New insights into the role of ppGpp and DksA through their effect on transcriptional regulation of housekeeping and colonization related genes of *Escherichia coli*

Anna Åberg

Department of Molecular Biology
Umeå University
Umeå 2008
Department of Molecular Biology
Umeå University
SE-901 87 Umeå
Sweden

Copyright © 2008 by Anna Åberg

ISSN 0346-6612


Printed by Print & Media, Umeå University, Umeå

Cover pictures: Electronmicrograph pictures of Escherchia coli expressing type 1 fimbriae (left) and flagella (right) structures (courtesy of Juan David Cabrer and Jorge Fernández). Pictures were taken at Servei de microscòpia electrònica de transmissió, Serveis centificotècnics, Universitat de Barcelona.
Till pappa

*Nothing in life is to be feared. It is only to be understood*

Marie Curie (1867-1934) Polish Scientist

*Success is going from failure to failure without a loss of enthusiasm*

Winston Churchill (1874-1965) English Statesman
Abbreviations used:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>iNTP</td>
<td>Initiator NTP</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>ncRNA</td>
<td>Non-coding RNA</td>
</tr>
<tr>
<td>NTP</td>
<td>Nucleoside triphosphates</td>
</tr>
<tr>
<td>ORF</td>
<td>Open-reading frame, coding sequence of a gene</td>
</tr>
<tr>
<td>ppGpp</td>
<td>Guanosine tetra- and pentaphosphate</td>
</tr>
<tr>
<td>RelA</td>
<td>ppGpp synthetase I</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNAP</td>
<td>RNA polymerase (α2ββ′ω-subunits)</td>
</tr>
<tr>
<td>rrrn</td>
<td>Genes encoding the rRNA</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>Rsh</td>
<td>RelA and SpoT homologues</td>
</tr>
<tr>
<td>SpoT</td>
<td>ppGpp synthetase II and ppGpp hydrolase</td>
</tr>
<tr>
<td>stable RNA</td>
<td>rRNA and tRNA</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>UPEC</td>
<td>Uropathogenic E. coli</td>
</tr>
</tbody>
</table>

*Science is a wonderful thing if one does not have to earn one's living at it*

*Albert Einstein (1879-1955) German Physicist*
# Table of contents

ABSTRACT .................................................................................................................. 7

PAPERS IN THIS THESIS ............................................................................................ 8

INTRODUCTION .......................................................................................................... 9

1. Sensing the environment .................................................................................................. 9

2. Transcription .................................................................................................................. 10
   2.1 The players in transcription ......................................................................................... 12
   2.2 Transcription step-by-step ............................................................................................ 13
      2.2.1 Promoter binding and initiation of transcription ......................................................... 14
      2.2.2 Elongation and termination ....................................................................................... 15
   2.3 Regulation of transcription ........................................................................................... 16
      2.3.1 Transcription factors ............................................................................................... 17
      2.3.2 Non-proteinaceous regulatory molecules ................................................................. 18

3. ppGpp - a stress alarmone .............................................................................................. 19
   3.1 Stringent response ........................................................................................................ 20
   3.2 DksA - a possible co-regulator ...................................................................................... 22
      3.2.1 Gre-factors .............................................................................................................. 23
      3.2.2 Regulation through the secondary channel ............................................................... 23
   3.3 Mechanism of regulation by ppGpp and/or DksA .......................................................... 25
      3.3.1 Direct negative regulation of transcription .............................................................. 25
      3.3.2 Direct and indirect positive regulation ....................................................................... 26

4. *Escherichia coli* - a model system ................................................................................ 27
   4.1 Strains of *E. coli* .......................................................................................................... 28
      4.1.1 Uropathogenic *E. coli* ............................................................................................ 28
   4.2 Type 1 fimbriae .............................................................................................................. 29
      4.2.1 General characteristics ............................................................................................. 30
      4.2.2 Genetic organization and regulation ......................................................................... 30
      4.2.3 Biofilm formation ..................................................................................................... 33

AIMS OF THIS THESIS ................................................................................................. 35

RESULT AND DISCUSSION ......................................................................................... 36

5. Global effect of ppGpp and/or DksA on gene expression in *E. coli* .................................. 36
   5.1 Effect of ppGpp and DksA on bacterial cell physiology ................................................. 37
   5.2 Effect of ppGpp and DksA on transcriptional stimulation and repression ..................... 37
6. Involvement of ppGpp in biofilm formation by commensal and pathogenic E. coli

6.1 ppGpp regulate expression of type 1 fimbriae at the phase variation level

6.2 ppGpp regulation of factors important for biofilm formation

7. Divergent effects of ppGpp and DksA on gene expression in vivo

7.1 Effect of DksA-deficiency on type 1 fimbriae and flagella expression

8. Dissecting the mechanism of regulation of fimB transcription by ppGpp and DksA

8.1 Direct effects of ppGpp and DksA on transcription from the fimB P2 promoter

8.2 Co-dependent versus independent effects of ppGpp and DksA on fimB transcription

9. Involvement of Gre-factors in differential regulation by ppGpp and DksA in vivo

9.1 Gre-factors stimulate fimB transcription in vivo and in vitro

9.2 Gre-factors are important for FliC expression in vivo

9.3 Competition for the secondary channel of RNAP

CONCLUDING REMARKS

ACKNOWLEDGMENTS

REFERENCES
Abstract

Bacteria have the ability to sense different environmental signals. When an environmental stress is detected, bacteria rapidly adjust their gene expression profile to be able to survive and thrive. The transduction of such environmental signals often requires the coordinated involvement of several factors that constitute complex regulatory networks. Hence, depending on the combination of signals, a unique gene expression profile required to adapt to the specific stress conditions is generated. Proteins are the best-known regulatory factors. However, non-proteinaceous molecules are also important in signal-responsive control of bacterial gene expression. Alarmones are low molecular weight non-proteinaceous regulatory factors which can characteristically be rapidly turned-over to mediate instant changes in gene expression. One such alarmone is the modified nucleotide ppGpp, which directly binds to RNA polymerase to alter its activity. The levels of this alarmone are expected to rapidly increase in response to any environmental stress that result in slow proliferation. DksA, a putative ppGpp co-regulator that likewise directly targets RNA polymerase, has been suggested to be required for both the positive and negative regulation mediated by ppGpp in *Escherichia coli*.

This thesis describes dissection of the role of ppGpp and DksA on transcriptional regulation, primarily using the *fim* genetic determinant that encodes for the type 1 fimbriae. Type 1 fimbriae are involved in adhesion to abiotic surface and initial adhesion/invasion of bladder cells, as well as in biofilm formation. We found that ppGpp regulates phase variation by increasing the sub-population of cells that express the fimbriae. The effect of ppGpp was ultimately traced to its role in transcription of the *fimB* gene that encodes a recombinase involved in the phase variation process (paper 1). In contrast, we unexpectedly found that lack of DksA causes an increase, rather than a decrease, in transcription from the *fimB* P2 promoter *in vivo*. However, *in vitro* transcription studies demonstrated that ppGpp and DksA, both independently and co-dependently, stimulate transcription from the *fimB* P2 promoter. These seemingly contradictory results from the *in vivo* and *in vitro* transcriptional studies were shown to be, at least in part, a consequence of the increased association of Gre-factors with RNA polymerase that can occur in the absence of DksA *in vivo* (paper 2).

The results outlined above have implications for the role of ppGpp and/or DksA in global gene expression. Using gene expression profile (*microarray analysis*) during the transition from logarithmic to stationary phase of *E. coli*, we found that while most of the genes regulated by ppGpp and DksA are regulated in the same direction by the two factors, many were not. In addition to the *fim* genes, genes involved in flagella functioning, taxis responses, and a few genes encoding different transport systems are also differentially regulated in ppGpp- and DksA-deficient strains *in vivo*. Our results clearly indicate that the effect of deficiencies in ppGpp and DksA is far more complex than phenotypic similarity of the corresponding mutants anticipated by the proposed concerted action of ppGpp and DksA on gene expression (paper 2 & 3).
Papers in this thesis

The results and discussion of this thesis is based on the following papers and manuscript labelled with their roman letters (I-III).

**Paper I**


**Paper II**


**Paper III**

Introduction

1. Sensing the environment

Organisms from all domains of life have the potential to sense their surroundings which provides them with the fundamental information required for development, adaptation and/or survival. Bacteria are arguably the most adaptive life forms, with an enormous diversity of species that are able to colonise the most adverse environments. Even though bacteria harbour very condensed genomes, they have developed an array of strategies for sensing, responding and adapting to different environmental conditions. The complex regulatory networks that bacteria have evolved allow them to modify their gene expression patterns in accordance with the environmental conditions encountered. These changes occur at the molecular, physiological and cellular levels to generate a phenotypic changes in entire clonal populations (201).

However, having a population of bacteria that all express the same genes/proteins is not always beneficial for survival. To promote phenotypic heterogeneity in an otherwise homogeneous genetic background, bacteria have developed another level of regulation known as phase variation. Phase variation results in differential expression for a specific genetic determinant in the clonal population to generate two sub-populations: one lacking or having decreased level of expression of the gene(s) under phase variation control and the other sub-population expressing the gene(s) fully. The key feature of phase variation is to switch from the “ON” to “OFF” phenotype and vice versa. Although a cell of the “OFF” phenotype does not express a given determinant, it has not lost the ability to switch to the “ON” orientation. The frequency of phase variation varies extensively in different genetic system, ranging from one cell in 10 per generation to one in 10,000. This gives rise to a heterogenic and dynamic changing population. It is believed that the phase variation event is random, since it is impossible to predict which cell in a population will undergo the switch. However, phase variation mechanisms have been shown to be very tightly regulated processes controlled by complex regulatory networks that can integrate several environmental signals. The frequency of the phase variation event can be modulated by involvement of accessory proteins or by changing the amount/activity of the proteins mediating the phase variation event (192, 193).

Phase variation control of gene systems has been extensively studied in pathogenic bacteria. Several surface structures like fimbriae, surface proteins, capsule and LPS are regulated by phase variation in bacteria species that live in symbiosis with other organisms. It is believed that phase variation represent a powerful adaptative strategy and that might promote the possibilities to evade the host immune response. Nevertheless, phase
variation has been described in non-pathogenic species and in bacteria not associated with other organisms (192, 193). Hence, it is possible that phase variation mediated control of gene expression is much more broadly dispersed than what has been documented until now. Studies of the mechanisms of phase variation will help to understand the importance of heterogeneity in a bacterial population during environmental survival.

2. Transcription

Transcription is defined as the DNA-dependent synthesis of RNA which is catalyzed by an enzyme complex known as RNA polymerase (RNAP) (40).

DNA (deoxyribonucleic acid) is a polymer of repeated deoxyribonucleotides and the universal molecule that contains the genetic information of most organisms. Each monomer (deoxyribonucleotide) consists of a deoxyribose sugar and a nitrogenous base. There are four different nitrogenous bases: adenine (A), cytosine (C), guanine (G) and thymine (T), defining the four different nucleotides present in the DNA. The sugars of the monomers are joined together by phosphodiester bonds forming long polymers. RNA is very similar in chemical structure to the DNA. The RNA (ribonucleic acid) is a polymer of ribonucleotides, which are formed by a ribose sugar and a nitrogenous base. In the case of the RNA, there are also four different bases, but in contrast to DNA, the uracil (U) nitrogenous base is used instead of the thymine. Generally, the RNA is a single-stranded molecule produced as a copy of one of the DNA strands (template strand) during transcription.

Different types of RNA products are produced by the RNAP during transcription. The messenger RNA (mRNA) is the molecule that is used as template to synthesize proteins during the process called translation that is catalyzed by the ribosomes (40). Two other types of RNA are produced by transcription that are not translated to protein but are involved in the translation of mRNA to protein: ribosomal RNAs (rRNA) that are components of the ribosome and transfer RNAs (tRNA) that are carriers of specific amino acids to the ribosome. In more recent years, it has been shown that other small non-coding RNAs (ncRNA) play additional different roles in gene regulation (174).

