UPPSALA
UNIVERSITET

UPTEC X 13 012
Examensarbete 30 hp
Juni 2013

Mutational analysis of the csgD mRNA leader:

search for a mode of regulation

Linnea Jonsäll

UPPSALA UNIVERSITET

Molecular Biotechnology Programme

Uppsala University School of Engineering

UPTEC X 13 012	**Date of issue 2013-06**

Author

Linnea Jonsäll

Title (English)

Mutational analysis of the csgD mRNA leader: search for a mode of regulation

Title (Swedish)

Abstract

The CsgD protein is the master regulator of a pathway leading to the formation of curli, in essence regulating the switch between a motile and a sessile lifestyle for bacteria. The 5'-UTR region of the csgD mRNA is a hotspot for multiple regulatory small RNAs (sRNA) involved in a complex regulatory network. Even though it is previously known how the interaction takes place it is unknown how sRNA binding affects the translational activity. In order to suggest a mode of regulation a mutational assay was performed by making changes in the csgD 5'-UTR and investigate what the translational effects were. Mutations in different regions are shown to affect the translation levels in various ways.

Keywords

Regulatory small RNA, enterobacteria, CsgD, OmrA, OmrB, mutational analysis, *in vivo* and *in vitro* translation assay

Supervisors

Gerhart Wagner
Uppsala University

Scientific reviewer

Magnus Lundgren
Uppsala University

Project name	Sponsors
Language **English**	Security
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **45**

Biology Education Centre Biomedical Center Husargatan 3 Uppsala
Box 592 S-75124 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 471 4687

Mutational analysis of the csgD mRNA leader: search for a mode of regulation

Linnea Jonsäll

Populärvetenskaplig sammanfattning

För encelliga organismer är den ständigt växlande omgivningen ett stort problem. För att organismen ska kunna överleva måste den kunna anpassa sig till olika temperaturer, pH och näringsämnen. Ett exempel på sådan anpassning är bakterien *E. coli* som antingen kan leva rörligt och simma omkring med hjälp av flageller, eller stillasittande i en biofilm där den håller sig fast med curli. Förmågan att kunna bilda biofilm och curli är nödvändig för *E. coli* eftersom det är en viktig virulensfaktor.

Cellen måste kunna reglera när den ska börja bilda curli, för om den väl har bytt livsstil till ett liv inkapslad i en biofilm är det svårt att byta tillbaka. Regleringen sker via transkriptionsfaktorn CsgD som indirekt inhibierar flageller och främjar curli. CsgD i sin tur regleras transkriptionellt via flera transkriptionsfaktorer och post-transkriptionellt av bland annat två regulatoriskta sRNA, OmrA och OmrB. Dessa sRNAn binder till mRNAt som kodar för CsgD så att det inte kan translateras. Bindningen sker dock långt ifrån den proteinkodande delen, och hur bindningen nedreglerar utryck av CsgD är okänt.

För att få en ledtråd till mekanismen görs en mutationsstudie där mutationer införs på valda ställen i *csgD* mRNA fuserat till läsramen för Green Fluorescent Protein (GFP). Hur regleringen via sRNA, eller translationseffektiviteten som sådan påverkas av mutationerna kan mätas genom hur mycket GFP som de olika mutanterna ger upphov till.

Under projektet påvisas effekter både på regleringseffektiviteten och de generella uttrycksnivåerna genom mutationer i vissa specifika delar av *csgD* mRNA.

Examensarbete 30 hp
Civilingenjörsprogrammet Molekylär bioteknik
Uppsala Universitetet, juni 2013

Table of contents

Abbreviations

BSA	Bovine serum albumin
Csg	Curli specific gene
DMF	Dimethylformamide
EDTA	Ethylenediaminetetraacetic acid
LA	Luria agar
LB	Luria broth
mRNA	Messenger RNA
Omr	OmpR-regulated sRNA
ORF	Open reading frame
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
RBS	Ribosome binding site
SD	Shine-Dalgarno
SDS	Sodium dodecyl sulfate
sRNA	Small (regulatory) RNA
TAE	Tris base, acetic acid and EDTA buffer
TBE	Tris base, borate and EDTA buffer
TF	Transcription factors

1. Introduction

1.1 Background

The complexities of bacterial regulatory processes have long been underappreciated. As single celled organisms, environmental conditions affect bacteria very directly. In response they have evolved an astonishing array of strategies to cope with harsh conditions such as heat and cold, high salinity, varying nutrient availability and changes in pH. With different prerequisites different life strategies become advantageous. As free living cells face new challenges they also might be forced to change their way of life.

Many species of bacteria, for example the well-known enterobacterium *Escherichia coli* which is the focus of this project, have the ability to transit between two very different mutually exclusive lifestyles. A motile lifestyle as a freely swimming bacteriuma or a sessile lifestyle as a stationary bacteriuma attached to a surface or embedded in a biofilm. Depending on the environment, these bacteria can shift behaviour to the most advantageous.

1.1.1 Motility and sessility

Bacterial biofilms serve as a protection for the inhabiting cells living embedded in it. Biofilms play an important role in the virulence for many bacteria, allowing the cells to attach to surfaces inside the host organism. They also protect the bacteria from antibiotics and attacks from the immune system. Biofilms consist of an extracellular matrix to which bacterial cells can attach. The matrix itself is formed cooperatively by the cells in the biofilm and excreted. The material in the matrix is largely exopolysaccharides such as cellulose. It is also made up of adhesive proteinaceous structures on the bacteria's cell surfaces such as adhesins and curli fimbriae.

Curli are a type of bacterial functional amyloid, a sort of protein structures on the outer surface of bacterial cells of many different species (Barnhart & Chapman, 2006). They are called curli due to the curled appearance of the protein fibres. Curli are used by the bacteria to attach to surfaces, and therefore may be an important factor for virulence. They are also an important part of biofilm matrix and are needed to attach the cells to the biofilm associated with a sessile lifestyle on a surface. (Holmqvist *et al*. 2012)

Flagella are propeller-like structures that the bacteria use for transporting themselves through media. Flagella are commonly associated with a motile lifestyle. As the cell prepares for the transition to sessile life flagella synthesis is inhibited. However, according to Mika and Hengge (2013) the flagella also play an important role in the initial steps of attachment to a surface.

Different lifestyles have different advantages and disadvantages, and must be regulated so that a bacterial population can optimize its behaviour. Once the adaptation to life in a biofilm has been made the cell can no longer easily switch back. A large number of small regulatory RNAs (sRNA) have been found to have an effect on the regulation of the expression of cellular components associated with each lifestyle. (Mika & Hengge, 2013)

1.1.2 Bacterial small RNAs

One growing interest in the field is bacterial regulation through bacterial small regulatory RNA (sRNA). These sRNAs are, as the name suggests, a type of RNA used by bacteria for regulation. sRNA are short RNA molecules usually about 50-300 nucleotides long. They typically do not contain open reading frames (ORF) and hence are not translated (Urban &

Vogel, 2007). sRNA regulation is very common in bacteria; in *E. coli* alone more than 80 sRNAs have been identified, affecting almost every physiological process. (Waters & Storz, 2009).

sRNAs regulate gene expression by several different mechanisms. Regulatory sRNA can act by interacting with target proteins to directly affecting protein activity. One mode of action is base pairing to messenger RNAs (mRNAs) and thus affects translation, or cause mRNA degradation. Base pairing can be complete or with limited complementarity.

The most studied group of sRNA is the trans-encoded sRNAs. These are small RNA molecules with act through imperfect base pairing to their target. The trans-encoded sRNAs, are transcribed from a segment of the DNA, distinct from that encoding the target mRNA, hence the name trans-encoded. Binding of a sRNA to an mRNA can have different effects on translational activity. Many sRNAs base-pair to the leader sequence of the mRNA, near or overlapping the ribosome binding site (RBS), and this interaction can block the ribosome binding and thus inhibit translation. The sRNA binding site on the mRNA can also be distant from the RBS but still have an effect of translation. Examples on such interactions are when binding of sRNA gives rise to or disrupts secondary structures, which might increase or decrease mRNAs stability. (Storz *et al.* 2011).

The RNA binding protein Hfq is commonly required for the function of the trans-encoded sRNAs in Gram-negative bacteria such as *E. coli*. (Storz *et al.* 2011). Hfq works as an RNA chaperon, ensuring the function of many sRNAs. Many of the sRNAs that are dependent on Hfq for proper function have inhibitory roles, to switch off expression of proteins which are no longer required. (Thomason *et al.* 2012). This protein will play an important part during the project.

1.1.3 Curli specific genes – csgD – and their regulation

One of the most exciting current research interests in this field concerns the transcription factor (TF) CsgD. This TF is a key regulator of many processes involved in the transition between motility and sessility in *E. coli*. CsgD is involved both in down-regulating production of flagella and enhancing the production of curli and cellulose. Thus curli synthesis is inversely correlated with flagella synthesis; making them mutually exclusive. These curli specific genes are involved in the synthesis of curli on many different levels. The csg genes are ordered into two operons. The *csg*BAC operon contains genes encoding the major curli subunits and proteins required for assembly of the curli fiber. The other operon, called *csg*DEFG, contains the gene encoding CsgD. Unlike many other TFs CsgD does not have any effect on its own expression. (Barnhart & Chapman, 2006). CsgD operates by directly binding to the promoter of the *csg*BAC operon and is therefore involved and required for the activation of the production of several important biofilm components (Boehm & Vogel, 2012).

In turn, such an important player must also be carefully regulated. CsgD is found in the centre of the complex pathway network that regulates the choice between motility and sessility. CsgD regulation is very complex and interestingly regulated at several different levels, both transcriptionally and post-transcriptionally, and is affected by signals both from within and outside of the cell. At least four different sRNAs have been shown to bind to and affect the translation levels from *csgD* mRNA. These sRNAs; McaS, RprA, GcvB and OmrA/OmrB (Jørgensen *et al*. 2012, Mika *et al*. 2012, Thomason *et al* 2012, Holmqvist *et al*. 2010) all have the same basic regulatory function, to basepair to the 5'-UTR of *csgD* mRNA and inhibit translation. Interestingly, the molecular mechanism by which these sRNAs repress CsgD

remains elusive. Some evidence regarding OmrA/B and McaS suggests that regulation is performed through direct interference with the initiation of translation. There are also indications which suggest that RprA and McaS are targeting secondary binding sites. (Boehm & Vogel, 2012).

In many ways this leader region resembles an RNA equivalent of the promoter region. (Boehm & Vogel, 2012). The leader sequence of the *csgD* mRNA has a long 5'-UTR region. This UTR region has a highly ordered secondary structure with multiple stem-loops; there are two regulatory modules. The region where the sRNAs bind is, interestingly, located far from the start codon in a genetically conserved region.

1.1.4 McaS

McaS is a 95 nt sRNA, which was the focus of attention for studies as early as ten years ago, however its function remained unknown. It is encoded in an intergenic region in the genome of *E.coli* and some related bacterial species, found between *abgR* and *ydaL*. Thomason *et al.* (2012) showed that McaS is a regulator that uses its three single stranded regions to regulate various pathways for biofilm syntesis. However, it has been shown that in the absence of McaS *csgD* is upregulated even if the interaction is not yet fully understood. (Boehm & Vogel, 2012). While McaS represses CsgD, and thus biofilm formation, it also activates FlhDC which the master regulator of a pathway that promotes formation of flagella. Contradictorily, it also is proposed to activate another pathway which gives rise to biofilm formation; however this is of a distinctly different kind than the CsgD-controlled biofilm. (Thomason *et al.* 2012, Boehm & Vogel, 2012).

1.1.5 GcvB

With the discovery of OmrA/B and McaS Jørgensen *et al*. (2012) suggested that the same region bound by these sRNAs might be a hotspot for Hfq-dependent sRNA binding. Thus the region was investigated with a reverse search strategy, aiming to isolate other sRNAs with regulatory function on *csgD*. GcvB, an sRNA previously known as a global regulator of amino acid transport, metabolism and synthesis, was found to interact with *csgD*. Another sRNA, RprA was also found in the same study. The predicted binding site for GcvB overlaps the region where OmrB is known to bind. Experiments indicate that GcvB is capable of inhibiting curli synthesis (Jørgensen *et al.* 2012). GcvB has been shown to impact as much as 1% of all mRNAs in *Salmonella* via its G/U-rich domain R1. (Sharma *et al.* 2011).

1.1.6 RprA

RprA is a slightly longer sRNA of 105 nt, found in another intergenic region between *ydiK* and *ydiL* in *E. coli*. Mika *et al*.(2012) suggest that RprA is a translational regulator of csgD. RprA is previously known to have an effect on the general stress sigma factor σ^S (RpoS). RpoS in turn is a known regulatory TF with effect on stationary phase gene expression. Among others, RpoS affects *csgD* translation, which is one of the clues to why RprA might be an interesting candidate for post-transcriptional regulation of *csgD*. RprA expression is also activated by a pathway leading to biofilm maturation, the RcsC/RcsD/RcsB two component pathway. In the study csgD was identified as a direct target for RprA and a RprA-*csgD* interaction is proposed. (Mika *et al*. 2012, Boehm & Vogel, 2012).

