.

Samples for enzyme assays were prepared harvesting 5 mL culture which was centrifuged at 6 000 rpm for 10 minutes. The pellet was resuspended in assay buffer followed by another centrifugation before storing the pellet at -80 °C. Samples for mRNA studies were prepared by mixing 0.5 mL culture with 1.0 mL RNAprotect Cell Reagent from QUIAGEN followed by thorough vortexing and incubation for five minutes in room temperature. The samples were pelleted by centrifugation at 5000 rpm for 10 minutes at 4 °C, the supernatant was discarded and the pellets were frozen at -80 °C.

Expression of recombinant proteins in low-scale

Single colonies from overnight streaks from glycerol stocks were inoculated in 14 mL Hi+YE basic broth containing appropriate antibiotic in 250 mL baffled Erlenmeyer flasks and grown overnight at 30 °C, 225 rpm. OD600 was measured to inspect growth before adding 6mL Hi+YE broth containing appropriate antibiotics and m-Toluate to a final concentration of 0.5 mM. Cultures were incubated at 30 °°C, 225 rpm for 3 hours before harvest as described below for SDS-PAGE and Western Blot. Transcript samples were also taken from the cultures using RNAprotect as described above.

Preparation of cell extract by sonication

Sonication lysis of cells utilizes the cavitations created by localized areas of high- and low pressure created by ultrasonic waves (Brown and Audet, 2008).

Sonication of samples for enzyme assays was performed with the samples kept on ice, using the tapered microtip of a Branson Sonifier 250 for 3 minutes at 30 % duty control and output control 3 for 3 minutes. Samples for Western blots were sonicated 90 seconds at 90 second intervals five times.

β-lactamase assay

Cell pellets were resuspended in 5 mL β-lactamase assay buffer, sonicated as described, and centrifuged for 20 minutes at 8 000 rpm and 4 °C. Samples were kept on ice and total protein concentration of the supernatant was determined using the Bradford method. Reaction kinetics of the supernatant was measured in 5 duplicates for each sample in a 96-well UV Transparent Nunc plate with the following reaction mixture:

Sample supernatant	20 µL
Bla assay buffer	170 µL
Penicillin G stock solution (10 mg / mL assay buffer)	10 µL

The penicillin was added before 30 seconds shaking at 200 rpm and immediately reading absorbance at 240 nm every 30 seconds for 15 minutes at room temperature. Expression values were calculated as the average of the duplicates' slopes in the linear range of the penicillin degradation, divided by total protein concentration of the sample.

Phosphoglucomutase assay

Phosphoglucomutase (celB) activity was measured as the increase in NADPH concentration due to the increased availability of glucose-6-phosphate (G6P) substrate from glucose-1-phosphate and glucose-1,6-bisphosphate (G1,6BP) conversion by phosphoglucomutase, following reactions (1) and (2) (Worthington, 2011). Reaction (2) is catalyzed by G6P dehydrogenase.

$$G1P + celB\text{-}P \leftrightarrow G1,6BP + celB \leftrightarrow G6P + celBP \quad (1)$$

$$G6P + NADP^{+} \leftrightarrow 6\text{-Phosphoglucono-}\delta\text{-lactone} + NADPH + H^{+} \quad (2)$$

Master mix was made with the following recipe, volume denotes amount per well of reaction:

Imidazole-HCl (0.1M, pH 7.4)	80 µL
$MgCl_2$ (0.1M)	10 µL
$NADP^{+}$ (6.2 mM; Sigma N0505)	16.1 µL
G6P dehydrogenase (100 U/mL; Sigma G4134)	2µL
Glucose-1,6-bisphosphate (1.0 mM; Sigma G7137)	1.6 µL
Glucose-1-Phosphate (0.1 M; Sigma G7000)	4 µL

MQ-water	86.3 µL

Cell pellets were resuspended in 5 mL phosphoglucomutase assay buffer, sonicated as described, and centrifuged for 60 minutes at 8 000 rpm and 4 °C. The supernatant was transferred to new tubes, kept on ice, and total protein concentration of the supernatant was determined using the Bradford method. Reaction kinetics of the supernatant was measured in 5 duplicates for each sample in a 96-well Nunc plate, by adding 5 µL sample to 200 µL master mix, 30 seconds shaking at 900 rpm, and measuring absorbance at 340 nm every 30 seconds in a [NEW SPECTROMACHINE]. Expression values were calculated as the average of the duplicates' slopes in the linear range of the increase in, divided by total protein concentration of the sample.

Alkaline phosphatase assay

Cell pellets were resuspended in 5 mL alkaline phosphatase assay buffer, sonicated as described, and centrifuged for 60 minutes at 8 000 rpm and 4 °C. The supernatant was transferred to new tubes, kept on ice, and total protein concentration of the supernatant was determined using the Bradford method.

Reaction kinetics of the supernatant was measured in 5 duplicates for each sample in a 96-well Nunc plate. Stock substrate solution of 5 mg/mL *p*-Nitrophenyl phosphate was prepared and kept on ice covered by aluminium foil. 128 µL sample was added to each well and incubated for 10 minutes at 37 °C, before addition of 32 µL substrate stock solution to each well, mixing at 900 rpm for 30 seconds. Absorbance was measured at 405 nm every 12 seconds for 20 minutes. Expression values were calculated as the average of the duplicates' slopes in the linear range of the absorbance increase, divided by total protein concentration of the sample.

Bio-Rad Bradford assay

Coomassie Brilliant Blue G250 exists in a red- and a blue color conformation. When it binds to protein, the red form is converted to the blue form, with an absorption maximum at 595 nm

(Bradford, 1976). The increase in absorbance is linear between 0.05 mg/mL and 0.5 mg/mL protein when using the Bio Rad Protein Assay reagent (Laboratories).

The Bio Rad Protein Assay reagent was diluted 1:5 in MQ-water, and a standard curve of 0.1, 0.2, 0.3, 0.4 and 0.5 mg/mL BSA was prepared for each Bio Rad dilution made.

Samples were diluted until absorbance at 595 after reaction was within the linear range of the standard curve (usually 0.6 - 1.1), and measured in five duplicates by adding 10 µL diluted sample to each well of a 96-well Nunc plate before addition of 200 µL Bio Rad Protein Assay reagent. The plated were incubated for five minutes in room temperature before shaking at 900 rpm for 30 seconds, and the absorbance was measured at 595 nm.

RNA isolation and cDNA preparation

RNAprotect treated cell pellets were resuspended in 100 µL TE buffer with lysozyme and RNA was extracted using the RNAqueous® Kit from Ambion according to its manual, with the exceptions of using 600 µL wash solution 1, and eluting in 2 x 50 µL elution buffer.

RNA concentration was measured using a NanoDrop 1000 (Thermo Scientific), and DNA contamination was removed using the TURBO DNA-*free*™ Kit, according to its manual. cDNA was prepared from the samples using the First-Strand cDNA Synthesis Kit with the following reaction mixtures:

Bulk first-strand cDNA reaction mix	2.5 µL
DTT solution	0.5 µL
$Pd(N)_6$	0.5 µL
RNA sample	4 µL

Samples were incubated at 37 °C for 60 minutes, and kept at -20 °C.

Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR)

SYBR® Green is a fluorescent dye that emits substantially more fluorescence when bound to double stranded DNA (Wong and Medrano, 2005). Thus, the more dsDNA in the solution, the more fluorescence. DNA is duplicated in an exponential manner in the exponential phase of PCR (Mullis et al., 1986), and SYBR® Green is used to quantify the starting amount of a

specific DNA molecule during qRT-PCR (Mullis et al., 1986; Wong and Medrano, 2005). The PCR cycle, or point in time, at which the fluorescence from the sample is higher than the background fluorescence is called the cycle threshold value (C_T), and is used in calculations. In relative quantitation of mRNA transcript cDNA, primers specific for the gene in question are used, and the sample gene expression is measured compared to an internal reference here 16S RNA. The relative expression is calculated in this study using the comparative C_T method ($2^{-\Delta\Delta Ct}$) (Livak and Schmittgen, 2001).

Primers for the different genes of interest are shown in **table 2.1**, and were diluted to 2 mg/mL before the reaction. cDNA samples for endogenous control (16S) were diluted 1:400 and target cDNA was diluted 1:100. qRT-PCR strips were used, and the reaction mixture for each well was as follows:

Power SYBR® Green master mix	12.5 µL
Forward primer	2.5 µL
Reverse primer	2.5 µL
MQ-water	2.5 µL
cDNA sample	5 µL

Each sample was made in triplicates and a non-template control was included to exclude contamination. The strips were capped and spun down shortly before qRT-PCR was performed in an Applied Biosystems 7500 Real Time PCR sytem with the standard settings. C_T values were recorded and relative gene expression was calculated using the comparative C_T method.

SDS-PAGE and Western Blot

Proteins in solution may be separated based on their molecular weight using sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE), where proteins are denatured with DTT and treated with SDS to get a negative uniform charge to mass ratio (Weber and Osborn, 1969). Smaller weight proteins wander farther due to lower resistance from the gel matrix, and the proteins in the gel may be electrophoretically transferred to a nitrocellulose membrane for further analysis after separation (Towbin et al., 1979). This blotting is referred to as a Western blot. (Burnette, 1981), and allows the detection and quantification of specific proteins using antibodies against the protein itself, or against a protein fusion tag. For such

immunoblots, the nitrocellulose membrane is blocked by incubation in protein rich solutions to avoid unspecific binding of the antibody (Towbin et al., 1979), before incubation with the antibody against the protein- or tag of interest.

In direct detection systems, such as the HisProbe-HRP (Thermo Fisher Scientific #15165) where a his-tag antibody is fused to horseraddish peroxidase (HRP), the probe binds to the protein of interest and may be directly detected. In indirect systems the primary antibody is bound to the protein of interest on the membrane, followed by a detection by a secondary antibody, with affinity for the primary and is fused to for example HRP. The Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific #32209) allows chemiluminescent detection of HRP from the luminescence emitted at 425nm by cleavage of luminol in the substrate (Bhandari et al., 2010).

Procedure:

E. coli containing plasmids with β-lactamase were cultured as described for enzyme assays, and cells containing IFN-α2b constructs were cultured following the described Protein Expression in low-scale cultivation. The cultures were pelleted at 6000 rpm for five minutes at 4 °C (Eppendorf 5810R centrifuge) and the pellets were thoroughly resuspended in 1 mL 0.9 % NaCl per 100 mg pellet. 1 mL cell suspension was transferred to Eppendorf tubes before centrifugation at 13 000 rpm for three minutes and removal of supernatant. The pellets were kept at -80 °C overnight.

Pellets were resuspended in 500 µL 1x SDS running buffer, and sonicated as described above and total protein concentration of the cell lysate was determined by Bio-Rad Bradford protein assay. 50 µL cell lysate was combined with 25 µL 3x sample buffer in Eppendorf tubes, and boiled at 95 °C for five minutes before being stored in -20 °C. SDS-PAGE was performed using precast ClearPAGE™ SDS-gels (C.B.S. Scientific), 12 % for β-lactamase and 16 % for IFN-α2b and the Dual Cool Electrophoresis System (DCX-700, C.B.S. Scientific). Precision Plus Protein Dual Color (Bio-Rad) was used as a standard, and the gels were run with 150V for at least 90 minutes, or overnight at 35V until proper separation had been achieved.

The plastic casing around the gel was removed and gels were rinsed with MQ-water. The sponge, filter paper and gel were incubated in blotting buffer for approximately 30 minutes. Immobilon™ PVDF Transfer Membrane (Millipore) was incubated in methanol for 15 seconds, MQ-water for two minutes, and blotting buffer for 10 minutes, before assembly of

the Western cassette as described in the Dual Cool Electrophoresis System manual. The blotting was run at 100V for 1.5 - 3 hours before incubation in TBS for 5 minutes with gentle agitation. Blocking was performed by incubation in blotto buffer for 60 minutes with gentle agitation before washing 3 times with TBST incubation for 10 minutes, and once with TBS for 2 minutes with gentle agitation.

For β-lactamase, the membrane was incubated overnight at 4 °C with Anti-Beta Lactamase antibody (Abcam, ab12251) diluted 1:5000 in blotto buffer with gentle agitation before washing 3 times with TBST incubation for 10 minutes, and once with TBS for 2 minutes with gentle agitation. The membrane was then incubated with Polyclonal Rabbit Anti-Mouse Ig/HRP (Dako, P 0260) in blotto buffer for 1 hour, before repeating the washing steps. For IFN-α2b, the membrane was incubated in HISprobe-HRP (1:12500 in blotto), for 60 minutes with gentle agitation before repeating the washing steps.

Chemiluminescent detection was performed by incubating the membrane in a mix of 5 mL solution A, and 5 mL solution B of the Pierce ECL Western Blotting Substrate for 1 minute, before reading chemiluminescence with a Chemi Doc™ XR+ Imager (Bio-Rad) every 30 seconds for 15 minutes.

Study of mRNA decay by inducer wash-out method

m-Toluate, a commonly used inducer for the XylS/*Pm* system, is a small organic compound that has been shown to freely and rapidly diffuse in over the cell membrane of *E. coli* (Kucharova, unpublished). High-performance liquid chromatography showed no m-toluate present in cell extracts following filtration from induced culture and resuspension in fresh medium (Kucharova, unpublished), thus ending the inducer signal. By sampling RNA as described earlier from the resuspended cell culture at time intervals, the decay of transcript could be followed by plotting the relative amounts obtained by the comparative C_T method. Analysis of transcript from the K_m^R gene of the plasmid vector used showed that only mRNA transcribed from the *Pm* promoter was affected by the inducer wash-out (Kucharova, unpublished).

To analyse the decay of transcript, freshly streaked cells inoculated in 20 mL LB containing appropriate antibiotic were grown over-night in 125 mL Erlenmeyer flasks at 30 °C, 225 rpm. Cultures were inoculated to OD 0.05 in 20 mL LB containing appropriate antibiotic in 250

mL baffled Erlenmeyer flasks and incubated in a water bath at 30 °C, 225 rpm, before induction to 0.5 mM m-Toluate at OD 0.5. Cells were grown for 90 minutes to reach steady-state levels of transcript (Berg et al., 2009), before harvesting 5 mL culture by vacuum filtration with 0.2 µm pore size filters, washing of the filter with 10 mL PBS, and resuspension of the filter in 10 mL prewarmed LB without inducer in a 250 mL baffled Erlenmeyer flask. The cultures were kept at 30 °C, 225 rpm in a water bath and culture samples for RNA isolation were taken as described, at 0, 2, 4, 6, 8 and 10 minutes after resuspension of the cells. Transcript amounts were measured by qRT-PCR and calculated relative to the 0 minute sample following the comparative C_T method (Livak and Schmittgen, 2001).

2.3 Bacterial Strains, Plasmids and Oligonucleotides

2.3.1 Bacterial Strains and Plasmids

The *E. coli* strain and plasmid constructs used in this study are described in **Table 2-1**, selected plasmid maps are shown in chapter 2.3.3.

Table 2-1: Bacterial strains and plasmids

Strain or plasmid	Description	Source or reference
Strain		
Escherichia coli DH5α	General cloning strain, genotype: F^-, *end*, hsdR17, (r_k^-, m_k^+), *sup*E44, *thi*-1, λ-, *rec*A1, *gyr*A96, *rel*A1, ϕ80d*lacZ*ΔM15.	(Bethesda, 1986)
Plasmids		
Bla plasmids		
pBSP1	RK2-based expression vector containing *bla,* with its wild type signal sequence (AGT 2nd codon), as reporter gene under control of XylS/*Pm* system; *rrnBT1T2*; *tlPP; hok/sok;* M13 *ori*; *t*; *Pneo; trfA*; *oriV*; *oriT;* Km^r; 9.5kb.	(Kucharova, unpublished)
pBSP1-X	pBSP1 derivatives; X denotes *bla* 2nd codon synonymous mutations (TCT, AGC, TCC, TCA, TCG), or *bla* signal sequence synonymous mutations (C19, C20) inserted using inserts from annealed oligonucleotides shown in **Table 2-2**; Km^r, 9.5kb.	This study
pMS2	pBSP1 derivative containing the following synonymous codon mutations in *bla* signal sequence; 2nd: AGT→TCT, 7th: CGT→CGC; Km^r, 9.5kb	(Kucharova, unpublished)
pMS3	pBSP1 derivative containing the following synonymous codon mutations in *bla* signal sequence; 2nd: AGT→TCT, 9th: GCC→GCT; Km^r, 9.5kb	(Kucharova, unpublished)
pMS4	pBSP1 derivative containing the following synonymous codon mutations in *bla* signal sequence; 2nd: AGT→TCT, 9th: GCC→GCT, 18th: TGC→TGT; Km^r, 9.5kb	(Kucharova, unpublished)
pCSP1	pBSP1 derivative with CSP signal peptide (Sletta et al., 2004) substituted for *bla'*s signal peptide; Km^r, 9.5kb	(Kucharova, unpublished)
pTMB17	pBSP1 derivative with deletion of *bla'*s signal peptide; Km^r, 9.5kb	(Kucharova, unpublished)
celB plasmids		
pLB11	RK2-based expression vector containing *celB* as reporter gene under control of XylS/*Pm* system; *rrnBT1T2*; *tlPP; hok/sok; t*; *Pneo; trfA*; *oriV*; *oriT;* Km^r; 9.0kb.	(Lale et al., 2011)

Table 2-1: Continued

Strain or plasmid	Description	Source or reference
pLB11-X	pLB11 derivatives; X denotes *celB* 2nd codon mutations (CCT, CCA, CCG) from inserting annealed synthetic oligonucleotides listed in **Table 2-2**; Km^r; 9.0kb	This study
phoA plasmids		
pASP1	RK2-based expression vector containing *phoA* as reporter gene under control of XylS/*Pm* system; *rrnBT1T2*; *tlPP; hok/sok;* M13 *ori*; *t*; *Pneo; trfA*; *oriV*; *oriT;* Km^r; 10.0kb.	(Kucharova, unpublished)
pASP1-AAG	pASP1 derivative; *phoA* second codon mutated to its synonymous mutant AAG by inserting annealed synthetic oligonucleotide listed in **Table 2-2**; Km^r; 10.0kb.	This study
pCSP1phoA	pCSP1 derivative where the *NcoI - BstAPI* fragment encoding the *bla* reporter gene was replaced by a PCR-fragment encoding the mature *phoA* gene. Km^R; 10.0kb.	(Kucharova, unpublished)
apr^R plasmids		
pAR69	RK2-based expression vector containing *aac(3)-IV* with the 69 first nucleotides of *celB* as a 5' fusion partner as the reporter gene under control of XylS/*Pm* system; *rrnBT1T2*; *tlPP; hok/sok; Pneo; trfA*; *t*; *oriV*; *oriT;* Km^r; 8.2kb.	(Kucharova, unpublished)
pAR69-D11	pAR69 derivative with *celB69* sequence exchanged by mutated celB69-D11 annealed synthetic oligonucleotides listed in **Table 2-2**; Km^r, 8.2kb.	This study
IFN-α2b plasmids		
pAT63	RK2-based expression vector containing the codon optimized version of *ifn-α2b* under control of XylS/*Pm* system, with a c-*myc*-his6 affinity tag; *rrnBT1T2*; *tlPP; hok/sok;* M13 *ori*; *t*; *Pneo; trfA*cop271C; *oriV*; *oriT;* Amp^r; 8.8kb. (Published as pIFN30S)	(Sletta et al., 2007)
pAT64	pAT63 derivative with a 5' *pelB* signal peptide fusion to *ifn-α2b*; Amp^r; his6; 8.9kb. (Published as pIFN30SpelB)	(Sletta et al., 2007)
pIN69	pAT64 derivative with *pelB* exchanged by the first 69 nucleotides of *celB* as a 5' fusion partner to *ifn-α2b*; ; Amp^r; his6; 8.9kb.	(Kucharova, unpublished)
pIN69-D11	pIN69 derivative with *celB69* sequence exchanged by mutated celB69-D11 annealed synthetic oligonucleotides listed in **Table 2-2**; Amp^r; his6; 8.9kb.	This study

2.3.2 Synthetic Oligonucleotides, qRT-PCR- and Sequencing Primers

The synthetic DNA insert oligonucleotides shown in **Table 2-2** were used for cassette exchange cloning to create variants of the *bla*, *phoA*, and *celB* genes. Oligonucleotides were

designed using Clone Manager Suite (Sci-Ed Software), and the corresponding forward and reverse primers were annealed as described above before ligation. The randomized oligos were designed by Veronika Kucharova, and Their doping percentages are shown in **Table 2-3**. Primers used for sequencing or qRT-PCR are listed in **Table 2-4**.

Table 2-2: Oligonucleotides used to construct synthetic DNA inserts. Numbers in sequences correspond to base doping percentages shown in **Table 2-3**.

Name	Sequence (5' to 3' direction)	Length (bp)	T_m (°C)	Restriction enzymes	Manu-facturer
Bla oligos					
Bla-TC-5'spiked	TATGTC1AT2CA3CA4TT5CG1GT6GC6C T1AT2CC6TT4TT4GC7GC8TT4TG5CT1 CC1GT1TT5GC	69	-	*NdeI*, *NcoI*	Eurogentec
Bla.wt.rev	CATGGCGAAAACAGGAAGGCAAAATGCCG CAAAAAAGGGAATAAGGGCGACACGGAAA TGTTGAATACTCA	71	82	*NdeI*, *NcoI*	Sigma-Aldrich
Bla.TCT.fwd	TATGTCTATTCAACATTTCCGTGTCGCCC TTATTCCCTTTTTTGCGGCATTTTGCCTT CCTGTTTTCGC	69	90.7	*NdeI*, *NcoI*	Sigma-Aldrich
Bla.TCT.rev	CATGGCGAAAACAGGAAGGCAAAATGCCG CAAAAAAGGGAATAAGGGCGACACGGAAA TGTTGAATAGACA	71	92.7	*NdeI*, *NcoI*	Sigma-Aldrich
Bla.AGC.fwd	TATGAGCATTCAACATTTCCGTGTCGCCC TTATTCCCTTTTTTGCGGCATTTTGCCTT CCTGTTTTCGC	69	91.3	*NdeI*, *NcoI*	Sigma-Aldrich
Bla.AGC.rev	CATGGCGAAAACAGGAAGGCAAAATGCCG CAAAAAAGGGAATAAGGGCGACACGGAAA TGTTGAATGCTCA	71	93.2	*NdeI*, *NcoI*	Sigma-Aldrich
Bla.TCC.fwd	TATGTCCATTCAACATTTCCGTGTCGCCC TTATTCCCTTTTTTGCGGCATTTTGCCTT CCTGTTTTCGC	69	91.3	*NdeI*, *NcoI*	Sigma-Aldrich
Bla.TCC.rev	CATGGCGAAAACAGGAAGGCAAAATGCCG CAAAAAAGGGAATAAGGGCGACACGGAAA TGTTGAATGGACA	71	93.2	*NdeI*, *NcoI*	Sigma-Aldrich
Bla.TCA.fwd	TATGTCAATTCAACATTTCCGTGTCGCCC TTATTCCCTTTTTTGCGGCATTTTGCCTT CCTGTTTTCGC	69	90.7	*NdeI*, *NcoI*	Sigma-Aldrich
Bla.TCA.rev	CATGGCGAAAACAGGAAGGCAAAATGCCG CAAAAAAGGGAATAAGGGCGACACGGAAA TGTTGAAT TGACA	71	92.7	*NdeI*, *NcoI*	Sigma-Aldrich

Table 2-2: Continued

Name	Sequence (5' to 3' direction)	Length (bp)	T_m (°C)	Restriction enzymes	Manu-facturer
Bla.TCG.fwd	TATGTCGATTCAACATTTCCGTGTCGCCC TTATTCCCTTTTTTGCGGCATTTTGCCTT CCTGTTTTCGC	69	91.3	*NdeI*, *NcoI*	Sigma-Aldrich
Bla.TCG.rev	CATGGCGAAAACAGGAAGGCAAAATGCCG CAAAAAAGGGAATAAGGGCGACACGGAAA TGTTGAATCGACA	71	93.2	*NdeI*, *NcoI*	Sigma-Aldrich
Bla.C19.fwd	TATGTCTATTCAACATTTCCGTGTTGCCC TTATTCCCTTTTTTGCGGCATTTTGCCTT CCTGTTTTCGC	69	90.4	*NdeI*, *NcoI*	Sigma-Aldrich
Bla.C19.rev	CATGGCGAAAACAGGAAGGCAAAATGCCG CAAAAAAGGGAATAAGGGCAACACGGAAA TGTTGAATAGACA	71	92.3	*NdeI*, *NcoI*	Sigma-Aldrich
Bla.C20.fwd	TATGTCAATTCAACATTTCCGAGTCGCCC TTATTCCCTTTTTTGCGGCATTTTGCCTT CCTGTTTTCGC	69	90.4	*NdeI*, *NcoI*	Sigma-Aldrich
Bla.C20.rev	CATGGCGAAAACAGGAAGGCAAAATGCCG CAAAAAAGGGAATAAGGGCGACTCGGAAA TGTTGAATTGACA	71	92.3	*NdeI*, *NcoI*	Sigma-Aldrich
PhoA oligos					
phoA-AAG.fwd	TATGAAGCAAAGCACTATTGCACTGGCAC TCTTACCGTTACTGTTTACCCCTGTGACA AAAGC	63	85.9	*NdeI*, *NcoI*	Sigma-Aldrich
phoA-AAG.rev	CATGGCTTTTGTCACAGGGGTAAACAGTA ACGGTAAGAGTGCCAGTGCAATAGTGCTT TGCTTCA	65	88.3	*NdeI*, *NcoI*	Sigma-Aldrich
CelB oligos					
celB-CCT.fwd	TATGCCTAGCATAAGCCCATTTGCCGGCA AGCCGGTCGATCCG	44	91.8	*NdeI*, *RsrII*	Sigma-Aldrich
celB-CCT.rev	GTCCGGATCGACCGGCTTGCCGGCAAATG GGCTTATGCTAGGCA	43	89.6	*NdeI*, *RsrII*	Sigma-Aldrich
celB-CCA.fwd	TATGCCAAGCATAAGCCCATTTGCCGGCA AGCCGGTCGATCCG	44	93.2	*NdeI*, *RsrII*	Sigma-Aldrich
celB-CCA.rev	GTCCGGATCGACCGGCTTGCCGGCAAATG GGCTTATGCTTGGCA	43	91.1	*NdeI*, *RsrII*	Sigma-Aldrich
celB-CCG.fwd	TATGCCGAGCATAAGCCCATTTGCCGGCA AGCCGGTCGATCCG	44	94.1	*NdeI*, *RsrII*	Sigma-Aldrich
celB-CCG.rev	GTCCGGATCGACCGGCTTGCCGGCAAATG GGCTTATGCTCGGCA	43	92.0	*NdeI*, *RsrII*	Sigma-Aldrich

Table 2-2: Continued

Name	Sequence (5' to 3' direction)	Length (bp)	T_m (°C)	Restriction enzymes	Manu-facturer
celB69-mut.Fwd	TATG6668768188766681117667768 8766771678166778671611716881 8167867666177CG	73	-	*NdeI, NcoI*	Eurogentech
celB69-wt.Rev	CATGGCCAGGGCGTCGATATTGACAAGAC GGTCCGGATCGACCGGCTTGCCGGCAAAT GGGCTTATGCTGGGCA	74	X	*NdeI, NcoI*	Sigma-Aldrich
celB69-D11.fwd	TATGTCTAGCATAAACCCATTTACCGGCA AGCCGGTCGTTCCGGACTGTCTTTTCAAT ACCGACGCCCTGGC	73	X	*NdeI, NcoI*	Sigma-Aldrich
celB69-D11.rev	CATGGCCAGGGCGTCGGTATTGAAAAGAC AGTCCGGAACGACCGGCTTGCCGGTAAAT GGGTTTATGCTAGACA	74	X	*NdeI, NcoI*	Sigma-Aldrich

Table 2-3: Base composition in synthetic oligo mixtures from Sigma-Aldrich. Percent denotes probability of the given base being incorporated at a given position.