DNA is compacted into chromosomes by the help of proteins and RNA. The number of chromosomes varies considerably between different organisms. In most cases, bacteria have one circular chromosome, although many exceptions have been described. In addition, extra chromosomal DNA elements, called plasmids, can be present in the bacterial cells. The segments of DNA that carry the genetic information for proteins - the building blocks of all living organisms - are called genes (Fig. 1).
Genes contain all the information to produce proteins (via mRNA), or non-coding RNAs (tRNA, rRNA, ncRNA) (Fig. 1A). A gene is delineated by a region called promoter - where RNAP binds to initiate transcription - and a downstream sequence called a transcription terminator - where transcription ends. The sequence between the promoter and the terminator is transcribed to RNA. In the case of genes that encode for proteins, the sequence of the DNA that specify the amino acid sequence of the protein is denoted the coding sequence or open-reading frame (ORF). All coding sequences are preceded by a Shine-Dalgarno sequence that is recognized by the ribosomes to start the translation of the RNA to a protein. In bacteria, the genetic information in the genomes is found in a specialized organization, and the expression unit is defined as operon. Different operons are termed monocistronic when it contains a single ORF (Fig. 1A), bicistronic when encoding two ORFs and polycistronic operon when more than two ORFs are under the control of a single promoter (Fig. 1B).
2.1 The players in transcription

Transcription is the first step in the process of gene expression and is catalyzed by the enzymatic complex RNAP in all organisms (Eukarya, Archaea and Bacteria). The activity of the RNAP is a major target in signal-responsive regulation of gene expression. Bacteria, in contrast to other divisions, only have one RNAP for the production of all RNA products. The bacterial catalytic core RNAP enzyme, with a molecular mass of ~400 kDa, is build up by five subunits (α₂, β, β' and ω, see Fig. 2) and is evolutionary conserved in terms of its primary sequence, tertiary structure and function (210). The β- and β'-subunits play a role in binding DNA and NTP (nucleoside triphosphates) and catalyze the RNA synthesis at the active site. These two subunits form the two pincers of the “crab-claw” shape of the RNAP (210). The catalytic activity of the RNAP is dependent on a two-metal-ion mechanism and the stabilization of binding of two coordinated Mg²⁺ ions in the active site of RNAP is crucial for its activity (166). The two identical α-subunits have two functional domains (see below, Fig. 2), the carboxyl terminal domain (α-CTD) that binds to the UP element of the extended promoter and the amino terminal domain (αNTD) that interact with the β- or β'-subunits of the RNAP (63). The ω-subunit has a role in both the assembly of the RNAP core complex (114), and in subsequent sensitivity of RNAP to the regulatory effects of small molecules (206). For correct positioning, RNAP needs to recognize a promoter sequence. This process is mediated by the association of an additional subunit, namely the sigma factor (σ), that ensures promoter specificity of the resulting RNAP holoenzyme (29, 84, 188).

![Fig. 2. An illustration of the different subunits forming the RNAP holoenzyme (α₂, β, β', ω and σ) and its interaction with different elements of a promoter. The consensus sequence of the -35 and -10 promoter elements recognized by the σ⁷⁰-factor are shown below. The box next to -10 element indicates the extended -10 element. Adapted from (26).](image)

Having several σ-factors provides powerful means to redirect the expression of sets of genes by a process known as promoter recognition which is determined by the σ-factor. Most bacteria genomes encode more than one σ-factor. There seem to be an approximate correlation between the number of σ-factors and the physiological diversity of the bacteria. For example, the soil bacterium *Streptomyces coelicolor* harbour the largest number of genes...
encoding $\sigma$-factors, a total of 63 genes (66). In *Escherichia coli* there are seven $\sigma$-factors (reviewed in (84)) that are classified into two families, $\sigma^{70}$-like and $\sigma^{54}$-like (203). The most extensively studied and most abundant $\sigma$-factor is $\sigma^{70}$, also called $\sigma^{D}$, which is involved in the expression of most housekeeping genes. $\sigma^{38}$ ($\sigma^{7}$) is essential for some genes expressed upon entry into stationary phase. $\sigma^{32}$ ($\sigma^{H}$) is involved in expression of heat shock proteins and $\sigma^{24}$ ($\sigma^{E}$) in expression of genes whose products deal with misfolded proteins in the periplasm. $\sigma^{28}$ ($\sigma^{F}$) controls the expression of flagella and chemotaxis genes and $\sigma^{FecI}$ in expression of the *fec* operon. These $\sigma$-factors all belong to the $\sigma^{70}$-like family. $\sigma^{54}$ ($\sigma^{N}$) is involved in expression of some stress response genes and genes involved in nitrogen scavenging. The $\sigma^{54}$-factors have no primary sequence similarity with $\sigma^{70}$ and have a different domain organization (28). In this thesis only the $\sigma^{70}$-like family of $\sigma$-factors will be reflected upon.

Five DNA sequence elements of the promoter have been identified to be important for the promoter recognition by the RNAP holoenzyme (Fig. 2). The two major determinants are the -35 and the -10 elements, which are located approximately at 35 and 10 base pairs (bp) upstream of the transcriptional start site (+1 site), respectively. Consensus hexameric sequences of the -35 and -10 elements have been established and these regions are separated by a spacer consisting of 17 ± 1 bp (31, 44, 121). Other elements that can also be important for the expression of some genes are the UP element, the extended -10 element and the discriminator located between -10 and +1. The UP element is a DNA sequence of about 20 bp, located upstream of the -35 element of some promoters, which is bound by the $\alpha$ CTD of the RNAP (153). The extended -10 element is a 3-4 bp motif, which has been suggested to be important for a subset of promoters that lack or have poor consensus -35 or -10 hexamers (78, 115). The discriminator refers to a G+C-rich region between the positions -10 to +1 present in some promoters (186).

### 2.2 Transcription step-by-step

As illustrated in Fig. 3, transcription is a cyclic process and can roughly be divided into three major steps: promoter binding and initiation, elongation and termination. In brief, the first step in the cycle requires the presence of RNAP core enzyme with a promoter specific $\sigma$-factor (84). The RNAP holoenzyme slides along the DNA until it encounters an appropriate promoter sequence where it binds, opens the two DNA strands around the transcriptional start site to form the transcription bubble required for RNA synthesis to start (155). The RNAP first enters an abortive initiation cycle with the production of short RNA products. Once the RNA products exceed ~12 to 14 nucleotides in length, transition to elongation takes place and RNAP leaves the promoter (promoter escape) and the $\sigma$-factor is released. Transcription elongation will continue until the RNAP reaches a termination signal, at which
point RNAP is releases from the DNA and so becomes available for association with another σ-factor to restart the cycle once again (20).

Fig. 3. The transcription cycle consist of four main steps; 1 - Promoter recognition, 2 - Initiation, 3 - Elongation and 4 - Termination. Upon transcriptional termination, RNAP is released (5) and can bind to another σ-factor (6), thus the cycle continues.

2.2.1 Promoter binding and initiation of transcription

Transcription initiation is a multi-step process consisting of three main steps; promoter recognition, initiation and promoter escape (Fig. 4). Transcription initiation involves the interaction of the RNAP to the promoter element. As already mentioned in the preceding section s, this requires that the RNAP is associated with a specific σ-factor. The σ-factor has three main functions; identification of the promoter sequence, positioning of the RNAP holoenzyme and aiding the unwinding of the two DNA strands (203).

When the RNAP initially binds to the promoter, a RNAP-DNA complex is formed called the closed-complex (RP<sub>c</sub>). In this status the DNA is double-stranded and the RNAP is covering a region from --55 to +1 relative to the transcriptional start site (38, 107). The complex undergoes several conformational changes in both RNAP and DNA known as isomerization steps. Eventually, the open-complex (RP<sub>o</sub>) develops as the DNA strands get separated in the region --11 to +3, allowing a base on the template strand to pair with the initiating NTP (iNTP) (38, 100, 146). After the formation of the open-complex, incorporation of NTPs drives transcription forward. However, initially the RNAP synthesize short abortive products from most promoters (2-7 nucleotides long) before starting elongation (79). During this process, the RNAP maintains contact with the -35 element and moves back and forth by a process also known as “scrunching”. The energy that is stored in these
intermediates has been suggested to allow the RNAP to begin the transition into the elongating complex (92, 150).

![Diagram](image.png)

Fig. 4. Different steps in promoter recognition and transcription initiation. RNAP holoenzyme (R) binds to the promoter (P) and forms the closed-complex (RPc) with the DNA double-strand. After a series of isomerization steps, the DNA double-strand is opened forming the open-complex (RPo). As the initiating NTP (iNTP) pairs with the +1 (transcriptional start) this forms the initiating complex (RPinit). Several rounds of abortive initiation occur, producing short RNA products before the elongating complex (EC) leaves the promoter and the σ-factor disassociate from the RNAP. Adapted from (146).

### 2.2.2 Elongation and termination

When the RNAP leaves the promoter (promoter escape) and transcription elongation starts it has been assumed that the σ-factor is always released from the RNAP. However, more recently it has been shown that the release is not obligatory for escaping from the promoter and that σ-factor release occurs stochastically during the early stages of elongation (149). The elongating complex (EC) is much stable and processive as compared to the initiation complex. During elongation, the RNAP-DNA complex can change between an active (elongating competent) and inactive (backtracked/paused) state (67, 124). Pauses occurring during elongation determine the overall elongation rate. Unlike during initiation, the elongation complex does not disassociate from the DNA when these pausing events occurs. There are two mechanistically distinct classes of pause signals; hairpin-dependent (e.g. his leader pause) and hairpin-independent pauses (e.g. ops pause) (5, 11). During pausing of RNAP, regulatory interactions, such as binding of transcription elongation factors and the proper positioning of the ribosome take place (6, 151). The exact position and time that the RNAP spend in the paused state can determine the effect of RNAP pausing on gene expression. Furthermore, pausing also plays a role in folding of the mRNA (132, 133) and in the termination process (68).

Elongation continues until the RNAP encounters a termination signal. There are two different classes of signals known: the Rho-independent, also known as intrinsic terminator, and the Rho-dependent ((74) and references therein). The intrinsic terminator is composed of a G+C-rich stem-loop followed by a
series of U residues. The Rho-dependent termination requires the binding of the Rho-factor to a \textit{rut} (Rho utilization) site on the nascent transcript followed by the interaction with the RNAP. Thus like initiation, elongation and termination are dynamic processes controlled at different levels.

2.3 Regulation of transcription

The activity of RNAP can be regulated throughout the entire transcription cycle. Although it has been anticipated that most transcription regulation occurs at the level of initiation, there are also numerous factors influencing the elongation and termination steps of transcription. In this thesis, effectors that regulate the first steps of transcription (initiation and elongation) are primarily discussed.

When studying bacterial gene regulation, it is important to bear in mind that the intracellular levels of RNAP are limiting. During active growth, most of the RNAP in the cell is associated with the genes encoding stable RNAs (rRNA and tRNA) that are very highly expressed since they are crucial components of the translational machinery. Therefore, there is a limited number of RNAP available to transcribe the 4-5,000 genes present in the cell (84). We might consider that inside the cell there is a competition between the nearly 2,000 different promoter sequences for the accessible RNAP (156). Several factors might affect the affinity of the RNAP to a specific promoter: promoter sequence, \( \sigma \)-factor composition, small ligands, transcription factors and structure of the chromosome. Each of these factors potentially promotes variations in the level of transcription and the impact of each of them varies from gene (or sets of genes) to gene.

Variations in the promoter sequence offers one way of regulating transcription by promoting the recruitment of the RNAP. However, this can hardly be considered as a mechanism of modulation of gene expression in the short term, since an adjustment would involve alterations in the DNA sequence. Nevertheless, accumulation of mutations through generations might provide a more appropriately controlled promoter sequence for bacterial adaptation to certain conditions. Thus promoter sequence changes might play an important role in the control of gene expression on an evolutionary time scale.

