Mechanism of promoter recognition and transcription regulation by the $\sigma^S$ subunit of *Escherichia coli* RNA polymerase

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZÜRICH
for the degree of
DOCTOR OF SCIENCES

Presented by
STEPHAN LACOUR
Dipl. Bordeaux University
Born April 4, 1973
French Citizen

Accepted on the recommendation of
Pr. A. J.B. Zehnder, examiner
Dr. P. Landini, co-examiner
Dr. A. Kolb, co-examiner
Pr. H.C. Davies, chairman

Zürich, 2003
In nature, bacteria have to deal with changing growth environments and must face long periods without substrates. As a result, cells remain over long periods of time in stationary phase physiology, living on limited nutrient resources and often accumulating toxic metabolites. Several “survival” genes of stationary phase physiology can be induced for adaptation or maintenance, among them rpoS. The rpoS gene encodes the σ^S protein, an alternative sigma subunit of RNA polymerase widely conserved at least among gamma proteobacteria. The sigma subunits allow RNA polymerase (RNAP) recognition of specific sets of promoters. σ^S controls many stress responses, including adaptation to starvation, osmotic challenge, DNA repair and production of extracellular structures important for adhesion and virulence behaviours. However, despite its key role in the regulation of a broad number of genes, little is known about the mechanism of promoter recognition and transcription initiation by σ^S-RNAP (Eσ^S) or about promoter sequence determinants for σ^S recognition. Indeed, the consensus sequence CTATACT proposed for σ^S-dependent genes in Escherichia coli, does not strongly differ from the optimal recognition sequence TATAAT for the vegetative sigma factor σ^70, responsible of the expression of essential genes. Thus, the C nucleotide (−13C) placed immediately upstream of the consensus hexamer (TATACT) for the −10 promoter element is the only sequence feature specific for promoter recognition by Eσ^S. In addition to the −13C feature, there is evidence that promoter recognition by either Eσ^70 or Eσ^S can be modulated by factors other than DNA sequence (e.g. DNA curvature, additional DNA-binding proteins).

The aim of this project was to understand the requirements for promoter recognition and transcription initiation by the Eσ^S form of RNAP. The first objective of this work was to identify promoter determinants involved in specific recognition by the holoenzyme Eσ^S. To achieve this goal, we performed site-directed mutagenesis of a σ^S-dependent model promoter. Our promoter of choice was the E. coli aidB promoter (PaidB), a “bona fide” rpoS-dependent promoter as it was shown to be selectively transcribed by Eσ^S in vitro. PaidB is also an interesting model since it displays three
levels of expression: repressed (by Lrp), induced (by RpoS, in stationary phase) or activated (by Ada, in presence of DNA alkylating agents). Thus, PaidB is a suitable model to study both factor-dependent and independent transcription. In exponential growth conditions, PaidB is poorly transcribed by the Eσ\(^{70}\) form of RNA polymerase. However, exposure to a DNA alkylating agents, such as methyl methane sulfonate (MMS), results in transcription activation of aidB by the Ada regulatory protein as part of the adaptive response to DNA-alkylation. Ada activates aidB transcription by the Eσ\(^{70}\) form of RNAP and can also further increase the efficiency of transcription by Eσ\(^{8}\).

As a preliminary work, we tried to understand why Eσ\(^{70}\) is limited in transcription initiation at PaidB in absence of trans-acting factors such as the Ada protein. We found that, although Eσ\(^{70}\) can bind the aidB promoter, it is not able to form an open complex efficient in transcription initiation. We show that Eσ\(^{70}\) does not induce a sufficient DNA melting around the -10 region to allow transcription initiation. We verify that the -13C is an important feature for transcription by Eσ\(^{8}\) but we found that -13C is not essential for σ\(^{8}\) activity in the PaidB context and gives little contribution to promoter selectivity.