Mix no:	Base composition
1	79% T, 7% A, 7% G, and 7% C
2	80% T, 10% A and 10% C
3	80% A and 20% G
4	80% T and 20% C
5	80% C and 20% T
6	79 % C, 7 % A, 7 % T, and 7 % G
7	79 % G, 7 % A, 7 % T, and 7 % C
8	79 % A, 7 % T, 7 % G, and 7 % C

Table 2-4: Primers used for and sequencing and qRT-PCR. qRT-PCR primers were designed using the Primer Express software (Applied Biosystems), and the sequencing primer were designed using Clone Manager Suite (Sci-Ed Software)

Name	DNA Sequence (5' to 3')	Target gene
Sequencing primer		
PmUTR.fwd	AAGAAGCGGATACAGGAGTG	*Pm* promoter, 97 bp upstream of ATG
qRT-PCR primers		
bla.fwd	ACGTTTTCCAATGATGAGCACTT	*β-lactamase*
bla.rev	TGCCCGGCGTCAACAC	*β-lactamase*
RT-celB.Fwd	ACCAGCTTCAATGAAAACCACA	*celB*
RT-celB.Rev	CGCCCTTGCGGTAATCG	*celB*
phoA-73F	GCACCAGAAATGCCTGTTCTG	*phoA*
phoA-134R	CCGCCGGGTGCAGTAA	*phoA*
APR69-11F	CAGCGGTGGAGTGCAATG	*aac(3)-IV*
APR69-79R	GGTTGAGAAGCTGACCGATGA	*aac(3)-IV*
ifna2b324.fwd	CGAGACCCCGCTGATGAA	*Ifn-α2b,celB69-ifn-α2b*
ifna2b396.rev	CAGATACAGGGTGATACGCTGAAA	*Ifn-α2b,celB69-ifn-α2b*

2.3.3 Plasmid Maps

Plasmids described in **Table 2-1** are shown below. Plasmid maps were created using Clone Manager Suite (Sci-Ed Software), and edited using Inkscape (www.inkscape.org).

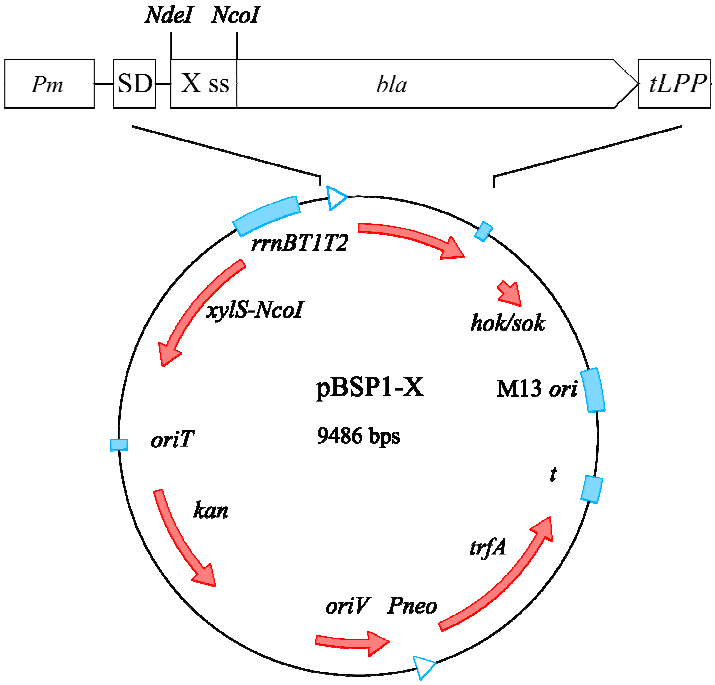


Figure 2-1: Map of pBSP1-X plasmids; X ss denotes the inserted signal sequence variant: wild-type, synonymous second codon mutations, or candidates identified from screening of randomized library; *bla*: *β-lactamase.* Oligos used for ligation of inserts are shown in **Table 2-2**. Elements of the plasmids are described in section 1.2.4.

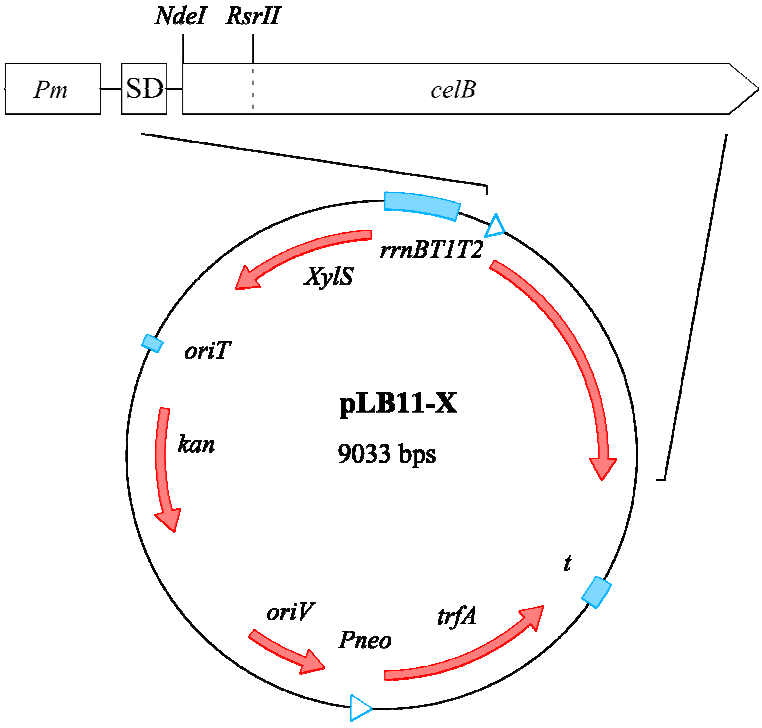


Figure 2-2: Map of pLB11-X plasmids; X denotes the 2nd codon of the inserted 5' coding sequence between *NdeI* and *RsrII*; *celB*: gene encoding phosphoglucomutase. Elements of the plasmids are described in section 1.2.4.

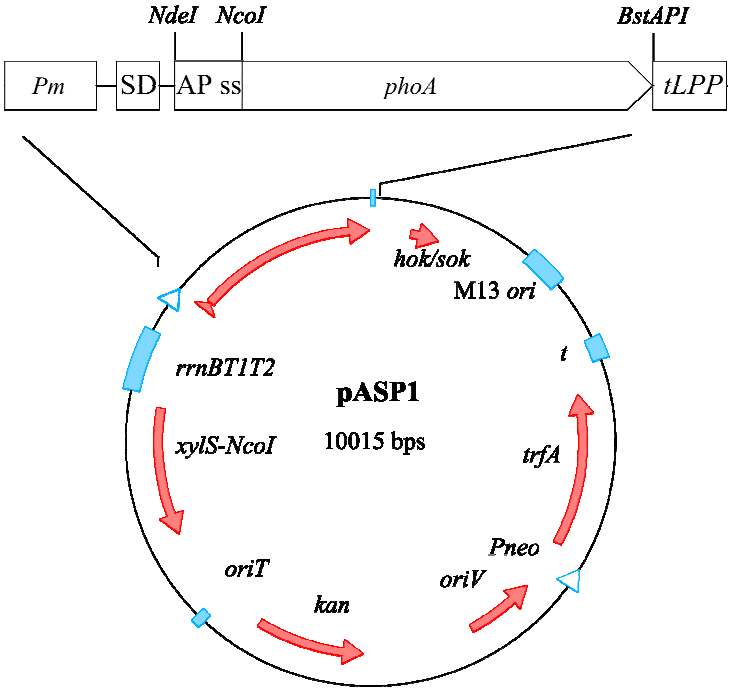


Figure 2-3: Map of pVK8 plasmid; *phoA*: gene encoding alkaline phosphatase (AP). The AP signal sequence (AP ss) was exchanged with annealed oligos described in **Table 2-2** to give the pVK8-AAG construct containing the 2nd codon synonymous mutation AAA→AAG. Elements of the plasmid are described in section 1.2.4.

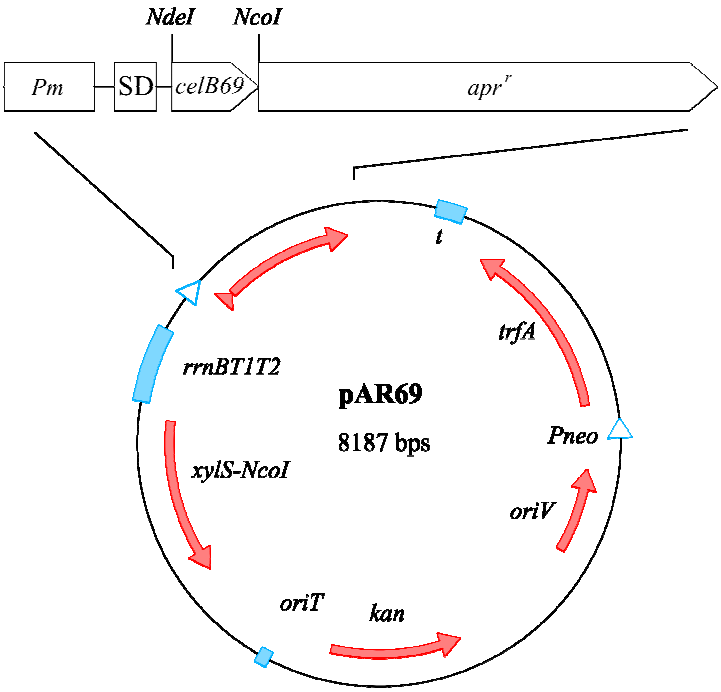


Figure 2-4: Map of pAR69 plasmid; *apr*^{r}: *aac(3)-IV* gene encoding 3-*N*-Acetyltransferase, giving apramycin resistance; *celB*$_{69}$; gene fusion of first 69 nucleotides of *celB* DNA. The randomized *celb*$_{69}$*-aac(3)-IV* library was created by exchanging the *celB69* gene fusion with celB69-mut.Fwd annealed to celB69.Rev, described in **Table 2-2**.

2.4 Bioinformatics Tools

Clone Manager Suite

Clone Manager Suite (Sci-Ed Software) allows the management and of DNA and amino acid sequences *in silico*. In this study it was used for planning of cloning, sequence analysis, sequence alignment and the generation of plasmid maps.

EMBOSS CAI web application

The open source software package ”European Molecular Biology Open Software Suite” (EMBOSS) consists of over 100 molecular biology and bioinformatics applications (Rice et al., 2000). Its codon adaptation index tool gives the option of reference codon tables from a variety of organisms, and is freely available as a web application at http://services.cbib.u-bordeaux2.fr/pise/cai.html. The application was used to analyze the CAI of the 5’ end of various transcripts in this study.

Python tAI Calculator

As no program for calculating the tRNA adaptation index of a sequence was freely available in a programming language known to the author, a script for calculating the tAI was written in Python 2.7 (Van Rossum, 2012). The script may be run in the Python distribution’s shell, IDLE, and allows the pasting of DNA sequences in caps (A, T, G, C) containing whitespace, that are dividable by 3. The tAI data is retrieved from a comma-separated file, tAI.csv, with *E. coli* tAI values for the individual codons obtained from (Tuller et al., 2010). The geometric mean is taken from the input sequence’s scores, and displayed. A comma-separated output file, output.csv, is created/overwritten for each run with the individual tAI values for all codons analyzed. The code of the tAI.py script and the tAI.csv contents is showed in appendix Appendix D. The script and the *E. coli* tAI table are included in the attached .zip-file, and are also freely available at http://folk.ntnu.no/adrianer/tAIcalculator/.

RBS calculator

The RBS calculator combines thermodynamic data from the major molecular interactions of translation initiation with an optimization algorithm to be to predict an RBS with a target translation initiation rate for a given gene (Salis et al., 2009). The model analyzes the UTR and the 5' sequence of the gene by calculating the attractive forces to the 16S rRNA, and its repulsion by mRNA secondary structures. The program also features a reverse engineering mode, where potential start-codons in an mRNA transcript may be analyzed for their translation initiation rate based on their putative RBS. This feature has been shown to give rates proportional to protein expression, and was used in this study to analyse various transcripts for their translation initiation efficiencies. The tool is freely available online at https://salis.psu.edu/software/forward.

Quikfold

The Quikfold server is part of the Mfold web server, which provides a range of DNA and RNA sequence structure prediction tools of varying complexity, all based on the minimum free energy for the folding of the input sequence (Zuker, 2003). Quikfold has a simple user interface and allows the prediction of several structures at once. It allows DNA or RNA folding and allows variation of, for example the maximum distance between paired bases. It returns a table of the smallest minimal free energies for given sequences, and provides models of the structures in pdf format. In this study, the RNA 3.0 folding energies and a maximum distance of 30 bases were used to analyze local secondary structures forming in the nascent transcript before the ribosome is able to bind. The tool is freely available online at http://mfold.rna.albany.edu/?q=DINAMelt/Quickfold.

Minitab®

Minitab® is a statistical analysis software distribution, used in this study under NTNU's license to calculate Pearson's correlations between gene expression and variations in features that might affect expression. The Pearson correlation reflects the degree to which two variables are related, ranging from +1 to -1, indicating a perfect positive or negative linear relationship, respectively (Rodgers and Nicewander, 1988). A correlation close to 0 means no correlation. Minitab also gives a P-value for the correlation, which denotes the probability of finding the given result if the correlation was 0, and a value of $P<0.05$ was chosen as the cut-off for statistical significance in this study.

3 Results

This study aimed to examine the effect of changes to the 5' coding sequence of recombinant mRNAs for efficient protein expression from pJBn derived expression vectors. The development of a tAI calculation tool used in the study is briefly described in chapter 3.1, followed by a description of the generation and characterization of a synonymous library in the 5' *bla* coding sequence in chapter 3.2. A finding that a different synonymous 2nd codon of *bla* had a profound influence on its expression is further explored by analysis of alternative 2nd codons of *bla* and two additional reporter genes in chapter 3.3. Lastly, the generation of a completely randomized library of a 5'-terminal *celB* gene fragment fused to the *aac(3)-IV* gene encoding apramycin resistance, and its screening for increased expression is described in chapter 3.4, along with the investigation of an identified *celB*-based fusion partner candidate's effect on the expression of a poorly expressed human protein.

3.1 Development of a tAI Calculation Tool

In the process of analyzing the sequences in this study *in silico*, a program for the calculation of the tRNA adaptation index (tAI) of a DNA sequence was developed in Python, as no tool to calculate the tAI of genes were available in a language known to the author. The script implements the tAI algorithm (Tuller et al., 2010) and obtains tAI data from a separate file, allowing the calculation of tAIs of other organisms by exchanging the data file. The program is relatively user-friendly, as it allows pasting of DNA sequences for analysis in a text based user-interface, and returns the given sequence's tAI. The tool is described in chapter 2.4, and its code is presented in Appendix D.

3.2 Selection for Optimized 5' Coding Sequences of the *bla Gene* by Directed Evolution

In previous work the effect of synonymous mutations of the 5' coding region of *bla* mRNA on its protein expression was investigated by the creation of a combinatorial mutagenesis synonymous codon library in the 69 nucleotide *bla* signal sequence (Kucharova, unpublished). As previously described, alterations in expression of the *bla* gene may be identified by screening the ampicillin tolerance of the host-cells, and mutants growing at increased ampicillin concentrations have higher expression. Several up-mutants were

identified from screening the library, and most of the highest growing ones (up to eight-fold higher than wild-type at 0.05 mM inducer) contained the 2nd codon mutation AGT→TCT on the DNA level, which in itself conferred an increase in resistance from 0.2 g/L to 1.0 g/L ampicillin.

3.2.1 Directed Evolution of the *bla* 5' Coding Sequence by Synonymous Mutations

To further increase the expression of the protein by synonymous mutations of the *bla* signal sequence, a directed evolution approach was applied in the design of a second library where the stimulating 2nd codon change was kept. A new wild-type expression vector for the library containing the TCT 2nd codon in the *bla* coding sequence, pBS_2P1, was constructed by annealing and ligation of synthetic oligonucleotides, and the library was created in the vector as described in chapter 2.2 . By transformation of *E. coli* DH5α, approximately 140 000 clones were obtained as the SII library.

The SII library was screened on a gradient of ampicillin concentrations from 0 to 6 g/L ampicillin supplemented with 0.5 mM inducer, as described in 2.2. The previously made SI library was also screened for comparison, and the survival rates of the *E. coli* DH5α host cells were determined. The SII library gave colonies growing up to 5.5 g/L ampicillin, while SI colonies were found to grow up to 2 g/L. The maximum growth of cells containing pBSP1 was 0.26 g/L. Survival frequencies from SI, SII, and pBSP1 wild-type are shown in **Figure 3-1**.

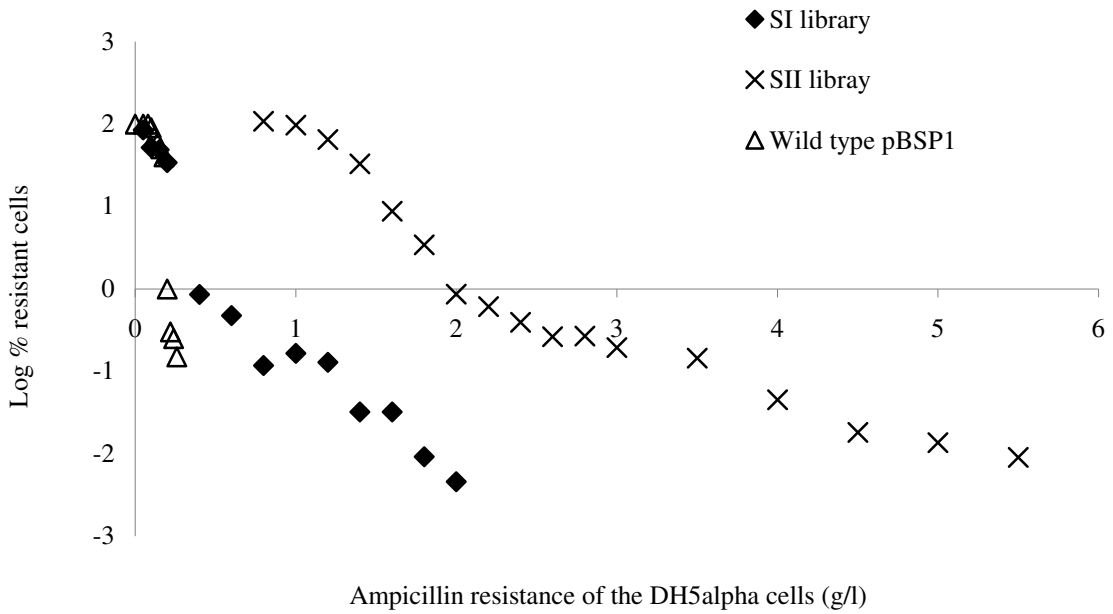


Figure 3-1: The SI and SII libraries were plated on solid LA with increasing ampicillin concentration at 0.05 mM m-Toluate induction, and the *E. coli* host cell's survival frequencies were determined as % survival at given concentrations compared to the growth on solid media containing only kanamycin. pBSP1 wild type's survival frequency had been previously determined (Kucharova, unpublished). Survival frequency is given as $\log_{10}$ of the % survival at the given ampicillin concentration. Data is shown in Appendix D.

Given the semi conservative mechanism of DNA-replication in *E. coli* (Meselson and Stahl, 1958), approximately 50 % of all the clones in the SII library should harbor the wild-type signal sequence, as the non-coding strands of the ligated inserts were designed to be complementary to the pBS$_2$P1*bla* signal sequence. To initially characterize the library, 12 individual colonies from the kanamycin plate from the SII library screening were randomly picked for sequencing. These showed that 9 out of 12 contained the wild type TCT signal sequence (alignment shown in **Table C-1**). Surprisingly, two of the three remaining mutants contained in-frame stop-codons as well as amino acid substitutions, and the third contained one amino-acid mutation.

Further characterization of the library was continued by plasmid isolation from 72 selected single colony *E. coli* host candidates, inoculated from different ampicillin concentrations between 1.2-5.5 g/L. The isolated plasmids were used to transform new *E. coli* DH5α cells for replica plating, described in chapter 2.2. This was done to confirm that the increased resistances did indeed arise from the synonymous mutations in the signal sequence, and not

from mutations in the chromosome of the host cells. Thirty-six of the retransformed candidates were found to grow on higher ampicillin concentration than cells harboring pBS_2P1, and subsequently their signal sequences were sequenced as described in chapter 2.2. Three of these candidates, termed C19, C20 and C22, were shown to only have synonymous mutations in the *bla* signal sequence, and C22 was found to be identical to MS3, a mutant identified in the SI library (**Table 3-1**). Three candidates harbored the wild-type pBS_2P1 *bla* signal sequence, and 30 had deletions or non-synonymous mutations in the signal sequence. The C19 and C20 *bla* signal sequence variants were commercially synthesized and cloned into pBS_2P1 to exclude plasmid mutations. The new constructs were confirmed by sequencing and used for further characterization by sequence analysis and expression experiments, as described below.

3.2.2 Sequence and Expression Analysis of Optimized 5' *bla* Synonymous Variants

The C19, C20 and C22 *bla* sequence variants, along with the best performing TCT second codon mutants from the SI library, MS2 and MS4, were initially characterized by replica plating and sequence analysis. The DNA sequence of their signal peptides is shown in **Table 3-1,** along with analysis data. All the mutations shown in the table are synonymous, and no other mutations were detected upstream or downstream of the displayed sequence. As there were no mutations downstream of codon 18 in any of the candidates, the sequence from the start codon up to codon 18 was used for calculation of tAI, CAI and translation initiation rate (T.I.R.) prediction, using the tools described in chapter 2.4. The T.I.R. was predicted including the UTR of the *Pm* promoter (appendix Appendix B) (Berg et al., 2009). The ampicillin tolerance values of the *E. coli* host cells were determined by replica plating in two biological occurrences at 0.05 mM inducer.

All analyzed *bla* signal sequence variants conferred higher ampicillin resistance than pBS_2P1, however none of the SII candidates were found to give resistance values exceeding those of the candidates already obtained in the SI library. The average number of point mutations in the shown SII library candidates was 1.33, compared to 3.33 for the analyzed SI library candidates. One of the variants obtained from the SII library had TCA as the second codon, and showed only a 0.05 g/L improvement in ampicillin tolerance over pBS_2P1. A trend was observed in the variants with mutations additional to the TCT 2nd codon in candidates from both libraries, where codons with low tAI had been inserted in the signal sequence, lowering

the total tAI of the analyzed sequence. The codon adaptation index (CAI) showed no correlation to increased ampicillin tolerance, and neither did the predicted translation initiation rate or the AT content of the *bla* signal sequences (not shown).

Table 3-1: *bla* signal sequence from selected SI and SII library candidates, along with ampicillin tolerance of the *E. coli* DH5α host cells and sequence analysis data. Candidates above the line were isolated in the SII library, MS2 and MS4 were only isolated in the SI library.

	DNA Sequence of the first 18 codons of the *bla* signal sequence	Ampicillin tolerance (g/L) [1] Induced	Ampicillin tolerance (g/L) [1] Un-induced	CAI[2]	tAI[2]	Codon change on DNA level (tAI)	T.I.R.[3]
pBSP1	ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTCGCC	0.20	0.020	0.600	0.165	-	2207
pBS_2P1	...TC.....	1.00	0.020	0.625	0.171	(AGT→TCT) (0.054875→0.10975)	2207
C19	...TC.....T.....	1.65	0.020	0.646	0.164	GTC→GTT (0.25→0.10975)	2207
C20	...TCA.....A.....	1.05	0.010	0.542	0.103	TCT→TCA (0.10975→0.125) CGT→CGA (0.5→0.00005)	4962
C22 (MS3)	...TC.....T.....	1.70	0.020	0.620	0.164	GCC→GTT (0.20→0.10975)	2207
MS2	...TC.....C.....	1.95	0.020	0.615	0.168	CGT→CGC (0.5→0.36)	2207
MS4	...TC.....T.....T.....	2.40	0.020	0.609	0.156	GCC→GTT (0.25→0.10975) TGC→TGT (0.125→0.054875)	2207

1) Induced determined to +/- 0.05 g/L, uninduced determined to +/- 0.001
2) Calculated as described in 2.4 from the first 18 codons.
3) Predicted as described in 2.4 from Pm UTR+coding sequence

The relative expression of *bla* from the candidates was further quantified by qRT-PCR to determine the accumulated transcript amounts, and by β-lactamase assays to determine the protein expression levels (**Figure 3-2**), as described in chapter 2.2. Expression vector pCSP1 containing the CSP signal sequence instead of the native *bla* signal sequence and a slightly different UTR (see appendix Appendix B), and pTMB17 without a signal sequence for *bla* were included as positive and negative controls, respectively.

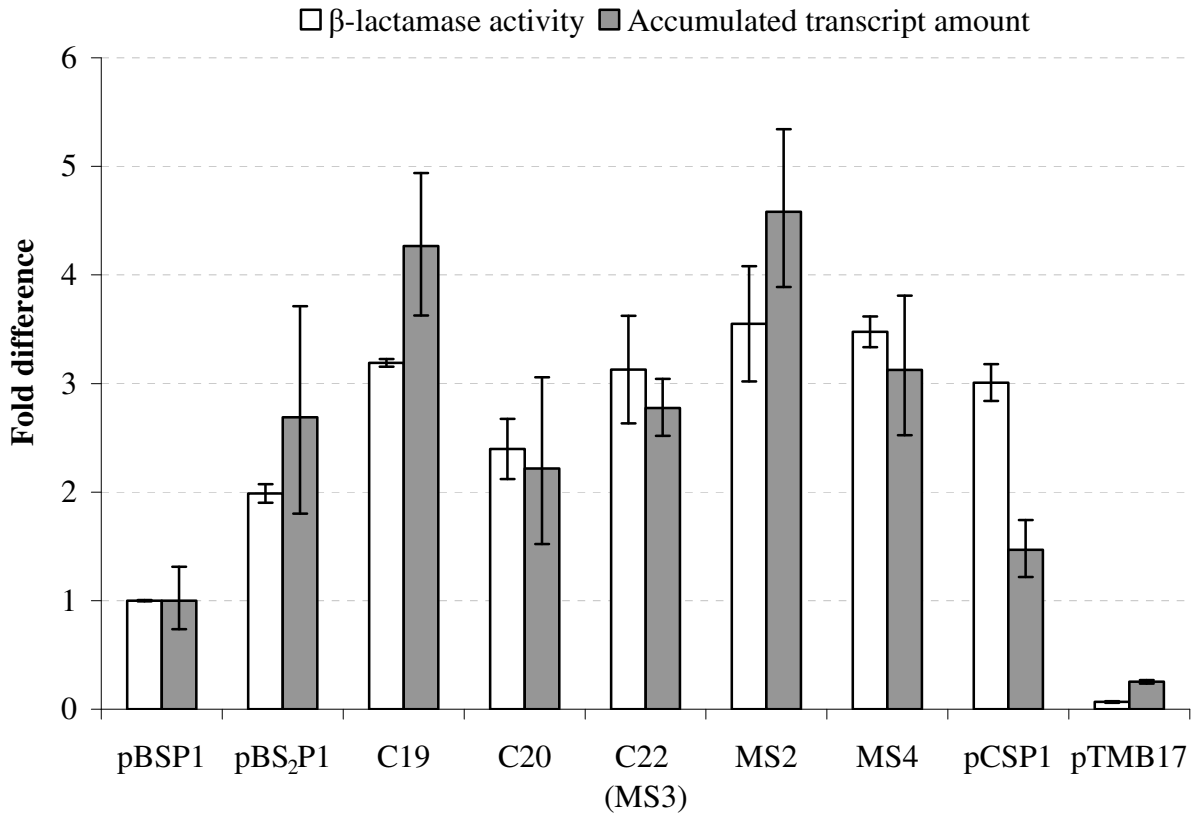


Figure 3-2: Relative expression and accumulated transcript-levels of the *bla* gene signal sequence variants at 0.1 mM m-Tol induction. pBSP1: native *bla* signal sequence (ss); pBS_2P1: *bla* ss with TCT second codon; C19, C20, C22 (MS3): pBS_2P1 vector containing *bla* ss variants identified in the SII library; MS2, MS4: pBS_2P1 vector containing *bla* ss variants identified in the SI library; CSP1: CSP instead of *bla* ss; pTMB17: deleted *bla* ss. β-lactamase protein expression was characterized by its cenzyme assay, performed in two biological replicas except for pTMB14; error bars indicate standard deviations between them. Accumulated transcript levels were determined by qRT-PCR, performed in two or three biological replicas; error bars indicate the accumulated incorporated standard deviation. Values are relative to pBSP1 which was arbitrarily set to 1. Data is shown in appendices E, F and I.