As mentioned earlier, most of the bacteria can express more than one \( \sigma \)-factor simultaneously leading to competition for binding of the different \( \sigma \)-factors to the RNAP (84). Depending on the specific \( \sigma \)-factor that binds to the core RNAP, expression of a specific subset of genes will be favored. Therefore, depending on the global composition of \( \sigma \)-factors present in the cell, a specific global gene expression profile would be generated (reviewed in (84, 203)). The global transcriptional capacity is also modulated by the
activity of anti-sigma factors that bind and sequester specific σ-factor so that they can not associate with RNAP (82).

The bacterial DNA in the cell is negatively supercoiled and its conformation changes according to environmental conditions and/or by the action of the so-called nucleoid-associated proteins (e.g. H-NS, Fis, IHF and HU). Most of these factors are very abundant and bind DNA non-specifically (8, 185). The conformation of DNA can affect both the accessibility of the RNAP to a promoter as well as the activity of the RNAP during transcription (185).

The role of the transcription factors and small ligands in transcription will be extensively discussed in the following sections.

2.3.1 Transcription factors

Classically, transcription factors (TFs) have been considered as proteins that bind to DNA. However, it is now evident that not all TFs directly bind DNA. For example, their action can be mediated by modifying the binding or changing the activity of other TFs or by interacting directly with the RNAP. Moreover, in the last years, studies have shown that many factors directly involved in regulating transcription are non-proteinaceous, like RNA products and small molecules.

In E. coli there are more than 300 predicted genes that encode for proteins that bind to DNA and either up- or down-regulate transcription (138, 181). The function of about half of those proteins have been verified experimentally. Most of them bind to sequence-specific regions on the DNA and control the expression of either a large number of genes (global regulator) or just one or few genes (specific regulator). In E. coli, about 50 % of all genes have been suggested to be regulated by seven global regulators: cyclic AMP receptor protein (CRP), fumarate reductase and nitrite reductase (FNR), integration host factor (IHF), factor for inversion stimulation (Fis), anaerobic respiratory control (Arc), nitrite regulation (Nar) and leucine regulatory protein (Lrp) (111). The TFs regulate gene expression in response to environmental changes. Sensing changes in environmental parameters might alter the amount, the availability and/or the activity of TFs and consequently induce important alterations in the expression profile of the bacterial cell. Certain conditions might induce the levels of TFs in the cell with consequent effects on gene expression. In some cases the TFs might either interact or not with other cellular factors which determine whether it is active or not. The availability of active TFs can also be altered as a consequence of environmental signals, following an alteration of the gene expression pattern. The alteration of the activity of the TFs can be mediated by binding to small ligands, like cAMP for the CRP protein (71), by covalent modifications, like phosphorylation by a sensor kinase in two-component systems, etc (172).
As the TFs bind to the DNA they can either increase (i.e. activator) or decrease (i.e. repressor) transcription. Different mechanisms have been described to promote activation or repression of transcription by studying how TFs affect specific genetic systems. Hence, the classical activation of transcription by improving the recruitment of the RNAP to the promoter can be achieved by several specific mechanisms: binding to the UP element (e.g. cAMP-CRP), binding to the -35 element (e.g. λ CI protein) or by changing the conformation of the promoter by binding to the -35 to -10 region (e.g. MerR family). Correspondingly, repression of transcription can occur by different and diverse mechanisms: steric hindrance for the RNAP to bind to the promoter (e.g. Lac repressor), by DNA looping that shuts transcription initiation (e.g. GalR), by modulation of an activator (e.g. CytR), etc (26). In addition to the enormous mechanistic diversity that exists, when analyzing how regulation occurs in the cell, it is important to have in mind that frequently the activity of a promoter is influenced by several environmental cues and regulated by more than one TF simultaneously. Moreover, it is common that global regulators and specific regulators work co-ordinately to regulate the expression of a genetic locus (111, 112). Furthermore, some TFs are regulating the expression and activity of other TFs, forming very complex regulatory networks.

2.3.2 Non-proteinaceous regulatory molecules

There are several non-proteinaceous molecules in bacteria that can regulate transcription and gene expression, both directly and/or indirectly, in different ways. Examples are small regulatory RNAs, modified nucleotides (cAMP, ppGpp and cyclic di-GMP) and quorum sensing molecules (AHLs and oligopeptides).

Small RNAs (sRNA) are non-coding RNAs of 40 to 400 nucleotides in length, which have the ability to regulate gene expression. In E. coli at least 50 sRNAs has been described so far, although just a few of them have been characterized. The expression of most of sRNA studied to date is induced in response to changes in environmental conditions. They regulate gene expression by using different mechanistic strategies (reviewed in (69, 113, 173, 175)), e.g.:

- by base-pairing with target mRNAs (both in cis and trans) affecting transcription termination, mRNA degradation and/or inhibiting translation.

- by interacting with proteins and thereby inhibiting or changing its activity.

- by directly interacting with the RNAP. For instance, the 6S RNA interacts with the RNAP changing its promoter preference. The 6S RNA is abundant in E. coli and other species, and its synthesis is induced upon entry in
stationary phase of growth. 6S RNA represses expression of several $\sigma^{70}$-RNAP promoters and induces expression of several $\sigma^{A}$-RNAP promoters. Its interaction with the active site of the RNAP mediates both the inhibition (by DNA blockage) and the activation of transcription (by template RNA synthesis) (195).

There are several low molecular mass molecules that play a role in bacterial gene regulation. The quorum sensing autoinducers (e.g. AHLs, AI-2, PQS and oligopeptides) modulate gene expression in response to population density. These molecules can regulate gene expression by interacting either with a transcription factor or with a response-regulator of a two-component system (197).

Several cytoplasmic second messengers are modified nucleotides. The cyclic dinucleotide 3'5'-cyclic diguanylic acid (c-di-GMP) is synthesized from GTP by proteins having diguanylate cyclase activity and degraded by proteins having phosphodiesterase activity. Intracellular c-di-GMP levels are altered by external signals (environmental changes) causing concomitant effects on gene expression. So far, the exact mechanism of how c-di-GMP alter gene expression is not known (152). Both quorum sensing and c-di-GMP co-ordinately regulate processes such as biofilm formation and virulence. Therefore, it has been proposed that the two regulatory pathways are linked (30). Cyclic-adenosine 3'5' monophosphate (cAMP) is synthesized from ATP by adenylate cyclases. cAMP binds to CRP (catabolite repressor protein) and activates the dimeric form of this global regulator. cAMP-CRP complex binds to the DNA, commonly at the promoter region, exerting its control of the gene expression (71). Another modified nucleotide involved in gene control is the guanosine-3',5'-bispyrophosphate (ppGpp). ppGpp is synthesized in response to different kinds of stress conditions and interacts directly with the RNAP (33, 35). This will be discussed more extensively below.

3. **ppGpp - a stress alarmone**

Guanosine tetra- and penta-phosphate, collectively called ppGpp or “magic spot”, are modified nucleotides that act as alarmones or stress signals in bacteria. ppGpp is very rapidly synthesized from GTP (or GDP) and ATP in response to any stress condition that will result in growth arrest (33). As early as 1969, Cashel and Gallant described that stimulation of intracellular level of ppGpp in *E. coli* in response to nutrient starvation has important consequences in the pattern of gene expression, termed the stringent response (32). Since then, it has been illustrated that ppGpp is also synthesized during many other stress conditions. ppGpp is not restricted to Gram-negative bacteria, it is also found in Gram-positive bacteria (85) and in the chloroplasts of plant cells where it also functions in stress related processes (179).
3.1 Stringent response

The hallmark of the stringent response in *E. coli* is the down-regulation of stable RNA synthesis (rRNA and tRNA) and ribosome production that occurs upon amino acid starvation (33, 157, 170). This is mediated by the small regulatory molecule, ppGpp (32). In *E. coli*, the synthesis of ppGpp occurs by two related proteins: RelA and SpoT (33, 208). The RelA protein mainly synthesizes ppGpp in response to amino acid starvation by recognizing and binding to stalled ribosomes, that have an uncharged tRNA bound in their A-site (33, 200). SpoT is a bifunctional protein that synthesizes ppGpp in response to other stress conditions than amino acid starvation (208). The mechanism leading to SpoT activation is not known (61, 200). Recently it was shown that SpoT is binding to CgtAc, a ribosome-associated protein in *E. coli*, but whether this interaction has any effect on the activity of the SpoT protein remains to be elucidated (204). Elevated levels of ppGpp during a long period time are deleterious for the cell physiology and therefore it is essential to hydrolyse it. In *E. coli*, the SpoT protein is responsible for the hydrolysis (33, 61). An *E. coli* strain that lacks both RelA and SpoT proteins do not synthesize ppGpp under any growth conditions and is denoted ppGpp0. These strains are unable to grow in minimal media as a consequence of extensive amino acid requirement (auxotrophy) (208).

There are several bacteria species that only harbour one copy of a RelA-SpoT homologue with both synthetase and hydrolase activity (116). These homologues are called Rsh (RelA-SpoT homologue) proteins and such bifunctional enzymes modulate the ppGpp levels through two distinct active sites (21). It has been shown that in some Gram-positive bacteria, the Rsh proteins do not require interaction with the ribosome to trigger their ppGpp synthetase activity. However, in the presence of stalled ribosomes the activity of the Rsh proteins increases (85). Rsh proteins have also been found in chloroplasts of plants cells and are responsible for accumulation of ppGpp due to stress (62, 179, 191). Notably, it has been found that there are some microorganisms that do not produce ppGpp, including some parasitic bacteria (*Treponema pallidum*, *Chlamydia* species, *Rickettsia prowazekii*) and some Achaea species (116).

ppGpp binds to the β- and β′-subunits of the RNAP core enzyme (34, 184). It has been shown that mutations within the genes *rpoB*, *rpoC* and *rpoD* that encodes for the RNAP subunits β, β′ and σ70 respectively, can render a RNAP that is unresponsive to the levels of ppGpp in the cell (33, 122). Cross-linking and co-crystallization studies of the RNAP of *Thermus thermophilus* and ppGpp has suggested that ppGpp interacts with the β- and β′-subunits in close proximity to the active site, coordinating the position of the Mg2+-ions (7, 34). The co-crystallization studies suggested that ppGpp can bind to the RNAP in two different orientations. The significance of these two binding positions for the mechanism of transcriptional regulation by ppGpp has not been determined (7). Recently, it has been suggested that the predicted binding
site for ppGpp within the RNAP of *T. thermophilus* may not be accurate for the *E. coli* RNAP. Substitutions of the predicted interacting amino acids residues in the core enzyme did not result in loss of ppGpp-mediated regulatory effects (205). Further studies are needed to elucidate the exact binding position of ppGpp within the RNAP. This knowledge will facilitate the understanding of the regulatory mechanism of ppGpp on transcription.

The interaction of ppGpp with the RNAP has pleiotropic effects on gene expression. Numerous genes are both repressed and induced in response to increasing concentration of ppGpp (see Fig. 5). Although the decrease of protein synthesis, as a consequence of repression of stable RNA, is perhaps the most predominant effect during the stringent response, ppGpp also regulates DNA replication, fatty acid synthesis and cell wall component expression, among others. There is also induction of amino acid biosynthesis, stress response related proteins, and alternative sigma factor such as RpoS, RpoE and RpoH (σ^S, σ^E and σ^H) (33, 50).

*Fig. 5. Synthesis of ppGpp in *E. coli* and consequential global effects on gene expression. ppGpp is synthesized from GTP/GDP and ATP by two parallel pathways depending on the stress condition encountered (33), the RelA- or the SpoT-dependent pathway. ppGpp is very rapidly hydrolysed by the SpoT protein. ppGpp binds to RNAP and redirects gene expression to survive the stress conditions. Adapted from (109).*
Furthermore ppGpp has been shown to play an important role in regulating genes involved in virulence and survival under host/environmental induced stress (21, 85). Hence, ppGpp regulates long-term persistence of *Mycobacterium tuberculosis* (41), quorum sensing and virulence in *Pseudomonas aeruginosa* (52), biofilm formation in *E. coli* and *Listeria monocytogenes* (1, 180), and has an effect on virulence in *Vibrio cholerae*, *Salmonella*, *Legionella*, *L. monocytogenes*, *Borrelia burgdorferi* and *Campylobacter jejuni* (60, 70, 119, 140, 165, 180).