1.1.7 OmrA and OmrB: The sRNAs investigated in this project

OmrA and OmrB (OmpR-Regulated sRNA A and B) are two redundant sRNAs which are closely related in sequence which likely emerged from a gene duplication event. These sRNAs were studied by Holmqvist *et al*. (2010). The authors found that overexpression of either sRNA caused a decrease in CsgD levels. Fig. 1 shows the interaction site between *csgD* and OmrA/B. OmrA interacts with direct antisense basepairing on the stem-loop furthest away from the start codon on the 5'-UTR of the *csgD* mRNA. The interaction is direct and interrupts the secondary structure of the stem-loop. However, this interaction is a long distance from any known ribosome binding site or standby site and the same is true for the start codon and related structures. The ribosome covers about 50 nt, but OmrA/B binds much further away than that and cannot be accounted for covering the RBS.

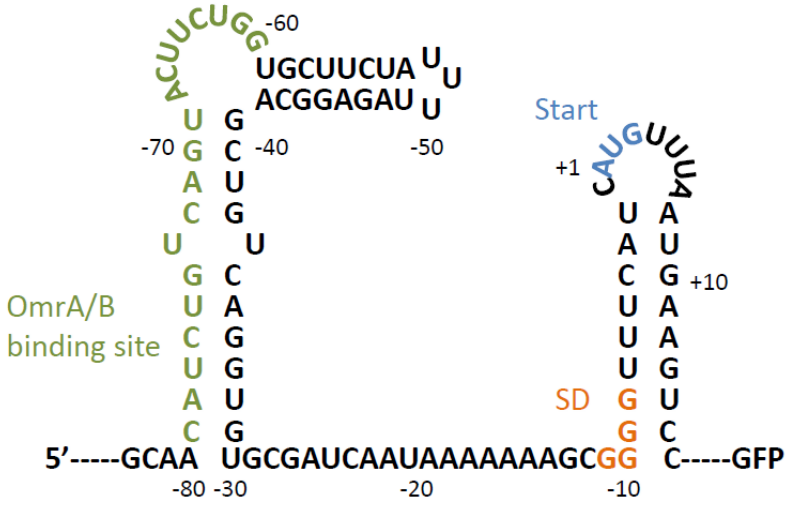


Figure 1. ***Secondary structure of csgD 5'-UTR*** *(in the figure the RNA sequence is shown fused to the sequence coding for GFP at the 3' end, as used in this project). The OmrA/B binding site is shown in green. As seen in the image the binding of OmrA/B is far from the Shine-Dalgarno (SD), the start codon and associated structures. Binding of OmrA/B unfolds the stem-loop, making it accessible.*

1.2 Outline of this project

The ultimate goal of this project is to propose a mechanism through which translation of CsgD protein is regulated by OmrA. It is previously known that the interaction between OmrA and the *csgD* mRNA takes place, what the interaction look like and what the effects are, but the mode of regulation is, as of yet, unknown. My project is a continuation on previous work, mainly by Erik Holmqvist. The project involves a mutational analysis of the *csgD* mRNA leader sequence. In order to propose a mode of regulation, a number of mutants were made, and the translational effect was studied. The experimental set up is inspired by recent research. First, different mutants were made by mutational polymerase chain reaction (PCR) on a plasmid containing a fusion of the *csgD* 5'-UTR and first several bases fused in-frame to GFP. The plasmids were transformed into *E. coli* strains unable to produce OmrA and OmrB, and in some cases also Hfq. The translation of the csgD::GFP fusion was monitored *in vivo* by measuring the fluorescence from GFP and *in vitro* by using an α-GFP antibody in Western blotting. Different conditions were tested, with and without the presence of the sRNA OmrA and the RNA binding protein Hfq.

2. Materials and methods

2.1 DNA sequences and primers

All primers used in this project can be found in the appendix section along with the full sequences of interest.

2.2 Strains and growth conditions

One Shot Top10 chemically competent *E. coli* (Invitrogen)
MC4100 relA+ omrAB- *E. coli*
MC4100 relA+ omrAB- hfq- (frt-Tet-frt) *E.coli*

Cells were grown on luria agar (LA) or in luria broth (LB) or M9 media (1x M9 salts, 0.40% glucose, 0.1 mM CaCL2, 2 mM MgSO4, 10 µg Thiamine, 1% Casamino acids). When appropriate the growth media was supplemented with antibiotics as a resistance marker. Antibiotics used are ampicillin [50 µg/ml], chloramphenicol [30 µg/ml] and Tetramycin. Incubation overnight in 37 °C or in room temperature over weekend (only possible when grown on LA plates).

2.3 Analysis

2.3.1 Analysis of DNA fragments

DNA fragments and PCR products were analyzed on 1% or 2% agarose gels (depending on product size) and run in tris base, acetic acid and EDTA buffer (TAE). The gel was stained with ethidium bromide and imaged on Gene Genius bio imaging system (SYNGENE).

2.3.2 Analysis of RNA fragments

RNA samples were analyzed on 4% acrylamide gel with 7.2 M urea and are run in tris base, borate and EDTA buffer (TBE). The gel was stained with StainAll in 10% dimethylformamide (DMF) diluted in dH_2O (1:1).

2.4 *In vivo* experiment assaying translational efficiency

2.4.1 Introducing mutations through mutational PCR

Plasmid pEH87 carrying a *csgD::GFP* fusion gene was used as a template and the mutations were introduced in the *csgD* leader region. PCR was performed using Phusion High-Fidelity DNA Polymerase (Thermo Scientific) with phosphorylated primers carrying mutations. The primers were phosphorylated using T4 Polynucleotide Kinase (Fermentas), following their standard protocol. The resulting PCR product was purified with a QIAquick PCR Purification Kit (Qiagen), DpnI treated to digest template plasmids and then ligated using Ready-To-Go T4 DNA ligase (GE Healthcare).

2.4.2 Amplification of plasmids

The mutant plasmids were introduced into One Shot Top10 chemically competent *E.coli* (Invitrogen), which were then incubated in 37 °C over night on LA with chloramphenicol [50 µg/ml]. Plasmids were extracted using QIAprep Spin Miniprep Kit and verified by sequencing performed by Uppsala Genome Center.

2.4.3 Creating chemically competent cells

Creating competent cells starting from overnight cultures of *E. coli* of strains MC4100 relA+ omrAB- and MC4100 relA+ omrAB- hfq- (frt-Tet-frt). The cells were first transferred to fresh

LB and were incubated in 37 °C, allowing the cells to reach optimal growth. The cells were pelleted by centrifugation and re-suspended in 2 ml 100 mM $CaCl_2$, 10mM Tris 7-9. The suspension was kept on ice for 30 minutes, then the pelleting and re-suspension steps were repeated in 600 µl 100 mM $CaCl_2$, 10mM Tris 7-9.

2.4.4 Transformation of strains

Transformation of competent cells was performed as follows. 20-50 ng plasmids were added to 100 µl of cell suspension. The mixture was kept on ice for five minutes, then heat-shocked in 42 °C for one minute. The cells were allowed to recover for 1.5 hours in 37 °C, and then they were plated on LA with ampicillin [50 µg/ml] and chloramphenicol [30 µg/ml]. The plates were incubate in 37 °C over night. A single colony from each plate was selected and re-streaked.

Transformation was performed twice. All cells were transformed with one of the plasmid types carrying the mutant *csgD::GFP* fusion gene. Each transformed cell had also received either a plasmid carrying a gene for expressing OmrA, OmrB or an empty control vector.

2.4.5 Measurement of fluorescence and cell density

Translation levels of *csgD::GFP* fusion mRNA was monitored as growth curves and fluorescence on a Tecan plate reader. Overnight culture in M9 media were diluted 1:100 to a final volume of 500 µl. 100 µl was added to one well on a 96 well clear bottom, black, assay plate with lid (Corning Incorporated), together with ampicillin and chloramphenicol as resistance markers. The plate reader is run for 16 hours overnight in 37 °C. A corresponding strain but without the ability to produce GFP was used as an autofluosescence control.

2.5 *In vitro* translation

2.5.1 Preparation of starting material by PCR

The starting material for the *in vitro* translation was made by PCR. Each plasmid carrying a *csgD::GFP* fusion was amplified by PCR, using a primer carrying the T7 promoter. The resulting plasmid fragment contained the fusion gene under control of the T7 promoter at the 5’ end. PCR was carried out with Phusion High-Fidelity DNA Polymerase (Thermo Scientific).

2.5.2 Preparing mRNA by i*n vitro* transcription

The DNA plasmid fragments were transcribed into mRNA by so called *in vitro* transcription in the following reaction mixture:

T7 Buffer x10	15 µl
DTT [0.5 M]	1.5 µl
Bovine serum albumin (BSA) x100	1.5 µl
Ribolock [40 u/µl]	1 µl
ATP [0.1 M]	6 µl
CTP [0.1 M]	6 µl
UTP [0.1 M]	6 µl
GTP [0.1 M]	6 µl
Spermidine [0.5 M]	0.3 µl
T7 RNA polymerase	3 µl
DNA template from PCR	95 µl
dH_2O	8.7 µl

The reaction mixture was incubated in 37 °C for two to four hours. The recommended mRNA amount is 25 µg of a 1 kb fragment per 100 µl according to the original protocol. However, only approximately 3-13 µg per 150 µl transcription reaction was generally used. The samples were treated with DNaseI (Fermentas) to remove remaining DNA template.

2.5.3 mRNA extraction

mRNA was extracted by phenol-chloroform extraction, using saturated phenol pH 7 and a mixture of chloroform and isoamyl alcohol (24:1). 2.5 volumes of -20 °C absolute ethanol are added, along with 12 mM of NaAC. The samples were either cooled to -80 °C for one hour or to -20 °C over night. The samples were pelleted by centrifugation 30 minutes in 4 °C, supernatant was replaced by -20 °C 70% ethanol and then centrifuged again. Supernatant was removed and the pellets were dried and re-suspended.

2.5.4 mRNA purification

Purification was carried out by gel electrophoresis. The gel was a 4% acrylamide gel with 7.2 M urea and is run in TBE. The samples were re-suspended in Urea Blue. mRNA bands were localized by UV shadowing and cut from the gel. The cut bands were submerged in elution buffer ($AcNH_4$ 500 mM, Ethylenediaminetetraacetic acid (EDTA) 0.1 mM, sodium dodecyl sulfate (SDS) 0.1%), and 1/5 volumes of phenol is added. The samples were put on a shaker in 4 °C over night. The following day the mRNA is extracted again as above and re-suspended in 25 µl of dH_2O.

2.6 *In vitro* translation assay

The translation assay was carried out using using PURExpress® *In Vitro* Protein Synthesis Kit (New England Biolabs) following their standard protocol. Different reaction volumes were used (5-25 µl), the mRNA concentration used is either 0.4 or 0.5 µM. The reaction was stopped by adding one volume of a mixture of β-mercaptoethanol and Western loading dye (1:10).

2.7 Western blotting

2.7.1 Gel electrophoresis

Gel electrophoresis on 10% acrylamide gel (Acrylamide 37.5:1), run in TGS running buffer (25 mM Tris base, 190 mM Glycine, 0.1% SDS).

2.7.2 Wet transfer

The bands were transferred to a blot membrane (Pall Corporation) by wet transfer. Transfer was done at 4 °C, with ice, on a stirrer in Transfer buffer (200 ml Methanol, 100 ml TGS x10, 700 ml dH_2O). Transfer was run over night at 35 mA.

2.7.3 Blocking and antibodies

Blocking was performed for one hour in blocking solution (Phosphate-buffered saline (PBS)-Tween (0.1%) with 3% BSA). Blocking solution is removed and replaced by fresh blocking solution with 1:5000 Anti-GFP-HRP antibody (Miltenyi Biotec) for another hour. The membrane was washed three times with PBS-Tween and twice with PBS.

2.7.4 Development of membrane

Membranes were developed using Amersham™ ECL plus Western Blotting Detection System (GH Healthcare) following their standard protocol.

2.7.5 Imaging

Images were made with BIORAD Imager, using settings for colorimetric and chemiluminescent measurements. The resulting images were analyzed with Imagelab.

3. Results

3.1 Mutations are introduced in the *csgD* 5'-UTR carried by a plasmid

To assay the mode of regulation of OmrA on *csgD* we wanted to study effects different mutations in the *csgD* 5'-UTR region would have on translation. To be able to measure the translation levels the Green Fluorescent Protein (GFP) was exploited. The gene encoding GFP was fused to the leader sequence of the *csgD* gene, so that the *csgD* Open Reading Frame (ORF) was in frame with the ORF encoding GFP. The fusion mRNA includes the entire *csgD* 5'-UTR region, along with the first several bases encoding CsgD as well as the full sequence encoding GFP. The construct is introduced into a plasmid as shown in fig. 2. When the *csgD::gfp* fusion was transcribed the mRNA had the sequence that binds OmrA/B and thus the fusion mRNA was regulated like *csgD*, but translated into GFP. Thus translation levels could easily be measured in the form of fluorescence (this will tell how much GFP is translated). We also used a specific α-GFP antibody in Western blotting. The effect of the different mutations was compared to how much GFP we got when GFP is fused to a wild-type *csgD* sequence. Mutations were introduced into the construct by PCR with primers containing the desired mutations. The primers can be found in appendix A.