The relative protein expression values between the candidates followed the relative ampicillin tolerance values reasonably well, except MS4 which did not show a correspondingly higher amount than the others (**Figure 3-2**, **Table 3-2**). The fold differences between candidates were lower for the determined active enzyme amounts than the ampicillin tolerance fold differences. In accordance with the ampicillin tolerance levels, the SII library candidates did not show improvement over the SI library candidates. The accumulated transcript values were also increased relative to pBSP1, corresponding to the increase in protein expression. pCSP1 showed a lower relative accumulated transcript than protein expression. pTMB17 protein expression was 7 % that of the wild-type pBSP1, and its transcript levels 27 %.

Western blot analysis of total protein from the candidates was performed, which confirmed the increases in β-lactamase protein production relative to the wild type pBSP1 (**Figure 3-3**). The protein amounts for the synonymous signal sequence candidates corresponded to the relative enzyme amounts from the assay. No protein was detected from the pTMB17 expression vector.

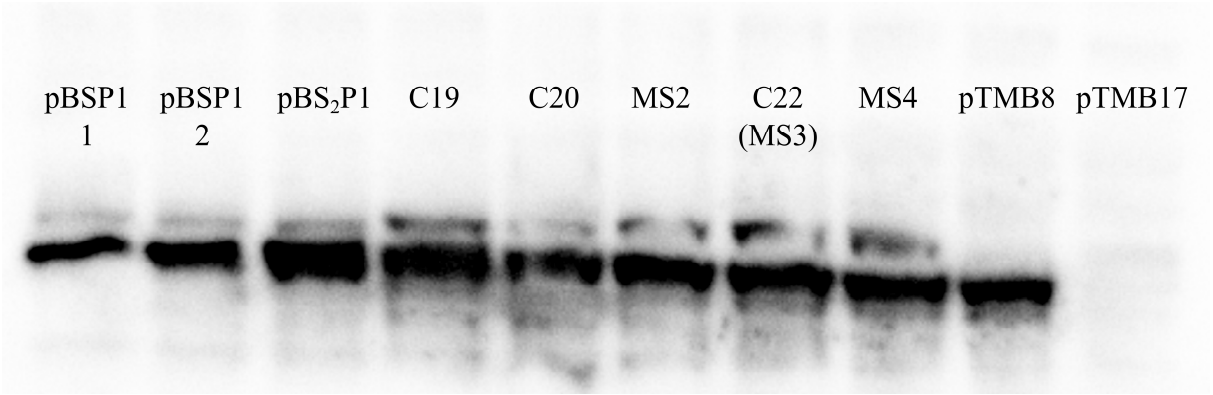


Figure 3-3: Western blot analysis of *bla* signal sequence synonymous signal sequence candidates from SI and SII libraries at 0.1 mM inducer. pBSP1 1and pBS_2P1 were prepared in a separate cultivation from pBSP1 2 and the other variants. 21.7 µg total protein was loaded in each well, and the experiment was performed as described in chapter 2.2.

To identify possible secondary structures in the transcripts that might affect the transcript stability or the translational initiation efficiency leading to the observed changes in expression, their RNA sequences were analyzed using Quikfold (Markham and Zuker, 2005). The structures were predicted from the constructs' UTRs to +69, including all of the *bla* signal sequence. However, for most constructs the sequence downstream of codon 18 did not affect the predicted structure or its stability, and structures excluding the last five codons are

shown (**Figure 3-4**). For C19 and CSP, the predicted conformation and stability were affected by including the last five codons, and the structure of C19 is displayed including its full *bla* signal sequence. The CSP construct is displayed including its UTR and codons up to +69.

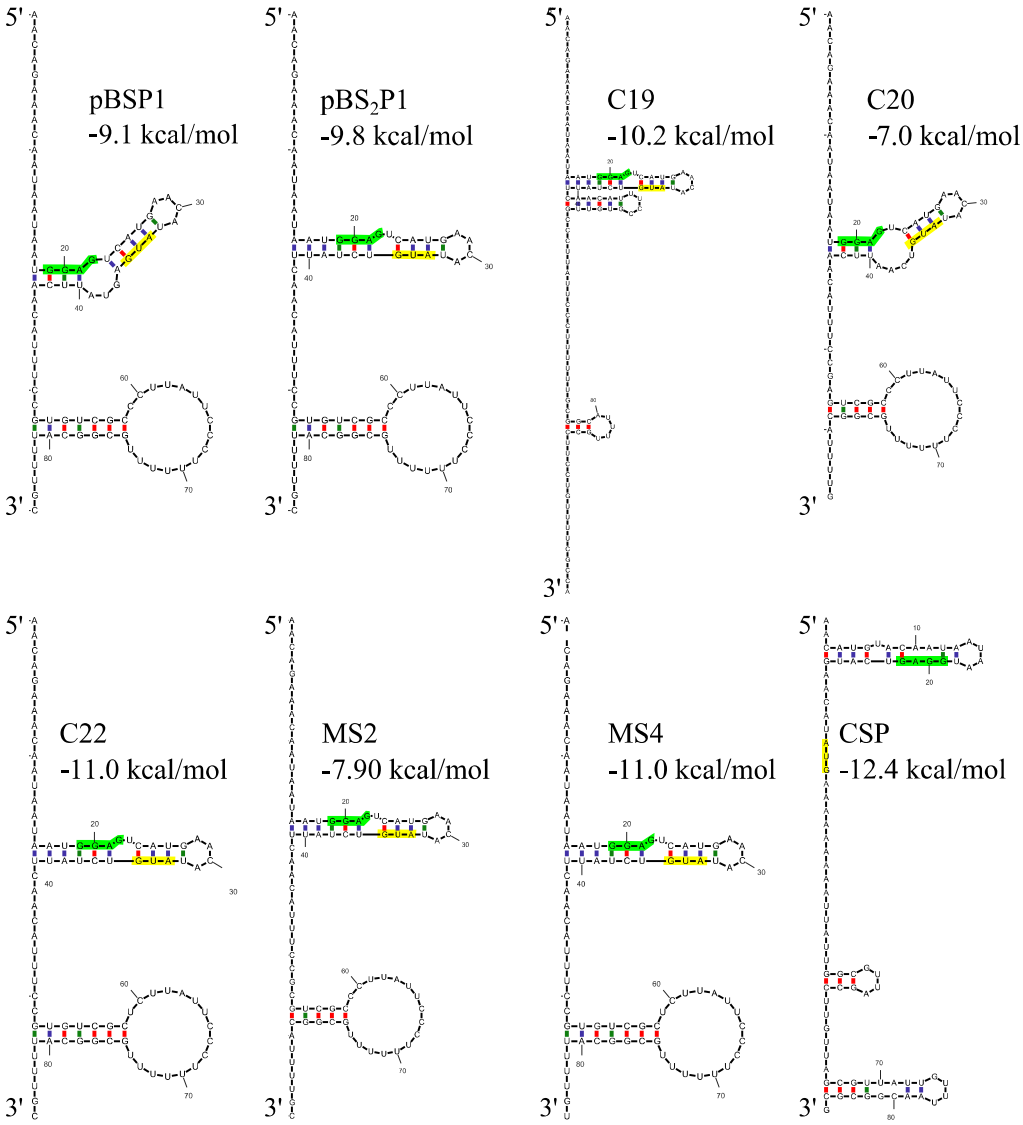


Figure 3-4: The most stable secondary structures of the different 5' synonymous *bla* transcripts predicted by Quikfold, as described in 2.4. See text for details on included sequence. pBSP1: native *bla* signal sequence (ss); pBS$_2$P1: *bla* ss with TCT second codon; C19, C20, C22: pBS$_2$P1 vector containing *bla* ss variants identified in the SII library; MS2, MS4: pBS$_2$P1 vector containing *bla* ss variants identified in the SI library; CSP: CSP instead of *bla* ss; SD sites are marked green, translational start sites are marked yellow. Red bonds indicate GC pairs. The analyzed sequences have U bases instead of T, as they were analyzed on the RNA level.

Most of the synonymous mutants were predicted to fold in similar conformations with a stem-loop forming around the TIR and one forming in the middle of the *bla* signal sequence. The predicted stabilities varied from -11.0 to -7.0 kcal/mol. The C19 *bla* synonymous ss variant was predicted to fold in a different manner, where two stem-loops formed in the TIR, and a smaller one in the middle of the ss. The predicted CSP structure had a stem-loop forming in the 5' terminal end and two towards the 3' terminal end of the analyzed sequence.

3.2.3 Investigation of mRNA Stability of Selected *bla* 5' Synonymous Variants

The C19 candidate was chosen, as the best performing *bla* signal sequence candidate from only the SII library, for further characterization by mRNA turnover analysis to investigate if increased mRNA half-lives could be the cause of the improved transcript accumulation values. Its mRNA degradation profile, along with that of pBSP1, was examined following the inducer wash-out protocol described in 2.2. The C19 synonymous *bla* signal sequence did not appear to confer increased stability on the *bla* transcript (**Figure 3-5**).

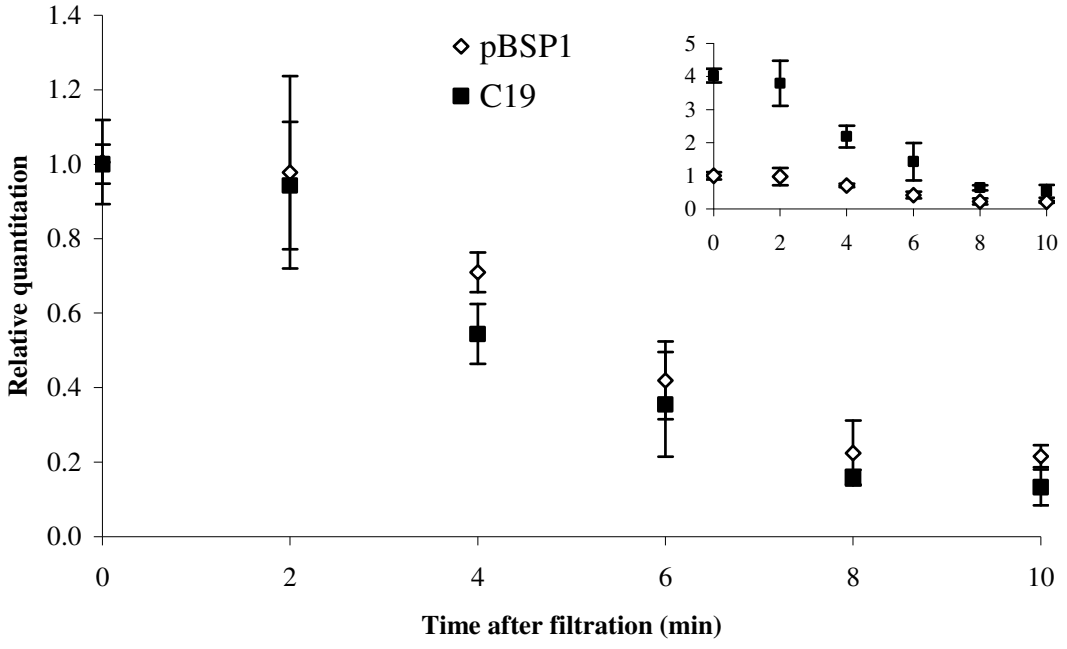


Figure 3-5: *bla* mRNA decay data for pBSP1 and signal sequence synonymous mutant variants C19 determined by the inducer wash-out method and qRT-PCR as described in chapter 2.2. Experiments were performed in three biological replicas; error bars indicate the standard deviation between them. C19 t=0 values were arbitrarily set to one to compare rate of decay; original values compared to pBSP1 t=0 are shown in the upper right corner. Data is shown in Appendix I.

3.3 2nd codon Synonymous Mutation Effects on Protein Expression

As the second codon was determined to be a major contributor to stimulation of the *bla* gene expression, this study was continued with investigation of all possible synonymous bla codons. To examine if a second codon synonymous change plays an important role in determining the expression level of other genes, two additional reporter genes, *phoA* and *celB*, were also examined with respect to the effect of synonymous 2nd codon usage.

3.3.1 β-lactamase

The second amino acid of β-lactamase is serine, and is encoded by six synonymous codons. The four remaining second codon variants of the *bla* signal sequence were constructed using the Bla.(AGC, TCC, TCA).fwd and .rev oligonucleotides (**Table 2-2**) as described in chapter 2.2, and confirmed by sequencing. Their host cells' ampicillin tolerances were determined by replica plating (2.2) at 0.05 mM induction in two biological replicas. The CAI and tAI of their 18 first codons were examined, as synonymous changes in this region was shown to affect expression for the SI and SII mutants (**Table 3-3**). The T.I.R. was found to be 2207 for all 2nd codon variants. The various 2nd codons conferred different ampicillin tolerance values to the host cells, and TCC was found to give the highest. AGT and AGC gave similar tolerance values. There was no correlation between CAI or tAI and tolerance values, and the AT content of the *bla* signal sequences did not correlate to the tolerance values either (not shown).

Table 3-3: Ampicillin tolerance of the *E. coli* DH5α host cells and sequence analysis data from the various 2nd codon synonymous *bla* signal sequence constructs.

2nd codon of *bla* signal sequence in pBSP1	Ampicillin tolerance (g/L) [1]		CAI[2]	tAI[2]	Codon change on DNA level (tAI)[3]
	Induced	Uninduced			
AGT	0.20	0.02	0.600	0.165	
TCT	1.00	0.02	0.625	0.171	AGT→TCT (0.054875→ 0.10975)
AGC	0.175	0.02	0.635	0.173	AGT→AGC (0.054875→0.125)
TCC	1.20	0.02	0.624	0.179	AGT→TCC (0.054875→0.25)

Table 3-3: Continued

2nd codon of *bla* signal sequence in pBSP1	Ampicillin tolerance (g/L) [1]		CAI[2]	tAI[2]	Codon change on DNA level (tAI)[3]
	Induced	Uninduced			
TCA	0.80	0.01	0.596	0.173	AGT→TCA (0.054875→0.125)
TCG	0.50	0.01	0.609	0.175	AGT→TCG (0.054875→0.165)

1) *Induced determined to +/- 0.025 g/L, uninduced determined to +/- 0.01*
2) *Calculated as described in 2.4 from the first 18 codons.*
3) *Predicted as described in 2.4 from Pm UTR+coding sequence*

To further characterize the *bla* synonymous 2nd codon variants, their relative expression of the *bla* gene was investigated by β-lactamase assays, and qRT-PCR analysis of the *bla* transcripts (**Figure 3-6**). The expression values followed the pattern of the ampicillin tolerance levels, except that TCC showed lower enzyme amounts than pBS_2P1. Transcript levels were also affected to a similar extent as the enzyme amounts.

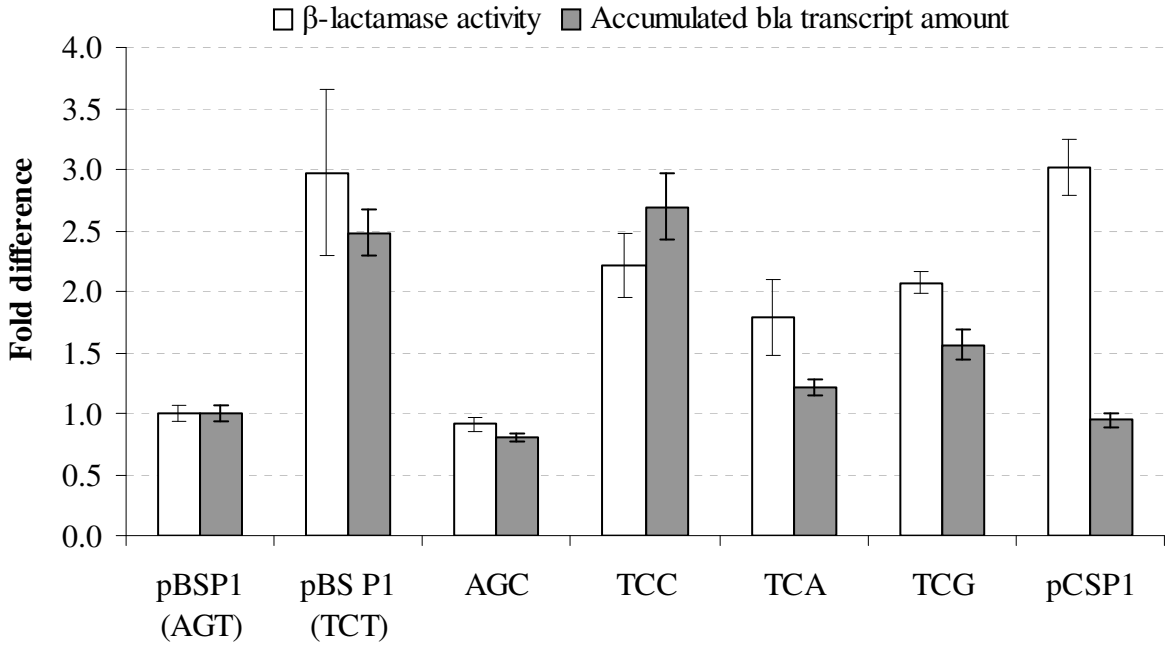


Figure 3-6: Relative β-lactamase expression and accumulated transcript-levels from the *bla* synonymous 2nd codon variants at 0.1 mM m-Tol induction. pBSP1: native *bla* signal sequence (ss), ATG as 2nd codon; pBS_2P1: *bla* ss with TCT second codon; AGC, TCC, TCA, TCG: pBS_2P1 vector with indicated 2nd codon; pCSP1: CSP instead of *bla* ss; β-lactamase protein expression was characterized by its enzyme assay, performed in at least two biological replicas; error bars indicate standard deviations between them. Accumulated transcript levels were determined by qRT-PCR in four technical replicas; error bars indicate their accumulated

incorporated standard deviations. Values are relative to pBSP1 which was arbitrarily set to 1. Data is shown in appendices E, F and I.

The β-lactamase protein expression was also examined by Western blot detection to confirm the relative protein expression values observed in the enzyme assay (**Figure 3-7**). Intensities of the protein bands appeared to correspond to the determined relative enzyme amounts, except for the higher amounts from the TCA and TCC constructs

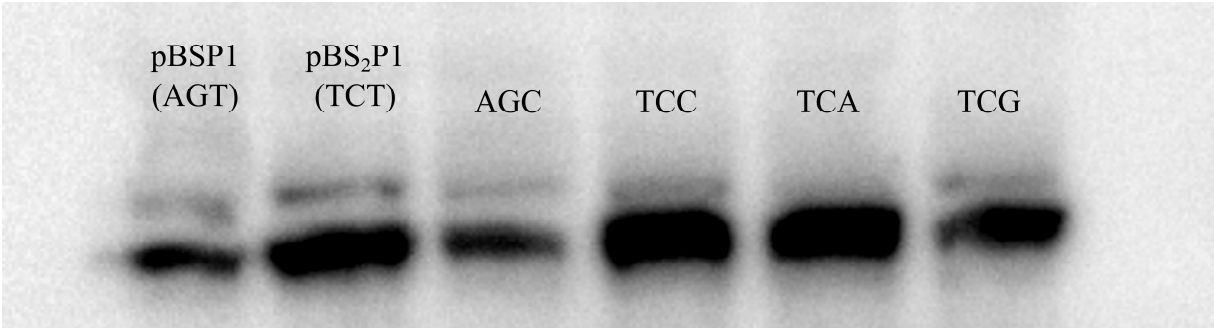


Figure 3-7: Western blot analysis of *bla* synonymous 2nd codon variants at 0.1 mM inducer. pBSP1: native *bla* signal sequence (ss), ATG as 2nd codon; pBS_2P1: *bla* ss with TCT second codon; AGC, TCC, TCA, TCG: pBS_2P1 vector with indicated 2nd codon; 21.7 µg total protein was loaded in each well, and the experiment was performed as described in chapter 2.2.

The secondary structure of the 2nd codon variants were predicted as described in chapter 2.2, to investigate if varying structures or stabilities could explain the variations in expression. Prediction of the UTR and the 69 nucleotides of the *bla* signal sequence revealed two stem-loops forming where the second one was identical to the second hairpin of pBSP1 (**Figure 3-4**). The second hairpin and the nucleotides downstream of it were not affected by the 2nd codon changes (data not shown), and thus the sequence downstream of the 6th codon was excluded to study only the structure and free energy of the varying hairpin. The most stable predicted secondary structures of the mRNA transcripts from transcriptional start site to the 6th codon are shown in **Figure 3-8**.

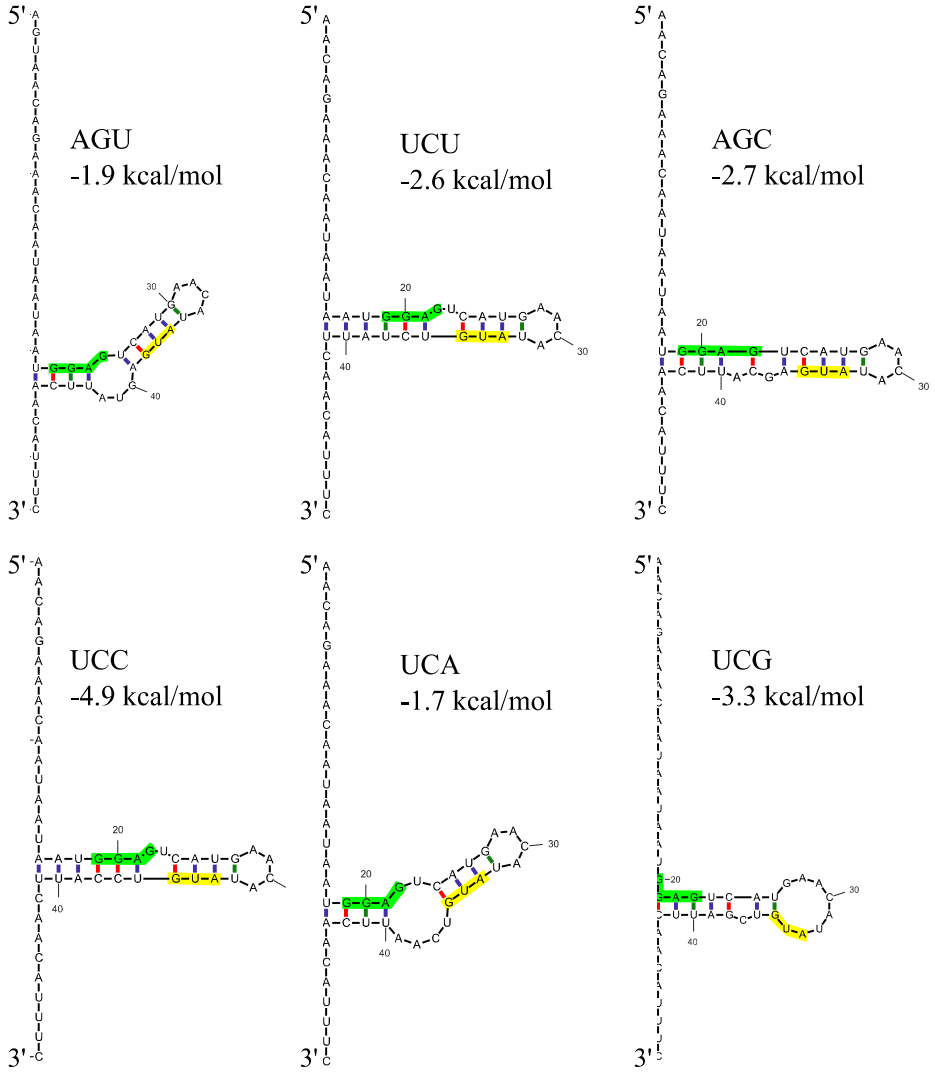


Figure 3-8: The most stable secondary structures of the synonymous 2nd codon *bla* transcripts predicted by Quikfold, as described in 2.4. See text for details on included sequence. AGU, UCU, AGC, UCC, UCA, UCG: pBSP1 variants containing indicated 2nd codon. SD sites are marked green, translational start sites are marked yellow. Red bonds indicate GC pairs. The analyzed sequences have U bases instead of T, as they were analyzed on the RNA level.

3.3.2 Alkaline Phosphatase

To investigate the effect of 2nd codon mutations on the expression level of the *phoA* reporter gene encoding alkaline phosphatase, its 2nd codon AAG's synonymous mutant, AAG, was created in the pASP1 plasmid using the phoAAAG-fwd and phoAAAG-rev oligonucleotides (**Table 2-2**), and the correct insertion was confirmed by sequencing. The expression levels of the two *phoA* 2nd codon variants were analyzed by alkaline phosphatase enzyme assays and qRT-PCR. The pCSP1phoA construct containing the CSP instead of phoA's native translocation signal sequence was included as a control. The results showed insignificant

differences in protein expression and accumulated transcript levels between the two *phoA* synonymous 2nd codon mutants and the pCSP1-phoA (**Figure 3-9**).

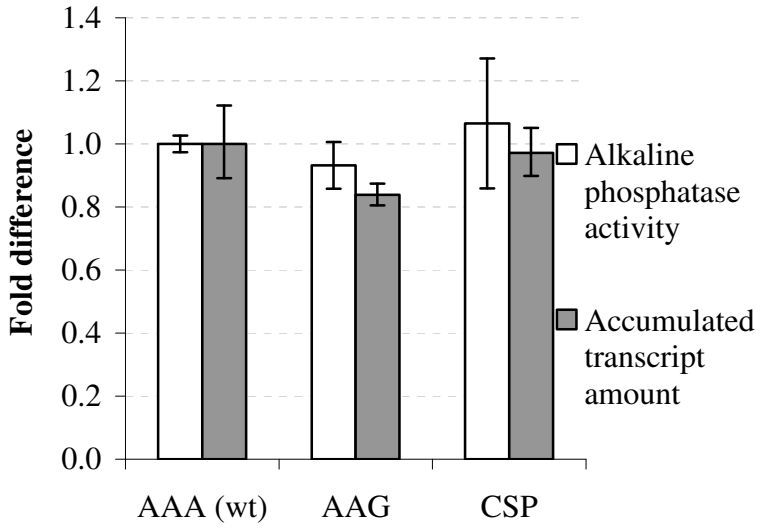


Figure 3-9: Relative alkaline phosphatase expression and accumulated transcript-levels from the synonymous 2nd codon *phoA* variants at 0.1mM m-Tol induction. AAA, AAG: pASP1 vector with indicated 2nd codon; CSP: pCSP1-phoA, CSP-*phoA* fusion. Alkaline phosphatase expression was characterized by its enzyme assay in two biologicalreplicas; error bars indicate standard deviations between them. Accumulated transcript values were determined by qRT-PCR, performed in three technical replicas; error bars indicate their incorporated standard deviations. Values are relative to AAA, which was arbitrarily set to 1. Data is shown in appendices E, G and I.

CAI, tAI and Quikfold analysis of the transcripts was performed as described for the *bla* second codon variants. Lower CAI, tAI and T.I.R. and secondary structure stability was observed for the AAG 2nd codon variant (**Table 3-4**, **Figure 3-10**).

Table 3-4: Sequence analysis data of the synonymous 2nd codon *phoA* variants.

Construct	CAI[1]	tAI[1]	Codon change on DNA level (tAI)	Translation initiation rate[2]
PASP1	0.552	0.222		2526
pASP1-AAG	0.518	0.208	AAA→AAG (0.75→0.24)	2207

1) Calculated as described in 2.4 from the first 18 codons.
2) Predicted as described in 2.4 from Pm UTR+ coding sequence

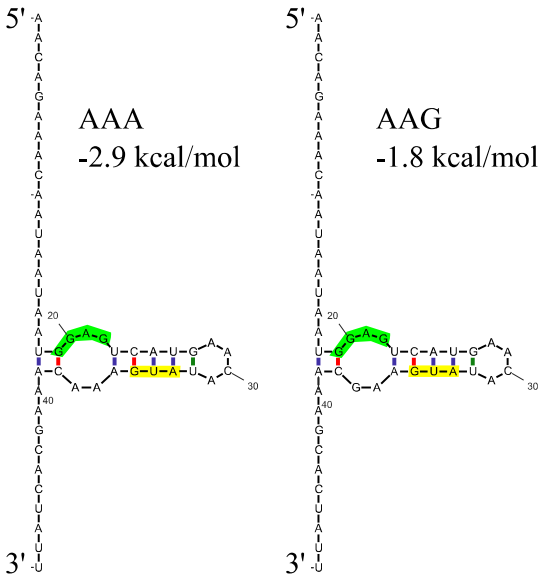


Figure 3-10: The most stable secondary structures of the synonymous 2nd codon *phoA* transcripts predicted by Quikfold, as described in 2.4. AAA, AAG: pASP1 variants containing indicated 2nd codon. SD sites are marked green, translational start sites are marked yellow. Red bonds indicate GC pairs. The analyzed sequences have U bases instead of T, as they were analyzed on the RNA level.