### 3.2 DksA - a possible co-regulator

DksA is a 151 amino acid long (17 kDa) polypeptide encoded by a non-essential gene found in *E. coli* and related bacteria. DksA was originally found as a multicopy suppressor of the temperature-sensitive phenotype of *dnaKJ* mutants in *E. coli* (91). Studies of *dksA* over-expression and *dksA* deletion strains suggest a role for DksA in a variety of cellular processes, including cell division, stringent response, quorum sensing and virulence in *E. coli*, *P. aeruginosa*, *Salmonella* and *Shigella flexneri* (76, 89, 117, 187, 198). Several studies suggested that DksA acts as a co-regulator for the ppGpp-dependent regulation of genes in *E. coli*. A *dksA* mutant strain shows an auxotrophy phenotype that can be suppressed by a *rpoB* mutation (βT563P) which also suppresses the auxotrophy of a ppGpp0 strain (122). Moreover, the effect of a *dksA* mutation on *rpoS* induction is similar to that observed in the absence of ppGpp (76). Furthermore, it has been shown that DksA enhances the effect of ppGpp on both the negatively (134) and positively (135) regulated genes, both *in vivo* and *in vitro*. However, it is notable that DksA is not as widely distributed among bacterial species as ppGpp (136, 137).

Conversely, when studying the phenotypes of a *dksA* and ppGpp0 mutant strains some differential phenotypes have been observed (2, 25, 110). The amino acid requirements are not exactly the same for the ppGpp0 and the *dksA* mutant strains (25) and the auxotrophic phenotype of the ppGpp0 strain can not be restored by over-expressing DksA in all *E. coli* K-12 strain backgrounds (110). Moreover, several genes in *E. coli* are differentially affected in ppGpp0 and *dksA* mutant strains *in vivo*, including genes for adhesion, chemotaxis and motility. It has been shown that DksA and ppGpp may exert independent effects on transcription both *in vivo* and *in vitro* (2, 110, 143). These conclusions will be further discussed in the Result and Discussion section of this thesis.

The crystal structure of DksA reveals a prominent coiled-coil domain in the N-terminus and a conserved Zn-finger motif in the C-terminus (137, 196). DksA is structurally very similar to the transcription elongation factor GreA, although they do not share any sequence homology (Fig. 6, (169)). DksA co-purifies with the RNAP. Although no co-crystallization structure between DksA and RNAP has yet been reported, it has been suggested that DksA interacts with
RNAP directly by protruding the coiled-coil N-terminus into the secondary channel, thereby stabilizing the interaction between RNAP and ppGpp (137).

3.2.1 Gre-factors

The Gre-factors are proteins that interact with the RNAP during transcription and suppress RNAP arrest or pausing by inducing cleavage of the RNA transcript within the active site of the enzyme (55, 123). In *E. coli* there are two genes coding for the Gre-factors, *greA* and *greB*. Although, neither *greA* nor *greB* are essential for viability, several phenotypes has been associated with *gre* mutant strains, including sensitivity to divalent metal ions, salt and temperature (171, 176). In *Pseudomonas*, a *greA* mutant strain is unable to grow in minimal media (108). The Gre-factors are evolutionarily conserved and found in more than 60 bacterial species. Members of the Gre-family, which includes DksA and Gfh1, share a conserved two-domain structure, the globular C-terminal domain and an extended coiled-coil N-terminal domain (105, 137, 169, 178).

The Gre-factors bind to the RNAP with the C-terminal domain near the secondary channel and the N-terminal coiled-coil domain protrudes into the channel so that its tip come in close proximity to the active site (106, 130, 167). The acidic residues, at the tip of the coiled-coil domain of the Gre-factors help to coordinate the Mg$^{2+}$ ions in the catalytic center of the RNAP, thereby inducing the endonucleolytic cleavage reaction (Fig. 6). The induced cleavage reactivates the stalled elongation complex formed as a result of backtracking or arrest of the RNAP (106, 167). The Gre-factors are also involved in earlier steps of the transcription, since they facilitate promoter escape and entrance to the elongation phase by decreasing the production of abortive RNAs (54, 80, 176). A recent microarray study in *E. coli* has shown the importance of GreA in gene regulation in vivo. The stimulatory effect of GreA on gene expression occurs by facilitating the promoter escape event that depends on the ability to stimulate transcript cleavage. On the other hand, GreA-mediated negative regulation appears to be indirect through some unknown factors (171). The cleavage reaction catalyzed by the Gre-factors removes miss-incorporated nucleotides and, therefore, the Gre-factors also contribute to the proofreading and fidelity of transcription (55, 209).

3.2.2 Regulation through the secondary channel

The secondary channel of the RNAP links the external milieu to the active site, acting as an entry port for NTPs and regulatory factors (123). The Gre-factors are members of a family of secondary channel regulators that share a common tertiary structure but have different effects on transcription regulation (45, 123). DksA influences ppGpp-mediated regulation by destabilizing open-complex formation (134, 135). GreA and GreB, as already
mentioned, facilitate promoter escape and rescue of stalled/arrested RNAP elongation complexes by inducing endonucleolytic cleavage (55). Another member, Gfh1 from the *Thermus* genus, does not stimulate endonucleolytic cleavage of RNAP but instead inhibits the GreA-mediated effect on RNA cleavage and RNA synthesis (104, 178). When the crystal structure of GreA and Gfh1 were compared, a striking difference in the orientation of the N-terminal and C-terminal domain was revealed. In Gfh1, the N-terminal domain is flipped ~170° as compared to GreA (178). Interestingly, the Gfh1 can adopt two different conformations (inactive and active) with the switch to the active conformation being dependent on low pH (105). Superimposition of the structures of DksA and GreA illustrated that the position of the acidic residues of the coiled-coil domain of DksA is likely oriented differently as compared to GreA (137).

It has been suggested that all these regulators by interacting with the secondary channel might have overlapping or antagonistic effects on gene expression. Potrykus *et al* showed that GreA and DksA have antagonistic effects on *rrnB* transcription *in vitro* and *in vivo* (143). On the other hand, Rutherford *et al* showed that while GreA had only modest effects on transcription, GreB could fulfill some of the regulatory roles of DksA on *rrnB* transcription *in vitro* (154). The presence of multiple secondary channel regulators in the cell implies that competition for binding to the secondary channel could potentially occur. Such competition would be influenced by

![Fig. 6. Structure of the RNAP core enzyme with the main channel containing the DNA (solid line) and the secondary channel marked in white. The dotted line shows the synthesized RNA exiting through the RNA exit channel and the two Mg²⁺ ions in the active site are marked by circles. To the right are shown factors predicted to enter (NTP and ppGpp) or bind to the secondary channel (DksA and GreA/GreB) in *E. coli*. The β-subunit is shown in dark grey and the β'-subunit in light grey. Adapted from (123).](image-url)
both the intracellular concentration of the various factors as well as their intrinsic binding affinities towards RNAP.

It has been shown in *E. coli* that the intracellular concentration of DksA, GreA and GreB are relatively constant through growth, with DksA being more abundant than either GreA and GreB (154). DksA and GreB were shown to have similar affinities for the binding to the RNAP, which apparently are much higher than the affinity shown by GreA. Due to the differences in relative concentration, it is assumed that a substantial proportion of the RNAP would be preferentially occupied with DksA. Whether that is valid for all growth phases and conditions remains to be elucidated. The knowledge that the structure of Gfh1 changes from an active to an inactive form upon pH alterations suggests that under certain conditions conformational changes of factors that interact with the secondary channel might alter their specific affinity for RNAP and, therefore, make important adjustments in the amount of RNAP interacting with each specific factor. Another possibility that has not been considered in the literature is that the affinity of RNAP for the different secondary channel TFs might be altered when RNAP binds a specific promoter sequence or interacts with a particular σ-factor. These interactions could potentially change the conformation of the secondary channel of RNAP holoenzyme and thereby interactions of regulatory factors through this part of the enzyme.

### 3.3 Mechanism of regulation by ppGpp and/or DksA

ppGpp and DksA can be classified as global transcriptional regulators because transcriptomic and proteomic studies have shown that they influence the expression of numerous bacterial genes. They activate and repress transcription from a number of promoters, both directly and indirectly (50, 110, 139, 161, 182). In most cases, the two factors seem to regulate transcription co-dependently. However, there are cases where apparently they affect transcription by independent mechanisms (2, 110).

#### 3.3.1 Direct negative regulation of transcription

Negative regulation mediated by ppGpp and DksA on rRNA production has been extensively studied over the last years (136). Moreover, direct inhibitory effects of ppGpp have also been reported for other promoters (13, 33, 142). Several mechanisms have been suggested to explain the direct negative effects of ppGpp on transcription.

- The *rrn* promoters form intrinsically unstable open-complexes and are very sensitive for further destabilization (13, 134, 136). Direct binding of
ppGpp and/or DksA to RNAP further destabilizes their already intrinsically unstable open-complexes leading to down-regulation.

- ppGpp also has direct negative effects on promoters with stable open-complexes. In these cases it has been proposed that ppGpp disturbs the promoter escape process to exert its negative effect (95, 142).

- the amount of iNTP have shown to be important for some promoters regulated by ppGpp, suggesting that competition between ppGpp and iNTPs for the active site could account for negative ppGpp-dependent effects (88, 208).

- ppGpp has also been demonstrated to induce pausing during the elongation step of transcription (88, 189).

These different mechanisms are not exclusive and could function simultaneously in the fine regulation of different genetic systems.

### 3.3.2 Direct and indirect positive regulation

ppGpp has been shown to be a positive regulator of several genes in vivo (33). The mechanism for the direct positive regulation by ppGpp has not been as extensively studied as those of direct negative regulation. For many years, no clear experimental evidences of a direct positive regulation of transcription by ppGpp could be obtained. Therefore, it was thought that the positive regulation by ppGpp occurred only by an indirect mechanism. However, when it was discovered that the DksA protein promotes ppGpp-mediated effect on transcription, direct positive effects of ppGpp has been revealed in vitro in the presence of DksA. These studies were first done with the amino acid biosynthesis genes (135). Direct effects of ppGpp on the phage promoter λ.PaQ has also been shown (141).

Promoters that are positively regulated by ppGpp and/or DksA often form very stable open-complexes. The destabilization mediated by ppGpp and DksA of the open-complex formation might actually help promoter escape and thereby transcription initiation from these promoters (14). Alternatively, it has been proposed that ppGpp/DksA lower the transition state energy required for formation of a rate-limiting intermediate in the pathway to open-complex formation thereby stimulating promoter output (136).

Superimposed on direct effects, it has been suggested that ppGpp could indirectly activate gene expression by altering the availability of the RNAP in the cell. Since ppGpp down-regulates transcription of the rrn genes, in the presence of ppGpp more σ^{70}.RNAP holoenzyme would be accessible for other promoters (12, 211).
ppGpp increases pausing of RNAP at all promoters tested. The ppGpp-induced pausing increases the time that the RNAP occupies the promoter. This increased residence time could facilitate productive interaction and binding of regulatory factors that could be important for stimulating transcription of some genes (22, 101).

ppGpp and/or DksA are also required for transcription from promoters recognized by alternative σ-factors (e.g. σ^32, σ^35, σ^54 and σ^7) which might be explained by three possible mechanisms (17, 39, 64, 65, 86, 103):

1. ppGpp/DksA directly affects the interaction of the σ-factor with the RNAP core enzyme thereby affecting the assembly of the alternative holoenzymes.

2. The promoters that are regulated by alternative σ-factors have kinetic properties which make them sensitive to the presence of ppGpp and DksA.