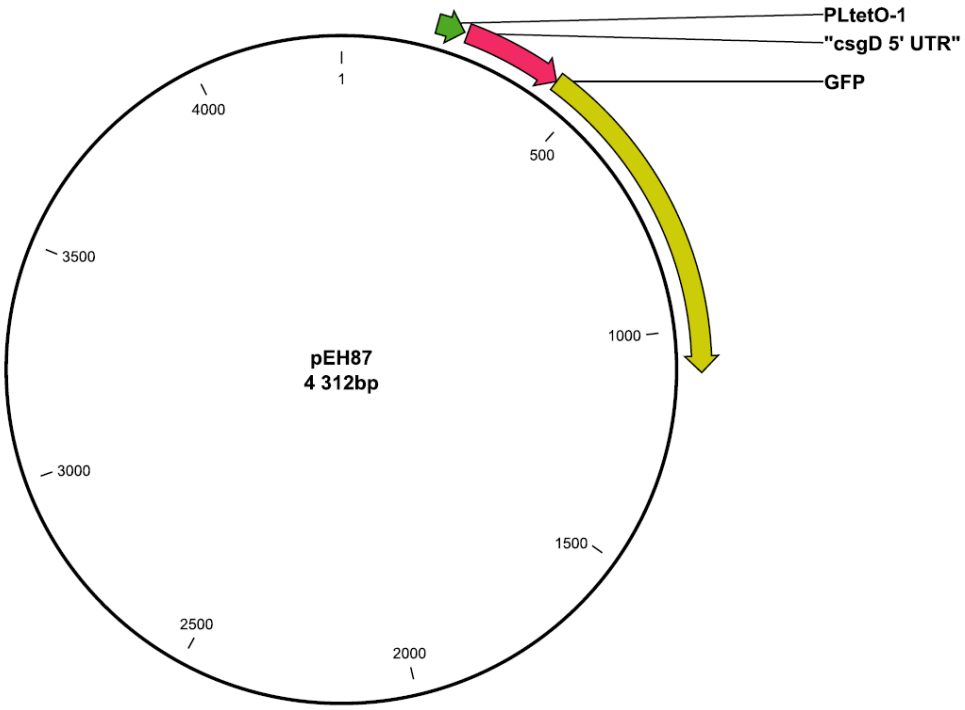


Figure 2. ***Plasmid pEH87 construct.*** *The construct used includes the PLtetO-1 promoter and the 5'-UTR from the gene encoding CsgD fused in frame to a gene encoding GFP. The plasmid is also carrying a resistance marker for chloramphenicol (not shown).*

3.2 Initial study: Mutants identified by FACS

The original plan was to assay a multitude of mutants all over the *csgD* 5'-UTR which potentially have effects on regulation or have a translational effect per se. These promising mutants had been revealed in a previous study, prior to my project. This study was a massive functional mapping of the *csgD* 5'-UTR. In this mutational mapping a large number of mutations in *csgD::gfp* fusions were made by error-prone PCR, and were carried on plasmids

into living *E. coli* cells. The cells were then sorted by fluorescent activated cell sorting (FACS) and sequenced. Mutations significantly enriched in one category (high or low translation) would be interesting for further study. (Holmqvist *et al*. 2013). In the very beginning of this project I worked with a few of these interesting mutants, which had been left uncharacterized after a previous project.

3.2.1 Mutants subject to study

The first mutants to be tested were designed by Erik Holmqvist. Most of these had mutations centred on the stem-loops in *csgD* 5'-UTR. Some of the mutant plasmids had already been created and transformed into the appropriate. The rest (the ones designated pLJXXX) were mutated and transformed by me, but used primers that had previously been designed and ordered. The mutants from the first part of the project were tested in an Δ*omrA/B* strain, unable to produce OmrA/B. These were transformed with plasmids carrying OmrA, OmrB or an empty control vector. This gives rise to three different backgrounds; No OmrA/B, only OmrA and only OmrB. The cells where then transformed with a plasmid carrying mutated versions of the *csgD::GFP* construct. Fig. 3 shows the placements of the mutants within the *csgD* 5'-UTR, and table 1 gives more detailed information for each mutation.

The effects of the mutations are assayed on a Tecan plate reader, measuring the fluorescence and OD_{600}. The fluorescence is normalized to the OD_{600} which gives a value for how much GFP each cell is producing. The translation levels for the mutants are compared to the wild-type.

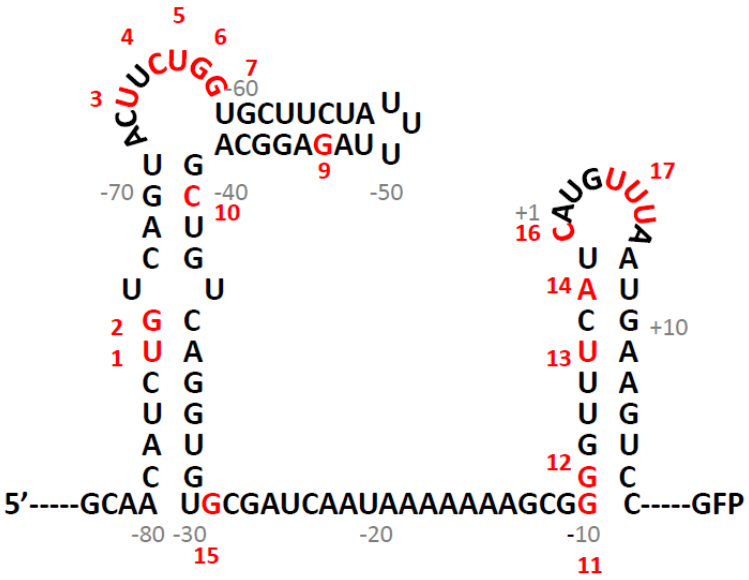


Figure 3. ***Initial trial mutations and their location*** *on the csgD 5'-UTR. The mutations are mostly concentrated on the stem-loops. Red in the image indicates a mutation, however note that the figure only shows the placement of the mutations. Refer to table 1 for full information about the mutations.*

Table 1. ***Mutants for the initial study****, designed by Erik Holmqvist. Mutant 8 is excluded from further study since sequencing showed that the mutation had not been successful.*

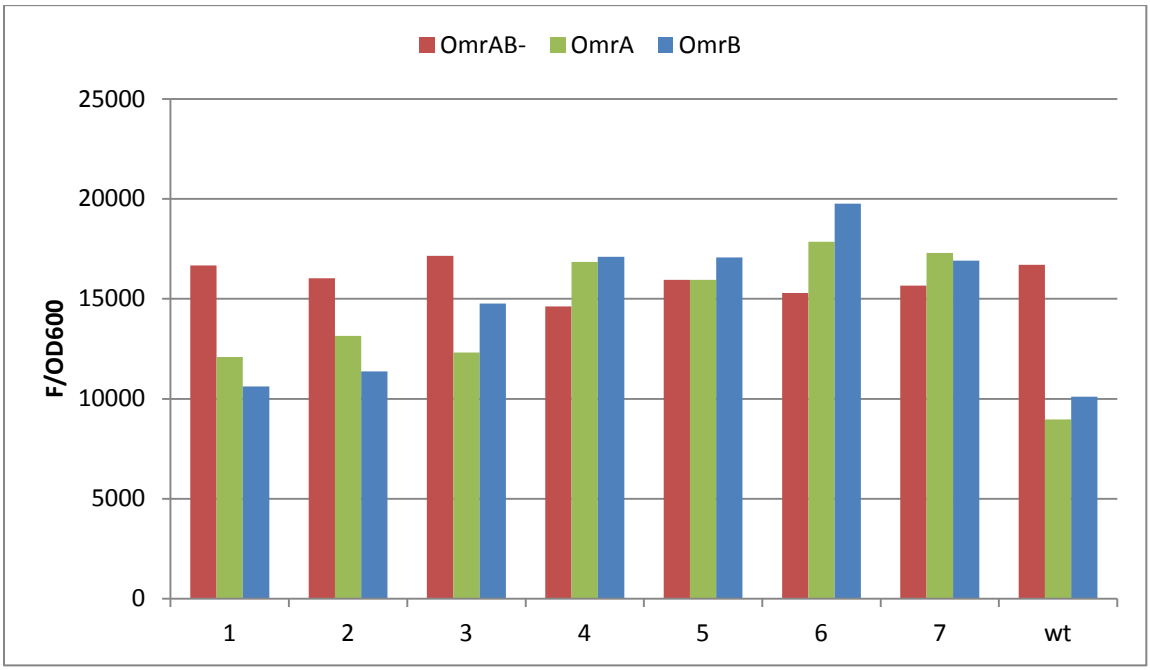
Mutation name	Mutation	Plasmid
1	-75 T:C	pEH211
2	-74 G:A	pEH212
3	-66 T:A	pEH213
4	-64 C:A	pEH214
5	-63 T:A	pEH215
6	-62 G:A	pEH216
7	-61 G:T	pEH217
8	-62 G:C	
9	-47 G:A	pEH218
10	-40 C:A	pLJ001
11	-10 G:A	pEH219
12	-9 G:A	pEH220
13	-6 T:C	pLJ002
14	-3 A:G	pLJ003
15	-29 G:C	pLJ004
16	-1 C:G	pEH221
17	+4,5,6 UUU=>AAA	pLJ005
wt	Wild-type (No mutation)	pEH87

3.2.2 Observations from fluorescence measurements for the initial set of mutants

As seen in fig. 4 and in table 2, several of the mutations showed very promising results in these preliminary measurements. The fluorescence levels in absence of OmrA/B were strongly affected for in particular mutation 11, 13 and 14. Interesting to note is that all of them are in the same stem-loop as the start site, and thus are likely to cause structural changes making the Shine-Dalgarno (SD) or start codon more accessible. 4, 5, 6, 7 and 10 seem to have a promising effect on the efficiency of regulation in the presence of OmrA/B; these mutations might cause loss of regulation. 4, 5, 6 and 7 are all located in or near the site where OmrA/B initiates binding to *csgD*. These mutants in particular would have been very interesting for further study.

However, this part of the project was abandoned for another set of mutations. Even though I did not do much work on these mutants myself, it would be an intriguing project to pursue. Since the results using the OmrA and OmrB plasmids were clearly similar in most cases OmrB was excluded for the remainder of the project.

a)



b)

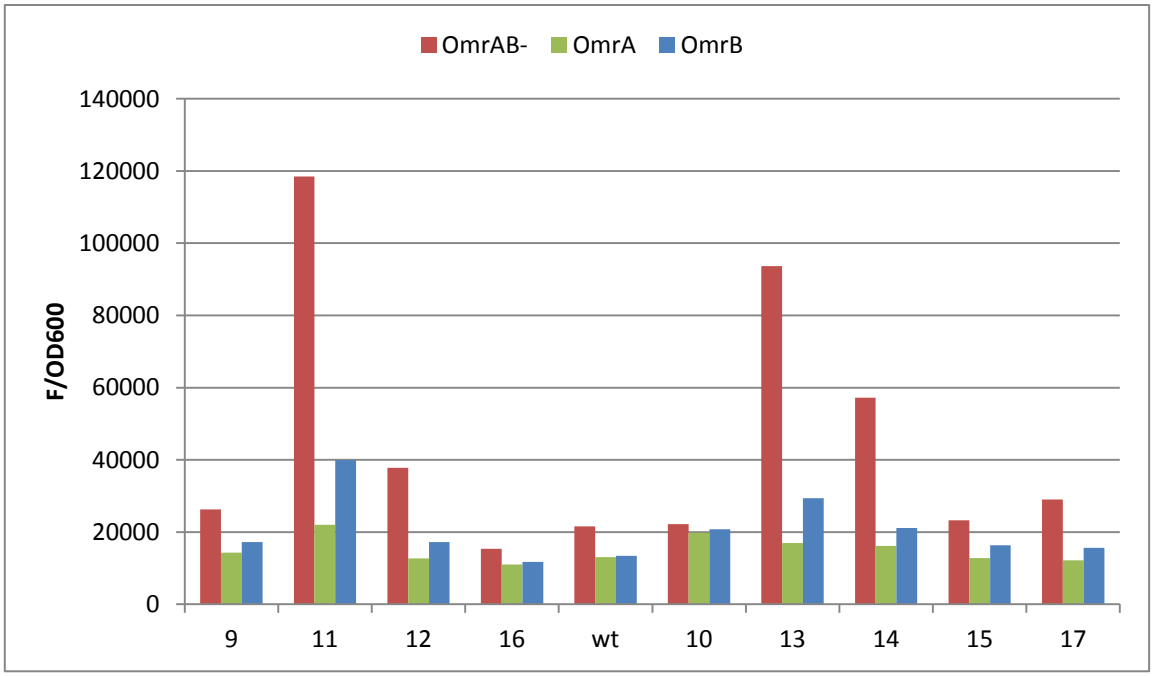


Figure 4. ***The initial GFP fluorescence measurement experiment.***
Fluorescence was measured during 16 hours of growth in Tecan plate reader. The graphs show the measured fluorescence from GFP, standardized with respect to OD600 (cell density). The experiment was performed in three different backgrounds; in cells without OmrA/B, in cells with OmrA and in cells with OmrB. The experiment was performed with technical triplicates, but no biological replicates. The figures show the mean of the technical triplicates. The figures are shown on different scales since 4b includes some mutants with very high translation. a) Shows mutants 1-7, and the wild-type. b) Shows mutants 9-17 (in the order they were created) and the wild-type.