3.3.3 Phosphoglucomutase

The 2nd codon's effect on gene expression was also investigated for phosphoglucomutase (*celB*). Its 2nd codon synonymous mutants (CCA, CCG, CCT) were created in the pLB11 plasmid using the celB.X.fwd and celB.X.rev oligonucleotides (**Table 2-2**), and the correct substitutions were confirmed by sequencing. The expression levels from the constructs were analyzed by phosphoglucomutase enzyme assays and qRT-PCR analysis. The results showed that CCA and CCT gave slightly higher expression than CCC, while CCG gave lower expression (**Figure 3-11**). Relative enzyme amounts and accumulated transcript values varied accordingly.

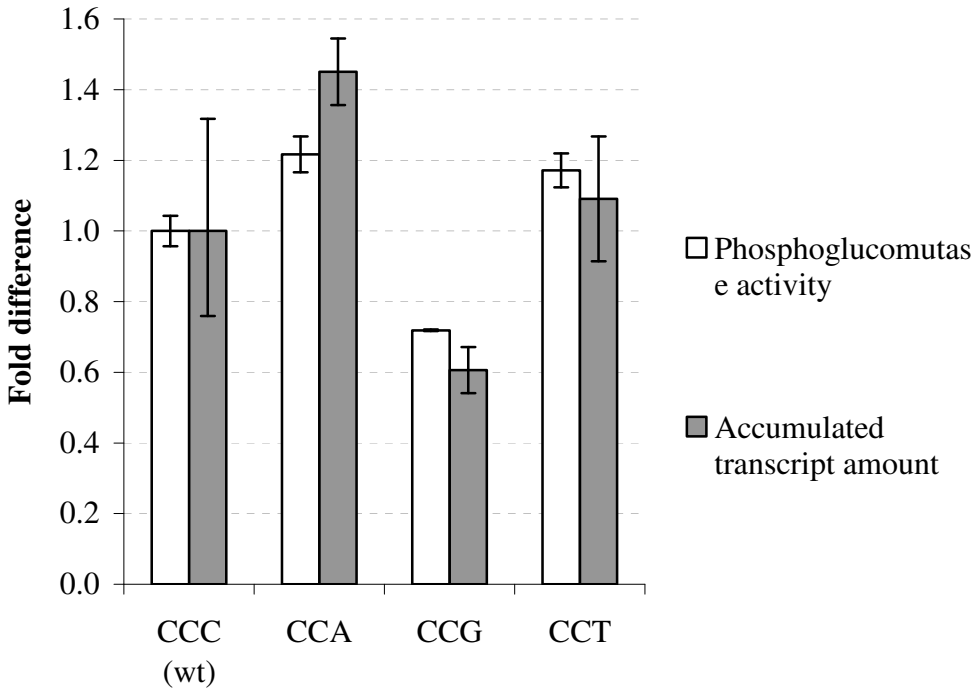


Figure 3-11: Relative phosphoglucomutase expression and accumulated transcript-levels from the synonymous 2nd codon *celB* variants at 0.1 mM m-Tol induction. CCC, CCA, CCG, CCT: pLB11 vector with indicated 2nd codon. Phosphoglucomutase expression was characterized by its enzyme assay in two biological replicas; error bars indicate standard deviations between them. Accumulated transcript values were determined by qRT-PCR, performed in two biological replicas; error bars indicate the standard deviation between them. Values are relative to CCC, which was arbitrarily set to 1. Data is shown in appendices E, H and I.

CAI, tAI and Quikfold analysis of the transcripts were performed on the transcript sequences as described for the *bla* second codon variants. Predicted secondary structures are shown in **Figure 3-12**, and transcript data are shown in **Table 3-5.** The CAI and tAI did not correlate with expression levels, but there was an indication of positive correlation between T.I.R. and relative phosphoglucomutase amounts (Pearson: 0.913, P=0.087). The AT content did not correlate to the changes in expression (not shown), and no change in predicted secondary structures or their stabilities were observed by the changed 2nd codons.

Table 3-5: Sequence analysis data of the synonymous 2nd codon *celB* variants.

Construct	CAI[1]	tAI[1]	Codon change on DNA level (tAI)	Translation initiation rate[2]
pLB11 (CCC)	0.530	0.204		3461
pLB11-CCA	0.557	0.204	CCC→CCA (0.125→0.125)	3461
pLB11-CCG	0.597	0.207	CCC→CCG (0.125→0.165)	2764
pLB11-CCT	0.546	0.195	CCC→CCT (0.125→0.054875)	3461

1) Calculated as described in 2.4 from the first 18 codons.
2) Predicted as described in 2.4 from Pm UTR+ coding sequence

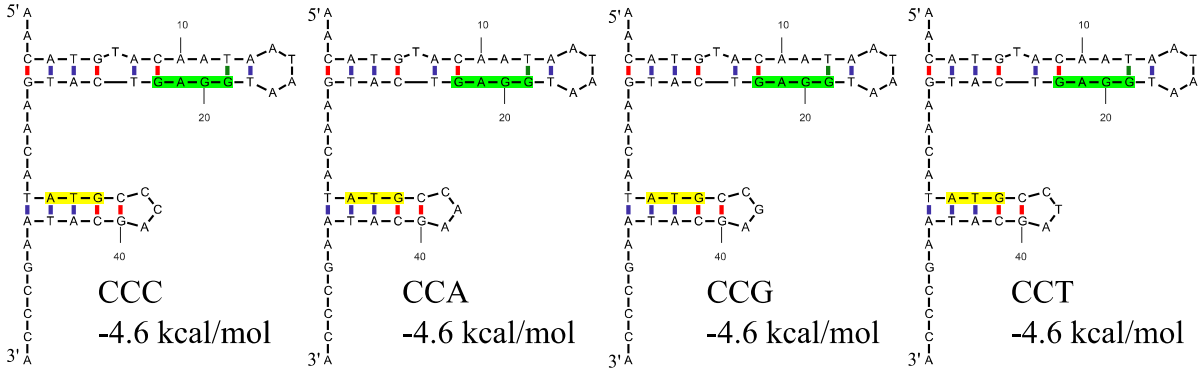


Figure 3-12: The most stable secondary structures of the synonymous 2nd codon *celB* transcripts predicted by Quikfold, as described in 2.4. CCC, CCA, CCG, CCT: pLB11 variants containing indicated 2nd codon. SD sites are marked green, translational start sites are marked yellow. Red bonds indicate GC pairs. The analyzed sequence has U instead of T as it was analyzed on the RNA level.

3.4 Randomization of a 5' Fusion Partner Sequence and its Application for Effective *ifn-α2b* Expression

As the SII *bla* 5' coding region synonymous library did not give great increases in expression, a different approach to optimizing 5' coding sequences of heterologous genes was taken. It was recently observed that fusions of short 5' terminal fragments of the highly expressed *celB* gene to the 5' coding region of the difficult-to-express *ifn-α2b* gene allowed high-level expression of the protein (Kucharova, unpublished). In order to increase the fusion partner's positive effect on expression, optimization of its coding sequence was performed by a randomized library of the $celB_{69}$ sequence fused to the reporter gene *aac(3)-IV*, encoding Aminoglycoside-(3)-acetyltransferase IV. As the gene confers apramycin resistance to the host-cells, the approach allowed screening for high expression mutants of the fusion partner by plating on solid media containing increasing concentrations of the antibiotic.

3.4.1 Combinatorial Mutagenesis of a Short 5' Terminal *celB* Fusion Partner for Increased Recombinant Gene Expression

A wild-type $celB_{69}$-*aac(3)-IV* fusion was established in the pAR69 plasmid, and shown to confer a resistance of 0.100 g/L apramycin (Kucharova, unpublished). The library was constructed and used to transform *E. coli* DH5α cells as described in chapter 2.2, and approximately 141 000 clones were obtained and pooled. To initially characterize the library, 12 individual colonies from a control kanamycin plate were randomly picked for sequencing, which showed that five contained the wild type $celB_{69}$ gene fragment fusion and 7 contained amino acid substitutions (**Table C-2**).

To isolate fusion partners giving increased expression, the library was screened on a gradient of apramycin concentrations as described in chapter 2.2 and few fp_{69}-*aac(3)-IV* host-cells grew at high concentrations (**Figure 3-13**). Twenty-four individual colonies growing at 0.12 to 0.70 g/L apramycin, termed D1-D24, were picked for sequencing. Fourteen unique $celB_{69}$ based fusion partners were identified (**Table 3-6**), D9 and D18 were identical in their fusion partner sequence, as were D20 and D22. The variants' AT contents were calculated and the nucleotide point mutation average for the displayed variants was 5.5, the average amino acid change was 4.23. The apramycin tolerance levels of the *E. coli* host cells were determined by

replica plating at 0.05 mM inducer levels after transformation by the constructs harboring the unique fusion partners. The concentrations used in the replica plating ranged from 0.1 to 0.85 g/L apramycin, as no constructs had been isolated from higher than 0.70 g/L. However, six constructs were found to confer strong growth at 0.85 g/L in the replica plating, and thus their upper tolerance levels were not determined. The tolerance levels of all variants were higher than what they had been isolated from, and a clear trend was seen where most of the fusion partners had an amino-acid change in the second codon from proline to serine. A correlation was also found between rising AT content of the fusion partner and ampicillin tolerance (Pearson 0.648, P=0.009), indicating a higher AT content positively affecting expression.

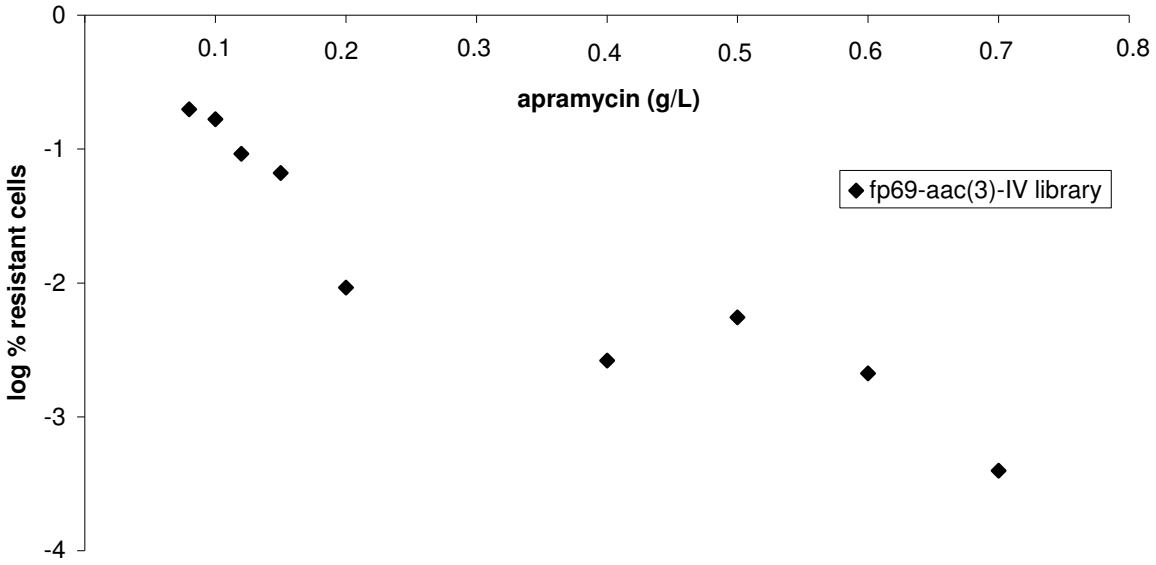


Figure 3-13: The *fp*$_{69}$*-aac(3)-IV* library was plated on solid LA with increasing apramycin concentration at 0.05 mM m-Toluate induction, and the *E. coli* host cell's survival frequencies were determined as % survival at given concentrations compared to the growth on solid media containing only kanamycin. Survival frequency is given as $\log_{10}$ of the % survival at the given ampicillin concentration. Data is shown in Appendix D.

Table 3-6: DNA and amino acid sequence of *celB*$_{69}$ based fusion partners of the up-mutants identified from screening of the *fp*$_{69}$*-aac(3)-IV* library.

	DNA Sequence	Translated amino acid sequence	Apramycin tolerance (g/L)	% AT[2]
***celB*$_{69}$**	ATGCCCAGCATAAGCCCATTTGCCGGCAAGCCGGTCGATCCGGACCGTCTTGTCAATATCGACGCCCTGGC	MPSISPFAGKPVDPDRLVNIDAL	0.100	39.4
D2	...T.T.CT..T.....	.ST.....	>0.850[1]	43.7
D3	...T.......T.T.A.....T.....	.S..IT.S.....	>0.850[1]	45.1
D4	.....A.......T.......A....T.........T.......G.....C.....	....I..T....V..G.L.....	>0.850[1]	45.1
D5	...T.....................C...........T....C..G.....	.S......R...Y.E.....	0.675	42.3
D7	...T...CT.A.........C.......G..............T...T..T.....	.STK.....R......FF.....	>0.850[1]	43.7
D9/D18	...T...TA....T.....	.SI.I.....	>0.850[1]	45.1
D10	...A.A.......A...T.A.T.....	.T..N.YS.....	0.725	45.1
D11	...T.T.......A.......A..............T.......T.....T......C.....	.S..N..T....V..C.F.T...	>0.850[1]	47.9
D13	...A.A....A.....A.A..A.....................A...A.....A.....	.T.K.QIT.......H..K....	0.700	46.5
D15	...T.........T.....................C.....	.S..I.......H.....	0.725	42.3
D17	...T..........TG...G.....................A.....	.S...AC.......E.....	0.550	42.3
D19	...T............A..........T.......C...AC..A....C.....	.S...Q...M..HHE.P.....	0.850	43.7
D20/D22	...T.....	.S.....	0.750	40.8
D21	...T.........A..T.....	.S..NL.....	0.850[1]	43.7

1) Strong growth at 0.850 g/L apramycin

2) AT content of the celB$_{69}$ *based fusion partner*

The *celB*$_{69}$ gene fusion adds 23 amino acids (celB$_{23}$), to the AAC(3)-IV protein and it is possible that the various amino acid composition of the screened N-terminal tags could have altered the conformation of the proteins, which could affect their enzymatic activity. Thus the increased apramycin resistance observed for some of the *celB*$_{69}$ based fusion partners could be due to a more effective enzyme instead of the desired increase in expression. In order to identify fusion partners that increase the expression levels, the 14 unique mutants were evaluated by relative quantitation of their *aac(3)-IV* transcript amounts. The qRT-PCR results showed one variant, D11, giving approximately 4.5-fold increase in the amount of transcript compared to the wild-type *celB*$_{69}$ sequence (**Figure 3-14**). Three variants showed reduced accumulated transcript values, while nine showed moderate increases. There was no significant correlation between the accumulated transcript amounts and the AT content.

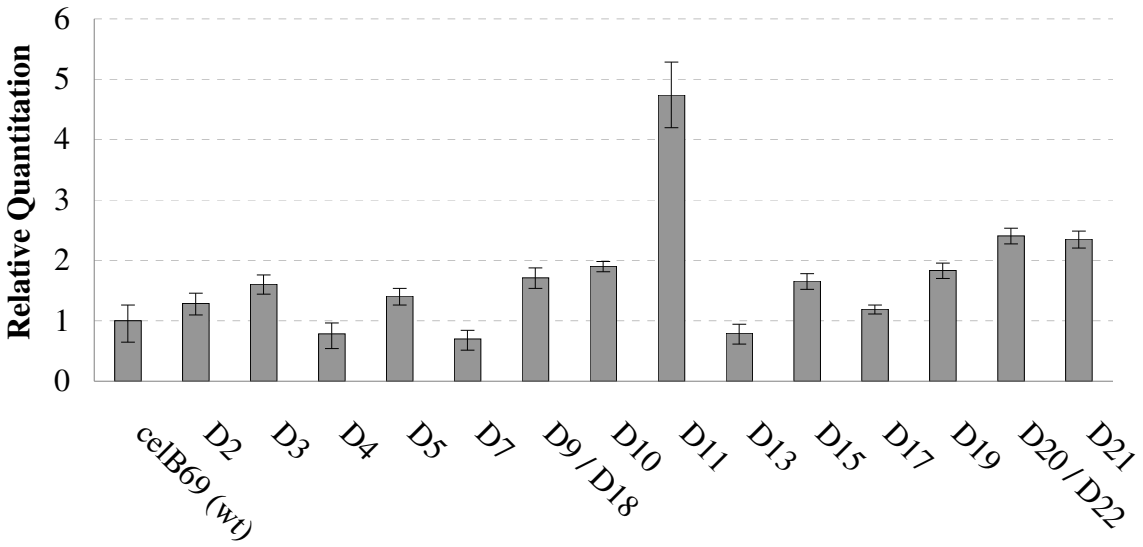


Figure 3-14: Accumulated *aac(3)-IV* transcript amount relative to *celB*$_{69}$*-aac* of the mutants identified by screening the *fp*$_{69}$*-aac(3)-IV* library, determined by qRT-PCR. All experiments were performed in three technical replicas except D11, which was analyzed in two biological replicas to verify the increase in accumulated transcript. D11's error bars indicate standard deviation between the two biological replicas, other variants' error bars indicate the incorporated standard deviations from their technical replicas. Data is shown in Appendix I.

3.4.2 Increased Expression of Human Interferon α2b by Application of an Optimized celB23 based N-terminal Fusion Partner

Given the positive effect on *aac(3)-IV* transcript levels observed from the D11 $celB_{69}$ based fusion partner, the 5' terminal $celB_{69}$ fusion to *ifn-α2b* in the pIN69 expression vector was replaced by the synthetic insert of celB69-D11.fwd and celB69-D11.rev, as described in chapter 2.2, to create pIN69-D11. IFN-α2b protein expression levels from the two vectors, in addition to expression levels from pAT64, containing the pelB translocation peptide fusion, and pAT63, without a peptide fusion, were compared by Western blot analysis, showing a clear increase in intensity for the D11 fusion partner compared to the wild type $celB_{69}$ (**Figure 3-15**). D11-IFN-α2b also showed stronger intensities than the high-level expression proven pelB-IFN-α2b fusion (Sletta et al., 2007).

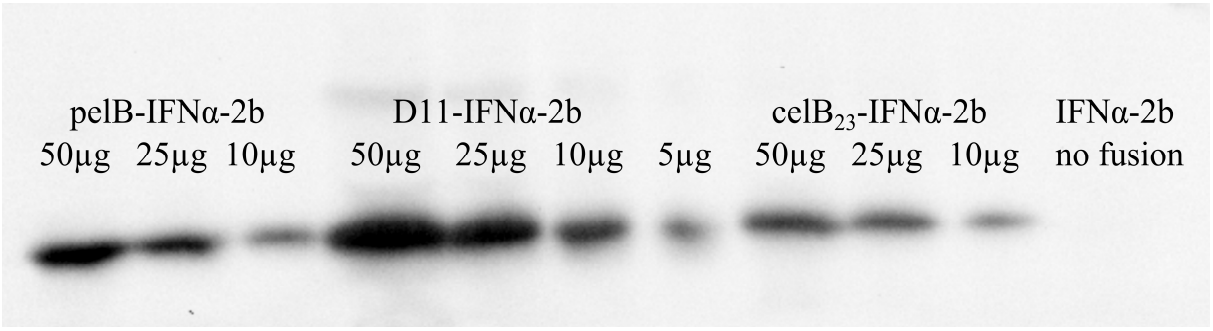


Figure 3-15: Chemiluminescent Western blot detection of IFN-α2b with the N-terminal pelB, D11 and celB23 fusion peptides, and IFN-α2b without an N-terminal fusion. The pelB fusion and IFN-α2b without a fusion peptide were included as controls. Total protein from cell extracts was loaded in dilution series, in amounts shown in the figure for all fusion-IFN-α2b samples. 50 µg total protein from sample IFN-α2b without fusion partner was loaded.

Relative quantification of expressed IFN-α2b was determined using Image Lab 4.0 software (Bio-Rad) by analysis of the intensity of the bands, showing a three-fold increase in protein from the D11 fusion partner compared to the wild-type celB23 partner (Table 3-7). The D11-IFN-α2b also produced 1.5 times more protein than the pelB-IFN-α2b fusion.

To investigate the sequence features of the *pelB*, *celB*, D11-*ifn-α2b* gene fusions, as well as of the original *ifn-α2b* 5' coding sequence, their CAI, tAI and translation initiation rates were examined as described in chapter 2.4 (**Table 3-7**). Sequence from transcriptional start site to +69 was included in the analysis for T.I.R, and from +1 to +69 for tAI and CAI. No correlation was found between the tAI or CAI and the relative quantification values. The

variation in predicted T.I.R. values of the transcripts matched the detected proteins' relative quantification values, except for the *ifn-α2b* transcript without a fusion partner.

Table 3-7: mRNA sequence analysis and relative quantification of protein expression from Western blot analysis of the IFN-α2b fusion constructs.

	CAI[1]	**tAI[1]**	**Translation initiation rate**	**Relative quantification[2]**
pelB-ifn-α2b	0.422	0.208	4743	1.84
celB$_{69}$-ifn-α2b	0.575	0.227	3461	1
celB$_{69}$-D11-ifn-α2b	0.642	0.210	5679	3.76
ifn-α2b (no fusion)	0.958	0.327	4335	Not detectable

1) Calculated as described in 2.4 from the first 69 codons.
*2) Calculated relative to celB$_{69}$-ifn-α2b from band intensities of the 10 µg total protein bands of the Western Blot analysis (**Figure 3-15**) using Image Lab 4.0 software (Bio-Rad Laboratories).*

The secondary structures of the *ifn-α2b* transcripts were predicted by Quikfold as described in chapter 2.4, and the most stable predicted secondary structures are shown in **Figure 3-16**. The UTR and 69 nucleotides following the translational start codon were included to be able to compare the structures. The *ifn-α2b* transcript without a fusion partner showed no structures that included the SD or the translational start site, unlike the others. The number of- and stability of predicted structures forming in the translation initiation region was lowered in the D11 *celB$_{69}$* based fusion partner compared to the wild-type *celB$_{69}$* fusion partner.

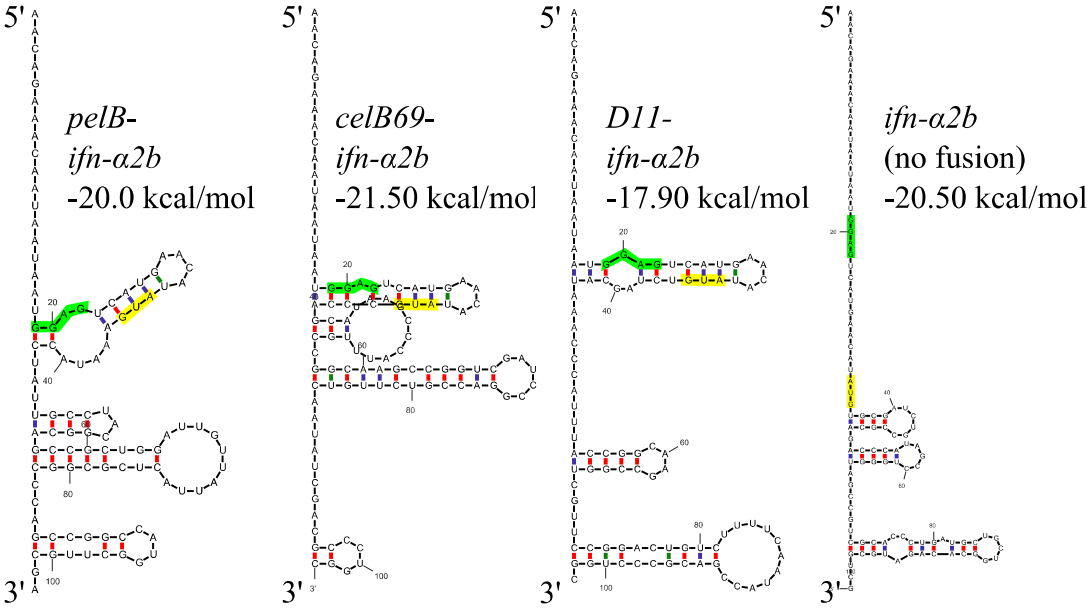


Figure 3-16: The most stable secondary structures of the fusion partner-*ifn*-α2b transcripts, predicted by Quikfold as described in 2.4. *pelB-ifn-α2b*: *pelB* fusion partner; *celB*$_{69}$*-ifn-α2b*: wild type *celB*$_{69}$ fusion partner; *D11-ifn-α2b*: D11 *celB*$_{69}$ based fusion partner; *ifn-α2b*: *ifn-α2b* gene expressed without a fusion partner. SD sites are marked green, translational start sites are marked yellow. Red bonds indicate GC pairs. The analyzed sequences have U bases instead of T, as they were analyzed on the RNA level.

4 Discussion

The 5' terminal end of the mRNA coding sequence has been shown to affect the translational efficiency of the transcript by taking part in secondary structures and by its codon usage (Kudla et al., 2009; Tuller et al., 2010; Goltermann et al., 2011), and optimization of these features may thus lead to increased protein expression. The aim of this study was to investigate the effect of changes to the 5' coding sequence of recombinant mRNAs on protein expression in order to improve heterologous protein production. This was examined by three different approaches; combinatorial synonymous mutagenesis of the 5' end of the *bla* gene, introduction of 2nd codon synonymous mutations in three reporter genes, and lastly combinatorial mutagenesis of a 5' fusion partner leading to high-level expression of a medically important protein. The results obtained by the three codon optimization strategies will be discussed in the following chapters.

4.1 Selection for Enhanced 5' *bla* Coding Sequences by Directed Evolution

By changing the 5' coding region synonymously with respect to its codon usage, the transcript's propensity to form secondary structures in the translation initiation region may be altered while retaining the original amino-acid sequence. This is an advantage over mutations that change the peptide sequence as such changes may lead to altered biological activity of the produced protein.

A synonymous library, termed SII, was constructed in the 5' coding sequence of the *bla* gene, conserving a positive 2nd codon mutation effect identified in a previous library (SI library). Screening of the SII library yielded three synonymous variants, termed 19, C20 and C22, conferring 8.25, 5.25 and 8.5 fold higher ampicillin tolerance than the *bla* wild-type, respectively (**Table 3-1**). The improvement over pBS$_2$P1 containing the AGT→TCT 2nd codon mutation was 1.65, 1.05 and 1.70-fold, respectively. C22 had been identified in the SI library, and the SII candidates did not confer higher ampicillin tolerance than the best mutants from the SI library, which were also characterized in this study. The identification of relatively few synonymous candidates from the SII library indicated that besides the dramatic effect from the 2nd codon change, additional alteration of the *bla* 5' coding region does not lead to big improvements in expression. Besides, 83% of the sequenced SII candidates had

non-synonymous mutations or deletions in the mutated *bla* signal sequence, indicating that the selection pressure from the library screening leads to selection for any *bla* gene variant that allows the host-cells to survive.

Characterization of expression levels by qRT-PCR analysis and β-lactamase enzyme assays showed that all the tested 5' synonymous *bla* mutants conferred increased levels of both accumulated transcript and active protein (**Figure 3-2**). Western blot analysis also confirmed increases in protein amounts (**Figure 3-3**). One possible explanation for the increases in the expression levels of the synonymous candidates might be an increased translational initiation efficiency of the transcript. That could in turn lead to better resistance from RNaseE degradation for the transcript, due to closer ribosomal spacing (Deana and Belasco, 2005). The changes in the sequence could also lead to higher stability of the transcripts by affecting the secondary structures of their 5' ends (Deana et al., 2008). To investigate if increases in mRNA stabilities could be causing the higher accumulated transcript levels, the transcript degradation profile of the best performing SII candidate, C19, was examined using the inducer wash-out method. When compared with the pBSP1 wild-type, it was clear that the stability of the C19 transcript was not increased (**Figure 3-5**), and it appeared to that of the original *bla* gene. The stability of the C22 candidate's transcript was also determined to be similar to the wild type *bla* transcript (Kucharova, unpublished).