3. The repression of *rrn* transcription in the presence of ppGpp and DksA increases the availability of pool of free core RNAP in the cell, thereby facilitating holoenzyme formation with alternative σ-factors.

As with negative regulation by ppGpp, these mechanisms are not mutually exclusive and more than one can act simultaneously.

4. **Escherichia coli - a model organism**

*E. coli* is a rod-shaped Gram-negative bacterium that was first discovered by Theodor Escherich in 1885. *E. coli* is classified as a γ-proteobacteria and part of the *Enterobacteriaceae* family. It is about 2 micrometer long and 0.5 micrometer wide. *E. coli* is commonly found in the lower intestine of warm-blooded animals and it is part of the normal flora of the gut. Most of the strains are harmless and provide benefits to the host by producing vitamin K₂ (16) or by preventing the colonization of pathogenic bacteria within the intestine, among other mechanisms (81). *E. coli* is a model organism commonly used in microbiology and genetics to study bacterial physiology, metabolism, gene regulation, signal transduction, and cell wall structure and function.

Wild-type *E. coli* has meager growth requirements and can synthesize all the macromolecular components of the cell from glucose. The bacterium can grow both in the presence (aerobic) and absence (anaerobic) of oxygen. Like other bacteria, *E. coli* has the ability to respond to environmental signals such as chemicals, pH, temperature, osmolarity, etc.
4.1  Strains of *E. coli*

*E. coli* strains are divided into several different serotypes based on their O-(somatic), H- (flagella) and K- (capsule) antigens. Practically, *E. coli* can roughly be divided into three sub-groups of strains: commensal, pathogenic and laboratory strains.

The commensal strains colonize the infant gastrointestinal tract within 40 hours after birth and may persist for life time. *E. coli* is the predominant facultative organism in the human gastrointestinal tract but represent only 1% of the total bacterial mass (183). Commonly, the commensal strains are considered as non-pathogenic.

The pathogenic *E. coli* strains can cause diverse diseases. Although they are mostly known as the causative agents of gastroenteritis, urinary tract infection and neonatal meningitis, they can also cause peritonitis, mastitis, septicemia and pneumonia (183). Based on their serological characteristics and virulence properties, pathogenic *E. coli* strains are classified into different virotypes. The virotype causing urinary tract infections are uropathogenic *E. coli*. The differences between commensal and pathogenic *E. coli* strains are evident when comparing their genome sequences. Pathogenic *E. coli* have acquired additional virulence genes by horizontal gene transfer. The “extra” DNA detected in the genomes of pathogenic strain is often located in discrete regions of the genome that are called pathogenicity islands. Interestingly, regulation of these virulence factors often involves transcription factors that are encoded by the backbone chromosome.

The laboratory strains are either commensal or virulent/pathogenic isolates that have been used as model organisms in the research laboratories for many years. After cultivation under the laboratory conditions, most strains loose some of their original phenotypes. The most commonly used laboratory strains of *E. coli* belong to the K-12 serotype.

During my thesis work I used laboratory K-12 strains and uropathogenic isolates of *E. coli*. The characteristic phenotypes of the former are described in the following section.

4.1.1  Uropathogenic *E. coli* - UPEC

Uropathogenic *E. coli* (UPEC) strains can colonize the urethra (urethritis), the bladder (cystitis) and/or kidneys (phyleonephritis) and are the causal agent of about 90% of all the urinary tract infections each year. UPEC express a number of different virulence determinants such as adhesins, iron uptake systems and cytotoxins (87) that promote colonization of the host tissues and successful establishment of infections (Fig. 7).
Fig. 7. Illustration of the different virulence and colonization factors that UPEC strains may express and that are important for colonization and establishment of infection. Adapted from (87).

4.2 Type 1 fimbriae

During the infection process, the adhesins expressed by UPEC aid colonization of different tissues and organs of the urinary tract by recognizing different surface receptors on the endothelial cells. The ability of individual UPEC isolates to colonize a specific tissue within the urinary tract depends on which kind of adhesins can be expressed (120).

Fimbriae, also called pili, are proteinaceous hair-like appendages found on the surface of the bacterial cell. The word “fimbriae” and “pili” comes from the Latin word for “thread” or “hair” (23). They range from 3-10 nanometers in diameter and can be several micrometers in length. They are built up by several different subunit types and have a specific adhesin protein at their tip. Bacteria use fimbriae to attach to each other, to biotic or abiotic surfaces and to host cells. In general, most Gram-negative bacteria produce fimbrial structures and similar systems have been described for some Gram-positive bacteria. The structure and assembly of the fimbrial subunits vary between different fimbriae systems and bacterial species. Different strains of E. coli have the genes and the ability to express several types of fimbriae structures. The studies presented in this thesis have been focused on better understanding of how the type 1 fimbriae are regulated in UPEC strains.
4.2.1 General characteristics

Type 1 fimbriae are the most common fimbriae in species of the Enterobacteriaceae family. About 70 to 80% of all wild-type E. coli strains have the genes to express these fimbriae. Type 1 fimbriae are involved in attachment to biotic and abiotic surfaces (144) and are important for the initial steps of biofilm formation (158, 159). Moreover, they are a crucial virulence factor of UPEC because they mediate initial adhesion and invasion of bladder cells by binding to mannose-containing receptors on the uroepithelium (83, 120, 162). UPEC can evade the human host responses by invading the epithelial cells of the bladder where they form intracellular bacterial communities (90). In addition, type 1 fimbria are required for maturation of such intracellular bacterial communities formed within epithelial cells that provides a protective mechanism against the host immune response (207).

Type 1 fimbriae are about 7 nanometers in diameter and 0.2 to 2 micrometers in length. The assembly of the fimbriae structures occurs by the “chaperone/usher-pathway” (118). The fimbriae consist of several non-covalently linked subunits, a major rod (FimA), a tip fimbriullum (FimF and FimG) and an adhesin (FimH) (Fig. 8). The adhesin recognize mannose-containing receptors and the binding of the fimbriae can be blocked by D-mannose or α-methylmannosides (120).

4.2.2 Genetic organization and regulation

The genes encoding for the synthesis of type 1 fimbriae are clustered in the fim determinant, which consist of a polycistronic operon containing the structural genes (fimAICDFGH) and two monocistronic operons (fimB, fimE) encoding for the recombinases involved in the regulation of the expression of the structural genes (see Fig. 8).

Transcription of the polycistronic fim operon is regulated by phase variation and by promoter activity. The phase variation is mediated by inversion of a 314 bp DNA fragment that contains the promoter for the polycistronic fim operon located just upstream of the major subunit gene, fimA (Fig. 9 and (56, 99)). Depending on the orientation of this invertible DNA fragment, the promoter is positioned to direct transcription of the structural fim genes (“ON” orientation) or not (“OFF” orientation). The “switching-mechanism” is mediated by the site-specific recombinases FimB and FimE (Fig. 9). FimB mainly mediates the switch from OFF-to-ON, and FimE the switch from ON-to-OFF (58, 98). In an E. coli K-12 population, grown in standard laboratory conditions, only about 10% of the cells have the main fim promoter in the ON-orientation, therefore, most cells in the population are afimbriated under these conditions.
Fig. 8. The genetic organization of the fim determinant of E. coli. The genes encoding the different subunits of the fimbriae are color-coded to match the fimbriae illustration shown to the right of the figure.

The phase variation event is further affected by DNA-binding factors as H-NS (heat-stable nucleoid-structuring protein), IHF (Integration host-factor) and Lrp (Leucine-responsive protein) (18, 51, 57, 59, 94, 125, 129). The expression of type 1 fimbriae is regulated in response to environmental stress conditions such as high osmolarity, pH and temperature, and expression is induced upon entry into stationary phase (49, 57, 160). The environmental-responsive regulation is a consequence of both regulation of the expression of the recombinases and by directly influencing the phase variation event (Fig. 9).

As previously indicated, fimB and fimE are transcribed by their own separate promoters and several factors regulate the expression of these factors (Fig. 9). Interestingly, the number of factors described to affect expression of fimB is much higher than the numbers of factors affecting fimE. Transcription of fimB is regulated by the alarmone ppGpp, the alternative σ-factor RpoS, the heat-stable nucleoid-structuring protein H-NS, the sialic acid (Neu5Ac) regulator NanR and the GlcNAc-6P-responsive regulator NagC (1, 47, 49, 163, 164). The expression of fimE is orientation dependent, meaning that it is expressed when the fim switch is in the ON-orientation (102). In addition, fimE is also negatively regulated by the global regulator H-NS (128, 129).
Fig. 9. Summary of the factors involved in regulating the expression of type 1 fimbriae. The expression of type 1 fimbriae is regulated at two levels, promoter activity (main promoter located upstream of the \textit{fim} operon) or by phase variation (invertible region between arrowheads). The phase variation event is mediated by the recombinases FimB and FimE and further fine-tuned by the binding of accessory factors like Lrp and IHF. The expression of FimB and FimE is affected by several TFs and this in turn will have an effect on phase variation. There are several environmental conditions that have been suggested to influence the expression of the fimbriae.
4.2.3 Biofilm formation

A biofilm is an organized community of microorganisms (e.g. bacteria, archaea, protozoa, fungi and algae) covered by a protective and adhesive matrix. Biofilms are usually found on solid substrates but can also form floating mats on liquid surfaces. Biofilms are commonly found on rocks and pebbles at the bottom of most water-streams where they serve as an important component in the food-chains. In recent years, it has been shown that biofilms causes serious troubles in industry by clogging pipes and causing corrosion, in food industry and in water distribution systems. From a medical perspective, biofilms formed on catheters and implants might be the origin of severe diseases (75, 199, 212). Another relevant feature of the biofilms is the observed increased resistance to: antibiotics, environmental stress and host defenses (3, 48, 77). Bacteria living in a biofilm usually have different properties from free-floating (planktonic) bacteria of the same species, as has been shown by different approaches, including microarray analysis studies (46, 158).

![Fig. 10.](#) The different steps and factors involved in biofilm formation of *E. coli*. Adapted from (168, 194).

In *E. coli*, biofilm formation proceeds in a biological circle consisting of five steps (see Fig. 10):

1. **Initial adhesion** by planktonic (free-swimming) cells to a surface.
2. Transition to a irreversible state known as **surface attachment**.
3. **Microcolony formation**, early development of the biofilm consisting of a few layers of cells.
4. **Maturation** of the biofilm with the characteristic pedestal formations.
5. **Dispersion** of the cells into the surrounding environment i.e. return to the planktonic state.
The initiation of a biofilm is regulated by environmental signals as nutrients, temperature, osmolarity, pH, iron and oxygen (126, 168). Nutrient availability also influences the thickness of the biofilm and dispersion (168). Therefore, it can be proposed that the starvation response pathway plays an overall role in the control of biofilm formation.

Several surface structures of *E. coli* contribute in biofilm formation ((194) and Fig 10). Flagella and motility have been shown to promote cell-to-surface contact and promote the bacterial spread along surfaces and in dispersal from the biofilm (144). Type 1 fimbriae are critical for the initial stable cell-to-surface attachment in a range of different media and biofilm growing systems (15, 144, 148, 158). Ag43, a self-recognizing adhesin mediating auto-aggregation, induces microcolony formation (42, 97). Both type 1 fimbriae and Ag43 has been shown to play a role in the intracellular bacterial communities formed inside the uroepithelium during urinary tract infection (4, 207). However, expression is controlled so that co-expression is likely to be mutually exclusive (159). Curli are heteropolymeric proteinaceous filamentous appendages that influence adherence properties of biofilm-forming *E. coli* (127). Over-expression of the curli subunit, CsgA, resulted in a constitutive biofilm forming phenotype of the cells (202) with the development of a mature biofilm with mushroom shaped pillars and medium filled channels (147). The extracellular polysaccharide produced by *E. coli* necessary for the establishment of the complex three dimensional structure and depth of the biofilm (43).
Aims of this thesis

The main aim of my thesis work was to elucidate the role of small regulatory molecules in transcriptional regulation of genes in \textit{E. coli}, in particular in control of genes involved in colonization and survival of uropathogenic \textit{E. coli} strains.