Table 2. ***Fluorecense ratio.*** *The ratio between the fluorescence for the mutants in OmrAB- background and the mutants in OmrA respective OmrB backgrounds.*

	Fluorescence(OmrAB-)/ Fluorescence(OmrA+)	Fluorescence(OmrAB-)/ Fluorescence(OmrB+)
wt	1,76	1,65
1	1,38	1,57
2	1,22	1,41
3	1,39	1,16
4	0,87	0,85
5	1,00	0,93
6	0,86	0,77
7	0,91	0,93
9	1,83	1,52
10	1,12	1,07
11	5,39	2,97
12	2,96	2,19
13	5,51	3,19
14	3,54	2,71
15	1,82	1,42
16	1,40	1,31
17	2,38	1,86

3.3 The A-stretch and the possibility of Hfq binding

Shortly into the project my supervisor Gerhart Wagner suggested that we should shift focus to another part of the *csgD* 5'-UTR. The RNA binding protein Hfq is known to bind a so called ARN-motif (where A stands for adenine, R for Purine (adenine or guanine) and N represents any base) on RNA. A motif that fit this description is found in the *csgD* 5'-UTR, in the single-stranded region between the two major stem-loops. This region, the 'A-stretch', consists of a 10 nt long stretch which only contains the bases A and one single U (at the 'N position' in the ARN motif). It was unknown what kind of effects mutations would have on regulation and translation levels. Wagner designed four mutations in the A stretch which all were supposed to disrupt the ARN-motif in different ways. One of the mutants had the A stretch shortened by three nucleotides and the others were of original length but had different sequences. A mutant with the entire A-stretch deleted which was not fully characterized after a previous project by Erik Holmqvist was also included in the assay. All the mutants are shown in table 3 and fig. 5. The advantage of only mutating the sequence in contrast to deleting it is that the original distance between the stem-loops stays untouched. However, both mutations and deletions might lead to changes in the mRNA secondary structure, which might affect the availability of for example the SD region or the start codon.

Table 3. ***The second set of mutations.*** *Mutants designed by Gerhart Wagner. Also includes ΔA by Erik Holmqvist.*

Mutation name	Mutation	Plasmid
MS	-22,21 AU:C, -19,18 AA:G, -16,15 AA:C	pLJ006
M1	-21 U:C, - 18 A:C, -16 A:C	pLJ007
M2	-20 A:C, -19 A:C, -15 A:C	pLJ008
M3	-21 U:C, -19 A:G, -17 A:C, -15 A:C	pLJ009
ΔA	A-stretch deletion	pEH105
wt	Wild-type (No mutation)	pEH87

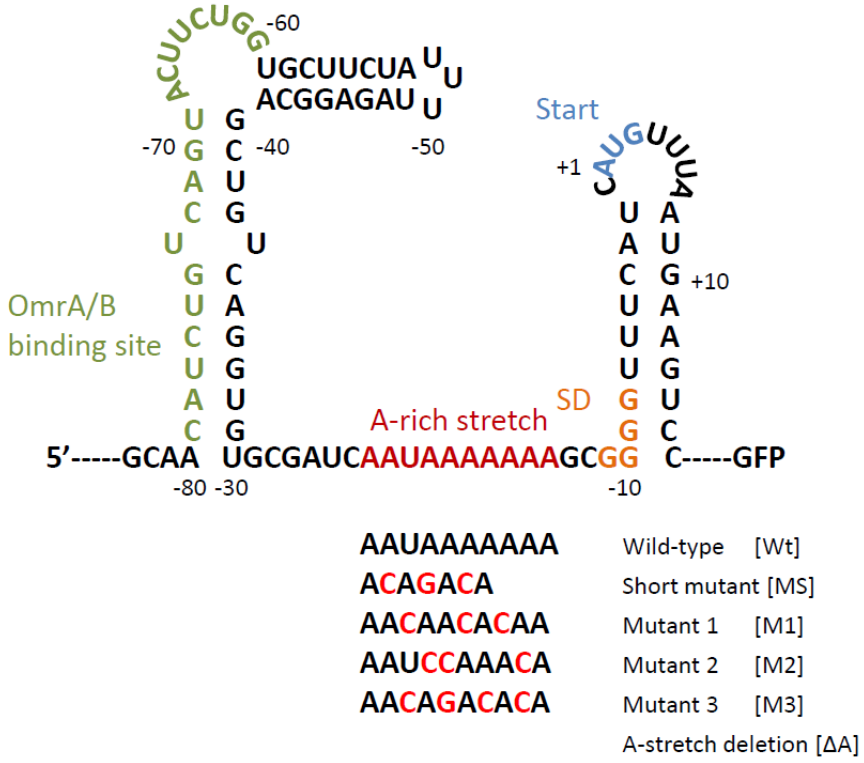


Figure 5. ***A-stretch mutations and their location on the csgD 5'-UTR.*** *The A-rich sequence is located in between the two major stem-loops, in fugure shown in red. Four mutations designed to disrupt potential Hfq binding are designed for this sequence, along with a fifth mutant with the entire A-rich sequence deleted.*

3.3.1 *In vivo* study of the A-stretch mutants

According to a secondary structure analysis performed by Cédric Romilly (not shown), the mutations do not seem to affect the secondary structure of the *csgD* 5'-UTR. Without structural changes around the start codon it is unlikely that the mRNA is simply more available for translation. Any effects on translational efficiency might therefore have other causes.

To assay which effect OmrA has on *csgD* translational activity it is necessary to test translational activity in the presence and absence of OmrA. It is also interesting to study the effect that the RNA binding protein Hfq has on regulation. To do so, one *E. coli* strain with OmrA/B deleted, and another with both OmrA/B and Hfq deleted were used. The strains were transformed with a plasmid carrying either OmrA or an empty control vector. Thereafter they

were also transformed with one plasmid carrying the *csgD::gfp* fusion, including mutations. The result is that each mutation is analyzed in each of the four different cell backgrounds: Without Hfq and OmrA/B, with Hfq but without OmrA/B, without Hfq but with OmrA and finally with both Hfq and OmrA present. Again, the new mutations effect on translational efficiency was assayed by measuring the fluorescence in a plate reader.

3.3.2 Observations from fluorescence measurements for the A-stretch mutants

These fluorescence measurements clearly show two things in particular. First, in the Hfq deletion strain regulation is almost completely lost, indicating that Hfq is necessary for translational regulation by OmrA. Secondly, it also shows that all of the A-stretch mutants kept the translational regulation in presence of both Hfq and OmrA. As seen in fig. 6 and table 4 below, the ratio of the GFP translated without and with OmrA is very close to 1 when Hfq is absent. However, when Hfq is present the translation of GFP decreases three- to fivefold compared to the translation without OmrA. These ratios give an indication of how well OmrA regulate a certain *csgD* 5'-UTR mutant. In these experiments it seems like the regulation of the *csgD* 5'-UTR sequence with mutation MS is slightly more effective than the others, while the ΔA mutation seems to be less efficient.

Table 4. ***Relative regulatory efficiency by OmrA on csgD in the absence and presence of Hfq.*** *The ratio between the fluorescence without and with OmrA is listed for every mutant.*

Hfq-	**Fluorescence(OmrA-)/ Fluorescence(OmrA+)**
MS	1,13
M1	1,10
M2	1,12
M3	1,14
ΔA	1,13
wt	1,05
Hfq+	**Fluorescence(OmrA-)/ Fluorescence(OmrA+)**
MS	5,50
M1	4,51
M2	4,18
M3	4,22
ΔA	3,57
wt	4,87

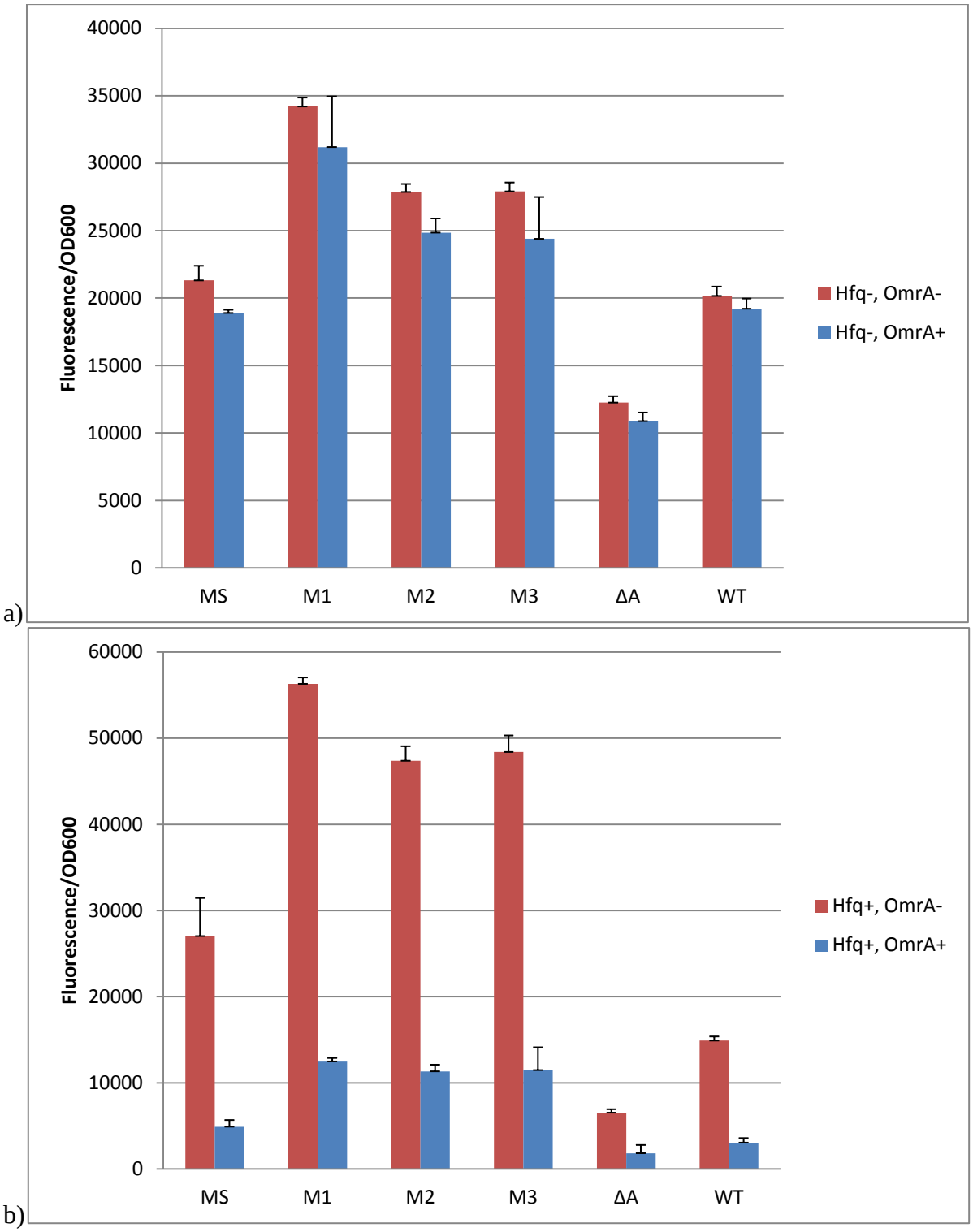


Figure 6. ***Fluorescence measurement experiment on A-stretch mutants.*** *16 hour run on Tecan plate reader in 37 °C. Biological quadruple replicates were used. The graphs show the mean of the biological replicates. The error bars show the standard deviation between the replicates.a) Measurement on the Hfq deletion cell background. Cells with and without the plasmid carrying OmrA are used.b) Measurement in a background with Hfq. Cells with and without the plasmid carrying OmrA are used.*

3.4 *In vitro* translation assay on the A-stretch mutants

However, an assay on living cells will always be affected by interference from other components than the ones we are currently interested in. To further analyze the mutants' effects on translational efficiency a cell-free *in vitro* translation assay was also performed. In the in vitro translation assay only *csgD::gfp* fusion mRNA and proteins required for translation will be present, effectively removing most of the noise seen in living systems. By performing this analysis with and without addition of Hfq and/or OmrA/B, different conditions are surveyed.

The starting material for the *in vitro* translation assay was PCR product, prepared with a forward primer carrying the strong and specific T7 promoter. The PCR product was transcribed *in vitro* as described in Materials and Methods. It is possible to directly use PURExpress® *In Vitro* Protein Synthesis Kit (New England Biolabs) on isolated DNA, for example PCR product, but in this case we decided against that. The reasoning behind is that since we are interested in the translational efficiency, the translational efficiency could also vary and interfere with the readings. The assay is performed as described and then is analyzed on Western blots. An α-GFP antibody is used to detect and visualize the GFP protein, assisted by chemiluminescence.

3.4.1 Optimization of the *in vitro* experimental set up

The first trials using *in vitro* translation were unsuccessful and did not yield any results. Before the real experiments could begin the system was tested and optimized. As seen in fig. 7b the detection limit for this antibody is between 100-500 µM of GFP protein. This was tried out by blotting a dilution series of purified GFP. An α-HIS antibody (targeting a tag of histidine residues) is also used on the same dilution series for comparison, producing a similar result as seen in fig. 7a.

During this project the *in vitro* translation assays were generally carried out in smaller volumes than recommended by the manufacturer. This was also tested beforehand to make sure that the smaller reaction volume would not impact on performance. The same reagent concentrations were used for five different reaction volumes between 5 µL and 25 µL. This was tested for two different proteins. First, in fig. 7c a csgD-3xFLAG tagged protein was used (see Appendix A, plasmid pEH110), which is a different construct lacking GFP. In fig. 7d the original *csgD::GFP* wild-type construct (from pEH87) was used. In both cases the reaction volume does not have any noteworthy effect on translation.