The sequences of the different 5' *bla* coding regions were subjected to bioinformatics analyses to identify reasons for the increased expression values. The RBS calculator predicted no change in T.I.R. for most synonymous mutants (**Table 3-1**), and the changes in the predicted secondary structures did not appear to allow better ribosomal access to the transcript as stabilities and structures did not correspond to expression levels (**Figure 3-4**). Interestingly, the CSP sequence, which served as a positive control in the expression experiments, had a Quikfold predicted structure with no hairpins forming in the translation initiation and its T.I.R. was predicted to 4279, compared to the 2207 predicted for most synonymous candidates (**Table 3-1**). It also showed a 2 to 1 translation to transcription ratio, indicating that the bioinformatics tools could be used for translational efficiency analysis. The fact that the translation to transcription ratio for the *bla* synonymous mutants were similar, which was supported by the bioinformatics tools, indicated that it was the transcription of the genes that was improved by the synonymous mutations of the 5' *bla* sequence.

An explanation for the increased accumulated transcript values might be that the mutations directly affect transcriptional efficiency. Features of the DNA sequence were suggested to cause transcriptional stalling up to 108 nucleotides downstream of the transcriptional start-sites of the *lacZ*, *cspA* and *tnaA* σ^{70} promoters (Hatoum and Roberts, 2008). If such a feature is present in the *bla* 5' coding region, it might have been changed by the synonymous mutations, leading to less frequent stalling and thus more efficient transcription as stalling may lead to termination (Weisberg and Gottesman, 1999).

The increase in transcriptional efficiency may also be explained by changes in promoter escape efficiencies by the RNA polymerase. Changes in the DNA region corresponding to the UTR of the *Pm* promoter was shown to strongly stimulate transcription of the *bla* gene (Berg et al., 2009). However, the effect was shown to be dependent on the 5' coding region of the transcript as well, as *celB* and *ifn-α2b* transcript accumulation rated did not increase to the same extent when expressed with the strong UTRs (Berg et al., 2012). As there is an interplay between the UTR and the coding region determining transcriptional efficiency, the efficiency might be affected by the *bla* 5' coding region changes.

All the synonymous candidates with the beneficial TCT 2nd codon had obtained codons with lowered tAI, and showed higher expression than the pBS_2P1 containing only the TCT change (**Table 3-1**). Lowered tAI of the 5' coding region may cause slower translation of the 5' end of the transcript and provide a more even distribution of ribosomes on the transcript, preventing ribosome "traffic jams" and increasing the rate of successful translations (Tuller et al., 2010). The authors discovered an evolutionary conserved "ramp" of low tAI in the 5' coding region of the genes of several genomes, and also showed a correlation between such a ramp and the fitness of *E. coli* overexpressing GFP (Kudla et al., 2009; Tuller et al., 2010). Such an effect in the analyzed *bla* candidates could be leading to higher fitness of the host-cells as the burden of expressing the protein at the given expression level is reduced, and thus allow further improvement from the already strong expression from the 2nd codon change.

The combinatorial mutagenesis of the *bla* signal sequence pointed out that the 2nd codon affects gene expression to a large extent. The next investigation was therefore focused on the characterization of the discovered effect.

4.2 2nd codon Synonymous Mutation Effects on Protein Expression

To study the 2nd codon's effect on gene expression, variants were constructed of three reporter genes *phoA*, *celB* and *bla,* with all synonymous mutations in their 2nd codons, encoding lysine, proline and serine, respectively. The effect on expression was most apparent in *bla* (**Figure 3-6**), where the relative expression could be improved three-fold, as opposed to *phoA* where only small changes in expression were observed (**Figure 3-9**). *celB* showed slight variations in expression from approximately 0.6- to 1.3-fold changes in transcript and active enzyme values (**Figure 3-11**).

As predicted from the *bla* library screenings, the TCT 2nd codon gave the best expression of the serine variants. TCC gave the second highest, followed by TCG and TCA, and AGC gave as low expression as the wild-type AGT. Western blot of the samples (**Figure 3-7**), showed a slightly weaker than expected band for TCG, which had about the same level of active enzyme as TCC and TCA. There was no correlation between the CAI, tAI, the stability of the predicted secondary structures, or the predicted T.I.R to the accumulated transcript levels or the relative active enzyme amount, indicating a codon-specific effect. A study analyzing all possible 2nd codons in the *lacZ* and the 3A' assay system also indicated a codon-specific effect on expression, that was not dependent on mRNA stability or secondary structures (Stenstrom et al., 2001). The study included codons for all amino acids, but also identified differences between synonymous 2nd codons. The authors identified high adenine content in position 1 and 2 as an important factor for high expression, and AGT and AGC as giving the highest expression for serine, with an average of 5.44- and 4.90-fold protein amounts, respectively, relative to the lowest expressing TCC. When comparing this study with the characterization of the *bla* 2nd codons where AGT and AGC gave the lowest expression, and TCC gave the second highest, it appears that the observed effects from the 2nd codon are gene specific.

The expression levels from the AGT and AGC variants were similar (**Figure 3-6**), indicating a tRNA specific effect as the two are both decoded by Ser3 tRNA (Stenstrom et al., 2001). TCT and TCC, which gave similar high expression in this study, are both decoded by Ser5, which further strengthens the notion of tRNA specific effects affecting the expression. This was also the case for the *phoA* gene, as the AAA and AAG codons encoding lysine are

decoded by the same tRNA and gave the same expression. Similarly, a study expressing human immunoglobulin α also showed no change in expression between AAA and AAG in the 2nd codon (Bivona et al., 2010). In contrast, major differences in expression were found between AAA and AAG in other studies (Looman et al., 1987; Stenstrom et al., 2001), further indicating that the 2nd codon effect is gene specific as well as tRNA specific. The comparison with the well expressed protein containing CSP fused to the mature AP enzyme indicates that the original alkaline phosphatase has a well-adapted translocation signal peptide, and there is a possibility that small changes from the synonymous mutations might have been over-shadowed by its already effective expression.

For the *celB* 2nd codon variants there was no correlation between expression and CAI or predicted secondary structure stability (**Figure 3-11**, **Table 3-5**). However, there was an indication of a positive correlation between T.I.R. and relative enzyme amount (Pearson: 0.913, P=0.087), as the CCG 2nd codon gave a lower predicted T.I.R. and expression than the others. A lower rate of translation could lead to lower mRNA stability, as the decrease in ribosomal density on the transcript allows RNaseE easier access (Deana and Belasco, 2005), which could explain the lowered accumulated transcript and enzyme amount. A tRNA specific effect was not present in this case, as the CCA and CCG codons are both decoded by the Pro3 tRNA. The contribution from the 2nd codons on expression did not correlate with Stenstrom et al., 2001, as they identified CCA giving the lowest relative expression, further supporting that the 2nd codon effect is sequence context specific.

In contrast to the three-fold changes in expression of active β-lactamase observed in this study (by varying the serine 2nd codon) the protein amounts stayed the same when expressing human immunoglobulin α with the six variants as the 2nd codon (Bivona et al., 2010). This, along with the discrepancies between this study and others (Looman et al., 1987; Stenstrom et al., 2001; Bivona et al., 2010), strongly suggests that the effect on expression from synonymously mutating the 2nd codon is gene dependent, and probably influenced by the downstream nucleotide sequence. As transcript levels are affected by the 2nd codon, the observed effects may arise due to increased transcriptional efficiency as described for the *bla* synonymous library. There also appears to be tRNA specific effects for some codons or genes, but not for others, as was previously observed (Looman et al., 1987; Stenstrom et al., 2001).

4.3 Randomization of an N-terminal CelB Fusion Peptide for Increased Protein Expression

The features causing the improved expression of the synonymous mutations were hard to pinpoint, and did not show sufficient potential to improve expression of poorly expressed heterologous genes. It was recently shown that the first 69 nucleotides of the *celB* greatly stimulates expression of *ifn-α2b* when fused 5' to its coding sequence (Kucharova, unpublished), and this fusion was chosen for optimization instead of proceeding with synonymous mutations. To further improve the $celB_{69}$ gene fusion tag, a combinatorial mutagenesis library was created in the $celB_{69}$-*aac(3)-IV* gene fusion and screened for increased apramycin resistance. 14 unique $celB_{69}$ sequences conferring higher apramycin resistance to the cells were identified (**Table 3-6**), indicating their possible positive effects on expression. Interestingly, also in this library the 2nd codon appeared to have an important effect on expression, as most of the candidates were changed from CCC to TCC, changing the amino acid from proline to serine.

In the previously mentioned study investigating the effect of the 2nd amino acid on the expression of human immunoglobulin α, a ten-fold variation in expression was observed by its exchange with various amino acids (Bivona et al., 2010). Both proline and serine as the 2nd residues were found to give similar high expressions in the tested protein. This again indicates that the 2nd codon effect is gene specific, as there was a positive selection for serine in the sequence variants based on $celB_{69}$, which has a native proline 2nd residue. The Bivona et al. study also identified a bias towards serine and alanine as the second amino acids of highly expressed genes of *E. coli*, which might explain the increase in expression observed in this study as the new fusion may be more adapted to high-level expression. There was also a statistically significant correlation between increased AT content of the fusion partner and higher apramycin tolerance of the host (**Table 3-6**, **Figure 3-14**), indicating higher translational efficiency due to lowered stability of secondary structures in the TIR (Krishna Rao et al., 2008; Kudla et al., 2009)

As different amino acids were incorporated in the peptide sequences, the increased resistance of the screened candidates might have been due to altered enzymatic activities. In order to identify sequence variants positively affecting the overall expression level and not the enzyme activity, analysis of accumulated transcript was carried out. qRT-PCR of the *aac(3)-IV* gene

showed that one of the fusion partners, D11, gave an almost five-fold increase in accumulated transcript compared to the wild type *celB*$_{69}$ (**Figure 3-14**). This finding indicated that the new gene fusion had been optimized with respect to increased gene expression. When fused in frame with *ifn-α2b*, the D11-*ifn-α2b* gene fusion gave three-fold higher IFN-α2b protein levels than the *celB*$_{69}$-*ifn-α2b*, and 1.5 times as high as the *pelB- ifn-α2b* fusion control included (**Figure 3-15**, **Table 3-7**).

To investigate the reasons for the variation in expression from the fusions, the UTR and the 5' coding sequences of the *ifn-α2b* gene and the *pelB-, celB*$_{69}$*-,* and D11-*ifn-α2b* fusions were analyzed for their CAI, tAI, T.I.R. and predicted secondary structures (**Table 3-7**, **Figure 3-16**). The identified high tAI and CAI values of the *ifn-α2b* fragment are due to previous codon optimization of the gene (Sletta et al., 2007). Interestingly, there were no predicted secondary structures in the 5' end of the *ifn-α2b* transcript, unlike all other transcripts analyzed in this study. As secondary structures in the 5' terminal end of the transcript lead to stabilization of the mRNA (Deana et al., 2008), it may be tempting to speculate that the *ifn-α2b* transcript without a fusion partner may be highly unstable due to the absence of 5' stem-loops, which could in turn lead to its low protein expression. The low stability of the *ifn-α2b* transcript with no fusion partner was confirmed by an inducer wash-out experiment that compared its stability to a *celB*$_{75}$-*infa2b* fusion, showing half-lives of two and six minutes, respectively (Kucharova, unpublished). The results are likely to be the same for the *celB*$_{69}$ fusion, as the extra nucleotides of the *celB*$_{75}$ are in the end of the fusion, thus probably not interfering with the secondary structures of the transcript's 5' terminal end.

A possible explanation for the three-fold increase in protein amount for the D11 fusion partner is that the predicted secondary structure of the wild type *celB*$_{69}$-*infa2b* transcript was stronger and had more hairpins forming close to the translation initiation region than the D11 fusion partner (**Figure 3-16**). This may lead to increased access to the SD sequence and the translational start site for the ribosome, thus allowing more efficient translation initiation. This could in turn further explain the increase in accumulated transcript the D11 variant causes on *aac(3)-IV* (**Figure 3-14)**, as the closer spacing of ribosomes on the transcript would protect it from degradation (Deana and Belasco, 2005). The RBS calculator also predicted that the D11-*infa2b* has a higher translation initiation rate than *celB*$_{69}$-*ifn-α2b*, further corroborating the possible stabilization by increased translation initiation.

5 Concluding Remarks

From the results obtained in this study and current literature, it is clear that the 5' coding region's effect on gene expression is very complex, and that it can affect expression in a myriad of ways. The *bla* library screenings showed that the gene expression was dramatically increased by altering the 2nd codon synonymously, and that it could be further increased by additional synonymous mutations. The main contributor to the stimulation was found to be an increase in transcriptional activity, and not increased transcript stability. The second codon variation of three reporter genes demonstrated that the sequence context at the 5' terminal end plays a crucial role in how changes in the sequence affect the expression, as the results contradicted some literature findings while supporting others.

Synonymous optimization however, did not show a great potential or a clear feature to be used for optimization of poorly expressed heterologous genes, as there was only a 3.5-fold increase in active protein amounts. Combinatorial mutagenesis optimization of an N-terminal celB fusion partner that had been shown to improve IFN-α2b expression to high-level production from undetectable levels, led to a three-fold further improvement in protein.

Based on the described findings, the effects of changes to the 5' terminal coding sequence appear to be gene specific and dependent on the adjacent genetic context. This makes it difficult to pinpoint which features should be addressed in a genetic optimization. However, as intertwined and complex these regulatory mechanisms may be, the increased expressions of the *bla* library candidates show that already established expression may be increased by simple synonymous changes of the 5' coding region, optimizing the gene sequence to the promoter and the UTR sequence. The successful identification of an improved fusion partner for IFN-α2b shows that optimization of the coding region of fusion partners is a powerful tool for enhancing the expression of poorly expressed genes.

6 Future Perspectives

More research is needed to further elucidate how the features of the 5' coding region affect gene expression. As the features are clearly influenced by their adjacent genetic context, it would be interesting to investigate the interplay between the UTR and the 5' coding region by combinatorial mutagenesis of both regions simultaneously, in for example the *bla* reporter gene system. By rationally designing the UTR and 5' coding region to maximize and minimize secondary structures and T.I.R., more insight in their contribution of *bla* transcript translation could be gained. Changing the signal sequence of *bla* to only low or high tAI codons would also be interesting, as there was a tendency of lowered tAI in the coding region of the *bla* variants.

In the applied aspect of utilizing the obtained results, it would be interesting to further develop the identified D11 fusion partner variant for use in heterologous expression. As there was a significant correlation between the fusion partner's AT content and survival of the apramycin fusion partners, further optimization could involve maximizing the AT content of D11 by synonymous mutations. The development of the candidate would also require the introduction of a protease cleavage site to be able to obtain the protein of interest without the fusion and retain its native N-terminal end, and several such cleavage sites are available (Terpe, 2003). It would also be interesting to examine the D11 fusion partner's effect on protein solubility, and measures could be taken to improve it (Fox et al., 2001; Terpe, 2003).

To further investigate the molecular aspects causing the improvement of D11 over wild-type, an inducer wash-out experiment could be performed to compare their stabilities. More bioinformatic analyses should also be done to the up-mutants of the *fp-aac(3)-IV* library to further examine reasons for their increased translational efficiencies.

References

Akcakaya, H., A. Aroymak, et al. (2007). "A quantitative colorimetric method of measuring alkaline phosphatase activity in eukaryotic cell membranes." Cell Biol Int **31**(2): 186-190.

Angov, E. (2011). "Codon usage: nature's roadmap to expression and folding of proteins." Biotechnol J **6**(6): 650-659.

Aune, T. and F. Aachmann (2010). "Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed." Applied Microbiology and Biotechnology **85**(5): 1301-1313.

Baker, T. A. and S. H. Wickner (1992). "Genetics and enzymology of DNA replication in Escherichia coli." Annu Rev Genet **26**: 447-477.

Bakke, I., L. Berg, et al. (2009). "Random mutagenesis of the PM promoter as a powerful strategy for improvement of recombinant-gene expression." Appl Environ Microbiol **75**(7): 2002-2011.

Barrick, D., K. Villanueba, et al. (1994). "Quantitative analysis of ribosome binding sites in E.coli." Nucleic Acids Res **22**(7): 1287-1295.

Belasco, J. G. (2010). "All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay." Nat Rev Mol Cell Biol **11**(7): 467-478.

Belasco, J. G., G. Nilsson, et al. (1986). "The stability of E. coli gene transcripts is dependent on determinants localized to specific mRNA segments." Cell **46**(2): 245-251.

Berg, L., V. Kucharova, et al. (2012). "Exploring the 5′-UTR DNA region as a target for optimizing recombinant gene expression from the strong and inducible Pm promoter in Escherichia coli." J Biotechnol **158**(4): 224-230.

Berg, L., R. Lale, et al. (2009). "The expression of recombinant genes in Escherichia coli can be strongly stimulated at the transcript production level by mutating the DNA-region corresponding to the 5'-untranslated part of mRNA." Microb Biotechnol **2**(3): 379-389.

Bethesda (1986). BRL pUC host: E. coli DH5α competent cells. Bethesda Research Lab Focus, Bethesda Research Laboratories. **8:** 9.

Bhandari, A., W. Kim, et al. (2010). "Luminol-Based Enhanced Chemiluminescence Assay for Quantification of Peroxidase and Hydrogen Peroxide in Aqueous Solutions: Effect of Reagent pH and Ionic Strength." Journal of Environmental Engineering-Asce **136**(10): 1147-1152.

Bivona, L., Z. Zou, et al. (2010). "Influence of the second amino acid on recombinant protein expression." Protein Expr Purif **74**(2): 248-256.

Blatny, J. M., T. Brautaset, et al. (1997). "Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon." Appl Environ Microbiol **63**(2): 370-379.

Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." Anal Biochem **72**: 248-254.

Brautaset, T., R. Standal, et al. (1994). "Nucleotide sequence and expression analysis of the Acetobacter xylinum phosphoglucomutase gene." Microbiology **140 (Pt 5)**: 1183-1188.

Brown, G. W. and D. S. Ray (1992). "Purification and characterization of DNA ligase I from the trypanosomatid Crithidia fasciculata." Nucleic Acids Res **20**(15): 3905-3910.

Brown, R. B. and J. Audet (2008). "Current techniques for single-cell lysis." J R Soc Interface **5 Suppl 2**: S131-138.

Browning, D. F. and S. J. Busby (2004). "The regulation of bacterial transcription initiation." Nat Rev Microbiol **2**(1): 57-65.

Burnette, W. N. (1981). ""Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate--polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A." Anal Biochem **112**(2): 195-203.

Callaghan, A. J., M. J. Marcaida, et al. (2005). "Structure of Escherichia coli RNase E catalytic domain and implications for RNA turnover." Nature **437**(7062): 1187-1191.

Carrier, T. A. and J. D. Keasling (1999). "Library of synthetic 5' secondary structures to manipulate mRNA stability in Escherichia coli." Biotechnol Prog **15**(1): 58-64.

Castellani, A. and A. J. Chalmers (1913). Manual of tropical medicine, Baillière, Tindall and Cox.

Celesnik, H., A. Deana, et al. (2007). "Initiation of RNA decay in Escherichia coli by 5' pyrophosphate removal." Mol Cell **27**(1): 79-90.

Chan, C. L. and C. A. Gross (2001). "The anti-initial transcribed sequence, a portable sequence that impedes promoter escape, requires sigma70 for function." J Biol Chem **276**(41): 38201-38209.

Chou, C. H., A. A. Aristidou, et al. (1995). "Characterization of a pH-inducible promoter system for high-level expression of recombinant proteins in Escherichia coli." Biotechnol Bioeng **47**(2): 186-192.

Cohen, S. N., A. C. Chang, et al. (1973). "Construction of biologically functional bacterial plasmids in vitro." Proc Natl Acad Sci U S A **70**(11): 3240-3244.

Crick, F. (1970). "Central dogma of molecular biology." Nature **227**(5258): 561-563.

Davis, J. H., A. J. Rubin, et al. (2011). "Design, construction and characterization of a set of insulated bacterial promoters." Nucleic Acids Res **39**(3): 1131-1141.

de Boer, H. A., L. J. Comstock, et al. (1983). "The tac promoter: a functional hybrid derived from the trp and lac promoters." Proc Natl Acad Sci U S A **80**(1): 21-25.

Deana, A. and J. G. Belasco (2005). "Lost in translation: the influence of ribosomes on bacterial mRNA decay." Genes Dev **19**(21): 2526-2533.

Deana, A., H. Celesnik, et al. (2008). "The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal." Nature **451**(7176): 355-358.

deSmit, M. H. and J. Vanduin (1990). "Secondary Structure of the Ribosome Binding-Site Determines Translational Efficiency - a Quantitative-Analysis." Proc Natl Acad Sci U S A **87**(19): 7668-7672.

dos Reis, M., L. Wernisch, et al. (2003). "Unexpected correlations between gene expression and codon usage bias from microarray data for the whole Escherichia coli K‐12 genome." Nucleic Acids Res **31**(23): 6976-6985.

Durland, R. H., A. Toukdarian, et al. (1990). "Mutations in the trfA replication gene of the broad-host-range plasmid RK2 result in elevated plasmid copy numbers." J Bacteriol **172**(7): 3859-3867.

Emory, S. A., P. Bouvet, et al. (1992). "A 5'-terminal stem-loop structure can stabilize mRNA in Escherichia coli." Genes Dev **6**(1): 135-148.

Escherich, T. (1885). "Die darmbakterien des neugeborenen und säuglings." Fortschr. Med. **3**: 515-522; 547-554.

Feng, L., W. W. Chan, et al. (2000). "High-level expression and mutagenesis of recombinant human phosphatidylcholine transfer protein using a synthetic gene: evidence for a C-terminal membrane binding domain." Biochemistry **39**(50): 15399-15409.

Ferrer-Miralles, N., J. Domingo-Espin, et al. (2009). "Microbial factories for recombinant pharmaceuticals." Microb Cell Fact **8**: 17.

Fournier, B., A. Gravel, et al. (1999). "Strength and regulation of the different promoters for chromosomal beta-lactamases of Klebsiella oxytoca." Antimicrob Agents Chemother **43**(4): 850-855.

Fox, J. D., R. B. Kapust, et al. (2001). "Single amino acid substitutions on the surface of Escherichia coli maltose-binding protein can have a profound impact on the solubility of fusion proteins." Protein Science **10**(3): 622-630.

Franch, T., A. P. Gultyaev, et al. (1997). "Programmed cell death by hok/sok of plasmid R1: processing at the hok mRNA 3'-end triggers structural rearrangements that allow translation and antisense RNA binding." J Mol Biol **273**(1): 38-51.

Frost, L. S., R. Leplae, et al. (2005). "Mobile genetic elements: the agents of open source evolution." Nat Rev Microbiol **3**(9): 722-732.

Fukagawa, Y., T. Takei, et al. (1980). "Inhibition of beta-lactamase of Bacillus licheniformis 749/C by compound PS-5, a new beta-lactam antibiotic." Biochem J **185**(1): 177-185.

Ghim, C. M., S. K. Lee, et al. (2010). "The art of reporter proteins in science: past, present and future applications." BMB Rep **43**(7): 451-460.

Goltermann, L., M. B. Jensen, et al. (2011). "Tuning protein expression using synonymous codon libraries targeted to the 5 ' mRNA coding region." Protein Engineering Design & Selection **24**(1-2): 123-129.

Gonzalez de Valdivia, E. I. and L. A. Isaksson (2004). "A codon window in mRNA downstream of the initiation codon where NGG codons give strongly reduced gene expression in Escherichia coli." Nucleic Acids Res **32**(17): 5198-5205.

Goodman, H. M. and A. Rich (1963). "Mechanism of Polyribosome Action during Protein Synthesis." Nature **199**: 318-322.

Grossman, A. D., J. W. Erickson, et al. (1984). "The htpR gene product of E. coli is a sigma factor for heat-shock promoters." Cell **38**(2): 383-390.

Gustafsson, C., S. Govindarajan, et al. (2004). "Codon bias and heterologous protein expression." Trends in Biotechnology **22**(7): 346-353.

Guzman, L. M., D. Belin, et al. (1995). "Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter." J Bacteriol **177**(14): 4121-4130.

Hacker, J. and G. Blum-Oehler (2007). "In appreciation of Theodor Escherich." Nat Rev Micro **5**(12): 902-902.

Han, Q., Q. Zhao, et al. (2005). "Molecular recognition by glycoside pseudo base pairs and triples in an apramycin-RNA complex." Angew Chem Int Ed Engl **44**(18): 2694-2700.

Hansen, M. J., L. H. Chen, et al. (1994). "The ompA 5' untranslated region impedes a major pathway for mRNA degradation in Escherichia coli." Mol Microbiol **12**(5): 707-716.

Hansted, J. G., L. Pietikainen, et al. (2011). "Expressivity tag: a novel tool for increased expression in Escherichia coli." J Biotechnol **155**(3): 275-283.

Hasan, N. and W. Szybalski (1995). "Construction of lacIts and lacIqts expression plasmids and evaluation of the thermosensitive lac repressor." Gene **163**(1): 35-40.

Hatoum, A. and J. Roberts (2008). "Prevalence of RNA polymerase stalling at Escherichia coli promoters after open complex formation." Mol Microbiol **68**(1): 17-28.

Hauser, P. S. and R. O. Ryan (2007). "Expressed protein ligation using an N-terminal cysteine containing fragment generated in vivo from a pelB fusion protein." Protein Expr Purif **54**(2): 227-233.

Hochuli, E., H. Dobeli, et al. (1987). "New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues." J Chromatogr **411**: 177-184.

Hsu, L. M., I. M. Cobb, et al. (2006). "Initial transcribed sequence mutations specifically affect promoter escape properties." Biochemistry **45**(29): 8841-8854.

Huang, C. J., H. Lin, et al. (2012). "Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements." J Ind Microbiol Biotechnol.

Itakura, K., T. Hirose, et al. (1977). "Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin." Science **198**(4321): 1056-1063.

James, A. L., J. D. Perry, et al. (1996). "Evaluation of cyclohexenoesculetin-beta-D-galactoside and 8-hydroxyquinoline-beta-D-galactoside as substrates for the detection of beta-galactosidase." Appl Environ Microbiol **62**(10): 3868-3870.

Jana, S. and J. K. Deb (2005). "Strategies for efficient production of heterologous proteins in Escherichia coli." Appl Microbiol Biotechnol **67**(3): 289-298.

Jiang, T., B. Xing, et al. (2008). "Recent developments of biological reporter technology for detecting gene expression." Biotechnol Genet Eng Rev **25**: 41-75.

Julian, P., P. Milon, et al. (2011). "The Cryo-EM structure of a complete 30S translation initiation complex from Escherichia coli." PLoS Biol **9**(7): e1001095.

Junod, S. W. (2007). Celebrating a Milestone: FDA's Approval of First Genetically-Engineered Product. Update Magazine, U.S. Food and drug administration.

Kaczanowska, M. and M. Ryden-Aulin (2007). "Ribosome biogenesis and the translation process in Escherichia coli." Microbiol Mol Biol Rev **71**(3): 477-494.

Kamionka, M. (2011). "Engineering of therapeutic proteins production in Escherichia coli." Curr Pharm Biotechnol **12**(2): 268-274.

Kane, J. F. (1995). "Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli." Curr Opin Biotechnol **6**(5): 494-500.

Kim, K. H., J. K. Yang, et al. (2008). "From no expression to high-level soluble expression in Escherichia coli by screening a library of the target proteins with randomized N-termini." Methods Mol Biol **426**: 187-195.

Kolaj, O., S. Spada, et al. (2009). "Use of folding modulators to improve heterologous protein production in Escherichia coli." Microb Cell Fact **8**(1): 9.

Krishna Rao, D. V., J. V. Rao, et al. (2008). "Optimization of the AT-content of codons immediately downstream of the initiation codon and evaluation of culture conditions for high-level expression of recombinant human G-CSF in Escherichia coli." Mol Biotechnol **38**(3): 221-232.

Kudla, G., A. W. Murray, et al. (2009). "Coding-Sequence Determinants of Gene Expression in Escherichia coli." Science **324**(5924): 255-258.

Laboratories, B.-R. Bio-Rad Protein Assay.