Specific aims:

- To determine whether ppGpp regulatory networks are involved in biofilm formation and type 1 fimbriae expression (Paper I).
- To establish the regulatory mechanism of ppGpp-mediated regulation of type 1 fimbriae (Paper I).
- To resolve the molecular mechanism of how ppGpp stimulates the expression of the \textit{fimB} gene (Paper II).
- To investigate the participation of DksA in the ppGpp-mediated regulation of type 1 fimbriation (Paper II).
- To clarify the discrepancy observed between the ppGpp- and DksA-mediated regulation of \textit{fimB} (Paper II).

\textit{Science never solves a problem without creating ten more}  
\textit{George Bernard Shaw (1856-1950) Irish Dramatist}
Result and Discussion

In this section I present and discuss the main results obtained during my thesis work. Data presented in the Papers of the thesis are referred to by their roman letters (I-III) and corresponding figure numbers.

5. **Global effect of ppGpp and/or DksA on gene expression in *E. coli***

A commonly used strategy to elucidate whether a specific gene is under the control of a particular regulatory factor is to determine the expression profile in cells deficient for that regulatory factor (mutants). In order to analyze redistribution of the global gene expression patterns in *E. coli* mutants’ deficient for the global transcriptional regulators ppGpp and DksA, we used a transcriptomic approach. The previously described identical and opposing effects of ppGpp and DksA on gene expression (2, 110, 135) primed us to analyze and compare the difference in gene expression pattern in single ppGpp<sup>0</sup> and ΔdksA mutant strains as well as in the double ppGpp<sup>0</sup> ΔdksA mutant strain. Transcriptomic analyses of the effects of ppGpp on gene expression in *E. coli* and other species have been performed previously, but most of those studies were performed under growth conditions where the levels of ppGpp were induced by amino acid starvation (RelA-dependent induction) (24, 50, 53, 145). Our work presumably represents the first study of *E. coli* K-12 gene expression in a ppGpp<sup>0</sup> compared to a Wt strain under conditions commonly used to study general effects by ppGpp, i.e. entry into stationary phase after growth in LB-media. Moreover, the gene expression profile of DksA- and ppGpp/DksA-deficient strains was also analyzed.

We could observe an extensive rearrangement of gene expression in all three mutant strains (ppGpp<sup>0</sup>, ΔdksA and ppGpp<sup>0</sup> ΔdksA) as compared to the Wt strain. About 20 % of all ORFs represented in the microarray were found to be significantly altered in the different mutants. The analysis focused on ORFs whose mRNA levels were altered more than 2-fold in the mutant strains as compared to the Wt. Among these, about 7 % of the total numbers of ORFs were affected in the single mutant strains and about 13 % in the double mutant strain (Paper III, Fig. 1A-B). This is consistent with previous work by others obtained from transcriptomic studies analyzing the effect of ppGpp and DksA on gene expression in *E. coli* and *Salmonella* (50, 139, 161, 182). Interestingly, considering the known co-dependent effect of ppGpp and DksA on gene expression very few ORFs were found to be altered in all three mutant strains as can readily be visualized when the array data was plotted in Venn diagrams (Paper III, Fig. 1A-B, Fig. 3).
5.1 Effect of ppGpp and DksA on bacterial cell physiology

In order to get an overview of the role of ppGpp and DksA in bacterial cell physiology, the ORFs with >2-fold altered expression were organized into categories according to the function of the protein that they encode. From this analysis, it was evident that both ppGpp and DksA repress expression of ORFs that encode proteins involved in transcription, purine and pyrimidine synthesis, protein synthesis, protein fate and cell envelope biogenesis, while stimulating expression of ORFs that encode proteins involved in energy metabolism, transport/binding, central intermediary metabolism and amino acid biosynthesis (Paper III, Fig. 1C and Table 2). All these alterations in gene expression will result in a decrease of the rate of translation and change in energy metabolism, thereby preparing the cell for adaptation to the slower growth condition, and are in full agreement with previous reports describing the general effect of ppGpp and DksA on bacterial physiology (33, 50, 139, 161, 182).

Interestingly, in most of the functional groups, the number of ORFs affected in the double mutant strain (ppGpp\(^{0}\) ΔdksA) was higher than the number of ORFs altered in each single mutant strain (Paper III, Fig. 1C). To rationalize these observations, two explanations might be proposed. Firstly, some ORFs are slightly affected in the single mutant strains (less than 2-fold) but are synergistically affected by the absence of both factors. Secondly, ORFs that are independently regulated by ppGpp or DksA, by one but not the other, would also be found among the genes affected in the double mutant strain. When analyzing the expression levels of the different ORFs altered in the double mutant strain, we found ORFs whose expression pattern exemplifies both of these proposed explanations (Paper III, Table 3 and S1).

5.2 Effect of ppGpp and DksA on transcriptional repression and stimulation

To our surprise the diversity of expression patterns when looking at the different ORFs was large, suggesting a remarkable complexity of the effect of ppGpp and DksA on transcription. This was further exemplified when analyzing the number of up- versus down-regulated ORFs in the different mutant strains (Paper III, Fig. 3A-B). The result showed that; i) a larger number of genes were affected in the double mutant strain than the single mutants; ii) the number of ORFs affected in all three mutant backgrounds was very limited; and iii) an surprising number of ORFs were only affected in either the ppGpp- or DksA-deficient strains. A significantly higher number of ORFs were found to be up-regulated in the DksA-deficient strain as compared to the number of down-regulated ORFs. These results might suggest that DksA play a more prominent role in repression than in stimulation of gene expression. As will be
discussed later, up-regulation of gene expression in the DksA-deficient strain does not always imply that DksA has a repressing effect on transcription. It could also be explained by an increased availability of the secondary channel of RNAP due to the physical lack of DksA (Paper II, Fig. 8 and Paper III, Fig. 7). The differences in the numbers of up- and down-regulated ORFs and differential regulation of individual ORFs suggested that the mechanism of repression and stimulation might be diverse and that the requirement for ppGpp and DksA in these processes might be dissimilar.

6. Involvement of ppGpp in biofilm formation by commensal and pathogenic *E. coli*

The establishment of biofilm communities is a multi-factor process in which the commitment to grow as a biofilm can be instigated in response to several environmental inputs and physiological stresses (48, 93, 168). However, which regulatory factors are involved in transduction of these signals is not fully understood. Recently it was shown that uropathogenic *E. coli* form complex bacterial communities with biofilm-like traits during intracellular growth in umbrella cells of the bladder, and therefore that biofilm formation is likely a key feature during its pathogenesis (4, 90). The involvement of ppGpp in regulating biofilm formation and survival has previously been suggested for *E. coli* K-12 strains (10).

To study the involvement of ppGpp in biofilm formation, mutations of the two ppGpp synthetases, RelA and SpoT, were introduced into two uropathogenic *E. coli* isolates, J96 and 536. We found that the biofilm-forming abilities of J96 and 536 were severely affected in the absence of ppGpp (Paper I, Fig. 1B). Very little biofilm formation was observed in the ppGpp^0 derivatives. A similar result was observed when using a non-pathogenic K-12 strain (Paper I, Fig. 1C).

The next obvious question was at what level does the regulation occur? There are several factors important for the biofilm formation in *E. coli* (Fig. 10). By blocking type 1 fimbriae mediated adhesion by addition of mannosides, the biofilm formation of the UPEC strains and K-12 strain were diminished, illustrating the importance of the type 1 fimbriae on the biofilm formation of these strains (Fig. 10 and Paper I, Fig. 1B-C). In the presence of mannosides, no difference on biofilm formation was observed between the Wt and ppGpp^0 derivatives, suggesting that ppGpp regulates the expression of the type 1 fimbriae. This was corroborated by analyzing a *fimH* mutant strain (lacking type 1 adhesin), again no difference between the Wt and the ppGpp^0 derivatives were observed (Paper I, Fig. 1C). Furthermore, when phenotypically analyzing the expression of type 1 fimbriae in both UPEC and K-12 strains by mannose-sensitive yeast agglutination (MSYA), a distinct
difference was observed, illustrating that the ppGpp\textsuperscript{0} derivatives did not express any fimbriae on their surface.

6.1 ppGpp regulates the expression of type 1 fimbriae at the phase variation level

It was of interest to investigate at what level the type 1 fimbriae system was regulated by ppGpp. Transcription of the polycistronic fim operon, encoding for the synthesis of type 1 fimbriae, is regulated by both phase variation and by promoter activity (Figs 8-9). The phase variation is catalyzed by two site-specific recombinases, FimB and FimE. Type 1 fimbriae synthesis is regulated in response to environmental stress conditions such as high osmolarity, pH and temperature, and expression is induced upon entry into stationary phase (49, 57, 160). The environmental regulation is consequence of both regulation of the expression of the recombinases and by directly influencing the phase variation event (Fig. 9).

The effect of ppGpp on fim expression was determined by transcriptional studies using reporter fusions and mutants for ppGpp synthesis. A PCR-based method was used to determine the phase variation state of the fim promoter region. The transcriptional profile of fim\textsubscript{A}-lacZ (major subunit of the structural genes) showed maximal expression during entry into stationary phase (Paper I, Fig. 2A), where it has been reported that levels of ppGpp are highest during growth in LB-media (9, 185). Expression of the fim\textsubscript{A}-lacZ fusion was severely down-regulated in the ppGpp-deficient strain (Paper I, Fig. 2A). Dissecting the mechanism for ppGpp-regulation on expression of type 1 fimbriae revealed that it operates exclusively through modification of the phase variation state of the fim promoter region (Paper I, Fig. 3). First, in the absence of phase variation, ppGpp has very little effect on transcription from the fim promoter per se, but a phase variable strain that lacked ppGpp had >8-fold reduced expression of a fim\textsubscript{A}-lacZ reporter (Paper I, Fig. 2 and Table 1). Second, a phase variable strain lacking ppGpp had a reduced population of cells with the fim promoter in the ON orientation (ON-cell) as measured by a PCR based approach (Paper I, Fig. 3). Third, a concomitant increase in the number of ON-cells was observed by artificially increasing the ppGpp levels in the cell (Paper I, Fig. 4F).

We could trace the effect of ppGpp on phase variation to the expression of the fim\textsubscript{B} gene that encodes the OFF-to-ON FimB recombinase. The expression of fim\textsubscript{B} was >3-fold higher in ppGpp proficient strains as compared to ppGpp-deficient strains whereas the ON-to-OFF FimE recombinase was only slightly affected (Paper I, Table 2). Furthermore, increasing the intracellular levels of ppGpp had an immediate effect on fim\textsubscript{B} expression. In further support, artificial elevation of FimB levels bypasses the need for ppGpp in type 1 fimbriae expression and biofilm formation (Paper I, Fig. 4G and data not shown).
6.2 ppGpp regulation of factors important for biofilm formation

Type 1 fimbria is not the only surface structure that is important for biofilm formation in E. coli. Flagella, Ag43 and curli are others envelope structures that are important at different steps in the biofilm formation process (194) and Fig. 10.

Ag43 is a self-recognizing adhesin that causes bacterial cells aggregation in liquid culture when highly expressed (96, 131). Magnusson et al showed that the auto-aggregation phenotype of E. coli was decreased in ppGpp0 strains and furthermore, the Ag43 protein level was also reduced in this strain (110). Similarly, an effect of ppGpp on ag43 expression was observed in our transcriptomic study (data not shown), with a clear decrease of ag43 transcript in the ppGpp0 strain as compared to the Wt. Ag43, as for type 1 fimbriae, are regulated both by phase variation and at the promoter activity level by several regulators and environmental conditions (73, 190). Results in our research group indicate that phase variation of Ag43 is strongly affected by ppGpp (J.D. Cabrera et al, unpublished data).

