UNIVERSITY OF TURKU

Transcriptional regulation in *Spirochaeta africana*

Verneri Nissilä

Master's thesis

Molecular biosciences (Biochemistry)

Department of Life Technologies

18 August, 2023

Turku

Master's thesis

Subject: Molecular biosciences (Biochemistry)

Author: Verneri Nissilä

Title: Transcriptional regulation in *Spirochaeta africana*

Supervisor: Georgi Belogurov and Vilma Trapp

Number of pages: 74

Date: 18 August, 2023

The genome sets the limits of how far the cells can differentiate or adapt; the extent to which either actually happens is the territory of transcription. The molecular mechanisms of transcription are well known for *E. coli* but recent data have shown that possibly important differences exist between the transcription machineries of bacteria. In this work, a hitherto unexplored transcriptional apparatus was investigated. Specifically, the *Spirochaeta africana* (*sfc*) RNAP and associated proteins were expressed heterogeneously and purified, and transcription templates were constructed that included a Broccoli sequence nested downstream from either the *sfc* rRNA promoter or the consensual Gre promoter. Transcription reactions were assembled *in vitro* and supplied with the fluorogen DFHBI-1T whose complex with the Broccoli transcript is fluorescent. The fluorescence from these assembled reactions was followed between 20-620 seconds post-initiation. The effect of two transcription factors (CarD and DksA) and a small molecule (guanosine tetraphosphate) were explored in different reactions. Factor independent transcription was stronger from the metabolic promoter (140% difference); CarD increased transcription from the rRNA promoter but decreased it from the metabolic promoter (+190% vs -40%). DksA decreased transcription from the rRNA promoter by 15% while no effect was verified for the metabolic promoter. With both promoters, equimolar amounts of CarD and DksA produced effects that were closer to the effect of CarD than that of DksA (total effects +155% vs -57%). The modulatory effect of ppGpp on CarD and DksA was studied in the context of the metabolic promoter; the effect of CarD was noted to be accentuated (+8%). Together, these data indicate three things. Firstly, the studied rRNA promoter is rate-limited by the formation of an open complex and that the metabolic promoters open complex is more stable. Secondly, an indication was found that the stress regulation of rRNA promoters in *S. africana* does not necessarily rely on DksA as in *E. coli* but perhaps CarD; similarity or the lack of it of the mechanism to that in play in other CarD-encoding bacteria, like *Mycobacterium tuberculosis*, remains to be verified. Finally, there might be a CarD-ppGpp axis that together is stronger than either of the components alone.

Keywords: transcription, biophysics, transcriptional regulation, spirochaetes

CONTENTS

Abbreviations

D276	aspartic acid at position 276
DFHBI-1T	(*Z*)-4-(3,5-difluoro-4-hydroxybenzylidene)-1,2-dimethyl-1H-imidazol-5(4H)-one
E	RNA polymerase core enzyme
EC	Elongation complex
eco	*Escherichia coli*
Eσ	RNA polymerase holoenzyme complex
FLAP	fluorescent light-up aptamer
NCR	Nonconserved region
ppGpp	guanosine tetraphosphate
QS	quorum sensing
r-protein	ribosomal protein
R599	arginine at position 599
RNAP	RNA polymerase
RP_c	closed promoter complex
RP_{itc}	initial transcribing complex
RP_o	open promoter complex
rRNA	ribosomal RNA
sc	*Streptomyces coelicolor*
sfc	*Spirochaeta africana*
S_N2	bi-molecular nucleophilic substitution
TSS	transcription start site
UP element	upstream element
$\alpha_2\beta\beta'\omega$	strings of this type refer to the RNAP subunit makeup
αCTD	C-terminal domain of the α subunit
αNTD	N-terminal domain of the α subunit
σ^{70}	sigma subunit with a molecular weight of 70 kDa
$\sigma_{1.1}$	subregion 1.1 of the sigma subunit

I INTRODUCTION

This introduction examines the structural biology that underlies transcription in bacteria. An attempt is made to represent information in a way that would allow the reader to comprehend some of the chief layers of transcriptional regulation in bacterial cells; this perspective is later much needed in the evaluation of the results of the experimental part of this work.