Lale, R., L. Berg, et al. (2011). "Continuous control of the flow in biochemical pathways through 5' untranslated region sequence modifications in mRNA expressed from the broad-host-range promoter Pm." Appl Environ Microbiol **77**(8): 2648-2655.

Landick, R., J. Carey, et al. (1985). "Translation activates the paused transcription complex and restores transcription of the trp operon leader region." Proc Natl Acad Sci U S A **82**(14): 4663-4667.

Laursen, B. S., H. P. Sorensen, et al. (2005). "Initiation of protein synthesis in bacteria." Microbiol Mol Biol Rev **69**(1): 101-123.

Lederberg, J. and E. L. Tatum (1946). "Gene recombination in Escherichia coli." Nature **158**(4016): 558.

Lisser, S. and H. Margalit (1993). "Compilation of E. coli mRNA promoter sequences." Nucleic Acids Res **21**(7): 1507-1516.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods **25**(4): 402-408.

Looman, A. C., J. Bodlaender, et al. (1987). "Influence of the codon following the AUG initiation codon on the expression of a modified lacZ gene in Escherichia coli." EMBO J **6**(8): 2489-2492.

Mackie, G. A. (1998). "Ribonuclease E is a 5'-end-dependent endonuclease." Nature **395**(6703): 720-723.

Makrides, S. C. (1996). "Strategies for achieving high-level expression of genes in Escherichia coli." Microbiol Rev **60**(3): 512-538.

Markham, N. R. and M. Zuker (2005). "DINAMelt web server for nucleic acid melting prediction." Nucleic Acids Res **33**(Web Server issue): W577-581.

Marques, S., M. Manzanera, et al. (1999). "The XylS-dependent Pm promoter is transcribed in vivo by RNA polymerase with sigma 32 or sigma 38 depending on the growth phase." Mol Microbiol **31**(4): 1105-1113.

McDowall, K. J., S. Lin-Chao, et al. (1994). "A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage." J Biol Chem **269**(14): 10790-10796.

Mergulhão, F. J. M., D. K. Summers, et al. (2005). "Recombinant protein secretion in Escherichia coli." Biotechnology Advances **23**(3): 177-202.

Meselson, M. and F. W. Stahl (1958). "The Replication of DNA in Escherichia Coli." Proc Natl Acad Sci U S A **44**(7): 671-682.

Miksch, G., F. Bettenworth, et al. (2005). "Libraries of synthetic stationary-phase and stress promoters as a tool for fine-tuning of expression of recombinant proteins in Escherichia coli." J Biotechnol **120**(1): 25-37.

Mitarai, N., K. Sneppen, et al. (2008). "Ribosome Collisions and Translation Efficiency: Optimization by Codon Usage and mRNA Destabilization." J Mol Biol **382**(1): 236-245.

Mullis, K., F. Faloona, et al. (1986). "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction." Cold Spring Harb Symp Quant Biol **51 Pt 1**: 263-273.

Najjar, V. A. (1948). "The isolation and properties of phosphoglucomutase." J Biol Chem **175**(1): 281-290.

Narhi, L. O., T. Arakawa, et al. (2001). "Asn to Lys mutations at three sites which are N-glycosylated in the mammalian protein decrease the aggregation of Escherichia coli-derived erythropoietin." Protein Eng **14**(2): 135-140.

Nirenberg, M., T. Caskey, et al. (1966). "The RNA code and protein synthesis." Cold Spring Harb Symp Quant Biol **31**: 11-24.

Nishikubo, T., N. Nakagawa, et al. (2005). "Improved heterologous gene expression in Escherichia coli by optimization of the AT-content of codons immediately downstream of the initiation codon." J Biotechnol **120**(4): 341-346.

Nudler, E. and M. E. Gottesman (2002). "Transcription termination and anti-termination in E. coli." Genes Cells **7**(8): 755-768.

Park, Y. S., S. W. Seo, et al. (2007). "Design of 5'-untranslated region variants for tunable expression in Escherichia coli." Biochem Biophys Res Commun **356**(1): 136-141.

Pedersen, M., S. Nissen, et al. (2011). "The functional half-life of an mRNA depends on the ribosome spacing in an early coding region." J Mol Biol **407**(1): 35-44.

Pines, O. and M. Inouye (1999). "Expression and secretion of proteins in *E. coli*." Mol Biotechnol **12**(1): 25-34.

Pribnow, D. (1975). "Bacteriophage T7 early promoters: nucleotide sequences of two RNA polymerase binding sites." J Mol Biol **99**(3): 419-443.

Ramos, J. L., S. Marques, et al. (1997). "Transcriptional control of the Pseudomonas tol plasmid catabolic operons is achieved through an interplay of host factors and plasmid-encoded regulators." Annual Review of Microbiology **51**: 341-373.

Rice, P., I. Longden, et al. (2000). "EMBOSS: the European Molecular Biology Open Software Suite." Trends Genet **16**(6): 276-277.

Ringquist, S., S. Shinedling, et al. (1992). "Translation initiation in Escherichia coli: sequences within the ribosome-binding site." Mol Microbiol **6**(9): 1219-1229.

Rittie, L. and B. Perbal (2008). "Enzymes used in molecular biology: a useful guide." J Cell Commun Signal **2**(1-2): 25-45.

Roberts, J. W., S. Shankar, et al. (2008). "RNA polymerase elongation factors." Annual Review of Microbiology **62**: 211-233.

Roberts, R. J. (2005). "How restriction enzymes became the workhorses of molecular biology." Proc Natl Acad Sci U S A **102**(17): 5905-5908.

Rodgers, J. L. and W. A. Nicewander (1988). "Thirteen Ways to Look at the Correlation Coefficient." The American Statistician **42**(1): 59-66.

Salis, H. M., E. A. Mirsky, et al. (2009). "Automated design of synthetic ribosome binding sites to control protein expression." Nat Biotech **27**(10): 946-950.

Sandkvist, M. and M. Bagdasarian (1996). "Secretion of recombinant proteins by Gram-negative bacteria." Curr Opin Biotechnol **7**(5): 505-511.

Satakarni, M. and R. Curtis (2011). "Production of recombinant peptides as fusions with SUMO." Protein Expr Purif **78**(2): 113-119.

Sharma, U. K. and D. Chatterji (2010). "Transcriptional switching in Escherichia coli during stress and starvation by modulation of σ70 activity." FEMS Microbiology Reviews **34**(5): 646-657.

Sharma, V., K. Park, et al. (2009). "Shape separation of gold nanorods using centrifugation." Proc Natl Acad Sci U S A **106**(13): 4981-4985.

Sharp, P. M. and W. H. Li (1987). "The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications." Nucleic Acids Res **15**(3): 1281-1295.

Shine, J. and L. Dalgarno (1974). "The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites." Proc Natl Acad Sci U S A **71**(4): 1342-1346.

Sletta, H., A. Nedal, et al. (2004). "Broad-host-range plasmid pJB658 can be used for industrial-level production of a secreted host-toxic single-chain antibody fragment in Escherichia coli." Appl Environ Microbiol **70**(12): 7033-7039.

Sletta, H., A. Tondervik, et al. (2007). "The presence of N-terminal secretion signal sequences leads to strong stimulation of the total expression levels of three tested medically important proteins during high-cell-density cultivations of Escherichia coli." Appl Environ Microbiol **73**(3): 906-912.

Son, Y. J., C. K. Kim, et al. (2008). "Effects of beta-mercaptoethanol and hydrogen peroxide on enzymatic conversion of human proinsulin to insulin." J Microbiol Biotechnol **18**(5): 983-989.

Sonoda, H. and A. Sugimura (2008). "Improved solubilization of recombinant human growth hormone inclusion body produced in Escherichia coli." Biosci Biotechnol Biochem **72**(10): 2675-2680.

Sorensen, H. P. and K. K. Mortensen (2005). "Advanced genetic strategies for recombinant protein expression in Escherichia coli." J Biotechnol **115**(2): 113-128.

Srivastava, P., P. Bhattacharaya, et al. (2005). "Overexpression and purification of recombinant human interferon alpha2b in Escherichia coli." Protein Expr Purif **41**(2): 313-322.

Steitz, J. A. and K. Jakes (1975). "How Ribosomes Select Initiator Regions in Messenger-Rna - Base Pair Formation between 3' Terminus of 16s Ribosomal-Rna and Messenger-Rna during Initiation of Protein-Synthesis in Escherichia-Coli." Proc Natl Acad Sci U S A **72**(12): 4734-4738.

Stenstrom, C. M., H. Jin, et al. (2001). "Codon bias at the 3'-side of the initiation codon is correlated with translation initiation efficiency in Escherichia coli." Gene **263**(1-2): 273-284.

Studier, F. W. and B. A. Moffatt (1986). "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes." J Mol Biol **189**(1): 113-130.

Tang, L., R. Jiang, et al. (2011). "Enhancing the recombinant protein expression of halohydrin dehalogenase HheA in Escherichia coli by applying a codon optimization strategy." Enzyme and Microbial Technology **49**(4): 395-401.

Terpe, K. (2003). "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems." Appl Microbiol Biotechnol **60**(5): 523-533.

Terpe, K. (2006). "Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems." Appl Microbiol Biotechnol **72**(2): 211-222.

Towbin, H., T. Staehelin, et al. (1979). "Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications." Proc Natl Acad Sci U S A **76**(9): 4350-4354.

Tuller, T., A. Carmi, et al. (2010). "An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation." Cell **141**(2): 344-354.

Valla, S., D. H. Coucheron, et al. (1989). "Cloning of a gene involved in cellulose biosynthesis in Acetobacter xylinum: complementation of cellulose-negative mutants by the UDPG pyrophosphorylase structural gene." Mol Gen Genet **217**(1): 26-30.

Van Rossum, G., et al (2012). "Python Programming Language – Official Website." Retrieved May 1st, 2012, from www.python.org.

Vee Aune, T. E., I. Bakke, et al. (2010). "Directed evolution of the transcription factor XylS for development of improved expression systems." Microb Biotechnol **3**(1): 38-47.

Wagner, L. A., R. F. Gesteland, et al. (1994). "An efficient Shine-Dalgarno sequence but not translation is necessary for lacZ mRNA stability in Escherichia coli." J Bacteriol **176**(6): 1683-1688.

Wahab, S. Z., K. O. Rowley, et al. (1993). "Effects of tRNA(1Leu) overproduction in Escherichia coli." Mol Microbiol **7**(2): 253-263.

Warner, J. R. and P. M. Knopf (2002). "The discovery of polyribosomes." Trends Biochem Sci **27**(7): 376-380.

Weber, K. and M. Osborn (1969). "The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis." J Biol Chem **244**(16): 4406-4412.

Weisberg, R. A. and M. E. Gottesman (1999). "Processive antitermination." J Bacteriol **181**(2): 359-367.

Welch, M., S. Govindarajan, et al. (2009). "Design parameters to control synthetic gene expression in Escherichia coli." PLoS One **4**(9): e7002.

Weston, A., M. G. Brown, et al. (1981). "Transformation of Escherichia coli with plasmid deoxyribonucleic acid: calcium-induced binding of deoxyribonucleic acid to whole cells and to isolated membrane fractions." J Bacteriol **145**(2): 780-787.

Wild, J. and W. Szybalski (2004). "Copy-control tightly regulated expression vectors based on pBAC/oriV." Methods Mol Biol **267**: 155-167.

Winther-Larsen, H. C., J. M. Blatny, et al. (2000). "Pm promoter expression mutants and their use in broad-host-range RK2 plasmid vectors." Metab Eng **2**(2): 92-103.

Wong, M. L. and J. F. Medrano (2005). "Real-time PCR for mRNA quantitation." Biotechniques **39**(1): 75-85.

Wood, K. V. (1995). "Marker proteins for gene expression." Curr Opin Biotechnol **6**(1): 50-58.

Worthington, K., Worthington, V. (2011). "Worthington Enzyme Manual." Retrieved 04.11.2012, from http://www.worthington-biochem.com/PGM/default.html.

Wu, C., V. R. Nerurkar, et al. (2008). "Effective modifications for improved homologous recombination and high-efficiency generation of recombinant adenovirus-based vectors." J Virol Methods **153**(2): 120-128.

Wu, J. and M. Filutowicz (1999). "Hexahistidine (His6)-tag dependent protein dimerization: a cautionary tale." Acta Biochim Pol **46**(3): 591-599.

Xu, F. and S. N. Cohen (1995). "RNA degradation in Escherichia coli regulated by 3' adenylation and 5' phosphorylation." Nature **374**(6518): 180-183.

Zagursky, R. J. and M. L. Berman (1984). "Cloning vectors that yield high levels of single-stranded DNA for rapid DNA sequencing." Gene **27**(2): 183-191.

Zhang, W., W. Xiao, et al. (2006). "mRNA secondary structure at start AUG codon is a key limiting factor for human protein expression in Escherichia coli." Biochem Biophys Res Commun **349**(1): 69-78.

Zuker, M. (2003). "Mfold web server for nucleic acid folding and hybridization prediction." Nucleic Acids Res **31**(13): 3406-3415.

Zuker, M. and P. Stiegler (1981). "Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information." Nucleic Acids Res **9**(1): 133-148.

List of appendices

Appendix A Mastercycler® ”ANNEAL” program

Lid 103 °C

1. 95 °C 3 min
2. 90 °C 1 min
3. 85 °C 2 min
4. 80 °C 5 min
5. 75 °C 5 min
6. 70 °C 10 min
7. 65 °C 10 min
8. 60 °C 5 min
9. 55 °C 5 min
10. 50 °C 5 min
11. 45 °C 2 min
12. 40 °C 2 min
13. 35 °C 1 min
14. 30 °C 1 min
15. 25 °C 1min
16. 20 °C 1 min
17. HOLD 4 °C

Appendix B UTR sequences from *Pm* promoter

pBSP1 derivatives, pASP1 derivatives, pAR69 derivatives, pAT63 derivatives, pIN69 derivatives:

```
                                          +1
5’–AACAGAAACAAUAAUAAUGGAGUCAUGAACAU AUG–‘3
```

pLB11 derivatives, pCSP1 derivatives:

```
                                          +1
5’–AACATGTACAATAATAATGGAGTCATGAACAT AUG–‘3
```

+1 denotes translational start site, sequence shown from transcriptional start site.

Appendix C Sequence alignments

Table C-1: Sequenced *bla* signal sequence of 12 randomly picked clones from kanamycin plate of SII library screening aligned to wild-type signal sequence of pBSP1-TCT. Sequencing was performed by Eurofins MWG Operong using the PmUTR.fwd primer shown in **Table 2-4.**

	DNA SEQUENCE	AMINO ACID SEQUENCE
TCT	TATGTCTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTCGC	MSIQHFRVALIPFFAAFCLPVF
R1	.....	.....
R2	.....	.....
R3	.....	.....
R4	.....	.....
R5	......A..C..T.....A..G.....T..A..C..........C..T.....A..A..A..A..A..	...H.L..........*...L
R6	.....	.....
R7	......A..C..T.....A..A..T..T..A..C..T.......C..G.....A..A..A..A.....	...H.L..........*....
R8	.....	.....
R9	.....T.....	.............V.....
R10	.....	.....
R11	.....	.....
R12	.....	.....

Table C-2: Sequenced *celB*$_{69}$ gene fusion sequence of 12 randomly picked clones from kanamycin plate of the *celB*$_{69}$*-aac(3)-IV* library screening aligned to wild-type *celB*$_{69}$. Sequencing was performed by Eurofins MWG Operong using the PmUTR.fwd primer shown in **Table 2-4.**

	DNA SEQUENCE	AMINO ACID SEQUENCE
CELB69	ATGCCCAGCATAAGCCCATTTGCCGGCAAGCCGGTCGATCCGGACCGTCTTGTC-AATATCGACGCCC-TGGC	MPSISPFAGKPVDPDRLVNIDAL
K1	.....-.....-....	.....
K2	.....-.....-....	.....
K3	.....C.....-....	.....QYRRP
K4	.....-.------.....-....	.....AL.M
K5	.....-.....-....	.....
K6	....G.....G.....A.....A.....A..-...T..T.....-....	.R.R.....E....I.FY..
K7	.....G.....A.....T.....T....---------.....T..-....	..R...I...S.....HRRPGHG
K8	.....C.....A.A.....A.....-.....A.....-....	.....YT.....N.....N..
K9	.....G.....CT.....--..C.....T...-....	.....G.A...ISTPW
K10	.....-.....-....	.....
K11	.....-.....C....	.....P
K12	.....-.....-....	.....

Appendix D Survival frequencies from library screenings

SII library

		Small 1	Small 2	Big w/o satellite cols	Avg big plate	% of kan50 plate	log %
Kanamycin 50ug/ml		221	-	-	22100	100 %	2
Amp g/L	**Dilution plated**						
0.8	1.00E+05	239	-	-	23900	108 %	2.034005627
1	1.00E+05	225	202	-	21350	97 %	1.985005606
1.2	1.00E+05	137	149	-	14300	65 %	1.810943764
1.4	1.00E+04	89 (smear)	726	-	7260	33 %	1.516544347
1.6	1.00E+04	178	208	-	1930	9 %	0.941165035
1.8	1.00E+04	76	75	-	755	3 %	0.533554678
2	1.00E+04	21	23	160	190	1 %	-0.065638673
2.2		-	-	134	134	1 %	-0.217287475
2.4		-	-	87	87	0 %	-0.404873021
2.6		-	-	58	58	0 %	-0.58096428
2.8		-	-	59	59	0 %	-0.573540262
3		-	-	43	43	0 %	-0.710923818
3.5		-	-	32	32	0 %	-0.839242295
4		-	-	10	10	0 %	-1.344392274
4.5		-	-	4	4	0 %	-1.742332282
5		-	-	3	3	0 %	-1.867271019
5.5		-	-	2	2	0 %	-2.043362278
6		-	-	0	0	0 %	

SI library

		Small 1	Small 2	Big w/o satellite cols	Avg big plate	% of kan50 plate	log %
Kanamycin 50ug/ml		218	-		21800	100 %	2
Amp g/L	**Dilution plated**						
0.05	1.00E+05	204	167		18550	85 %	1.92988742
0.1	1.00E+05	100	125		11250	52 %	1.712696029
0.15	1.00E+05	127	85		10600	49 %	1.686849372
0.2	1.00E+05	84	65		7450	34 %	1.533699779
0.4	1.00E+05	20	22	164	187	1 %	-0.066614887
0.6	1.00E+04	10	-	106	103	0 %	-0.325619269
0.8	1.00E+04	1	-	41	25.5	0 %	-0.931916313
1	10000	-	-	36	36	0 %	-0.782153993
1.2		-	-	28	28	0 %	-0.891298462
1.4		-	-	7	7	0 %	-1.493358454
1.6		-	-	7	7	0 %	-1.493358454
1.8		-	-	2	2	0 %	-2.037426498
2		-	-	1	1	0 %	-2.338456494
2.6		-	-	0	0	0 %	-

WILD TYPE pBSP1

Ampicillin (g/l)	Log %
0.00	2.00
0.05	2.00
0.08	2.00
0.10	1.95
0.12	1.85
0.14	1.78
0.16	1.70
0.18	1.60
0.20	0.00
0.22	-0.52
0.24	-0.60
0.26	-0.82

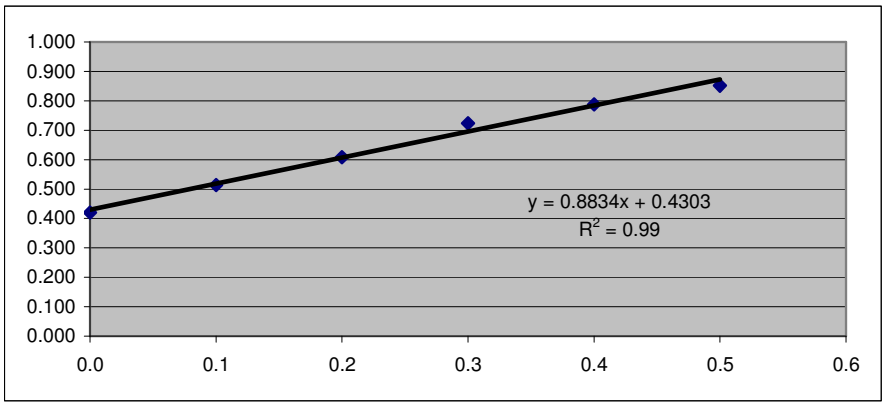
fp $_{69}$ ***-aac(3)-IV*** **library**

Apramycin concentration (g/L)	Small 1 (10^-5)	Small 2 (10^-5)	Big (10^-3)	Avg big plate	% of kan50 plate	log %
0 (Kanamycin 50mg/ml)	1264	1289	-	126400	100 %	2
0.08	2	1	453	251.00	0.199 %	-0.702073352
0.1	2	2	233	211.00	0.167 %	-0.777464619
0.12	1	1	150	116.67	0.092 %	-1.034800284
0.15	2	0	51	83.67	0.066 %	-1.179194607
0.2	0	0	35	11.67	0.009 %	-2.034800284
	Big (10^-3)	**Big (10^-3)**	**Big (10^-3)**			
0.4	2	5	3	3.33	0.003 %	-2.578868329
0.5	5	12	4	7.00	0.006 %	-2.256649034
0.6	3	3	2	2.67	0.002 %	-2.675778342
0.7	-	1	0	0.50	0.000 %	-3.40277707
0.8	-	-	0	0.00		
0.9	-	-	0	0.00		
1.0	-	-	0	0.00		

Appendix E BSA standard curves

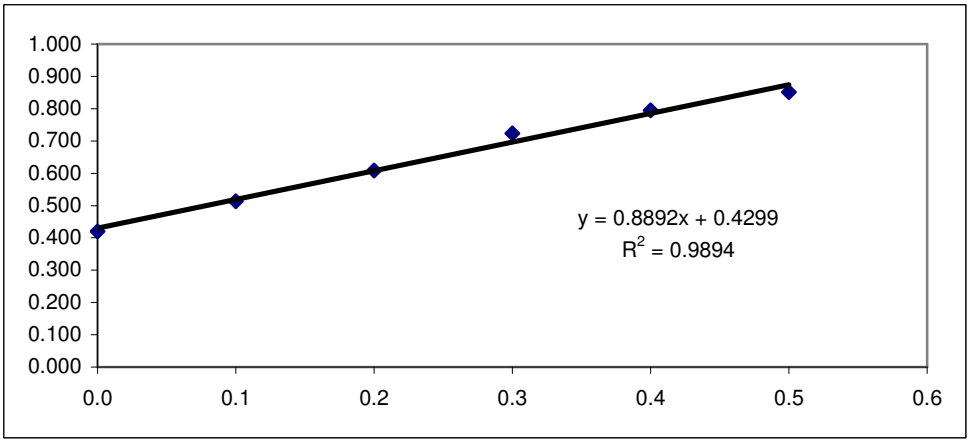
BSA standard curve for 3/3-2011, 21/3-2011 assays

						Average	Stand. Dev.
0.0	0.4105	0.4357	0.418	0.4257	0.4135	0.421	0.010
0.1	0.4993	0.5186	0.5174	0.5071	0.5263	0.514	0.011
0.2	0.5982	0.6233	0.6123	0.6025	0.6073	0.609	0.010
0.3	0.7289	0.7263	0.7384	0.7005	0.7255	0.724	0.014
0.4	0.782	0.7736		0.7907	0.8067	0.788	0.014
0.5	0.8577	0.8504	0.8388	0.8549	0.8548	0.851	0.007



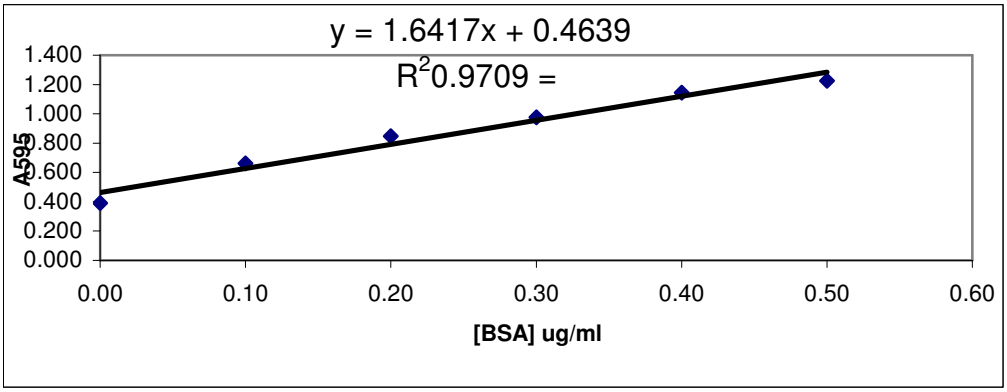
Biorad BSAstandard curve for 4/5-2011 assay

						Average	Stand. Dev.
0.0	0.4105	0.4357	0.418	0.4257	0.4135	0.421	0.010
0.1	0.4993	0.5186	0.5174	0.5071	0.5263	0.514	0.011
0.2	0.5982	0.6233	0.6123	0.6025	0.6073	0.609	0.010
0.3	0.7289	0.7263	0.7384	0.7005	0.7255	0.724	0.014
0.4	0.782	0.7736	0.8223	0.7907	0.8067	0.795	0.020
0.5	0.8577	0.8504	0.8388	0.8549	0.8548	0.851	0.007



Biorad measurment standard curve 12/5-2011

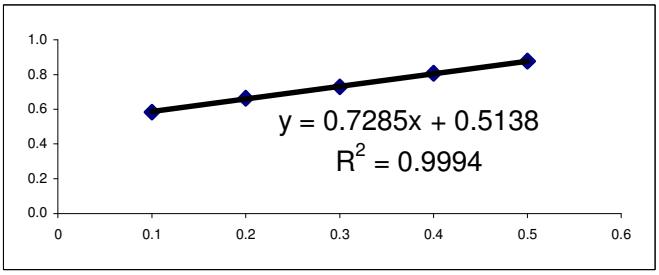
							Average	Stand. Dev.
0.0	0.3837	0.3894	0.3929	0.391	0.3917	0.3963	**0.391**	**0.004**
0.1	0.6498	0.6591	0.669	0.663	0.6688		**0.662**	**0.008**
0.2	0.8558	0.8474	0.8432	0.872	0.8546	0.8165	**0.848**	**0.018**
0.3	0.996	0.9632	0.9862	0.934	1.0109	0.9639	**0.976**	**0.028**
0.4	1.1518	1.135	1.0976	1.095	1.1879	1.2002	**1.145**	**0.044**
0.5	1.2066	1.2181	1.228	1.204	1.2155	1.2769	**1.225**	**0.027**



BSA (mg/mL)

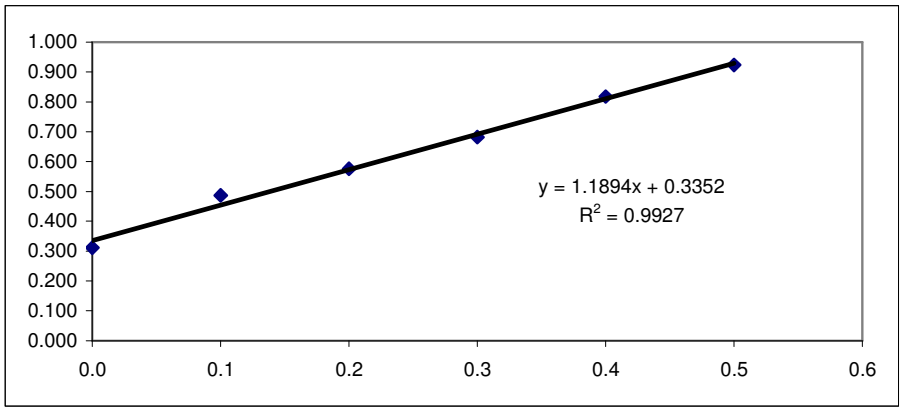
BSA standard curve for bla assays 29/10-2011, 30/10-2011

	1	2	3	4	5	6	AVG	STDEV
0	0.391	0.372	0.361	0.384		0.395	**0.381**	**0.014**
0.1	0.540	0.627	0.580	0.584	0.607	0.570	**0.585**	**0.030**
0.2	0.659	0.679	0.681	0.667	0.662	0.635	**0.664**	**0.016**
0.3	0.750	0.735	0.722	0.731	0.704	0.735	**0.729**	**0.015**
0.4	0.785	0.799	0.807	0.829	0.857	0.765	**0.807**	**0.033**
0.5	0.872	0.866	0.879	0.874	0.869	0.904	**0.877**	**0.014**