The flagella system consists of numerous operons harbouring several open-reading frames. The operons are expressed in a temporal order and are divided in early, middle and late operons (36). The expression of the middle ORFs are regulated by the FlhDC regulators and σ70-RNAP holoenzyme while the late ORFs are regulated by FlgM anti-sigma factor and σ70(σFliA)-RNAP holoenzyme. Contradictory results of the effect of ppGpp on flagella expression have been published (50, 110). Magnusson et al also showed that the ppGpp-deficient strain was deficient in motility on plates (110). Conversely, Dufree et al showed that the flagella and chemotaxis genes were down-regulated in presence of high levels of ppGpp (50). They also suggested that ppGpp/DksA directly regulate expression of FlhDC (master regulators of middle genes). There are some major differences in the experimental set up’s that could in part be the explanation for the differences in the results obtained.

Our experimental results showed that ppGpp act as a positive regulator of the transcription of the genes involved in flagella biosynthesis and chemotaxis (Paper III, Fig. 5A). Furthermore, we could observe a clear down-regulation of FlIC (major subunit) and FliA (σF, regulating the late ORFs) and several other flagella ORFs in the ppGpp0 strains as compared to the Wt (Paper III, Figs 5-6). Our data indicated that ppGpp primarily regulates the early and middle operons (Paper III, Fig. 5), i.e. ORFs that are transcribed by σ70-RNAP holoenzyme. The studies performed by Dufree et al were performed during amino acid starvation conditions in mid-log phase of growth. Furthermore, Dufree et al did not use a ppGpp-deficient strain and the studies followed the expression during 30 minutes after inducing high levels of ppGpp by amino
acid starvation (50). It might be that under these conditions, some additional regulator is expressed resulting in the differential effects observed. It remains to be elucidated whether the effect we observe on flagella expression in the ppGpp-deficient strain is direct or indirect, and at what level it occurs.

Metabolic stresses and many environmental conditions that reduce bacterial growth are cues for rapid induction of the intracellular levels of ppGpp (33). During maturation of the intracellular bacterial communities of uropathogenic E. coli, it was described that after an initial fast growing phase, a significant decrease in the growth rate occurs with the concomitant appearance of biofilm traits (90). Our results provide a possible regulatory mechanism for the changes in the expression pattern during maturation of the intracellular E. coli communities. Under this scenario, reduced growth rate would presumably induce elevated levels of ppGpp, which in turn would stimulate the expression of type 1 fimbriae, Ag43 and flagella, and trigger the successful establishment of an intracellular biofilm-like community.

Nevertheless, knowing that all factors can not be expressed at the same time shows the importance of additional level of regulation of expression. Both the fim operon and ag43 are regulated by phase variation while flagella operons are temporally regulated. A reciprocal regulation of motility, adhesion and aggregation is logical and cross-regulation between the fimbrial gene products and both flagella genes and ag43 has been shown (27, 37, 72). This suggests that very rarely will there be a bacterial cell in a population expressing all three factors simultaneously. This illustrates one example where having a population of bacteria expressing the same genes/proteins is not always beneficial for the survival.

7. Divergent effects by ppGpp and DksA on type 1 fimbriae and flagella in vivo

During the analysis of the array data, a number of the ORFs that were differentially altered in at least two of the three mutant strains as compared to the Wt were identified (i.e. up-regulated in one and down-regulated in the other). About 10-20 % of the ORFs affected in the three mutant strains fall into this group that are differentially by ppGpp and DksA in vivo. These ORFs were grouped into different classes depending on their pattern of expression in the different mutant strains (Paper III, Fig. 4). There were very few ORFs that were repressed by ppGpp (i.e. up-regulated in the ppGpp- strain) and stimulated by DksA (i.e. down-regulated in the ΔdksA strain). Most of the differentially affected ORFs were activated by ppGpp (i.e. down-regulated in the ppGpp- strain) and repressed by DksA (i.e. up-regulated in the ΔdksA strain) (Paper III, Fig. 4). These results suggest that the concerted action by ppGpp and DksA seems to depend on whether a promoter is positively or negatively regulated by ppGpp. Many of the ORFs differentially regulated by
ppGpp and DksA (Paper III, Fig. 4B) encode proteins involved in type 1 fimbriae expression, flagella biosynthesis and chemotaxis. The differential effect of ppGpp and DksA on these genetic determinants was further analyzed.

7.1 Effect of DksA-deficiency on type 1 fimbriae and flagella expression

The observed effect of ppGpp on biofilm formation and the discovery of DksA as a possible co-regulator of ppGpp-mediated regulation prompt us to analyze the involvement of DksA in biofilm formation. In contrast to the ppGpp\textsuperscript{0} strain, the \(\Delta dksA\) strain showed an increase in biofilm formation that could effectively be abolished by the addition of mannosides (Paper II, Fig. 3C). This suggests that the increase in biofilm formation was due to elevated levels of type 1 fimbriae in the \(\Delta dksA\) strain (Paper II, Fig. 3C). This was further corroborated by analyzing the MSYA phenotype of the mutant strain which showed elevated levels of type 1 fimbriae on the surface of the \(\Delta dksA\) mutant. The findings were also corroborated by studying the \textit{fimA-lacZ} reporter fusion. We concluded that in the \(\Delta dksA\) strain there is an increase in type 1 fimbriae expression by increasing the percentage of ON-cells (Paper II, Fig. 3A-B). When dissecting the mechanism of hyperfimbriation in the DksA-deficient strains it was revealed that the observed effect of DksA on the phase variation occurred by enhanced transcription of the \textit{fimB} gene (Paper II, Fig. 3). Recently, the negative effect of DksA on type 1 fimbriation in \textit{E. coli} was also evidenced by MSYA and electron microscopy by Magnusson \textit{et al} (110).

Similar to the type 1 fimbriae expression, genes of the flagella operons were also shown to be differentially regulated by ppGpp and DksA in the array data (Paper III, Fig. 5). The effect of DksA was confirmed by analyzing the protein levels of FlIC (major subunit) and FlIA (\(\sigma^F\), regulator). A clear up-regulation of both proteins could be observed in the DksA-deficient strain (Paper III, Fig. 6). Furthermore, many of the ORFs important for the chemotaxis response were also up-regulated in the DksA-deficient strain. We tested the chemotactic response of the three mutant strains to L-serine, maltose and glucose, as compared to the Wt strain (Paper III, Fig. 5B). The DksA-deficient strains clearly showed an increased chemotactic response as compared to the Wt, whereas the ppGpp\textsuperscript{0} strain did not respond at all. At what level ppGpp and DksA affect the expression of the flagella and chemotaxis operons is still not known.
8. Dissecting the mechanism of regulation of 
*fimB* transcription by ppGpp and DksA

In order to elucidate the molecular mechanism behind the differential regulation of ppGpp and DksA on the transcription of *fimB*, we dissected the independent and co-dependent effects that these factors have on *fimB* expression. It has been suggested that *fimB* transcription is controlled by three promoters and that the expression is influenced by several regulatory factors (Fig. 9), i.e. the alternative σ-factor RpoS, the heat-stable nucleoid-structuring protein H-NS, the sialic acid (Neu5Ac) regulator NanR and the GlcNAc-6P-responsive regulator NagC (47, 49, 163, 164). Our results illustrate that *fimB* is mainly transcribed from the P2 promoter (Paper I, Fig. S1) and primer extension analysis showed that both ppGpp and DksA affect transcription from the predominant P2 promoter both in the commensal and pathogenic strain (Paper I, Fig. 4D-E and Paper II, Fig. 5C). Additionally, the effect of ppGpp and DksA was determined in mutants of the previously known regulators. However, none of those factors were found to be involved in the ppGpp- or DksA-mediated regulation of *fimB* expression (Paper I, Table 2 and Paper II, Table 2). Together, these results suggested that the effect by ppGpp and DksA could occur either directly at the *fimB* promoter or through a so far unknown factor.

8.1 Direct effects of ppGpp and DksA on transcription from the *fimB* P2 promoter

To investigate the possible direct effects of ppGpp and DksA on *fimB* transcription, DNA templates spanning the *fimB* P2 promoter were tested in a reconstituted *in vitro* transcription system. *In vitro* transcription is an approach used to determine direct effects of regulatory factors on transcription. In this assay, transcription is quantified after combining the specific template DNA with RNAP holoenzyme and the regulators to be tested. By varying the concentration of the components and the timing or the order of addition of the different components it is possible to determine regulation at different steps.

To elucidate whether ppGpp and DksA have a direct effect on *fimB* transcription, multiple round *in vitro* transcription assays were performed in the presence of σ70-RNAP holoenzyme and increasing concentrations of ppGpp or DksA. Surprisingly, transcription from the *fimB* P2 promoter was stimulated by the presence of either factor separately (Paper II, Fig. 6A-B) and an additive effect on *fimB* P2 transcription could be observed in the presence of both factors (Paper II, Fig. 6C). These results contrast the observed negative effect of DksA *in vivo* and the proposed concerted action of ppGpp and DksA on transcription (Paper II, Fig. 1C).
To further test the idea that ppGpp can stimulate fimB transcription independently of DksA in vivo, ppGpp levels were increased by inducing a plasmid expressing a truncated version of RelA. This allele do not require the association to a stalled ribosome to induce the synthesis of ppGpp (177). The expression of fimB was increased both in presence and absence of DksA as measured by the expression of a chromosomal fimB-lacZ allele in a ppGpp\textsuperscript{0} strain (Paper II, Fig. 8C). In agreement with the in vitro data, the increase was not as extensive in the absence of DksA, indicating that for full expression of fimB both ppGpp and DksA are required, although ppGpp can independently stimulate transcription both in vivo and in vitro. The in vitro and in vivo data suggested that ppGpp and DksA have both independent and co-dependent effects on fimB transcription.

8.2 Co-dependent versus independent effects of ppGpp and DksA on fimB transcription

The apparent co-dependent and independent effects of ppGpp and DksA on transcription of fimB in vivo and in vitro prompted us to try and dissect these different effects in vitro. This involved analysis of the effects of ppGpp and DksA on the association of RNAP to the promoter DNA, on open-complex stability, and on the rate of formation of full-length transcript from the fimB P2 promoter.

DksA has been shown to facilitate the binding of RNAP to promoter DNA in vitro (139, 143). To explore if such a mechanism could explain the stimulatory effect of DksA on fimB transcription, we performed gel-shift assays with a DNA template containing the fimB P2 promoter, a constant amount of RNAP and increasing concentrations of DksA (Paper II, Fig. 7A). DksA dramatically enhanced the ability of RNAP to bind to the fimB P2 promoter. On the other hand, ppGpp did not show any effect on the binding of RNAP to the promoter DNA (Paper II, Fig. S1).

The molecular mechanism(s) by which ppGpp and DksA have their co-dependent effects on transcription from σ\textsuperscript{70}-promoters are still not very well known. One feature commonly studied is the concerted action of ppGpp and DksA on destabilizing open-complexes promoters (134, 135). For promoters with very unstable open-complexes (e.g. rrn) this destabilization will result in repression of transcription, while for promoters with very stable open-complexes this destabilization might help RNAP escape the promoter and consequently promote productive transcription (14). The fimB P2 promoter forms very stable complexes (Paper II, Fig. 7B) and, ppGpp and DksA did not show any effect independently on the open-complex stability. However, together they decrease the half-life from >200 min to ~30 min (Paper II, Fig. 7B).
The rate of formation of full-length transcripts was tested in an in vitro transcription assay, since it has been postulated that ppGpp might have an effect on abortive initiation, promoter escape and transcriptional pausing (14, 22, 101). Interestingly, ppGpp showed a repressive effect on the rate of formation of full-length transcripts both in presence and absence of DksA (Paper II, Fig. 7C-D, 8D-E). Despite the repression of the rate of formation of full-length transcripts, the maximal amount of transcript from the fimB P2 promoter was increased in the presence of ppGpp and even further increased when both regulatory molecules were provided simultaneously.