In general, four levels of regulation are represented:

In the first chapter, the overall structure of the RNA polymerase (that is, the enzyme catalysing transcription) is presented. While the chemistry of nucleotide polymerisation is conserved across all life, functionally important differences have been identified in bacterial RNA polymerases. In the second chapter, the specificity factors are reviewed. It is shown that specificity in transcription is due to a dissociable σ subunit that guides the core polymerase to the correct genomic location; the structural elements responsible for these interactions are reviewed. In the third chapter, the structures cognate to the sigmas' structures are reviewed. Emphasis is placed on showing that it is the sum of the promoter elements and the holoenzyme that together determine the kinetics of transcription from a DNA sequence. In the fourth chapter, proteins that influence the different kinetic steps, either promoting or inhibiting them, are reviewed to show how the cells utilise different mechanisms to rearrange the relative firing rates from different promoters.

1 The RNA polymerase core enzyme: structure and catalysis

The process of DNA-templated biosynthesis of RNA, known as transcription, is one of the central processes in the central dogma of molecular biology (Crick, 1970). Transcription, a complicated and tightly regulated process, is catalysed by the enzyme RNA polymerase (RNAP, or simply E); in the case of all cellular life forms, the RNAP is composed of multiple functional subunits. The bacterial polymerase, specifically, is composed of four functionally indispensable polypeptides $\alpha_2\beta\beta'$ and the somewhat less important ω subunit (Murakami, 2015). In concert with the RNAP enzyme itself, several auxiliary or accessory proteins are found; the most important of them is the dissociable σ subunit that confers the RNAP its capacity to initiate transcription from carefully regulated regions in the DNA, promoters. Further auxiliary proteins regulate the kinetics of transcription, providing the cells with a way to reorganise the preferred transcription strengths from different promoters. Analysis of the polymerases' global structure and movement during the catalytic cycle suggests that the polymerases across the spectrum of life are conserved (Cramer, 2002). However, while the polymerases share common fundamental

biochemistry and approximate steps in their movements, several factors contribute to transcriptions functional diversity and adaptability.

1.1 RNA polymerase complex catalyses transcription

The RNAP is a large enzyme with a distinctive, evolutionarily preserved three dimensional structure. Specifically, the *Thermus aquaticus* enzyme is 150 ånströms in length, 110 in width and 115 in height (Zhang *et al.*, 1999), and features pronounced crab claw-resembling protrusions. The active cleft of the polymerase is buried at the bottom of a cleft (measuring 70 ångstöms in depth and over 100 in width in another enzyme studied; Saecker *et al.*, 2012). The active site cleft is positively charged on the inside but negatively so on the outside (Nudler *et al.*, 2009). The outer surface is, under some experimental conditions at least, coated with up to 250 magnesium ions (Vassylyev *et al.,* 2002). Additionally, the conserved residues from the β and β' subunits chelate two catalytically necessary magnesium ions. This general architecture of the core enzyme is conserved across the multisubunit polymerases. However, studies from different species have revealed differences in both the basal choreography and the regulatory mechanisms of transcription (Chen *et al.*, 2021).

1.2 The catalytically necessary core components have conserved functions

As mentioned, the polymerase is composed of multiple subunits, with the most frequent composition of the core enzyme being $\alpha_2\beta\beta'\omega$. Of these subunits, all but ω are absolutely necessary for the polymerase function. Each has also been described to play a defined role in either the assembly or catalysis of the enzyme complex. Through sections 1.2.1 to 1.2.2, the functions of these essential subunits are reviewed; discussion on the ω and alternative dispensable subunits is left for section 1.3. The goal here is to establish the essentials of the role of each subunit. The variation they experience is treated in greater depth in section 1.3.

1.2.1 Two identical α subunits drive polymerase assembly

The two identical α subunits — sometimes annotated α^{I} and α^{II} — are crucial for RNA polymerase assembly. The dimerisation of these two subunits into α_2 is the first step in polymerase assembly, both *in vitro* and *in vivo* (Ebright and Busby, 1995). Early studies into the assembly suggested multiple sites used for dimerisation contacts (Blatter *et al.*, 1994; Ebright and Busby, 1995). Structural data suggests as follows. The dimer has a defined binding surface, located singly on the α^{I} subunit, for the β subunit that covers an area of about 1,620 $Å^2$. The α^{II} subunit, in a subsequent but clearly separate step, forms contacts to both β (approximately 240 $Å^2$) and β' subunits (approximately 960 $Å^2$; elegantly reviewed in Sutherland and Murakami, 2018). The assembly of the polymerase appears not dependent on the subunit sequences alone,