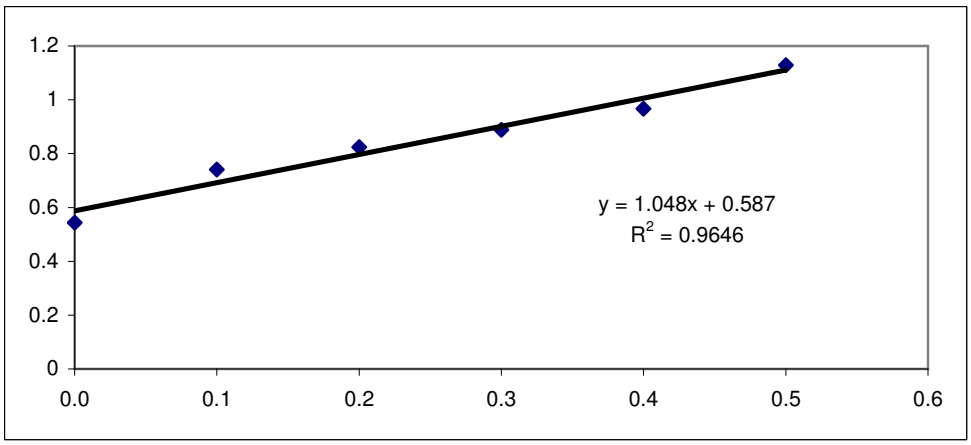
Biorad BSAstandard curve for 25/5-2011 assay

							Average	Stand. Dev.
0.0	0.299	0.3157	0.3052	0.3189	0.3129	0.3145	0.311	0.007
0.1	0.473	0.4948	0.503	0.4759	0.4922	0.4785	0.486	0.012
0.2	0.5474	0.5853	0.6066	0.5847	0.5646	0.5636	0.575	0.021
0.3		0.7017		0.6798	0.6591	0.6852	0.681	0.018
0.4			0.7999	0.8239	0.8431	0.8034	0.818	0.020
0.5		0.9168	0.8953	0.9527		0.9296	0.924	0.024



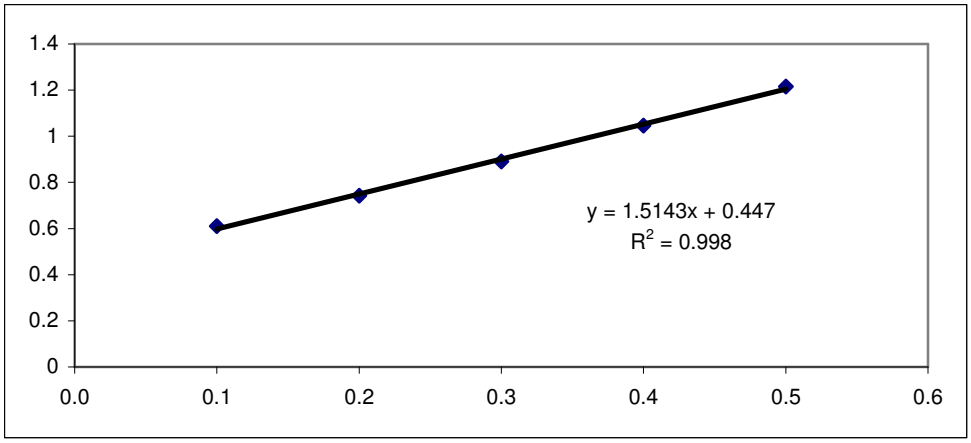
Biorad BSAstandard curve for 2/11-2011, 11/11-2011 assay

						Average	Stand. Dev.
0.0	0.5249	0.5489	0.5462	0.5557		0.543925	0.013
0.1	0.7238	0.7439	0.7534	0.7437	0.7395	0.74086	0.011
0.2	0.7842	0.8298	0.8439	0.8411	0.8226	0.82432	0.024
0.3	0.8786	0.9134	0.8845	0.8743	0.893	0.88876	0.015
0.4	0.9413	0.9906	0.9739	0.9445	0.9847	0.967	0.023
0.5	1.1352	1.1384	1.1031	1.1298	1.1382	1.12894	0.015



Biorad BSAstandard curve for 8/12-2011, 13/12-2011 assays

						Average	Stand. Dev.
0.1	0.6079	0.6158	0.6078	0.611	0.6096	0.61042	0.003
0.2	0.7326	0.7518	0.7401	0.7619	0.7267	0.74262	0.014
0.3	0.8737	0.9004	0.908	0.891	0.8837	0.89136	0.014
0.4	0.9906	1.0692	1.0921		1.034	1.046475	0.044
0.5	1.1635	1.196	1.2446	1.2362	1.2378	1.21562	0.035



Appendix F β-lactamase assay data

BLA assay 03.03.2011 0,05M inducer 2nd codons

	Slope from Bla assay				Average	Slope/ avg protein concentr.	STDEV
pBS2P1							
Parallell 1	102.70	110.60	98.96	110.30	**105.64**	**504.14**	**10.26**
Parallell 2	102.20	112.60	110.70	109.40	**108.73**		
pBSP1							
Parallell 1	30.97	30.85	33.61	33.81	**32.31**	**134.49**	**9.77**
Parallell 2	28.24	29.08	29.28	30.01	**29.15**		
CSP							
Parallell 1	88.82	91.52	98.02	99.06	**94.36**	**421.44**	**30.64**
Parallell 2	76.01	86.38	87.73	90.39	**85.13**		

pBSP1	BioRad measurements						Average	Stand. Dev.	Concentration (mg/ml)
Parallell 1	0.611	0.6376	0.613	0.622	0.6131		0.62	0.01	0.21
Parallell 2	0.6054	0.6307	0.6131	0.598	0.6194	0.6346	0.62	0.01	0.21
pBSP2									
Parallell 1	0.6238	0.6119	0.6114	0.645	0.641	0.6376	0.63	0.01	0.22
Parallell 2		0.6219	0.6354		0.6443	0.6416	0.64	0.01	0.23
CSP									
Parallell 1	0.6072	0.6264	0.6165	0.625	0.6073		0.62	0.01	0.21
Parallell 2		0.6227	0.6308	0.606	0.6125	0.6296	0.62	0.01	0.22

BLA assay 21.03.2011 0,05M inducer 2nd codons

pBS2P1	Slope from Bla assay				Average	Slope/ avg protein concentr.	STDEV
Parallell 1	106.00	99.67	109.90	110.70	**106.57**	**1327.95**	**136.32**
Parallell 2	106.10	112.50	120.60	153.80	**123.25**		
pBSP1							
Parallell 1	31.75	31.90	33.32	33.17	**32.54**	**368.87**	**84.29**
Parallell 2	44.19	45.44	44.78	45.89	**45.08**		
CSP							
Parallell 1	101.20	99.52	101.60	102.40	**101.18**	**1052.12**	**27.57**
Parallell 2	102.90	106.30	109.10	101.70	**105.00**		

pBS2P1	BioRad measurements						Average	Stand. Dev.	Concentration (mg/ml)
Parallell 1	0.4849	0.5107	0.5064	0.489	0.5051	0.5071	0.50	0.01	0.08
Parallell 2	0.5033	0.5112	0.5111	0.527	0.5054	0.5198	0.51	0.01	0.09
pBSP1									
Parallell 1	0.4805	0.5064	0.5015	0.508	0.5095	0.5084	0.50	0.01	0.08
Parallell 2	0.5451	0.5404	0.5516	0.55	0.5322	0.5449	0.54	0.01	0.13
CSP									
Parallell 1	0.5007	0.5122	0.5253	0.521	0.5276	0.5215	0.52	0.01	0.10
Parallell 2	0.5132	0.511	0.5057	0.515	0.5208	0.5288	0.52	0.01	0.10

25.05.2011 0.05mM inducer bla assay 2nd codons

Protein activity (slope x -1)

	1	2	3	4	5	6	AVG	STDEV	Slope / total protein
pBSP1	60.77	67.76	62.94	65.21	63.99	63.99	**64.11**	**2.33**	**329.94**
pBS2P1	92.480	85.840	90.090	92.22	95.830	94.490	**91.83**	**3.54**	**462.50**
AGC	37.090	38.300	37.710	37.940	38.370	40.620	**38.34**	**1.21**	**189.51**
TCC	84.450	93.060	84.190	92.540	91.900	90.200	**89.39**	**4.04**	**442.82**
TCA	60.080	66.230	57.120	59.430	58.200	51.650	**58.79**	**4.72**	**339.57**
TCG	56.700	58.650	60.040	64.300	62.730		**60.48**	**3.06**	**436.56**
CSP	105.200	104.400	106.200	111.200	109.400	109.500	**107.65**	**2.75**	**618.29**

Total Protein concentration (BioRad A595)

	1	2	3	4	5	6	AVG	STDEV	Protein concentration (mg/mL)
VK6 1	0.5768	0.5758	0.5605	0.5899	0.5569	0.5363	**0.5660**	**0.0189**	**0.1943**
AN1	0.5731	0.5861	0.563	0.5566	0.5831	0.5645	**0.5711**	**0.0118**	**0.1985**
AGC	0.5928	0.5547	0.5518	0.6094	0.5649	0.5796	**0.5755**	**0.0227**	**0.2023**
TCC	0.5742	0.5816	0.5301	0.6	0.5877	0.5765	**0.5750**	**0.0239**	**0.2019**
TCA	0.5511	0.5303	0.5541	0.5528	0.5193	0.5374	**0.5408**	**0.0142**	**0.1731**
TCG	0.5051	0.4903	0.4897	0.4944	0.5044	0.5145	**0.4997**	**0.0099**	**0.1385**
CSP	0.5438	0.5482	0.5466	0.5447	0.5309	0.5379	**0.5420**	**0.0065**	**0.1741**

BLA assay 04.05.2011 0,05M inducer 2nd codons

	Slope from Bla assay					Average	avg Slope/ protein concentr.	STDEV
pBS2P1								
Parallell 1	137.50	136.50	135.80	135.80	136.30	**136.38**	**769.73**	**42.62**
Parallell 2	148.00	145.30	148.80	148.30	147.10	**147.50**		
pBSP1								
Parallell 1	51.64	50.42	44.66	50.63	57.05	**50.88**	**338.88**	**12.34**
Parallell 2	46.56	49.18	47.10	50.15	48.24	**48.25**		
CSP								
Parallell 1	139.70	161.70	159.40	160.90	157.20	**155.78**	**1069.86**	**57.89**
Parallell 2	138.90	136.60	147.30	124.90	171.50	**143.84**		

	BioRad measurements					Average	Stand. Dev.	Concentration (mg/ml)	Average
pBS2P1									
Parallell 1	0.5684	0.5883	0.59	0.598	0.59	0.01	0.18	0.184	
Parallell 2	0.5872	0.609	0.606	0.605	0.60	0.01	0.19		
pBSP1									
Parallell 1	0.5256	0.5497	0.5319	0.564	0.54	0.02	0.13	0.151	
Parallell 2	0.5688	0.5912	0.5845	0.597	0.59	0.01	0.17		
CSP									
Parallell 1	0.5767	0.5836	0.5944	0.575	0.58	0.01	0.17	0.146	
Parallell 2	0.5452	0.546	0.5229	0.532	0.54	0.01	0.12		

BLA assay 12.05.2011 0,05M inducer 2nd codons

pBSP1	Slope from Bla assay						Average	Slope/ avg protein concentr.	STDEV
Parallell 1	29.420	27.730	29.300	23.950	29.470	26.350	**27.70**	**277.13**	**61.53**
Parallell 2	36.180	36.470	35.350	34.900	32.380	35.370	**35.11**		
pBS2P1									
Parallell 1	95.500	91.450	92.350		89.300	86.750	**91.07**	**987.96**	**151.82**
Parallell 2	97.340	92.260	98.000			86.450	**93.51**		
AGC									
Parallell 1	39.850	39.160	39.880	38.080	41.290	36.250	**39.09**	**264.56**	**8.51**
Parallell 2	34.440	37.200	37.830	35.670	35.820	34.950	**35.99**		
TCA									
Parallell 3	67.730	68.560	68.620	62.910	63.910	57.200	**64.82**	**554.12**	**158.52**
Parallell 4	70.510	72.560	70.560	71.560	75.040	73.210	**72.24**		
TCG									
Parallell 3	82.790	83.050	97.330	81.330	85.190	83.250	**85.49**	**591.88**	**101.64**
Parallell 4	72.540	75.440	75.910	73.650	76.660	77.040	**75.21**		

pBSP1	BioRad measurements						Average	Stand. Dev.	Concentration (mg/ml)
Parallell 1	0.6463	0.6494	0.6238		0.6617	0.7117	0.66	0.03	0.12
Parallell 2	0.6377	0.6	0.6407	0.668	0.672		0.64	0.03	0.11
pBS2P1									
Parallell 1	0.5865	0.626	0.6168	0.601	0.5546	0.6172	0.60	0.03	0.08
Parallell 2	0.6377	0.6141	0.6036	0.657	0.6659	0.651	0.64	0.02	0.11
AGC									
Parallell 1	0.7115	0.7148	0.7114	0.704	0.7229	0.7075	0.71	0.01	0.15
Parallell 2	0.6841	0.6848	0.7113	0.667	0.6607	0.6856	0.68	0.02	0.13
TCA									
Parallell 3	0.702	0.7241	0.7086	0.707	0.7038	0.6825	0.70	0.01	0.15
Parallell 4	0.6515	0.6467	0.6282	0.656	0.6403	0.6285	0.64	0.01	0.11
TCG									
Parallell 3	0.6998	0.6868	0.688	0.661	0.664	0.6526	0.68	0.02	0.13
Parallell 4	0.7116	0.7066	0.6949	0.702	0.6999	0.6932	0.70	0.01	0.14

29.10.2011 0.05mM inducer bla assay 2nd codons

Protein activity (slope x -1)

	1	2	3	4	5	AVG	STDEV	Slope / total protein
pBSP1	35.6	37.6	37.03		38.09	**37.08**	**1.08**	**205.26**
pBS2P1	98.63	110.5	85.99	97.53	100.7	**98.67**	**8.75**	**512.85**
TCC	87.66	90.89	90.93	90.63	93.46	**90.71**	**2.06**	**461.04**

Total Protein concentration (BioRad A595)

	1	2	3	4	5	AVG	STDEV	Protein concentration (mg/mL)
VK6 1	0.6214	0.6581	0.6516	0.6418	0.6541	**0.6454**	**0.0147**	**0.1806**
AN1	0.6668	0.6339	0.6694	0.6482	0.6515	**0.6540**	**0.0145**	**0.1924**
TCC	0.6398	0.6648	0.6419	0.6715	0.6677	**0.6571**	**0.0151**	**0.1968**

29.10.11 0,1mM inducer bla assay 29.10 Library clones

Protein activity (slope x -1)

	1	2	3	4	5	AVG	STDEV	Slope / total protein
pBSP1 1	#37,18	39.76	45.35	41.45	42.04	**42.15**	**2.34**	**264.57**
pBSP1 2	49.13	50.04	50.04	48.72	50	**49.59**	**0.62**	**271.60**
pBS2P1	110.1	102.1	112.1	113.2	112.3	**109.96**	**4.54**	**516.55**
C19	166.0	152.6	178.4	182.6	145.3	**164.98**	**16.06**	**844.85**
C20	131.7	133.8	125.9	131.4	137.5	**132.06**	**4.22**	**695.43**
MS2	169.5	172.1	165.4	164.8	172.0	**168.76**	**3.51**	**851.04**
C22	174.1	178.8	170.2	188.7	178.8	**178.12**	**6.92**	**744.89**
MS4	194.5	195.5	208.9	209.8	218.6	**205.46**	**10.28**	**905.05**
CSP	143.4	160.4	142.3	148.1	148.0	**148.44**	**7.18**	**774.19**

Total Protein concentration (BioRad A595)

	1	2	3	4	5	AVG	STDEV	Protein concentration (mg/mL)
pBSP1 1	0.6209	0.6436	0.6266	0.6448	0.6134	**0.6299**	**0.0139**	**0.1593**
pBSP1 2	0.6312	0.6453	0.6478	0.6579	0.6518	**0.6468**	**0.0099**	**0.1826**
pBS2P1	0.6372	0.6768	0.6885	0.6832	0.6587	**0.6689**	**0.0210**	**0.2129**
C19	0.6367	0.6457	0.6686	0.6705	0.6588	**0.6561**	**0.0146**	**0.1953**
C20	0.6446	0.6414	0.6648	0.6617	0.6482	**0.6521**	**0.0105**	**0.1899**
MS2	0.6561	0.6574	0.656	0.6442	0.6776	**0.6583**	**0.0121**	**0.1983**
C22	0.6686	0.6849	0.7048	0.7006	0.6811	**0.6880**	**0.0148**	**0.2391**
MS4	0.6698	0.6798	0.6939	0.675	0.6774	**0.6792**	**0.0090**	**0.2270**
CSP	0.661	0.663	0.6483	0.6318	0.6633	**0.6535**	**0.0136**	**0.1917**

30.10.11 0,1mM inducer bla assay Library clones

Protein activity (slope x -1)

	1	2	3	4	5	AVG	STDEV	Slope / total protein
pBSP1 1	60.77	67.76	62.94	65.21	63.99	**64.13**	**2.60**	**234.97**
pBSP1 2	55.09	55.21	54.21	53.51	55.71	**54.75**	**0.88**	**248.58**
pBS2P1	119.8	116.8	124.8	118.1	117.8	**119.46**	**3.17**	**494.81**
C19	178.8	171.2	185.1	189.4	185.9	**182.08**	**7.18**	**777.34**
C20	126.9	132.9	133.5	137.1	132.5	**132.58**	**3.66**	**531.97**
MS2	201.9	207.2	209.1	208.9	210.3	**207.48**	**3.31**	**949.19**
C22	183	182.3	182.9	186.5	185.5	**184.04**	**1.84**	**841.01**
MS4	194.4	208.9	207	207.3	202.3	**203.98**	**5.89**	**864.55**
CSP	154.4	138.4	160.2	163.3	156.3	**154.52**	**9.65**	**756.50**
TMB17	3.508	3.334	2.708	3.568	3.9	**3.41**	**0.45**	**16.26**

Total Protein concentration (BioRad A595)

	1	2	3	4	5	AVG	STDEV	Protein concentration (mg/mL)
pBSP1 1	0.6932	0.7381	0.7211	0.7167	0.6941	**0.7126**	**0.0171**	**0.2729**
pBSP1 2	0.6395	0.6827	0.6893	0.6734	0.6863	**0.6742**	**0.0182**	**0.2202**
pBS2P1	0.6839	0.7376	0.588	0.7247	0.7142	**0.6897**	**0.0538**	**0.2414**
C19	0.6433	0.7051	0.7068	0.6663	0.7007	**0.6844**	**0.0253**	**0.2342**
C20	0.7125	0.6895	0.6987	0.6922	0.6839	**0.6954**	**0.0098**	**0.2492**
MS2	0.6649	0.6663	0.7069	0.669	0.6581	**0.6730**	**0.0173**	**0.2186**
C22	0.6762	0.64	0.6877	0.6887	0.6735	**0.6732**	**0.0177**	**0.2188**
MS4	0.6974	0.6758	0.6783	0.6856	0.6913	**0.6857**	**0.0080**	**0.2359**
CSP	0.6649	0.6435	0.6793	0.6624	0.6629	**0.6626**	**0.0114**	**0.2043**
TMB17	0.6798	0.6664	0.6714	0.6609	0.6543	**0.6666**	**0.0087**	**0.2097**

Standard curve: y=ax+b x=(y-b) /a

Appendix G Alkaline phosphatase assay data

phoA assay 0,1mM 02.11.2011

Protein activity (slope x -1)

	1	2	3	4	5	AVG	STDEV	Slope / total protein
pASP1 1	1015.7	1007.4	1016.4	953.32	1010.6	**1000.68**	**23.91**	**8347.00**
pASP1 2	1046.3	1001	1003.8	923.14	995.75	**994.00**	**39.76**	**10859.63**
phoA AAG	1064	1012.6	1046.8	1064.4	975.13	**1032.59**	**34.36**	**8447.70**
CSP phoA	893.5	899.36	863.66	865.22	891.7	**882.69**	**15.12**	**11627.16**

Total Protein concentration (BioRad A595)

	1	2	3	4	5	AVG	STDEV	Protein concentration (mg/mL)
VK8 1	0.6932	0.7381	0.7211	0.7167	0.6941	**0.7126**	**0.0171**	**0.1199**
VK8 2	#0,6395	0.6827	0.6893	0.6734	0.6863	**0.6829**	**0.0060**	**0.0915**
phoA AAG	0.6839	0.7376	#0,588	0.7247	0.7142	**0.7151**	**0.0198**	**0.1222**
CSP phoA	0.6798	0.6664	0.6714	0.6609	0.6543	**0.6666**	**0.0087**	**0.0759**

phoA assay 0,1mM 11.11.2011; Assay = 5x dillution

Protein activity (slope x -1)

	1	2	3	4	5	AVG	STDEV	Slope / total protein
VK8 1	722.87	714.41	716.13	731.84	751.38	**727.33**	**13.50**	**19901.77**
VK8 2	726.87	698.54	719.2	735.39	734.57	**722.91**	**13.53**	**19918.34**
phoA AAG	729.22	686.61	686.69	702.79	722.42	**705.55**	**17.70**	**19604.74**
CSP phoA	711.65	663.03	694.42	649.62	687.25	**681.19**	**22.21**	**18306.78**

Total Protein concentration (BioRad A595)

	1	2	3	4	5	AVG	STDEV	Protein concentration (mg/mL)
VK8 1	0.7827	0.7921	0.8047	0.7812	0.7318	**0.7785**	**0.0248**	**0.1827**
VK8 2	0.7807	0.8072	0.7793	0.7733	0.7454	**0.7772**	**0.0197**	**0.1815**
phoA AAG	0.7634	0.8025	0.7841	0.7842	0.7437	**0.7756**	**0.0202**	**0.1799**
VK9	0.7764	0.7955	0.7841	0.7665	0.7874	**0.7820**	**0.0099**	**0.1860**

Appendix H Phosphoglucomutase assay data

December 8th, celB (phosphoglucomutase) assay 0,1mM induction, 2nd codon synonymous mutants

Protein activity (slope from assay)

	1	2	3	4	5	AVG	STDEV	Slope / total protein
celB wt	0.0036	0.0034	0.0034	0.0034	0.0032	**0.0034**	**0.00014**	**0.033**
CCA	0.0048	0.0048	0.0049	0.0048	0.0050	**0.0049**	**0.00009**	**0.041**
CCG	0.0030	0.0031	0.0030	0.0030	0.0030	**0.0030**	**0.00004**	**0.023**
CCT	0.0040	0.0041	0.0042	0.0041	0.0043	**0.0041**	**0.00011**	**0.037**

Total Protein concentration (BioRad A595)

	1	2	3	4	5	AVG	STDEV	Protein concentration (mg/mL)
celB wt	0.5949	0.5976	0.6088	0.609	0.6145	**0.6050**	**0.0074**	**0.1043**
CCA	0.6094	0.6281	0.6283	0.6269	0.6438	**0.6273**	**0.0109**	**0.1191**
CCG	0.6285	0.6492	0.6410	0.6437	0.6511	**0.6427**	**0.0080**	**0.1292**
CCT	0.5893	0.6136	0.6162	0.6223	0.6393	**0.6161**	**0.0161**	**0.1117**

December 13th, celB (phosphoglucomutase) assay 0,1mM induction, 2nd codon synonymous mutants

Protein activity (slope from assay)

	1	2	3	4	5	AVG	STDEV	Slope / total protein
celB wt	0.0035	0.0033	0.0034	0.0033	0.0031	**0.0033**	**0.00015**	**0.026**
CCA	0.0044	0.0045	0.0040	0.0037	0.0043	**0.0042**	**0.00033**	**0.030**
CCG	0.0026	0.0026	0.0026	0.0026	0.0025	**0.0026**	**0.00004**	**0.019**
CCT	0.0038	0.0037	0.0037	0.0037	0.0035	**0.0037**	**0.00011**	**0.031**

Total Protein concentration (BioRad A595)

	1	2	3	4	5	AVG	STDEV	Protein concentration (mg/mL)
celB wt	0.6241	0.6352	0.6465	0.6546	0.6515	**0.6424**	**0.0113**	**0.1290**
CCA	0.6311	0.6594	0.6694	0.6638	0.6528	**0.6553**	**0.0133**	**0.1376**
CCG	0.6512	0.6589	0.6635	0.6674	0.6482	**0.6578**	**0.0072**	**0.1392**
CCT	0.6051	0.6355	0.6281	0.6352	0.6295	**0.6267**	**0.0112**	**0.1187**

Appendix I qRT-PCR data

Bla 2nd codon qRT-PCR 27.05.2011

	Avg C_T Bla	Avg C_T 16S	ΔC_T Bla, 16S	ΔC_T StDev	$\Delta\Delta C_T$ Relative to VK6	RQ ($2^{-\Delta\Delta Ct}$)	Incorporated -	StDev +
pBSP1	22.762	12.462	10.300	0.101	0.000	**1.000**	**0.067**	**0.072**
pBS2P1	21.415	12.421	8.993	0.108	-1.307	**2.474**	**0.178**	**0.192**
AGC	23.442	12.822	10.621	0.054	0.321	**0.801**	**0.029**	**0.030**
TCC	21.818	12.943	8.876	0.141	-1.424	**2.683**	**0.251**	**0.276**
TCA	22.704	12.676	10.028	0.076	-0.272	**1.207**	**0.062**	**0.065**
TCG	22.448	12.786	9.662	0.115	-0.638	**1.556**	**0.119**	**0.129**
CSP	22.859	12.477	10.382	0.085	0.083	**0.944**	**0.054**	**0.057**

	Ct (bla)	Ct(16s)	Ct (bla)	Ct(16s)	Ct (bla)	Ct(16s)	Ct (bla)	Ct(16s)
pBSP1	22.76	12.453	TCC	21.8	12.957	CSP	22.828	12.484
pBSP1	22.73	12.381	TCC	21.796	12.912	CSP	22.936	12.469
pBSP1	22.71	12.566	TCC	21.677	12.901	CSP	22.814	12.414
pBSP1	22.85	12.447	TCC	22	13	CSP		12.541
Average	22.762	12.462	**Average**	21.8183	12.9425	**Average**	22.8593	12.477
StDev	0.065	0.077	**StDev**	0.13393	0.045347	**StDev**	0.06676	0.052211
pBS2P1	21.422	12.484	TCA	22.714	12.617			
pBS2P1	21.345	12.371	TCA	22.748	12.628			
pBS2P1	21.35	12.409	TCA	22.661	12.698			
pBS2P1	21.541		TCA	22.692	12.76			
Average	21.4145	12.4213	**Average**	22.7038	12.67575			
StDev	0.09138	0.0575	**StDev**	0.03665	0.066645			
AGC	23.415	12.805	TCG	22.453	12.845			
AGC	23.418	12.845	TCG	22.382	12.769			
AGC	23.42	12.834	TCG	22.36	12.748			
AGC	23.516	12.802	TCG	22.597	12.782			
Average	23.4423	12.8215	**Average**	22.448	12.786			
StDev	0.04921	0.0213	**StDev**	0.10697	0.041753			

Bla synonymous library qRT-PCR 02.11.2011

	Avg C_T Bla	Avg C_T 16S	ΔC_T Bla, 16S	ΔC_T StDev	$\Delta\Delta C_T$ Relative to VK6	RQ (2^-ΔΔCt)	Incorporated -	StDev +
pBSP1	23.592	13.441	10.150	0.243	0.000	1.000	0.155	0.184
pBS2P1	22.405	13.496	8.909	0.054	-1.242	2.365	0.087	0.090
C19	21.465	13.329	8.136	0.068	-2.014	4.040	0.186	0.195
C20	21.815	12.788	9.027	0.333	-1.123	2.178	0.450	0.566
MS2	21.470	13.216	8.254	0.072	-1.896	3.722	0.181	0.190
MS3	21.468	14.152	7.316	0.033	-2.834	7.130	0.162	0.165
MS4	21.819	13.392	8.427	0.213	-1.723	3.301	0.453	0.526
CSP	22.397	13.139	9.258	0.120	-0.892	1.856	0.148	0.161