Taken together, these in vitro experiments illustrated that ppGpp and DksA have independent and co-dependent effects on multiple steps of transcription from the fimB promoter, namely i) a ppGpp-independent effect of DksA on RNAP-DNA complex formation, ii) co-dependent effects on open-complex stability and iii) a DksA-independent effect of ppGpp on the rate of accumulation of full-length transcripts, which might be consequence of alterations in promoter escape, abortive initiation or elongation.

Notably, in no instance was DksA found to have a negative effect on fimB transcription in vitro. These seemingly contradictory results for the role of DksA on fimB transcription in vitro and in vivo could readily be explained if another factor(s) are operational in absence of DksA in vivo.

9. Involvement of Gre-factors in differential regulation by ppGpp and DksA in vivo

A possible explanation for the paradoxical in vitro versus in vivo discrepancy of the role of DksA in transcription from the fimB promoter might be that in the absence of DksA, the regulatory activity of some additional factor(s) is more efficient. DksA is a structural homologue to GreA and GreB, and all three proteins exert their regulatory effects through the secondary channel of the RNAP (19, 130, 137, 169, 178). The Gre-factors can reduce abortive initiation, facilitate promoter escape, suppress elongation pausing and arrest (19, 54, 80). We reasoned that the lack of DksA may facilitate interaction of the Gre-factors with the RNAP, which in turn may specifically affect the expression of some genes.

9.1 Gre-factors stimulate fimB transcription in vivo and in vitro

In order to determine if the Gre-factors could be responsible for the increased expression of fimB in the absence of DksA in vivo, we analyzed mutants for the Gre-factors in E. coli, over-expressed the Gre-proteins in vivo and performed in vitro transcription assays with purified GreA and GreB-His proteins.
Mutations of the \textit{greA} and \textit{greB} were introduced into ppGpp- and/or DksA-deficient strains harbouring a \textit{fimB-lacZ} reporter. Loss of GreA and GreB had no effect in the DksA-proficient strains, however, in the absence of DksA the loss of the Gre-factors counter-balance the observed increase of \textit{fimB} expression (Paper II, Fig. 8A). Similarly, when over-expressing the Gre-factors in presence and absence of DksA a significant increase on \textit{fimB} expression could be observed in the absence of DksA, independently of the presence or absence of ppGpp (Paper II, Fig. 8B).

The \textit{in vivo} data suggest that the Gre-factors can stimulate the expression of \textit{fimB} transcription in the absence of DksA. Therefore, we tested if the Gre-factor could relieve the ppGpp-mediated delay of the production of full-length transcripts, since the Gre-factors have been suggested to have an effect on abortive initiation, promoter escape, elongation pausing and arrest (19, 54, 80). Addition of the Gre-factors to the \textit{in vitro} transcription reaction relieved the ppGpp-mediated delay, both in the presence and absence of DksA suggesting that the Gre-factors regulate \textit{fimB} transcription directly by affecting the later stages of transcription initiation and/or elongation (Paper II, Fig. 8D). The Gre-factors additionally increased the maximal amount of full-length transcript produced from the \textit{fimB P2} promoter. The absolute maximal levels of transcripts were observed in the presence of all four factors, i.e. ppGpp, DksA, GreA and GreB (Paper II, Fig. 8E). This net effect on transcription is most likely a combination of all their independent and co-dependent effects on the different steps of transcription initiation (Fig. 11).

9.2 Gre-factors are important for FliC expression \textit{in vivo}

To analyze if the expression profiles of the flagella and chemotaxis operons observed in the DksA-deficient strain might, similarly to the \textit{fim} operon, be a consequence of the increased association of Gre-factors with the RNAP, the effect of \textit{gre} mutations on the expression of the flagella expression was studied. The amount of FliC (flagella major subunit) was determined by Western blot analysis of the different mutant strains (Paper III, Fig. 7). Intriguingly, in the \textit{ΔgreA ΔgreB} and the \textit{ΔdksA ΔgreA} strains the amount of FliC was severely down-regulated as compared to the Wt and \textit{ΔdksA} strains, suggesting the Gre-factors play an important role in the FliC expression even in the presence of DksA \textit{in vivo}. Whether this effect occurs directly at the \textit{fliC} promoter or indirectly by regulating the expression of another factor are speculations that will require further experimentation to be corroborated.

The effect of the Gre-factors has also been determined for the \textit{argI} transcription (data not shown), a gene that is positively regulated by both ppGpp and DksA (Paper II, Table 1). In this case, no effect of the Gre-factors on the expression of this gene was observed either in the presence or in the
absence of DksA. This suggests that the Gre-factors do not always alter the expression of promoters that are under control of DksA.

9.3 Competition for the secondary channel of RNAP

Notably, our studies suggest that the Gre-factors might have different role in regulation of ppGpp and/or DksA dependent genes in vivo depending on the promoter. For fimB transcription, the Gre-factors only have an effect in vivo in the absence of DksA (Fig. 11 and Paper II), for FliC the Gre-factors affect the expression even in the presence of DksA (Paper III), while for argl the Gre-factors are dispensable (Paper II, data not shown).

The intracellular concentrations of DksA, GreA and GreB, and their predicted affinities for the RNAP has suggest that most RNAP would be occupied by DksA rather than GreA and GreB (154). However, whether this scenario is true for all growth conditions remain to be examined. Data has been presented indicating that Gfh1 (belonging to the Gre-family) exhibits a pH-dependent switch between an active and inactive conformation (105). If this is also the case for the other Gre-family proteins needs to be further studied. I consider that our data suggest that there can be conditions where the affinity of the secondary channel factors could change to override their relative abundance. It is also plausible that interactions of the RNAP to a specific promoter sequence or with a particular σ-factor might alter the conformation of the secondary channel of the RNAP thereby changing the affinity of the RNAP for the different secondary channel TFs.

Fig. 11. Regulatory network controlling transcriptional regulation of the fimB promoter. Schematic illustration of the fimB regulatory region is shown with DNA-binding TFs H-NS, NanR and NagC depicted below the DNA and non-DNA-binding regulatory molecules that directly target RNAP (αββ'ωσ) depicted above the DNA; ppGpp as dots, DksA as red circles, GreA as orange diamonds and GreB as red diamonds. Stimulatory effects on transcription from the fimB promoter are exemplified as arrow-heads whereas repressive effects are indicated by dash-ended lines. Note, while DksA stimulates transcription, its absence would allow greater access of the Gre-factors to RNAP, which would also have stimulatory effects. Adapted from (Paper II, Fig. 9).
Concluding Remarks

The data presented in this thesis highlights new insights into gene regulation by ppGpp and DksA. Early studies have mostly implicated ppGpp in regulating house-keeping genes that will slow-down the translation machinery during amino acid starvation conditions (stringent response). A growing number of studies have showed that intracellular levels of ppGpp are likely to be induced in bacteria in response to any environmental stress condition that results in slow proliferation. Furthermore, it is now established that ppGpp also plays a central role in regulating genes involved in virulence and survival in the face of host induced stress.

For many years it was a struggle to corroborate direct effects of ppGpp on gene regulation in vitro. Although, genetic, cross-linking, and co-crystallization studies suggested that ppGpp interacts directly with the $\beta$- and $\beta'$-subunits of the RNAP within the active site cleft, very little or no effect of ppGpp could be observed in vitro compared to the major effects observed in vivo. It was suggested that DksA was the missing co-factor for ppGpp-dependent regulation. However, as shown in this thesis as well as by others, there are a number of genes regulated only by one of these factors, strongly suggesting that concerted action by ppGpp and DksA is not true for all genes regulated by ppGpp.

The discovery of the involvement of the Gre-factors in the ppGpp-mediated regulation introduces a new perspective to the importance of the secondary channel of the RNAP in gene regulation. It also illustrates a new dynamic view of the RNAP conformation that might change by association to different transcription factors consequently altering expression from certain susceptible promoters. It is still not known if the interaction of one factor with RNAP can affect the binding of another, or for how long one factor remains associated with RNAP. Nevertheless, it is plausible to speculate that there might be an active exchange of RNAP binding factors throughout the transcription process.

I propose that we are moving towards a new model of ppGpp-mediated regulation, which is far from as fully elucidated as one might hope after almost 40 years of research.

*Ju mer man tänker, ju mer inser man att det inte finns något enkelt svar*

Nalle Puh/ A.A. Milne

48
Acknowledgements

First, making sure that I will not forget anyone, I would like to say thanks to everyone that have supported and helped me during these 6 years of PhD studies at the Department of Molecular Biology, both within and outside of work.

Enormous gratitude goes to my “never-ending” smiling supervisor Carlos (the man with the voice), without you I would never have managed to get this far, your huge support and sometimes crazy ideas have made working with you a pleasure, I also want to thank you for being a good friend also outside work. To the people in Microbiology department in Barcelona: Juanda, Jorge (the two Spanish guys that have survived long visits in Umeå, thank you for the EM pictures that made the cover of my thesis look beautiful), Núria, Aitziber, Christina (thanks to you 3 for all the help with the microarray analysis), Mari (most happy and crazy person I have ever met) and the others; big thanks to you for giving me a nice time during my trips to Spain.

In the department I would like to start to say thanks to Vicky for your tremendous help and support, for always trying your best to answer all of my questions and teaching me so much, thanks to you I finally understood what all of my result meant. Big thanks also go to your group members who have taken me in as a “member”; to Linda and Pelle (kommer aldrig glömma våra champagne-kvällar), Lisandro (for being such a good friend), Ellinor and the others. I want to thank past and present members of the Friday journal club and people whom I have shared lab with during these years; Bernt Eric (my second supervisor) and Sun, thank you for allowing me to participate in many activities and the nice gatherings at your house; to Annika (for your wonderful book-reviews and stories of your night-dreams and for being who you are), Stina (for your smile), Jurate (for the nice dinners and “pyssel” at your place), Claudia (for following me to learn Salsa, among other things) and Monica, thank you for the trips and all fun stuff that we have been doing inside and outside the lab, you all mean a lot to me; Emma (for being my training partner and a good friend), Anna R (for bringing a breath of fresh air to the lab), Connie, Hyun-Sook, Barbro, Karolis, Tianyan, Takahiko, Soni, Pramod, Berit (for still sharing office with me), Jörgen, Edmund (maybe one day you will hear me sing), Jonas, Jonny, Teresa, Sakura etc. To the people in SvB group; Sven, Christer, Pär, Marie, Marija, Yngve, Ingela, Lelle, Betty, etc., for good times on trips to Germany and Japan, and for all the fun things that would never had happened at the department without you. I want thank the people of the “bunker/beer corner”; Jenny, Maria N, Petra E, Sofia, Sara C, Micke, Stefan, Jocke, Giovanni, Sara R, etc; and to the secretaries Marita, Berit, Helena and Ethel, to Jonny, to Siv, to Media, and to Marianella, without you the department would not function properly.
Jag vill också tacka mina släktingar och mina vänner som stöttat mig och försökt förstå vad det egentligen är jag har pysslat med på dagarna under dessa år; Petra och Helen med familjer, vi har hållt ihop länge nu och det kommer blir många år till, TACK för att ni finns. Alla er i Sävar med familjer; Jesper, Anette och Björn, Susanne och Hans, Sussie, vi har haft mycket roligt ihop under åren. Kusin Maria, du är underbar, allt stöd du har gett under det senaste året är helt otroligt, jag lovar att det kommer vara mera glada smilisar hädanefter ☺☺☺.

Dessan, min vän, du kommer för alltid stå mig otroligt nära, ditt stöd och din vänskap har hjälpt mig otroligt mycket genom åren, mer än jag kan beskriva i ord. TACK och åter TACK för att du finns, ÅLSKAR DIG, och din underbara familj, Maja (vi ska äta popcorn snart), Filip (längtar att få se dig gå) och Anders.

Jag vill avsluta med att tacka min familj så otroligt mycket, Mamma, ”Storebror” Robert och Pappa (saknar dig, hoppas du har det bra där uppe), ni betyder allt för mig, utan er skulle jag inte vara någonting, ÅLSKAR ER.
References