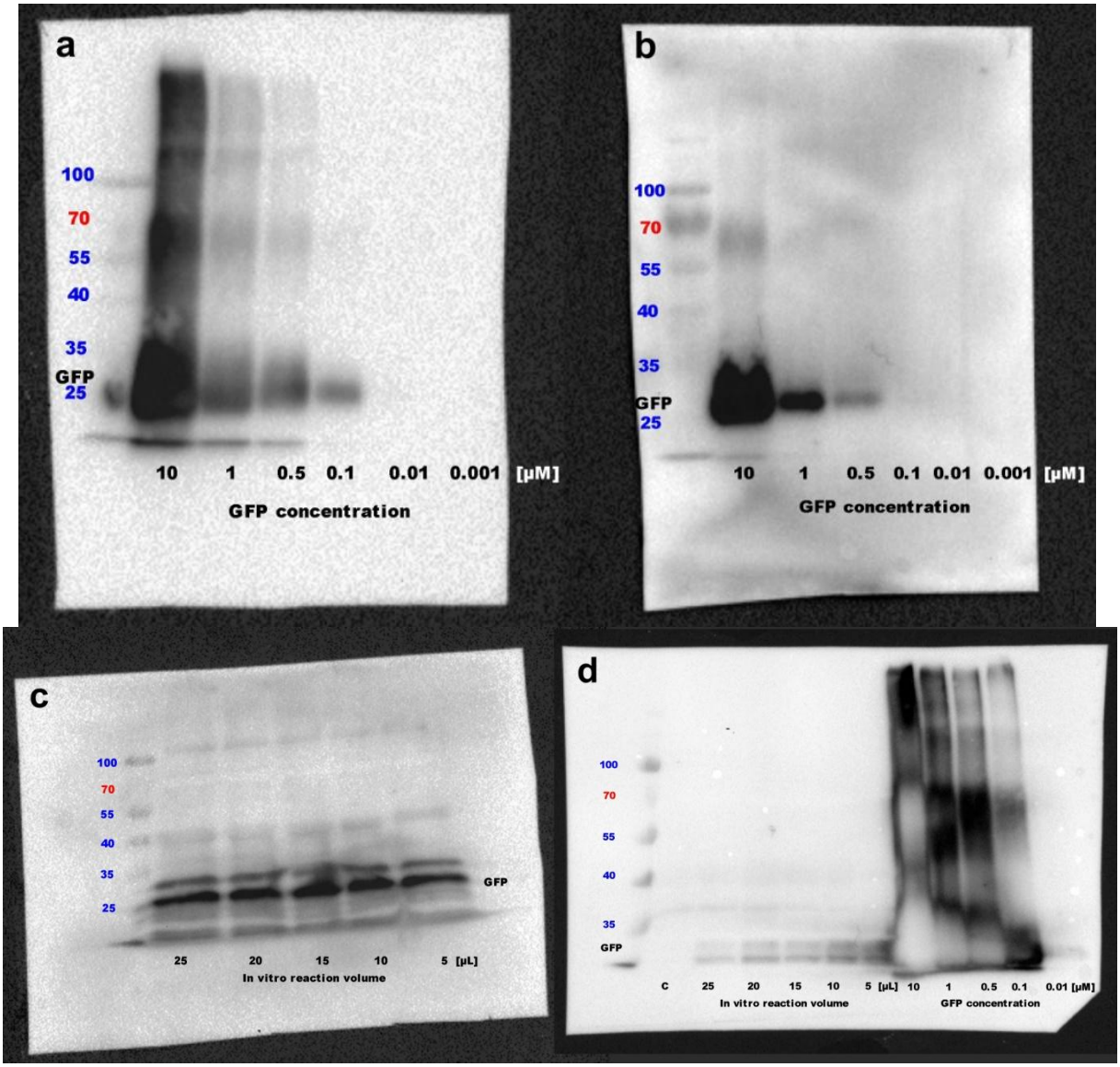


Figure7. ***Results from system optimization.*** *These tests were performed to assure that the antibody was effective and that small in vitro reaction volumes still produce reliable results.a) Different concentrations of GFP targeted by α-HIS antibody. b) Different concentrations of GFP targeted by α-GFP antibody. c) FLAG (pEH110) transcribed in vitro in different reaction volumes, targeted by α-FLAG antibody. d) GFP (pEH87) transcribed in vitro in different reaction volumes (left) and a GFP dilution series targeted by α-GFP antibody (right).*

3.4.2 Translation time course

When the system had been optimized the actual assay could begin. Fig. 8 shows the translation levels from 0.4 µM mRNA from all A-stretch mutants were analysed in 20 µL during a 120 minutes translation assay. Every 30 minutes a 5 µL sample was removed from each reaction.

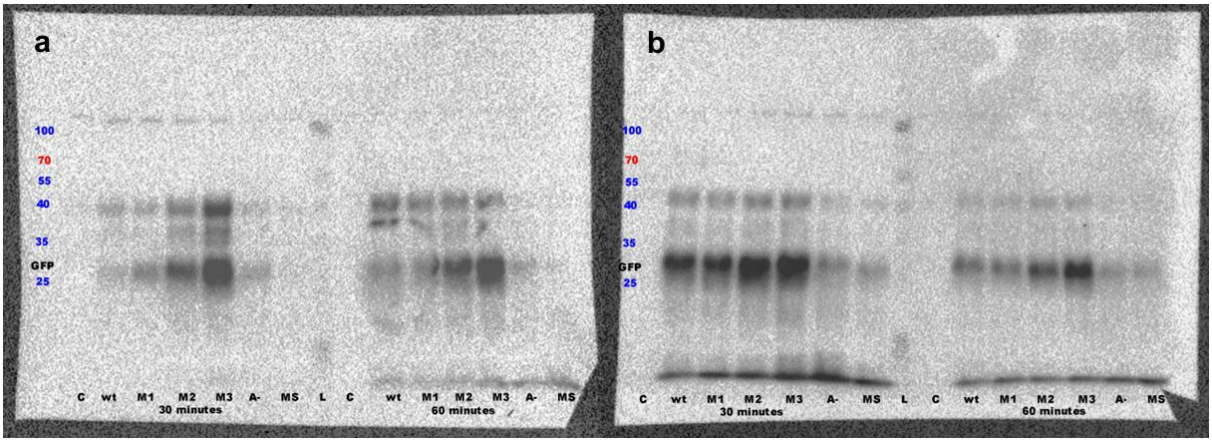


Figure 8. ***GFP translation time course.*** *Western blots showing the GFP expression during different time points in the in vitro translation assay. a) Includes time points 30 and 60 minutes. An unfortunate stain is faintly noticeable over the wt and M1samples for 60 minutes. b) Includes time points 90 and 120 minutes. All samples for 120 minutes seem affected by loading error.*

3.4.3 Observations made during *in vitro* translation

In the translation assay seen in fig. 9 a clear phenotype could be distinguished for the different mutants. They appear to roughly group into three categories, an inefficiently translated group (ΔA and MS), a medium group (wt and M1), and a highly expressed group (M2 and M3). Interesting to note is that not only does the expression levels differ greatly between the samples, but also the pattern of expression over time. Most of the mutants and the wild-type have a slow start, while two (M2 and M3) spike very early. With so few data points it is hard to tell if any of the samples reach a plateau during the first 90 minutes of this experiment. However, it seems like the fastest mutant M3 might already have reached its plateau early during the experiment.

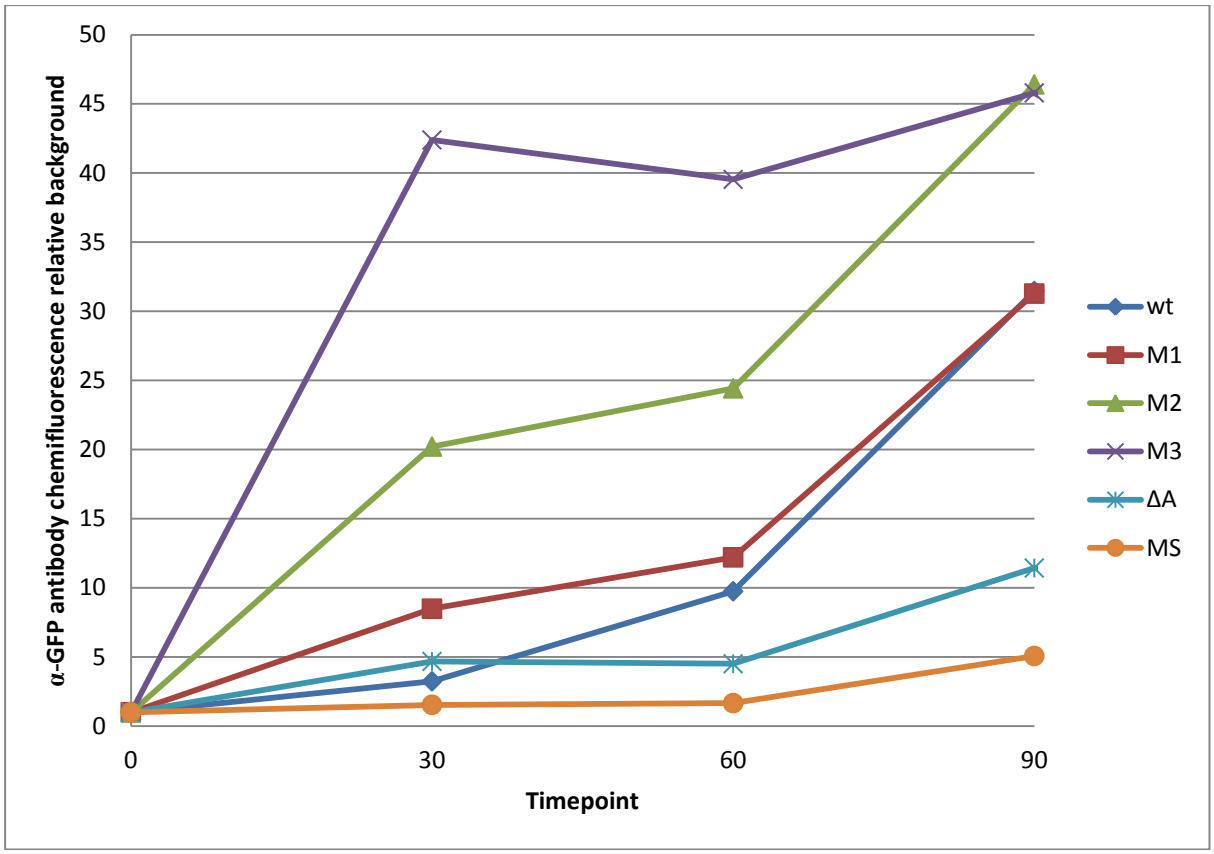


Figure 9. ***GFP translation time course plot.*** *The GFP levels from fig 8 plotted against time. GFP amount for time points 30, 60 and 90 minutes after start of incubation. 120 minutes was excluded from this graph. mRNA concentration was 0.4 µM.*

3.4.4 Effects of Hfq on translation of mutant mRNAs

The idea behind the experiments with the A-stretch mutants is that the A-stretch is a binding site for Hfq. Hence, the most interesting experiment to pursue was to see how the different mutant mRNAs would behave in the presence of Hfq. At this stage of the project both time and mRNA were starting to run out, and it was essential to save as much wild-type mRNA as possible. To optimize the Hfq amounts we decided to test different concentrations of Hfq on M1 and M2 mRNA. M1 was selected for its apparent similarity to the wild-type translational behaviour. M2 was selected for its high translation, and to have a control reaction. The reaction was stopped after 60 minutes for M2 and 75 minutes for M1. While we realize that this was not the perfect set up for a control experiment, it was essential to save the precious wild-type mRNA for the actual experiment rather than the optimization.

3.4.5 Optimisation in the presence of Hfq

This optimization experiment aimed at examining which concentration of Hfq would be optimal for a translation assay for all the A-stretch mutations. This experiment was performed in 10 µL 0.5 µM mRNA, only one end point measurement was made. The reaction was stopped after 75 minutes (M1) and 60 minutes (M2). Four different Hfq concentrations were tested, ranging from 0.17 to 3.6 µM. One reaction for each mutant was tested without Hfq as a control.