as some mutants with defective β' binding are still successfully assembled in living cells (Kimura and Ishihama, 1996). The carboxyterminal domain of the subunit (αCTD) forms weak connections to the AT-rich UP elements in the promoter DNA (Chen *et al.*, 2003). Structures have been published that indicate a different mechanism of recognition in the αCTD-DNA binding surface. These have led the researchers to propose different recognition requirements for two types of promoter elements (Savery *et al*., 1998): some promoter determinants (that is, those very close to the -35 element) likely require both of the αCTDs for activity, while others (distal to the -35 element) only one. Lastly, the aminoterminal domain (αNTD), makes contacts to auxiliary or accessory proteins, but their importance remains a matter of controversy (Egan *et al*., 2000). It appears to function as a docking surface alone as the ΔαNTD strains are viable but display defects in auxiliary protein-mediated regulation (Egan *et al.*, 2000); notice, however, that some mutant phenotypes are defective under normal growth conditions, also (Kannan *et al.*, 2001). Recent literature has identified transcriptional regulators that also associate to the αCTD (Ball and van Kessell, 2019), and the one cited here associates to both.

1.2.2 Alike β and β' subunits form the catalytic centre of the enzyme

The two largest subunits of the core polymerase, β (coded by the gene *rpoB* in *E. coli*) and β' (*rpoC*) constitute approximately 80% of the polymerase's mass. (The β subunit has a molecular weight of 150 kDa, while β', 155 kDa.) Their assembly and recruitment is, as noted in 1.2.1, dependent on the preceding correct dimerisation of the two α subunits. However, despite the larger subunits' requirement for α_2 for assembly, larger contact surfaces (approximately 7730 $Å^2$) form between them than from them to the α subunits. Together, these subunits form the three dimensional nexus where catalysis, or the polymerisation of nucleotides, happens. In a chemical sense, a normal S_N2 reaction occurs between the incoming nucleotide and the last nucleotide in the elongating chain. After the formation of the phosphodiester bond, a pyrophosphate is released as a side product. Reverse reactions in the active site are also possible. From a structural perspective, a group of three aspartic acids are important; they coordinate the two catalytically necessary magnesium ions in the active site. The catalytic chemistry, as has been mentioned, is conserved; therefore, a detailed treatment of the mechanism is not offered.

Interestingly, in cyanobacteria, a division of the *rpoC* into two smaller genes has happened (Xie *et al.*, 1989). Of these fragmented genes, the *rpoC1* gene corresponds to the aminoterminal portion of the parental gene, and is referred to as the γ subunit; the carboxyterminal part has maintained the name β'. (In chloroplasts, the *rpoC1* gene product is referred to as β' and the *rpoC2* product as β''.)

1.3 Polymerases have diversified during evolution

Two factors contribute to the structural diversity of bacterial RNA polymerases. Firstly, three dispensable, accessory subunits (δ, ω and ε) have been described, each with a limited prevalence and presumably lineage-specific functions. In *Bacillus subtilis*, for example, the δ subunit has been proposed to function in the recycling of the RNA polymerase, suggesting an interesting but not catalysis-related role for these types of units (Pei *et al.*, 2020). Secondly, the conserved regions of the polymerase are sometimes surrounded by inserts or interspersed with spacers, among which heterogeneity is notable. In addition to these, normal sequence variation happens in the regions that are not catalytically necessary. The presence of these factors is important to acknowledge as they underlie some of the lineage specific-variation in transcription in bacteria.

Spacers are more common than the accessory subunits. In a pair of papers, Lane and Darst (2010a; 2010b) showed that conserved regions in the core enzyme components are, in fact, frequently separated by spacers that can reach 500 amino acids in length; the functions of the vast majority of such spacers remain unexplored. Lane and Darst note that the inserted individual spacers are typically confined to narrow clades of bacteria but the diversity of the inserts, on the other hand, is high. At least some of the spacers can be crystallised independently of the rest of the polymerase (Chlenov *et al*., 2005), suggesting a rigid structural role or possibly a lineage-specific signalling functionality. In an illuminating example from *Escherichia coli*, certain intra-domain inserts in β and β' were noted to render the polymerase less temperature sensitive (Artsimovich *et al.*, 2003), suggesting a structural role in the first place; recently the same inserts were implicated in the direct recruitment of the transcriptional regulator TraR (Chen *et al.*, 2019), highlighting the below-surface complexity of the biology of these inserts. Additionally, some inserts have direct consequences for the catalytic properties of the polymerase. Windgassen and colleagues (2014) for example showed that Sequence insertion 3 influences the behaviour of the trigger loop, and, thus, the elongation phase of transcription as a whole.