Conflicting evidence exists on the role, for Eσ\(^{8}\)-mediated transcription, of upstream promoter elements, the so-called “UP element”. The UP element is found upstream of the -35 box at many bacterial promoters where it functions as a binding site for the α subunit of the RNA polymerase (α\(_{2}\)ββ'ω) and enhances transcription by recruiting the polymerase to the promoter independently or with the contribution of additional trans-acting factors. At the aidB promoter, a proximal UP element extends from -38 to -60 and overlaps the binding site for the Ada activator. Comparing aidB expression in a wild type, an ada and an rpoS background strains we could show that the UP element is a dispensable promoter element for σ\(^{8}\)-mediated transcription of PaidB. Surprisingly, despite the fact that cells were grown in the absence of exogenous DNA-alkylators, we found that disruption of the UP element as well as substitution of two crucial nucleotides (GC) in the Ada binding-site affect aidB transcription in vivo in an rpoS-independent but Ada-dependent fashion. Thus, our results suggest that Ada production is induced at the onset of stationary phase cells grown in rich medium, probably by endogenously produced DNA alkylating compounds. This slight
accumulation of the Ada regulator leads to weak but detectable expression of the aidB promoter in stationary phase even in an rpoS deficient strain.

We also investigated the role of possible additional regulators at the aidB promoter. Interestingly, the general regulator H-NS appeared to positively regulate aidB expression; the Crl protein, a positive modulator at some rpoS-dependent promoters, also stimulates transcription from PaidB.

Since the upstream promoter elements do not appear to strongly influence rpoS-dependent aidB transcription, we performed an extensive mutational analysis of the −10 region of the aidB promoter. We identified different promoter features important for activator-independent transcription by Es[]. We show that the presence of the −13C promoter element, proposed to be specific for Es[], indeed favours open complex formation by this form of RNAP. A so-called "TG motif", upstream of the −10 promoter element, was also found to be important for optimal σS-dependent transcription. We show that, in contrast to σ70, σS is able to recognise and use a TG motif at different locations from −16 to −14 position. Finally, we demonstrated that the cytosine at −12 position (−12C), in place of a T nucleotide canonical for σ70-dependent promoters (tATAAT), is the major determinant for σS-selectivity at PaidB. The presence of a −12C does not allow Es70 to form an open complex proficient in transcription initiation, ability that could be restored by a C to T substitution (change to the canonical −12T for Es70). In contrast, EsS was able to bind, open and efficiently initiate transcription independently of the nature (C or T) of the −12 base. As a consequence, we proposed the motif TG(N)CYATACT (Y= T or C; (N)= any or no nucleotide) as new consensus −10 sequence for σS promoter recognition and the motif TG(N)CCATAACT for stringent rpoS-dependent promoters.

We verified the presence of the specific features for EsS recognition (−13C, −12C and TG motif) among known σS-dependent promoters and assessed the frequency of appearance of the proposed pattern upstream of coding regions of the E. coli chromosome (putative promoters). We could conclude that σS-promoter-features identified at PaidB are common to a larger set of rpoS-dependent promoters.

To confirm our findings and to identify new members of the σS-regulon, we compared the expression profile at the onset of stationary phase (batch culture in LB
medium) of the strain MG1655 and its $rpoS$ derivative, using the gene array technology. Promoters of genes differently expressed in the microarray were determined by primer extension and the conservation of the proposed promoter elements assessed. A large number of $rpoS$-dependent genes newly identified through this approach have unknown function, thus suggesting that a significant part of the $rpoS$ regulon must yet be characterised. From the analysis of the promoter sequences we could conclude that presence of a $-12C$ nucleotide occurs with high frequency and could therefore be considered as a characteristic promoter feature for a subclass of adaptation genes expressed in early stationary phase in rich medium. We propose a different physiological role for the genes belonging to the stringent and non stringent $rpoS$-dependent promoters.

The consensus sequence here proposed can be used as a tool for the identification of $rpoS$-dependent genes also in other bacteria, especially Gram negative $\gamma$-Proteobacteria. Such motif could be employed for the design of promoters to be used in biotechnological applications for instance to achieve gene expression at the onset of stationary phase.