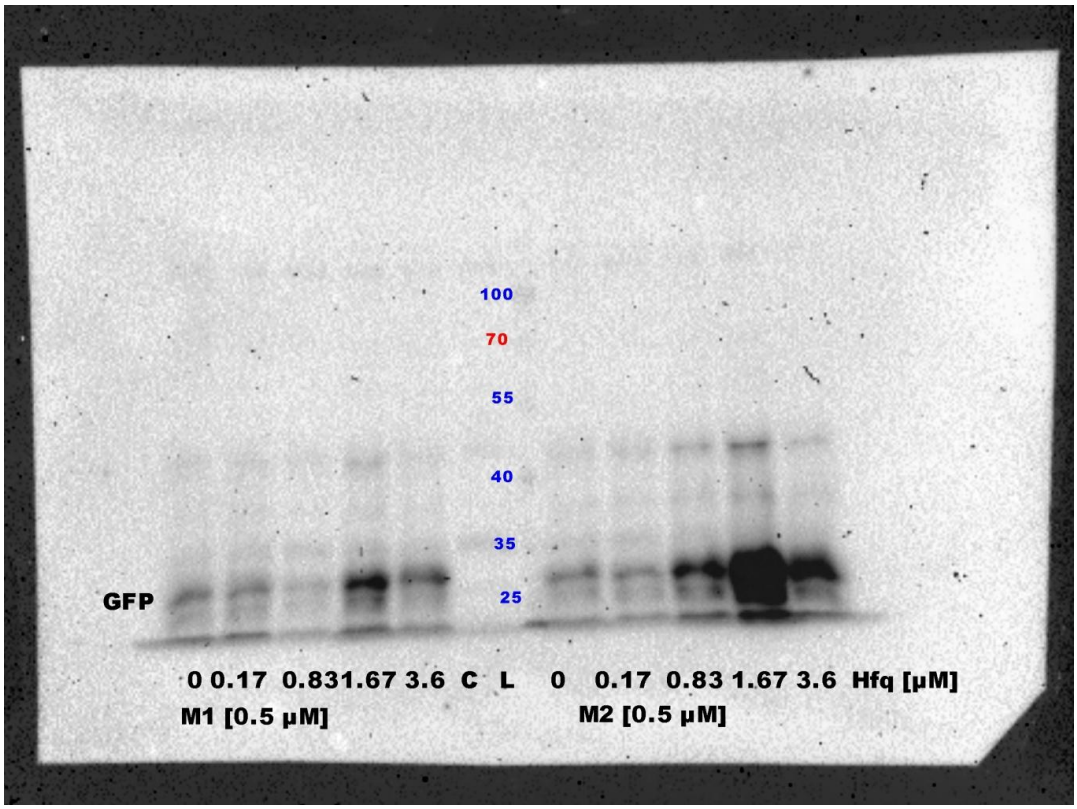


Figure 10. ***Translational effects by addition of Hfq.*** *Western blot showing the translational level for M1 and M2 under the influence of raising concentrations of Hfq. The translation is clearly peaking somewhere around 1.67 µM for both mutants.*

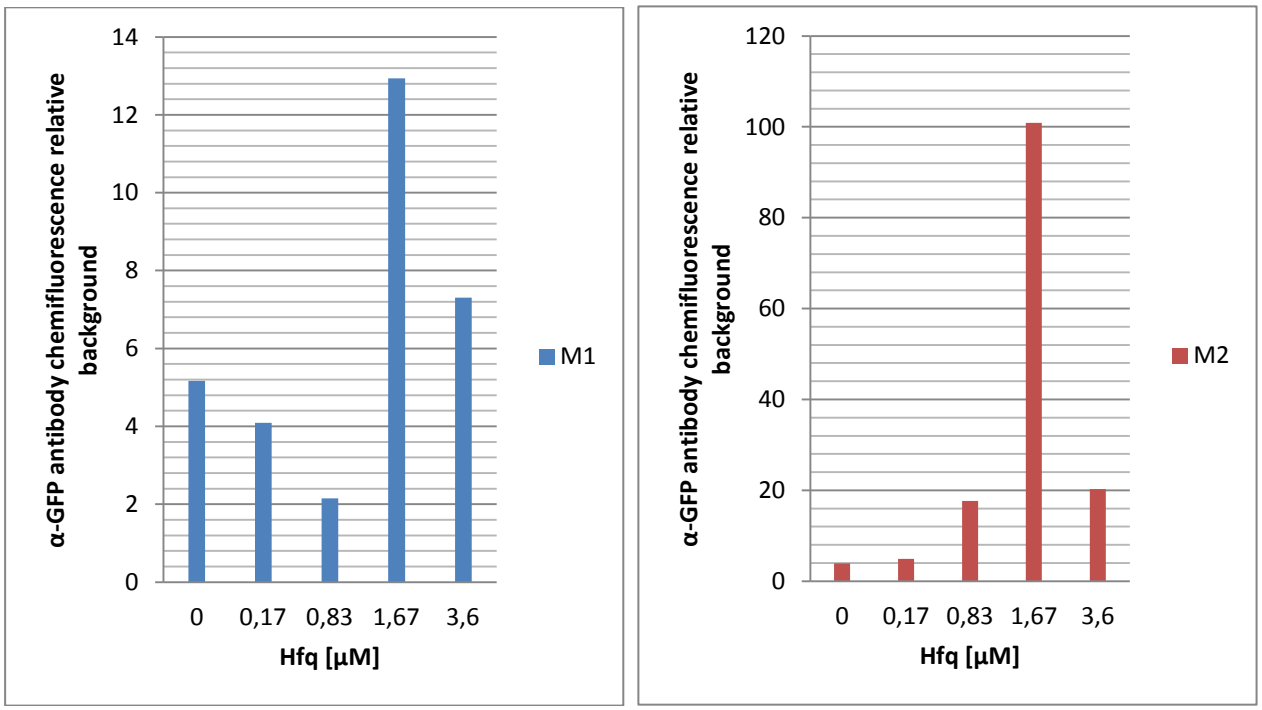


Figure11. ***Translational effects by addition of Hfq.*** *Graphs over the amount of translated GFP for different concentrations of Hfq.*

As shown in fig. 10 and 11 the translation is most efficient in an Hfq concentration of 1.67 µM. For M1 there is a clear effect, the expression increase close to threefold (2.5 times) compared to the expression in the absence of Hfq. However for M2 the expression bursts with the addition of Hfq, increasing 26 times. This dramatic effect is certainly interesting; the next step was to test all the different mutations and the wild-type, at the concentration of Hfq which gave the most interesting effect. In this case the most effective Hfq concentration is approximately three times as high as the concentration on mRNA in the experiment.

3.4.6 Translation time course in the presence of Hfq

In this last experiment in this project, 0.4 µM of all A-stretch mutants (except MS for which only 0.37 µM was possible) where assayed over a time course just as before, except that the latest time point 120 was omitted. The concentration of Hfq was chosen as 1.2 µM, three times higher than the mRNA concentration.

The signal from the Western blot in fig. 12 was very weak for unknown reason in this experiment. It is therefore hard to draw a conclusion from it, and even harder to compare to the corresponding experiment without Hfq.

Just as before, the translation seen in fig. 13 roughly clumps together in three translation level categories. However, in the presence of Hfq these groups have changed members. M3 still holds the position as the most highly translated mutant; with a quick start, a clear linear phase, and it also seem to plateau out rather early as before. Intriguingly, this time the wild-type is associated with M2 instead of M1 in the mid- translation group. M1 have instead moved down to MS and ΔA in the low- translation group. The difference between high and low translation has also increased.

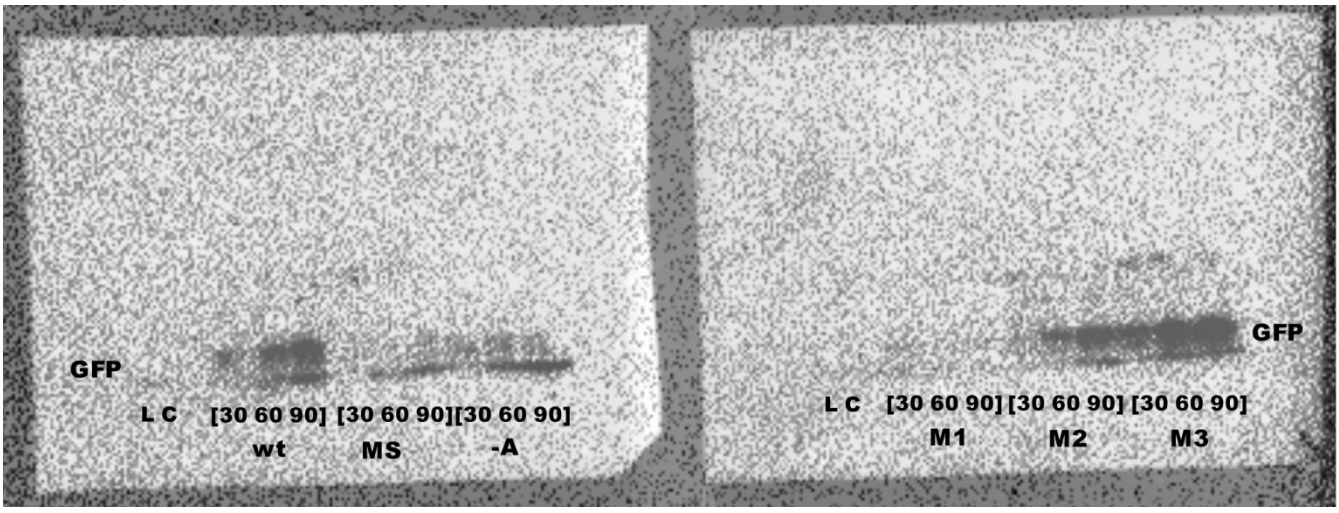


Figure 12. ***GFP translation time course in precence of Hfq.*** *In vitro translation assay on 0.4 µM mRNA in the presence of 1.2 µM Hfq. The translation was measured at three different timepoints for each of the six samples.*

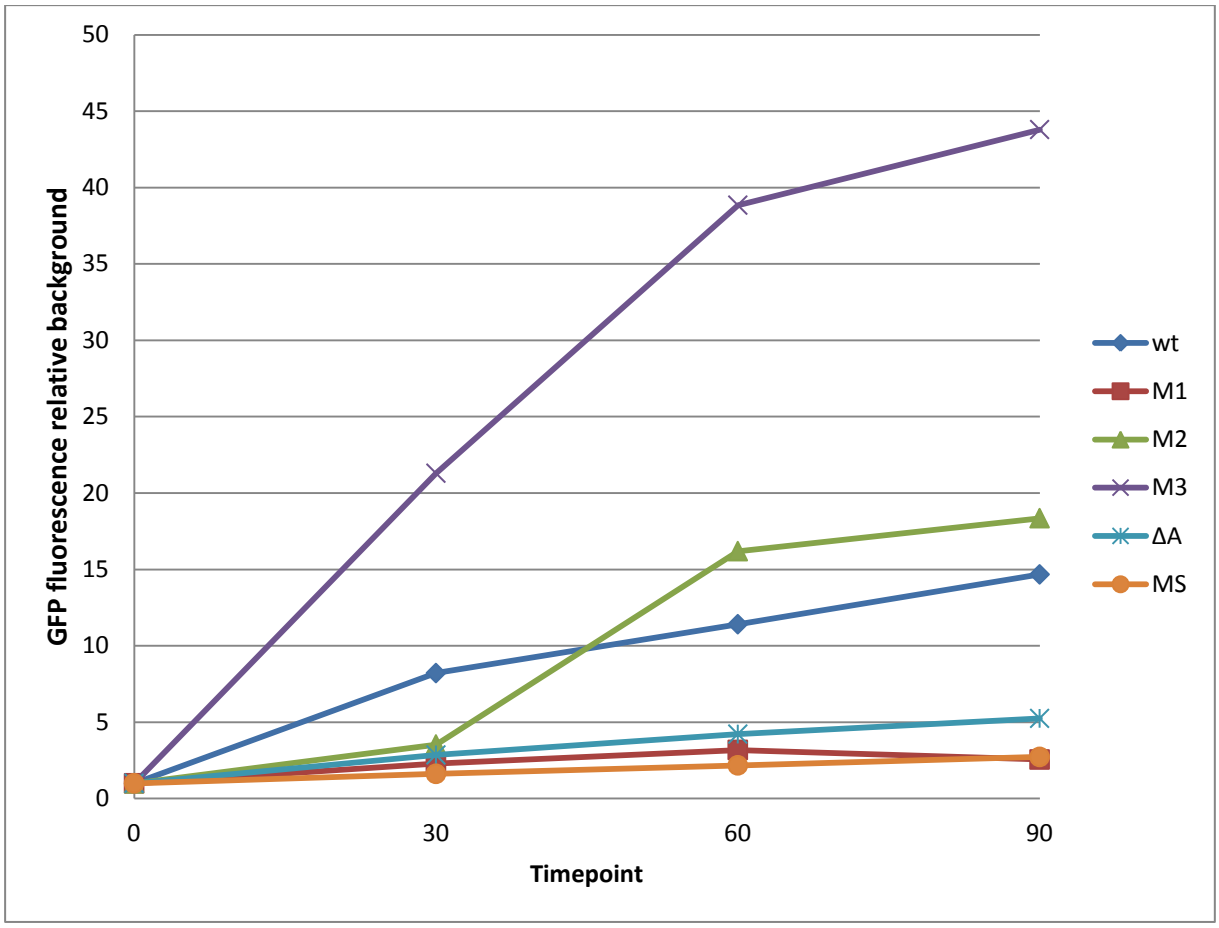


Figure 13. ***GFP translation time course in precence of Hfq plot.*** *The translation from fig. 12 plotted against time. GFP amount for time points 30, 60 and 90 minutes after start of incubation.*

4. Discussion

4.1 Current results

4.1.1 Observations concerning the initial study mutants

While these mutants had a minor role in the project, the results which are indicated after the initial measurements are among the most telling results from this project. By simple taking a quick look at the graphs in fig. 4a and 4b, and then compare the mutations differing the most from the wild-type to the locations of the mutations as seen in fig. 3 patterns start to emerge. As already discussed partially adjacent to the graphs in the previous section, mutations in the regulatory stem-loops have distinct phenotypes that differ from the wild-type.

Perhaps the most striking pattern is the loss of regulation for several mutants with mutations in or near the OmrA/B binding initiation site. Mutants 4, 5, 6 and 7 all have mutations in the actual binding initiation site in the top of the stem-loop, and all cause severe loss of regulation, interestingly causing higher translation in the presence of OmrA/B than without. Mutants 2 and 10 also show a significant loss of regulation, even though they are located further from the initiation site. Both these mutations are found on the same stem OmrA/B binds to; mutant 2 near the bulge on the same strand as the binding, and mutant 10 near the top of the stem, possibly affecting its stability.