In addition to the functionally imperative subunits $\alpha_2\beta\beta'$, at least three accessory subunits (separate polypeptides) have been characterised for the bacterial RNA polymerase, of which ω is usually designated a member of the core enzyme, also. The three subunits are different and poorly understood. They are not similar in sequence and are differently distributed; for the most part, their functions remain a mystery. In the following, what is known for two of these factors is briefly summed; of these, the first is present in the holoenzyme used in the practical part of this work, while the other is introduced to offer perspective to the possible roles these smaller units can play in transcription.

1.3.1 The ω subunit might offer structural rigidity for the polymerase

The ω subunit is not essential for transcription, thus separating it from other constituents of the core polymerase (reviewed by Kurkela *et al*., 2021). Its molecular weight is approximately 10 kDa, and it was originally thought of as a mere copurifying impurity. Early research noted that upon purification, most of the protein is enriched in the same fraction with the RNA polymerase (Zalenskaya *et al.*, 1990), which can be interpreted as an indication of a biologically meaningful interaction between the two; later research has determined binding to the β' subunit (see Vassylyev *et al.*, 2002 for ω in the *E. coli* holoenzyme). To the best of my knowledge, the ω subunit remains the only member of the core polymerase proteins for which a non-RNAP protein contact has not been identified; this is despite the fact that regulatory roles have been supposed as the primary functionality of omegas'. Tangible evidence has, thus far, remained elusive. Cheng and colleagues in 2010 noted that *Xanthomonas campestris* pv. *campestris* polymerase expressed without ω shows 70% lower activity than that expressed with ω. This, they concluded, was a sign that the subunit increases the stability of the core polymerase. They also suggest that the results indicate that, in this species at least, the ω subunit would also be essential for function. Alternatively, or perhaps additionally, it has been suggested to also promote the original assembly of the core. The model organism used in this study has an ω subunit on which no research has been done.

1.3.2 The *Bacillus subtilis* ε subunit resembles a viral protein

Encoded by the *ykzG* gene (Yang and Lewis, 2008), the ε subunit was recently characterised in *Bacillus subtilis*, whose core polymerase assembly was previously designated as $\alpha_2\beta\beta'\delta\omega_1\omega_2$, where the omegas' subscripts serve to identify two subunits (Wiedermannová *et al.*, 2014). Of these, the ω_2 was known to correspond to the ω subunits in other species; ω character was assumed for the other small subunit co-purifying with the polymerase as well. Keller and others (2014) showed, however, by crystallising the protein and examining it through X-ray crystallography, that its structure does not correspond to that of a typical ω subunit. Rather, Keller and colleagues noted a structural similarity to Gp2 group of viral proteins, which are thought to function in the suppression of the host cell transcription apparatus, contributing to its hijacking for the virus use. Transcriptomics, *in vitro* transcription studies or studies on the localisation of the *ΔykzG* were not different from those of the wild type; the authors thus concluded that while the subunit appears to always associate within the core polymerase, its importance, under the studied conditions at least, is minimal. A further avenue in ε research might be to evaluate how it performs in infected cells.