	Ct (bla)	Ct (16s)
pBSP1 1	23.70	13.693
pBSP1 1	23.87	13.579
pBSP1 1	23.66	13.336
pBSP1 2	23.524	13.335
pBSP1 2	23.434	13.330
pBSP1 2	23.362	13.375
Average	**23.592**	**13.441**
StDev	**0.187**	**0.156**
pBS2P1	22.455	13.465
pBS2P1	22.395	13.52
pBS2P1	22.364	13.503
Average	22.4046667	13.496
StDev	0.04626374	0.02816026

	Ct (bla)	Ct (16s)
C19	21.462	13.340
C19	21.512	13.274
C19	21.42	13.372
Average	21.46467	13.32867
StDev	0.046058	0.049973
C20	22.000	12.906
C20	21.682	12.458
C20	21.763	13.000
Average	21.815	12.788
StDev	0.165254	0.289627
MS2	21.519	13.207
MS2	21.444	13.278
MS2	21.448	13.163
Average	21.47033	13.216
StDev	0.042194	0.058026

	Ct (bla)	Ct (16s)
MS3	21.471	14.132
MS3	21.45	14.172
MS3	21.484	
Average	21.46833	14.152
StDev	0.017156	0.028284
MS4	21.765	13.202
MS4	21.849	13.359
MS4	21.843	13.614
Average	21.819	13.39167
StDev	0.046861	0.207933
CSP	22.42	13.262
CSP	22.394	13.129
CSP	22.378	13.026
Average	22.39733	13.139
StDev	0.021197	0.118317

Bla synonymous library qRT-PCR 08.12.2011

	Avg C_T Bla	Avg C_T 16S	ΔC_T Bla, 16S	ΔC_T StDev	$\Delta\Delta C_T$ Relative to VK6	RQ (2^-ΔΔCt)	Incorporated -	StDev +
pBSP1	21.966	12.904	9.062	0.175	0.000	1.000	0.114	0.129
pBS2P1	20.411	12.929	7.482	0.184	-1.581	2.991	0.358	0.407
C19	20.174	13.241	6.933	0.071	-2.130	4.376	0.210	0.221
C20	21.677	13.787	7.890	0.166	-1.172	2.254	0.245	0.275
MS2	21.266	13.074	8.192	0.061	-0.870	1.828	0.075	0.078
MS3	20.756	13.577	7.179	0.063	-1.883	3.689	0.157	0.164
MS4	20.995	13.495	7.501	0.075	-1.562	2.952	0.149	0.157
CSP	22.352	13.399	8.953	0.144	-0.110	1.079	0.102	0.113
pTMB18	24.964	13.916	11.048	0.058	1.986	0.252	0.010	0.010

	Ct (bla)	Ct (16s)		Ct (bla)	Ct (16s)		Ct (bla)	Ct (16s)
pBSP1	21.89	12.995	C19	20.137	13.260	MS3	20.801	13.567
pBSP1	22.00	12.717	C19	20.153	13.185	MS3	20.742	13.629
pBSP1	22.01	13	C19	20.232	13.279	MS3	20.724	13.534
Average	**21.966**	**12.904**	**Average**	20.174	13.24133	**Average**	20.75567	13.57667
StDev	**0.067**	**0.162**	**StDev**	0.050863	0.049702	**StDev**	0.040278	0.048232
pBS2P1	20.385	12.725	C20	21.827	13.897	MS4	20.925	13.475
pBS2P1	20.382	13.052	C20	21.598	13.769	MS4	20.996	13.524
pBS2P1	20.465	13.01	C20	21.605	13.694	MS4	21.065	13.485
Average	20.4106667	12.929	**Average**	21.67667	13.78667	**Average**	20.99533	13.49467
StDev	0.04707795	0.1779129	**StDev**	0.13024	0.102647	**StDev**	0.070002	0.025891
pTMB18	24.93	13.949	MS2	21.298	13.069	CSP	22.276	13.352
pTMB18	25	13.883	MS2	21.261	13.024	CSP	22.33	13.528
pTMB18	24.962		MS2	21.239	13.129	CSP	22.45	13.318
Average	24.964	13.916	**Average**	21.266	13.074	**Average**	22.352	13.39933
StDev	0.03504283	0.04666905	**StDev**	0.029816	0.052678	**StDev**	0.089062	0.112718

Bla synonymous library qRT-PCR 23.02.2012

	Avg C_T bla	Avg C_T 16S	ΔC_T, bla 16S	ΔC_T StDev	$\Delta\Delta C_T$ Relative to C19 0	RQ (2^-ΔΔCt)	Incorporated -	StDev +
pBSP1	23.051	14.533	8.517	0.285	0.000	1.000	0.179	0.218
pBS2P1	21.731	14.651	7.080	0.256	-1.437	2.708	0.440	0.526
C19	21.111	14.725	6.386	0.083	-2.131	4.380	0.244	0.258
MS3	21.628	14.661	6.967	0.197	-1.550	2.928	0.375	0.430
pBSP1	23.256	14.608	**C19**	21.144	14.801			
pBSP1	23.161	14.490	**C19**	21.093	14.727			
pBSP1	22.735	14.502	**C19**	21.096	14.646			
AVERAGE	**23.051**	**14.533**	**AVERAGE**	**21.111**	**14.725**			
STDEV	**0.277**	**0.065**	**STDEV**	**0.029**	**0.078**			
pBS2P1	21.602	14.855	**MS3**	21.734	14.766			
pBS2P1	21.860	14.517	**MS3**	21.725	14.612			
pBS2P1		14.581	**MS3**	21.426	14.605			
AVERAGE	**21.731**	**14.651**	**AVERAGE**	**21.628**	**14.661**			
STDEV	**0.182**	**0.180**	**STDEV**	**0.175**	**0.091**			

phoA 2nd codons qRT-PCR 23.11.2011

	Avg C_T pho	Avg C_T 16S	ΔC_T, phoA 16S	ΔC_T StDev	$\Delta\Delta$CT Relative to pASP1	RQ (2^-$\Delta\Delta$Ct)	Incorporated -	StDev +
pASP1 1	23.321	15.054	8.267	0.098	0.000	1.000	0.065	0.070
pASP1 2	22.447	14.483	7.964	0.146	-0.303	1.233	0.119	0.131
phoA AAG	22.517	14.527	7.991	0.059	-0.276	1.211	0.049	0.051
phoA CSP	22.594	14.617	7.977	0.156	-0.290	1.222	0.125	0.139

	Ct (phoA)	Ct(16s)		Ct (phoA)	Ct(16s)
pASP1 1	23.33	15.066	**phoA AAG**	22.501	14.494
pASP1 1	23.40	15.101	**phoA AAG**	22.566	14.572
pASP1 1	23.23	14.995	**phoA AAG**	22.485	14.514
Average	23.321	15.054	**Average**	22.51733	14.52667
StDev	0.081	0.054	**StDev**	0.042899	0.040513
pASP1 2	22.333	14.456	**phoA CSP**	22.767	14.622
pASP1 2	22.4	14.508	**phoA CSP**	22.483	14.580
pASP1 2	22.608	14.485	**phoA CSP**	22.531	14.648
Average	22.447	14.483	**Average**	22.59367	14.61667
StDev	0.143398	0.026057628	**StDev**	0.152018	0.034312

celB 2nd codons Transcript amounts, December 14th 2011

	Avg C_T pho	Avg C_T 16S	ΔC_T, phoA 16S	ΔC_T StDev	$\Delta\Delta C_T$ Relative to VK8	RQ (2^-$\Delta\Delta$Ct)	Incorporated -	StDev +
celB wt	22.431	14.446	7.985	0.115	0.000	1.000	0.077	0.083
celB CCA	23.180	15.663	7.517	0.047	-0.469	1.384	0.044	0.046
celB CCG	23.585	14.983	8.602	0.128	0.617	0.652	0.055	0.060
celB CCT	22.205	14.501	7.704	0.095	-0.282	1.216	0.078	0.083

	Ct (phoA)	Ct(16s)		Ct (phoA)	Ct(16s)
celB wt	22.52	14.42	celB CCG	23.477	15
celB wt	22.39	14.543	celB CCG	23.555	14.993
celB wt	22.38	14.374	celB CCG	23.723	14.956
Average	22.431	14.446	**Average**	23.585	14.983
StDev	0.075	0.087	**StDev**	0.125714	0.023643
	Ct (phoA)	**Ct(16s)**		**Ct (phoA)**	**Ct(16s)**
celB CCA	23.176	15.615	celB CCT	22.228	14.459
celB CCA	23.196	15.703	celB CCT	22.187	14.437
celB CCA	23.167	15.671	celB CCT	22.2	14.608
Average	23.17967	15.663	**Average**	22.205	14.50133
StDev	0.014844	0.044542115	**StDev**	0.020952	0.093029

celB 2nd codons Transcript amounts, December 21st 2011

	Avg CT celB	Avg C_T 16S	ΔCT, celB 16S	ΔC_T StDev	$\Delta\Delta C_T$ Relative to VK8	RQ (2^-ΔΔCt)	Incorporated StDev -	+
celB wt	23.284	14.737	8.547	0.247	0.000	1.000	0.157	0.187
celB CCA	22.446	14.501	7.945	0.058	-0.602	1.517	0.060	0.063
celB CCG	23.897	14.513	9.384	0.105	0.837	0.560	0.039	0.042
celB CCT	23.362	14.764	8.597	0.204	0.050	0.966	0.127	0.147

	Ct (celB)	Ct(16s)		Ct (celB)	Ct(16s)
celB wt	23.37	14.977	celB CCG	23.889	14.509
celB wt	23.29	14.716	celB CCG	24	14.549
celB wt	23.19	14.519	celB CCG	23.801	14.481
Average	23.284	14.737	**Average**	23.89667	14.513
StDev	0.090	0.230	**StDev**	0.099721	0.034176

	Ct (celB)	Ct(16s)		Ct (celB)	Ct(16s)
celB CCA	22.425	14.553	celB CCT	23.353	14.650
celB CCA	22.426	14.465	celB CCT	23.368	14.643
celB CCA	22.487	14.484	celB CCT	23.364	15.000
Average	22.446	14.50066667	**Average**	23.36167	14.76433
StDev	0.035511	0.046306947	**StDev**	0.007767	0.204123

qRT-PCR celB69 aac(3)-IV library 01.03.2012

	Avg CT aac	Avg C_T 16S	ΔCT aac, 16S	ΔC_T StDev	$\Delta\Delta C_T$ Relative to wt	RQ (2^-ΔΔCt)	Incorporated StDev -	+
pAR69 wt	23.158	10.305	12.852	0.435	0.000	**1.000**	**0.260**	**0.352**
D2	23.228	10.742	12.486	0.200	-0.367	**1.289**	**0.167**	**0.192**
D4	24.051	10.845	13.206	0.388	0.354	**0.782**	**0.185**	**0.242**
D7	23.574	10.201	13.373	0.337	0.521	**0.697**	**0.145**	**0.183**
D11	21.202	11.019	10.182	0.559	-2.670	**6.364**	**2.045**	**3.012**
D13	23.956	10.777	13.179	0.295	0.327	**0.797**	**0.147**	**0.181**
D19	23.160	11.183	11.977	0.100	-0.876	**1.835**	**0.123**	**0.131**
D20	24.397	12.812	11.585	0.078	-1.267	**2.407**	**0.127**	**0.134**

	Ct (aac)	Ct(16s)		Ct (aac)	Ct(16s)		Ct (aac)	Ct(16s)
pAR69 wt	23.114	10.723	**D7**	23.580	10.030	**D19**	23.252	11.225
pAR69 wt	23.355	10.259	**D7**	23.538	10.587	**D19**	23.109	11.141
pAR69 wt	23.004	9.934	**D7**	23.604	9.986	**D19**	23.118	#12,42
Average	**23.158**	**10.305**	**Average**	23.574	10.201	**Average**	23.160	11.183
StDev	**0.180**	**0.397**	**StDev**	0.033	0.335	**StDev**	0.080	0.059
D2	23.207	10.525	**D11**	21.768	11.076	**D20**	24.322	#14
D2	23.278	10.799	**D11**	21.129	10.832	**D20**	24.409	12.836
D2	23.199	10.903	**D11**	20.708	11.150	**D20**	24.461	12.788
Average	23.228	10.742	**Average**	21.202	11.019	**Average**	24.397	12.812
StDev	0.043	0.195	**StDev**	0.534	0.166	**StDev**	0.070	0.034
D4	24.083	10.696	**D13**	24.040	10.584			
D4	24.044	11.284	**D13**	23.829	#12,22			
D4	24.026	10.554	**D13**	24.000	10.970			
Average	24.051	10.845	**Average**	23.956	10.777			
StDev	0.029	0.387	**StDev**	0.112	0.273			

qRT-PCR celB69 aac(3)-IV library 11.03.2012

	Avg CT aac	Avg C_T 16S	ΔCT aac, 16S	ΔC_T StDev	$\Delta\Delta C_T$ Relative to wt	RQ (2^-ΔΔCt)	Incorporated StDev -	+
pAR69 wt	23.439	12.103	11.336	0.215	0.000	1.000	0.138	0.160
D3	22.792	12.141	10.651	0.142	-0.685	1.607	0.150	0.166
D5	23.219	12.376	10.843	0.141	-0.493	1.408	0.131	0.145
D9	23.225	12.666	10.559	0.142	-0.777	1.714	0.160	0.177
D10	22.500	12.090	10.409	0.065	-0.927	1.901	0.083	0.087
D11	21.010	11.849	9.161	0.088	-2.176	4.517	0.266	0.283
D15	23.678	13.071	10.608	0.113	-0.728	1.657	0.125	0.135
D17	22.517	11.433	11.084	0.091	-0.252	1.191	0.073	0.077
D21	22.931	12.828	10.103	0.086	-1.233	2.351	0.137	0.145

	Ct (aac)	Ct(16s)		Ct (aac)	Ct(16s)		Ct (aac)	Ct(16s)
pAR69 wt	23.400	12.140	D9	23.186	12.709	D15	23.639	13.128
pAR69 wt	23.633	12.199	D9	23.170	12.754	D15	23.612	13.084
pAR69 wt	23.283	11.969	D9	23.319	12.535	D15	23.784	13.000
Average	**23.439**	**12.103**	**Average**	23.225	12.666	**Average**	23.678	13.071
StDev	**0.178**	**0.119**	**StDev**	0.082	0.116	**StDev**	0.093	0.065
D3	22.819	#12,636	D10	22.533	12.092	D17	22.501	11.517
D3	22.650	12.179	D10	22.446	12.134	D17	22.478	11.366
D3	22.908	12.103	D10	22.520	12.045	D17	22.571	11.416
Average	22.792	12.141	**Average**	22.500	12.090	**Average**	22.517	11.433
StDev	0.131	0.054	**StDev**	0.047	0.045	**StDev**	0.048	0.077
D5	23.224	12.278	D11	#21,585	#12,173	D21	22.843	12.851
D5	23.202	12.314	D11	20.959	11.885	D21	23.000	#13,522
D5	23.231	12.537	D11	21.060	11.813	D21	22.949	12.805
Average	23.219	12.376	**Average**	21.010	11.849	**Average**	22.931	12.828
StDev	0.015	0.140	**StDev**	0.071	0.051	**StDev**	0.080	0.033

pBSP1 mRNA degradation data

0 min

	bla	16S	ΔCT, bla 16S	Average ΔCT, bla 16S	ΔΔCT toative to pBSP1 0	Fold difference to pBSP1 0min
21.des	21.359	13.061	8.298		-0.129	1.094
	21.684	13.136	8.548	8.427	0.121	0.920
	21.295	12.859	8.436		0.009	0.994
19.jan	22.09	13.735	8.355		0.043	0.970
	21.996	13.973	8.023	8.312	-0.289	1.222
	22.369	13.812	8.557		0.245	0.844
24.jan	20.471	13.123	7.348		-0.009	1.006
	20.601	13.235	7.366	7.357	0.009	0.994
	22.222	13.127				
AVERAGE	**21.565**	**13.340**	**8.116**		0.000	1.005
STDEV	**0.689**	**0.393**	**0.498**		**0.159**	**0.113**

pBSP1 mRNA degradation data

2 min

	bla	16S	ΔCT, bla 16S	ΔΔCT Relative to pBSP1 0	Fold difference to pBSP1 0min
21.des	22.133	13.71	8.423	-0.004	1.003
	22.132	14.149	7.983	-0.444	1.361
	22.149	14.178	7.971	-0.456	1.372
19.jan	21.593	12.414	9.179	0.867	
	21.729	13.18	8.549	0.237	0.848
	21.394	13.014	8.38	0.068	0.954
24.jan	21.148	13.348	7.8	0.443	0.736
	21.216	13.393	7.823	0.466	0.724
	21.003	13.374	7.629	0.272	0.828
AVERAGE	**21.611**	**13.418**	**8.193**	0.161	0.978
STDEV	**0.452**	**0.551**	**0.486**	**0.429**	**0.258**

4 min

	bla	16S	ΔCT, bla 16S	ΔΔCT Relative to pBSP1 0	Fold difference to pBSP1 0min
21.des	22.122	13.289	8.833	0.406	0.755
	22.147	13.353	8.794	0.367	0.776
	22.077	13.242	8.835	0.408	0.754
19.jan	22.133	13.198	8.935	0.623	0.649
	22.075	13.142	8.933	0.621	0.650
	22.162	13.367	8.795	0.483	0.715
24.jan	20.94	13.13	7.81	0.453	0.731
	21.104	13.124	7.98	0.623	0.649
	21.215	13.016	8.199	0.842	
AVERAGE	**21.775**	**13.207**	**8.568**	0.536	0.710
STDEV	**0.522**	**0.116**	**0.443**	**0.153**	**0.053**

6 min

	bla	16S	ΔCT, bla 16S	ΔΔCT Relative to pBSP1 0	Fold difference to pBSP1 0min
21.des	22.338	13.066	9.272	0.845	0.557
	22.314	13.019	9.295	0.868	0.548
	22.402	13.048	9.354	0.927	0.526
19.jan	22.604	12.796	9.808	1.496	0.354
	22.253	12.459	9.794	1.482	0.358
	22.427	12.424	10.003	1.691	0.310
24.jan	21.837	12.994	8.843	1.486	0.357
	21.796	12.908	8.888	1.531	0.346
	21.921	13	8.921	1.564	
AVERAGE	**22.210**	**12.857**	**9.353**	1.321	0.419
STDEV	**0.288**	**0.249**	**0.433**	**0.338**	**0.104**

pBSP1 mRNA degradation data

8 min

	bla	16S	ΔCT, bla 16S	ΔΔCT Relative to pBSP1 0	Fold difference to pBSP1 0min
21.des	23.701	13.649	10.052	1.625	0.324
	23.238	13.792	9.446	1.019	
	23.242	13.68	9.562	1.135	
19.jan	24.203	13.083	11.12	2.808	0.143
	23.121	13.499	9.622	1.310	
	23.27	13.019	10.251	1.939	0.261
24.jan	23	12.615	10.385	3.028	0.123
	22.449	13.219	9.23	1.873	0.273
	22.441	13.085	9.356	1.999	
AVERAGE	**23.185**	**13.293**	**9.892**	1.860	0.225
STDEV	**0.553**	**0.387**	**0.614**	**0.696**	**0.088**

10 min

	bla	16S	ΔCT, bla 16S	ΔΔCT Relative to pBSP1 0	Fold difference to pBSP1 0min
21.des	23.107	12.652	10.455	2.028	0.245
	23.143	12.642	10.501	2.074	0.238
	23.137	12.579	10.558	2.131	0.228
19.jan	23.245	14.261	8.984	0.672	
	23.507	14.29	9.217	0.905	
	23.383	14.324	9.059	0.747	
24.jan	22.943	13.188	9.755	2.398	0.190
	22.863	13.021	9.842	2.485	0.179
	22.901	13.196	9.705	2.348	
AVERAGE	**23.137**	**13.350**	**9.786**	1.754	0.216
STDEV	**0.217**	**0.742**	**0.619**	**0.752**	**0.030**

C19 mRNA degradation data

0 min

	bla	16S	ΔCT, bla 16S	ΔΔCT Relative to pBSP1 0	Fold difference to pBSP1 0min
21.des	19.151	12.767	6.384	-2.043	4.122
	19.142	12.761	6.381	-2.046	4.131
	19.137	12.808	6.329	-2.098	4.282
19.jan	19.098	12.487	6.611	-1.816	
	19.098	12.571	6.527	-1.785	
	19.084	12.676	6.408	-1.904	3.742
24.jan	18.248	12.861	5.387	-1.970	3.918
	18.250	12.809	5.441	-1.916	3.774
	18.183	12.891	5.292	-2.065	4.184
AVERAGE	**18.821**	**12.737**	**6.084**	-1.960	4.022
STDEV	**0.447**	**0.134**	**0.541**	**0.112**	**0.211**

C19 mRNA degradation data

2 min

	bla	16S	ΔCT, bla 16S	ΔΔCT Relative to pBSP1 0	Fold difference to pBSP1 0min
21.des	21.359	15.085	6.274	-2.153	4.449
	21.280	15.147	6.133	-2.294	
	21.281	15.108	6.173	-2.254	4.771
19.jan	19.176	12.292	6.884	-1.428	
	19.147	12.599	6.548	-1.764	3.396
	19.165	12.409	6.756	-1.556	2.940
24.jan	18.733	13.260	5.473	-1.884	3.691
	18.670	13.122	5.548	-1.809	3.504
	18.780	12.863	5.917	-1.440	
AVERAGE	**19.732**	**13.543**	**6.190**	-1.842	3.792
STDEV	**1.196**	**1.218**	**0.493**	0.335	**0.688**

4 min

	bla	16S	ΔCT, bla 16S	ΔΔCT Relative to pBSP1 0	Fold difference to pBSP1 0min
21.des	21.147	14.042	7.105	-1.322	2.501
	21.156	13.996	7.160	-1.267	2.407
	21.199	13.989	7.210	-1.217	2.325
19.jan	21.409		21.409	13.097	
	20.400	17.666	2.734	-5.578	
	20.321	13.000	7.321	-0.991	1.987
24.jan	20.676	13.522	7.154	-0.203	
	20.370	13.600	6.770	-0.587	
	20.259	13.687	6.572	-0.785	1.723
AVERAGE	**20.771**	**14.188**	**8.159**	0.127	2.189
STDEV	**0.454**	**1.446**	**5.174**	5.115	**0.324**

6 min

	bla	16S	ΔCT, bla 16S	ΔΔCT Relative to pBSP1 0	Fold difference to pBSP1 0min
21.des	21.469	14.034	7.435	-0.992	1.989
	21.508	14.036	7.472	-0.955	1.939
	21.524	14.023	7.501	-0.926	1.900
19.jan	22.475	12.601	9.874	1.562	
	22.332	12.704	9.628	1.316	
	23.034	12.581	10.453	2.141	
24.jan	20.298	12.712	7.586	0.229	0.853
	20.429	12.886	7.543	0.186	0.879
	20.210	12.867	7.343	-0.014	1.010
AVERAGE	**21.475**	**13.160**	**8.315**	0.283	1.428
STDEV	**1.015**	**0.661**	**1.272**	1.166	**0.567**

C19 mRNA degradation data

8 min

	bla	16S	ΔCT, bla 16S	ΔΔCT Relative to pBSP1 0	Fold difference to pBSP1 0min
21.des	22.772	13.721	9.051	0.624	0.649
	22.546	13.805	8.741	0.314	0.805
	23.000	13.745	9.255	0.828	0.563
19.jan	21.122	12.245	8.877	0.565	0.676
	21.156	12.185	8.971	0.659	0.633
	21.237	12.324	8.913	0.601	0.659
24.jan	21.132	12.844	8.288	0.931	0.524
	21.010	12.963	8.047	0.690	0.620
	20.924	12.932	7.992	0.635	0.644
AVERAGE	**21.655**	**12.974**	**8.682**	0.650	0.641
STDEV	**0.850**	**0.656**	**0.458**	0.172	**0.078**

10 min

	bla	16S	ΔCT, bla 16S	ΔΔCT Relative to pBSP1 0	Fold difference to pBSP1 0min
21.des	22.822	13.847	8.975	0.548	0.684
	22.593	13.568	9.025	0.598	0.661
	22.630	14.000	8.630	0.203	0.869
19.jan	21.509	15.469	6.040	-2.272	
	21.502	12.233	9.269	0.957	0.515
	22.773	12.665	10.108	1.796	0.288
24.jan	21.137	12.610	8.527	1.170	0.444
	21.266	12.475	8.791	1.434	0.370
	21.521	12.946	8.575	1.218	0.430
AVERAGE	**21.973**	**13.313**	**8.660**	0.628	0.533
STDEV	**0.709**	**1.024**	**1.096**	1.192	**0.192**

Appendix J tAIcalc.py code and tAI.csv

```
#OPEN tAI TABLE
f=open("tAI.csv")
tAI=[]
i=0
import math

def geomeanlog(numbers):      ##Geometric mean of tAIs from all codons using
    summ=0                    ##log transformation
    for n in numbers:
        summ+=math.log10(n)
    logG=summ/len(numbers)
    return 10**logG

##Split comma-separated values
for line in f:
    linje=line.split(";")
    tAI.append(linje)
    tAI[i][1]=tAI[i][1].rstrip()
    i+=1
f.close()

toggle=1                      ##Toggle for program loop
#while toggle==1:             ##Run program
    f=open('output.csv','w')  ##Open/create outputfile
    print 'Please remember to copy the output.csv file for every analysis
if you want to keep it.'
    sequ=raw_input("Type / Paste sequence in CAPS! Whitespace allowed,
exclude stop codon: ")

    seq=''                    ##Split DNA sequence, remove whitespace
    for i in range(0,len(sequ)):
        if sequ[i]=='A'or sequ[i]=='T'or sequ[i]=='C'or sequ[i]=='G':
            seq+=sequ[i]

    a=0
    b=3
    codons=[]
    values=[]
    strings=['Codon no;tAI\n']      #ASSIGN tAI to individual codons
    if (len(seq)%3)==0:
        while a<len(seq):
            codons.append(seq[a:b])
            a+=3
            b+=3
        for x in range(0,len(codons)):
            for i in range(1,65):
                if codons[x]==tAI[i][0]:
                    codons[x]=[codons[x],tAI[i][1]]
                    values.append(float(tAI[i][1]))
                    strings.append(str(x+1)+';'+str(values[x])+'\n')

        tai=geomeanlog(values)      ##Geometric mean of tAIs from all codons
        f.writelines(strings)       ##Write to outputfile
        f.close()

        print codons                ##Display results
        print "The tAI of the given sequence is: ",tai
        print "Would you like to analyze one more sequence?"
```

```
        ans=raw_input("yes / no: ")       ##Prompt for new run
        if 'yes' not in ans:

            f.close()
            toggle=0
    else:
        print "Invalid number of bases!"  ##If not divided by three, prompt
        print"Would you like to analyze on more sequence?"
        ans=raw_input("yes / no: " )
        if ans!='yes':
            toggle=0
            f.close()

print "Thank you for using this awesome program!"
f.close()
```

tAI.csv

```
CODON;TAI
AAA;0.75
AAC;0.5
AAG;0.24
AAT;0.2195
ACA;0.125
ACC;0.25
ACG;0.29
ACT;0.10975
AGA;0.125
AGC;0.125
AGG;0.165
AGT;0.054875
ATA;0.163204
ATC;0.375
ATG;1
ATT;0.164625
CAA;0.25
CAC;0.125
CAG;0.33
CAT;0.054875
CCA;0.125
CCC;0.125
CCG;0.165
CCT;0.054875
CGA;0.00005
CGC;0.36
CGG;0.125
CGT;0.5
CTA;0.125
CTC;0.125
CTG;0.54
CTT;0.054875
GAA;0.5
GAC;0.375
GAG;0.16
GAT;0.164625
GCA;0.375
GCC;0.25
GCG;0.12
GCT;0.10975
GGA;0.125
GGC;0.5
GGG;0.165
GGT;0.2195
GTA;0.625
GTC;0.25
GTG;0.2
GTT;0.10975
TAA;0.163204
TAC;0.375
TAG;0.163204
TAT;0.164625
TCA;0.125
TCC;0.25
TCG;0.165
TCT;0.10975
TGA;0.125
TGC;0.125
TGG;0.165
TGT;0.054875
TTA;0.125
TTC;0.25
TTG;0.165
TTT;0.10975
```