Another significant effect, distinct from the loss of regulation is the huge increase of translation in absence of regulatory sRNAs seen for several mutants. These mutants have mutations near the SD or start codon. The most striking effect is seen for mutation 11; showing almost a six-fold increase of translation compared to the wild-type. Mutant 11 mutates the G at -10 in the SD region to an A; changing the sequence from GGGG to GAGG, which is a much stronger SD. Mutants 13 and 14, placed near the top of the stem in this stem-loop also increase the translational activity, possibly by weakening the secondary structure and making the start codon more accessible. Smaller effects are also seen for mutations 12 and 17. Mutant 12 is also positioned in the SD region, changing the sequence to GGAG. Mutation 17 is located in the loop, just after the start codon, mutating the UUU triplet to AAA. Interesting to notice is even if a mutation causes higher translation in the absence of OmrA/B, the translation under regulation does not differ significantly from the wild-type. The strength of the regulation can handle the extra pressure of increased translational efficiency. While all these effects certainly may be trivial, changes in secondary structure making the SD and start codon more accessible or blocking the binding of OmrA/B, there might also be hidden secrets.

4.1.2 The A-stretch mutants and their behaviour *in vivo*

In sharp contrast to the first set of mutants as discussed above, the A-stretch mutants does not show any obvious phenotype that can be related to the character of the mutations. *In vivo* all of the mutants which kept the original length of the A-stretch show generally increased translation for all tested conditions. The regulation of these mutants does not seem to be affected; when keeping the general increase of translation in mind, the decrease after addition of OmrA is approximately on scale to that of the wild-type. This effect is more noticeable when Hfq is present.

The shortened mutant, mutation MS, is expressed at the same level as the wild-type when Hfq is absent, and about twice as high when Hfq is present and OmrA is absent. The regulation of MS seems to be slightly more effective than for the wild-type.

The A-stretch deletion mutant ΔA shows very low expression in general. It is hard to interpret since such a large deletion also results in great structural changes. For example, deleting the A-stretch puts the two stem-loops of *csgD* mRNA in close proximity. But it is also possible that the deletion of the A-stretch affects for example the suspected ability to bind Hfq. Interestingly ΔA is the only mutant to be more translated in the Hfq-free background; even though these cells have poor health and generally give low protein yield. The same observation is made for the wild-type; it too is more highly translated in the strain where Hfq is absent.

In general it is risky to draw conclusions from comparisons between the results from the ΔHfq background with the cells that produce Hfq. Since the ability to produce Hfq is very important for many pathways in the cell, not just regulation of *csgD* translation, the ΔHfq *E. coli* are sickly and slow-growing. This makes the different translation levels hard and uncertain to compare.

4.1.3 Comparison of *in vitro* translation with and without Hfq

When the plotted curves from fig. 9 and 13 are put next to each other for comparison, the different translation patterns are quite striking. In the absence of Hfq (fig.9) translation of most mutants is slow at first and then spikes up after the first hour. These late spikes might be an artefact, it seem like it could be a systematic error since the same pattern shows for all samples. After the addition of Hfq (fig. 13) the mutants start to plateau out after the first hour.

The A-stretch mutants where designed with diminished Hfq binding in mind. The most important motif to avoid was the ARN-motif discussed previously, this motif is avoided in all mutants. Another motif, the YAA-motif (consisting on the sequence YAAYAA where Y represents a pyrimidine (cytosine or uracil (or thymine in DNA)), and A is adenine) is also known to bind Hfq. Only mutant M1 has a motif which could fit (AACAAC), but in that case it is backwards. None of the mutants should actually bind Hfq at the mutated A-stretch, yet several seem affected by its presence.

Since the signal was very weak in the latter experiment (fig. 12) I do not think it is possible to compare the translation levels between fig. 9 and 13, it is also uncertain to draw any conclusions from the general translation levels. Even though it seems like the translation is at approximately the same level for the two samples, that is not necessarily the case. Since the mutants seem to reach a plateau already after 60 minutes in presence of Hfq it is possible that the system has started to run out of some vital components to keep up translation.

What happens to M1 with the addition of Hfq is uncertain. Its translation rate drops down from the same level as the wild-type to almost nothing. Since this builds on just one single experiment it is possible that what we are looking at is a failed reaction, however it does not have to be. M2 is an interesting mutant in this context, its translational level drops from high to similar to the wild-type. This is a strong effect, and hard to explain since Hfq should not be able to bind. It has one CAA in its A-stretch sequence, but that alone should not be enough for binding to occur.

Interesting to note is also the effects on M1 and M2 registered in the Hfq titration experiment as seen in fig. 10, 11a and 11b. It is very obvious that the addition of a certain amount of Hfq has a significant effect on translational efficiency. This huge effect is not as noticeable in the following translation course experiment in presence of Hfq. M1 and M2 certainly seem to be the most affected mutants, showing the most apperant changes compared to the wild-type. This is where the problem with the incomparable scales between fig. 9 and 13 becomes an

issue. We cannot be certain if the relative translation levels of M1 and M2 drop when Hfq is present, or if it just looks like it does because the wild-type and M3 spurts. It is definitely necessary to re-perform these experiments to be able to draw proper conclusions.

Mutant M3 is an outlier in both cases with a clearly higher translation compared to the wild-type. Its strong translation seems inexplicable when compared to the *in vivo* assay where the translation is at the same level as M1 and M2.

The two shortened mutants MS and ΔA appear more or less unaffected by the addition of Hfq. They have low translation without Hfq, and still hardly any translation at all, possibly even lower, with Hfq.

After the *in vitro* translation assay and the Hfq titration experiment it became evident that Hfq does impact the translation levels for the A-stretch mutants. Considering that these mutants were designed solely with Hfq binding in mind, this is a remarkable result. The most obvious reason would of course be that Hfq is able to bind somewhere else on the *csgD* mRNA.

4.1.4 Comparison of *in vivo* and *in vitro* results for the A-stretch mutants

One intriguing result to notice is that the results *in vitro* and *in vivo* does not fully correspond for the A-stretch mutants. If the graphs from fig. 6a and 6b are compared to the plots in fig. 9 and 13 several distinct differences emerge. First, *in vivo* the translation of M1, M2 and M3 are very high compared to the wild-type very even between the three mutants, with M1 as slightly more translated than the other two. *In vitro* the translation is clearly dominated by M3, this is especially apparent in fig. 13. M2 and M1 follow with lower translation. In fig. 9 M2 behaves similarly to M3, and M1 is similar to the wild-type. But with the addition of Hfq in fig 13 the mutants change behaviour, with M2 following the wild-type and M1 was being translated a very low rate. Similar patterns emerge for MS. *In vivo* MS is translated in larger quantities than the wild-type. *In vitro*, with and without the addition of Hfq it has very low translation.

ΔA on the other hand behaves similarly *in vivo* and *in vitro* with a very low rate of translation in both cases. This is an interesting result and does indicate that the A-stretch plays an important role for the *csgD* mRNA, although this role remains elusive.

One of the most interesting outcomes is the differences in the translation level patterns with and without Hfq *in vivo* and *in vitro*. *In vivo* the relative translation between the different mutants is quite similar both with and without Hfq, and with and without OmrA, although the relative amounts change the fluorescence levels are still in the same order, giving a similar profile. *In vitro* the translation patterns between the mutants change dramatically. Both *in vivo* and *in vitro* the addition of Hfq seems to increase the differences between high and low translation, without Hfq the translation levels are more even.

4.2 Future prospects

4.2.1 Experiments at different growth temperatures

As discussed by Holmqvist *et al.* (2010) it has previously been reported that many *E. coli* strains only express curli at temperatures below 30˚C. However, in this project all cell cultures were grown in 37˚C. The fluorescence measurements were also performed in 37˚C. It might have been interesting to assay if the translation and regulation patterns would have differed if the cells were grown in, for example, 28˚C as well.

4.2.2 Repeat *in vitro* experiments

In the *in vivo* part of the A-stretch mutant project several biological replicates were made. Four different replicates run at two different occasions, which gave a unanimous result and the possibility to make reliable error estimation. While four replicates might not be enough to give a statistically ensured answer, it is much more reliable than no replicates. The *in vitro* part was afflicted both by lack of time and materials, so no experiment was ever performed twice. This makes the *in vitro* results harder to interpret and although they seem interesting they may be unreliable.

4.2.3 Addition of OmrA/B to the *in vitro* translation assay

The natural experiment which would have been interesting to perform is so obvious that it almost should have been a mandatory part of the project if the amount of time had allowed for it. The focus of the project is the effects of OmrA/B on *csgD*, yet I did not have time to do any *in vitro* studies which included OmrA/B. This is a major flaw in this part of the project; it is of course hard to study regulation by OmrA/B in an assay where OmrA/B is not present. *In vitro* translation assays would be performed as before, with and without the addition of Hfq, and with and without OmrA/B. With proper planning and replicates it would give results comparable to those I received *in vivo*.

In the *in vivo* translation it seemed like regulation by OmrA was slightly more efficient than regulation by OmrB for the wild-type. According to Gerhart Wagner it has been observed that OmrA is generally more effective in vivo while OmrB is more effective in vitro. This would be interesting to study, and to find out the causes behind it. It might not hold true for all the mutants; several of the mutants from the initial study were more strongly regulated by OmrB. This might be a possible tool for investigating the underlying causes.

4.2.4 Further *in vitro* experiments with and without Hfq

While it is very interesting to compare the results between the *in vitro* translation assay with and without Hfq, it is not without risk. The two experiments in fig. 9 and 13 are done at different times, with the differences in performance which follows. For example the signal strength was very dissimilar between the experiments. The two experiments should be repeated in parallel so that all such differences are eliminated.

4.2.5 Hfq binding to *csgD* mRNA should be studied by footprinting and gel shift

With the interesting preliminary results when Hfq was added to the *in vitro* translation assay we would like to know more about the interaction between Hfq and *csgD*. It would of course also be necessary to assay whether Hfq actually do bind to the *csgD* 5'-UTR, and if so, where it binds. This could be done in two ways, preferably both. One method which would be suitable is a gel shift assay. The RNA with the addition of different concentrations is run on a non-denaturing gel. If Hfq has bound to a segment it will be heavier and shift position in the gel. Another method which would be suitable is a protein-RNA footprinting assay. In principle, this method is studying how the pattern of randomized cutting of the RNA changes with the addition of Hfq. Since the RNA will be protected by bound Hfq it cannot be cut where Hfq covers the sequence, and so the binding site can be extracted.

4.2.6 More mutations

The original plan in this project was to assay a large array of mutants, *in vivo* only. The mutants used in the first part of the project, distributed over the stem-loops of *csgD* 5'-UTR indicated a very promising prospect. Many of these mutants had very clear effects *in vivo* which cannot be denied. While many of these can be explained as trivial effects on the initiation of OmrA/B binding, or secondary structure changes near the start site, some are

harder to interpret. One should not discard results because they seem superficially simple. There are of course many other interesting mutations to be made and assayed apart from the ones from the initial part of this project, and even more could be identified by randomized mutations coupled with FACS and deep sequencing discussed earlier. All of these would be interesting subjects for translation assays *in vitro* as well as *in vivo*.

4.3 Conclusions and concluding remarks

The aim of this project was to propose a mechanism by which the expression of CsgD is regulated by OmrA/B. It has also come to centre on the question how Hfq is involved in this regulation.

The initial study with mutations on the two stem-loops in *csgD* 5'-UTR appears to confirm what has previously been shown in different studies; that mutations directly interfering with the binding initiation of OmrA/B have the most dramatic effect on regulation. And mutations in the stem-loop where the SD and start codon are located give rise to higher translation, likely by making the secondary structure looser and more accessible or in one case by improving the SD sequence.

The effect of mutations in the A-stretch are harder to interpret, not showing any real signs of loss of regulation, but instead affects translation levels quite dramatically. Deleting the entire A-stretch results in very low translation *in vivo* and *in vitro*. This of course indicates that the A-stretch is important, although the answer to why remains to be found. The *in vitro* translation assays with and without Hfq further complicates the image of regulation of *csgD* mRNA. An example is the contradictory observation that there a significant effect on translation in presence of Hfq for mutants which should not be able to bind Hfq. The observed effects are real, although inexplicable.

At present the data I have is not enough to propose a novel method of regulation for the *csgD*-OmrA/B interaction. However, it seems certain that Hfq plays an important role in this as in the *in vivo* experiments all mutants lost regulation when Hfq was absent. Many times the results seem almost contradictory, such as how the mutant with the strongest *in vivo* regulation by OmrA, MS, does not seem to be particularly affected by Hfq *in vitro*. The results indicate that the A-stretch is very important for *csgD*; and although the cause remains unknown to date, this certainly is the most intriguing part of the project.

Acknowledgements

I would like to thank my supervisor Gerhart Wagner for giving me the opportunity to work on my degree project in his group. It has been very educational and I am grateful for all the guidance and help. And of course Magnus Lundgren, who accepted to be my scientific reviewer, even though he was on parental leave.

I would also like to thank Mirthe Hoekzema and Cédric Romilly who has both supervised me in the lab during different parts of the project. Always available to help me out in times of need, and teaching me everything I needed to know. And also the rest of the GW group and all my friends in the lab, we had fun together and it is a great team.

Lars-Göran Josefsson also deserves a special mention and thanks for helping me with all the practical concerns surrounding the project.

And at last, my family (that includes you John), for some various proofreading and being there for me.

References

1. Barnhart MM and Chapman MR. 2006. **Curli Biogenesis and Function.** *Annual Review of Microbiology 60:* 131–147
2. Boehm A and Vogel J. 2012. **The *csgD* mRNA as a hub for signal integration via multiple small RNAs.** *Molecular Microbiology 84:* 1-5
3. Guillier M and Gottesman S. 2008**. The 5' end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator.** *Nucleic Acids Research 36:* 6781-6794
4. Holmqvist E, Reimegård J, Sterk M, Grantcharova N, Römling U, Wagner E G H. 2010. **Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis.** *The EMBO Journal 29:* 1840-1850
5. Holmqvist E, Reimegård J, Wagner E G H. 2013. **Massive functional mapping of a 5'-UTR by saturation mutagenesis, phenotypic sorting and deep sequencing.** *Nucleic Acids Research* doi: 10.1093/nar/gkt267
6. Jørgensen MG, Nielsen JS, Boysen A, Franch T, Møller-Jensen J, Valentin-Hansen P. 2012. **Small regulatory RNAs control the multi-cellular adhesive lifestyle of Escherichia coli.** *Molecular Microbiology 84:* 36-50
7. Mika F, Busse S, Possling A, Berkholz J, Tschowri N, Sommerfeldt N, Pruteanu M, Hengge R. 2012. 1. **Targeting of csgD by the small regulatory RNA RprA links stationary phase, biofilm formation and cell envelope stress in Escherichia coli.** *Molecular Microbiology 84:* 51-65
8. Mika F, Hengge R. 2013. **Small Regulatory RNAs in the Control of Motility and Biofilm Formation in *E. coli* and *Salmonella*.** *International Journal of Molecular Sciences 14:* 4560-4579
9. Sharma CM, Papenfort K, Pernitzsch SR, Mollenkopf HJ, Hinton JCD, Vogel J. 2011**. Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA.** *Molecular Microbiology 81:* 1144-1165
10. Storz G, Vogel J, Wassarman K M. 2011. **Regulation by Small RNAs in Bacteria: Expanding Frontiers.** *Molecular Cell 43:* 880-890
11. Thomason MK, Fontaine F, De N, Storz G. 2012**. A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*.** Molecular *Microbiology 84:* 17–35
12. Urban J and Vogel J. 2007. **Translational control and target recognition by *Escherichia coli* small RNAs *in vivo*.** *Nucleic Acids Research 35:* 1018-1037
13. Waters LS and Storz G. 2009. **Regulatory RNAs in Bacteria.** *Cell 134:* 615-628

Appendix

Appendix A

Overview of mutations in pEH87 plasmid

Mutations designed by Erik Holmqvist, used in the first part of the project.

Mutation name	Primers	Mutation	Plasmid name
1	631, 632	-75 T:C	pEH211
2	632, 633	-74 G:A	pEH212
3	636, 637	-66 T:A	pEH213
4	638, 639	-64 C:A	pEH214
5	640, 641	-63 T:A	pEH215
6	642, 643	-62 G:A	pEH216
7	648, 649	-61 G:T	pEH217
8	EHO-829, EHO-830	-62 G:C	
9	654, 655	-47 G:A	pEH218
10	EHO-831, EHO-832	-40 C:A	pLJ001
11	662, 663	-10 G:A	pEH219
12	664, 665	-9 G:A	pEH220
13	EHO-833, EHO-834	-6 T:C	pLJ002
14	EHO-835, EHO-836	-3 A:G	pLJ003
15	EHO-837, EHO-838	-29 G:C	pLJ004
16	817, 818	-1 C:G	pEH221
17	EHO-839, EHO-840	+4,5,6 UUU=>AAA	pLJ005

Mutations designed by Erik Holmqvist and Gerhart Wagner, used in the main part of the project.

Mutation name	Primers	Mutation	Plasmid name
ΔA		Δ(AAUAAAAAAA)	pEH105
Ms	MHO-060, MHO-061	-22,21 AU:C, -19,18 AA:G, -16,15 AA:C	pLJ006
M1	MHO-062, MHO-063	-21 U:C, - 18 A:C, -16 A:C	pLJ007
M2	MHO-064, MHO-065	-20 A:C, -19 A:C, -15 A:C	pLJ008
M3	MHO-066, MHO-067	-21 U:C, -19 A:G, -17 A:C, -15 A:C	pLJ009

Oligonucleotides

Name on tube	Description	Tm [°C]	Sequence [5'-3']
EHO-239	Carrying T7 promoter		gaaattaatacgactcactataggcagatgtaatccattagttttatatttta c
EHO-346	GFP (pEH87) Reverse		TCGCTATTATGCTTACTATTTATCGTCGTCA TCTTTG
EHO-715	FLAG (pEH110) Reverse		GATGCCTCTAGATTTAAATGCTCGAAT
EHO-829	csgD M8 Forward	70	CGTGCTTCTATTTTAGAGGCAGCTGTCAGG

EHO-830	csgD M8 Reverse	71	AGAAGTACTGACAGATGTTGCACTGCTGTG
EHO-831	csgD M10 Forward	63	ATGTCAGGTGTGCGATCAAT
EHO-832	csgD M10 Reverse	63	CTGCCTCTAAAATAGAAGCACCA
EHO-833	csgD M13 Forward	61	CTCATCATGTTTAATGAAGTCCATAGT
EHO-834	csgD M13 Reverse	62	ACCCCGCTTTTTTTATTGATC
EHO-835	csgD M14 Forward	62	GTCATGTTTAATGAAGTCCATAGTATTCA
EHO-836	csgD M14 Reverse	63	GAAACCCCGCTTTTTTTATTG
EHO-837	csgD M15 Forward	63	CGATCAATAAAAAAAGCGGG
EHO-838	csgD M15 Reverse	63	GACACCTGACAGCTGCCTCT
EHO-839	csgD M17 Forward	61	AAAAATGAAGTCCATAGTATTCATGGTC
EHO-840	csgD M17 Reverse	63	CATGATGAAACCCCGCTT
MHO-060	csgD A-stretch MS Forward	64,8	GCGGGGTTTCATCATGTTTA
MHO-061	csgD A-stretch MS Reverse	66,8	TGTCTGTGATCGCACACCTG
MHO-062	csgD A-stretch M1 Forward	67,5	CACAAGCGGGGTTTCATCA
MHO-063	csgD A-stretch M1 Reverse	66,8	TTGTTGATCGCACACCTGAC
MHO-064	csgD A-stretch M2 Forward	65,8	AAACAGCGGGGTTTCATCA
MHO-065	csgD A-stretch M2 Reverse	66	GGATTGATCGCACACCTGAC
MHO-066	csgD A-stretch M3 Forward	67	ACACAGCGGGGTTTCATCA
MHO-067	csgD A-stretch M3 Reverse	65	CTGTTGATCGCACACCTGAC

Plasmids

Plasmid name	**Description**
pEH87	pCsgD-GFP (csgD leader (EHO-355/EHO-356) cloned (NsiI/NheI) in frame with GFP in pXG-10)
pEH105	pEH87_Δ125-135 (Δ-14 to -23)
pEH110	pcsgD-3xFLAG
pEH211	-75 T:C
pEH212	-74 G:A
pEH213	-66 T:A
pEH214	-64 C:A
pEH215	-63 T:A
pEH216	-62 G:A
pEH217	-61 G:T
pLJ001	-40 C:A
pLJ002	-6 T:C
pLJ003	-3 A:G
pLJ004	-29 G:C
pLJ005	+4,5,6 UUU:AAA
pLJ006	-22,21 AU:C, -19,18 AA:G, -16,15 AA:C

Appendix B

The sequences as used in the *in vitro* part of this project. The T7 promoter is shown in lowercase letters, the A-stretch where mutants were introduced is blue, the SD sequence is green and the start site is red.

pEH87: wildtype

gaaattaatacgactcactataggCAGATGTAATCCATTAGTTTTATATTTTACCCATTTAGGGC
TGATTTATTACTACACACAGCAGTGCAACATCTGTCAGTACTTCTGGTGCTTCTAT
TTTAGAGGCAGCTGTCAGGTGTGCGATCAATAAAAAAAGCGGGGTTTCATCATGT
TTAATGAAGTCCATAGTATTCATGGTCATACATTATTGTTGGCTAGCAAAGGAGA
AGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATG
GGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTACATACGGAAAGC
TTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTT
GTCACTACTTTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCATATGAA
ACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACT
ATATCTTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAG
GTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGG
AAACATTCTCGGACACAAACTCGAGTACAACTATAACTCACACAATGTATACATC
ACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAAC
ATTGAAGATGGATCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTG
GCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGCTTTC
GAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCT
GCTGGGATTACACATGGCATGGATGAGCTCTACAAATAATGAATTCGAGCATTTA
AATCTAGAGGCATC

pLJ006 = pEH87_MS: Short Mutation

gaaattaatacgactcactataggCAGATGTAATCCATTAGTTTTATATTTTACCCATTTAGGGC
TGATTTATTACTACACACAGCAGTGCAACATCTGTCAGTACTTCTGGTGCTTCTAT
TTTAGAGGCAGCTGTCAGGTGTGCGATCACAGACAGCGGGGTTTCATCATGTTTA
ATGAAGTCCATAGTATTCATGGTCATACATTATTGTTGGCTAGCAAAGGAGAAGA
ACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGC
ACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTACATACGGAAAGCTTAC
CCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCA
CTACTTTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCATATGAAACGG
CATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATAT
CTTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTG
ATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAA
CATTCTCGGACACAAACTCGAGTACAACTATAACTCACACAATGTATACATCACG
GCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACATT
GAAGATGGATCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCG
ATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGCTTTCGAA
AGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCT
GGGATTACACATGGCATGGATGAGCTCTACAAATAATGAATTCGAGCATTTAAAT
CTAGAGGCATC

pLJ007 = pEH87_M1: Mutation 1

gaaattaatacgactcactataggCAGATGTAATCCATTAGTTTTATATTTTACCCATTTAGGGC
TGATTTATTACTACACACAGCAGTGCAACATCTGTCAGTACTTCTGGTGCTTCTAT
TTTAGAGGCAGCTGTCAGGTGTGCGATCAACAACACAAGCGGGGTTTCATCATGT
TTAATGAAGTCCATAGTATTCATGGTCATACATTATTGTTGGCTAGCAAAGGAGA

AGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATG
GGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTACATACGGAAAGC
TTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTT
GTCACTACTTTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCATATGAA
ACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACT
ATATCTTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAG
GTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGG
AAACATTCTCGGACACAAACTCGAGTACAACTATAACTCACACAATGTATACATC
ACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAAC
ATTGAAGATGGATCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTG
GCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGCTTTC
GAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCT
GCTGGGATTACACATGGCATGGATGAGCTCTACAAATAATGAATTCGAGCATTTA
AATCTAGAGGCATC

pLJ008 = pEH87_M2: Mutation 2
gaaattaatacgactcactataggCAGATGTAATCCATTAGTTTTATATTTTACCCATTTAGGGC
TGATTTATTACTACACACAGCAGTGCAACATCTGTCAGTACTTCTGGTGCTTCTAT
TTTAGAGGCAGCTGTCAGGTGTGCGATCAATCCAAACAGCGGGGTTTCATCATGT
TTAATGAAGTCCATAGTATTCATGGTCATACATTATTGTTGGCTAGCAAAGGAGA
AGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATG
GGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTACATACGGAAAGC
TTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTT
GTCACTACTTTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCATATGAA
ACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACT
ATATCTTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAG
GTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGG
AAACATTCTCGGACACAAACTCGAGTACAACTATAACTCACACAATGTATACATC
ACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAAC
ATTGAAGATGGATCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTG
GCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGCTTTC
GAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCT
GCTGGGATTACACATGGCATGGATGAGCTCTACAAATAATGAATTCGAGCATTTA
AATCTAGAGGCATC

pLJ009 = pEH87_M3: Mutation 3
gaaattaatacgactcactataggCAGATGTAATCCATTAGTTTTATATTTTACCCATTTAGGGC
TGATTTATTACTACACACAGCAGTGCAACATCTGTCAGTACTTCTGGTGCTTCTAT
TTTAGAGGCAGCTGTCAGGTGTGCGATCAACAGACACAGCGGGGTTTCATCATGT
TTAATGAAGTCCATAGTATTCATGGTCATACATTATTGTTGGCTAGCAAAGGAGA
AGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATG
GGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTACATACGGAAAGC
TTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTT
GTCACTACTTTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCATATGAA
ACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACT
ATATCTTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAG
GTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGG
AAACATTCTCGGACACAAACTCGAGTACAACTATAACTCACACAATGTATACATC
ACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAAC
ATTGAAGATGGATCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTG

GCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGCTTTC
GAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCT
GCTGGGATTACACATGGCATGGATGAGCTCTACAAATAATGAATTCGAGCATTTA
AATCTAGAGGCATC

pEH105: ΔA; A-stretch deletion
gaaattaatacgactcactataggCAGATGTAATCCATTAGTTTTATATTTTACCCATTTAGGGC
TGATTTATTACTACACACAGCAGTGCAACATCTGTCAGTACTTCTGGTGCTTCTAT
TTTAGAGGCAGCTGTCAGGTGTGCGATCGCGGGGTTTCATCATGTTTAATGAAGT
CCATAGTATTCATGGTCATACATTATTGTTGGCTAGCAAAGGAGAAGAACTTTTC
ACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAAT
TTTCTGTCAGTGGAGAGGGTGAAGGTGATGCTACATACGGAAAGCTTACCCTTAA
ATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTT
TGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCATATGAAACGGCATGAC
TTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCA
AAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCC
TTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCT
CGGACACAAACTCGAGTACAACTATAACTCACACAATGTATACATCACGGCAGA
CAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACATTGAAGA
TGGATCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGC
CCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGCTTTCGAAAGATC
CCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGAT
TACACATGGCATGGATGAGCTCTACAAATAATGAATTCGAGCATTTAAATCTAGA
GGCATC