2 The specificity subunit σ: partitioning the transcriptional space

While the multisubunit RNA polymerase is catalytically autonomous — that is, capable of processively synthesising RNA on its own — it does not have affinity to particular sequences in DNA or the capacity to melt double stranded DNA on its own. (It has even been suggested that the positively charged active cleft of the naked polymerase is actually detrimental to regulated transcription, as it would form a complex with *any* DNA sequence irrespective of actual sequence; Bae *et al.*, 2013; Mekler *et al.*, 2002.) To circumvent the lack of specificity and DNA-melting capacity, the core enzyme associates with a dissociable specificity (σ) subunit to form the RNA polymerase holoenzyme complex (hereafter also Eσ), a molecular assembly capable of initiation at specific loci in DNA (Burgess, 1969; Burgess *et al.*, 1969). The stoichiometry of RNA polymerases and σ subunits would indicate that naked polymerases are highly improbable in cells; the research of the role of active cleft in DNA recruitment has consequently been halted in the case of multisubunit polymerases. It is necessary to stress, nevertheless, that this is the standard for multisubunit enzymes alone: the well characterised viral T7 polymerase, for example, can autonomously associate to DNA (Skinner *et al.*, 2004); non-bacterial multisubunit polymerases have, for the most part, different recruitment strategies (see overview by Kornberg, 2007). The case of viral polymerases is evolutionarily understandable as the viral genome, whose replication is the sole target for the virus, contains only a limited number of genes, all of which are to be transcribed, nullifying any need to differentially recognise target genes.

The σ subunits in bacteria are diverse. Two types have been found, namely the σ^{70} and σ^{54} families of proteins. (The names of these families refer to the molecular weights of their representative members in *E. coli* where they were originally described.) This work concentrates on the σ^{70} family of proteins; a concise introduction to the σ^{54} family-regulated phenotypes is available in Kazmierczak *et al.*, 2005, for example.

2.1 Members of the σ^{70} family proteins have four conserved regions

Nomenclature of sigmas is layered but not too complicated. Straightforwardly, four domains have been identified: σ_1, σ_2, σ_3 and σ_4 (Gross and Gruber, 2003). Finer details are described by referring to subregions (for example, σR3.1). For the most part, the subregions follow the numbering in the parental domain they constitute. Thus, σR4.1 and σR4.2 together constitute the fourth domain, or σ_4. The conservation patterns of these different structural elements are different (Lonetto *et al.*, 1992). Consequently, the members of the σ^{70} family have been divided into four groups, each with a typical makeup of a member. In this work, I refer to the members of such groups as “group one protein”, for example; when referring to parts of a σ factor, I have

adopted a nomenclature where no distinction is made between the domains and subregions. To give an example, in this work I shall refer to $\sigma_{1.1}$ while denoting the subregion σR1.1 found in the domain σ_1.

To a great extent, the structure of sigmas corresponds to their physiological functions (Gross and Gruber, 2003; Kazmierczak *et al*., 2005), supporting their classification into separate groups. Group one proteins are referred to as the full-length sigmas: their structure is composed of the domains σ_1, σ_2, σ_3 and σ_4. The molecular weight of these sigmas is approximately 70 kDa (Paget, 2015); each species is thought to only encode one σ of this group, namely the housekeeping σ factor. They are the only group of essential sigmas. Group two proteins resemble group one proteins a lot, but do not have σ_1. These sigmas are nonessential and frequently respond to different nutritional stresses. Group three proteins contain the domains σ_2, σ_3 and σ_4; examples of their regulated functions include sporulation and flagellum biosynthesis. Group four proteins include a hugely diverse set of sigmas that respond to various stresses and external signals. They are composed of only two domains: σ_2 and σ_4. They represent the lightest sigmas in bacteria, with molecular weights of around 20 kDa (Paget, 2015). **Figure 2.2** represents graphically the four groups and their canonical domains.

In the following, the basics of each of the domains and their most important subregions are summarised. For references, refer to the review by Paget (2015).

2.1.1 The σ_1 domain is only found in the housekeeping sigmas

The first domain in sigmas is σ_1; its only subregion is $\sigma_{1.1}$. Another subregion, $\sigma_{1.2}$, has been recognised but is annotated as a part of the σ_2 domain. This difference is rooted in the evolutionary diversity of sigmas: the $\sigma_{1.2}$ subregion is present in practically all characterised σ^{70} family members, while only the canonical housekeeping σ^{70} proteins are thought to have the subregion $\sigma_{1.1}$. (There is, in actual fact, a limited number of exceptions from the general rule; see Newlands *et al.*, 1993 for instance.) In any case, the presence of a $\sigma_{1.1}$ is taken to signify the housekeeping character of a family σ^{70} member. It has been suggested that it promotes the housekeeping sigmas to adopt very compact forms when they are not bound to the core polymerase. This might represent a mechanism to inhibit nonproductive binding of sigmas to DNA.

Also interestingly, the $\sigma_{1.1}$ is very acidic in nature; the same is true for nucleic acids. As was noted earlier, there have been speculations about the role of the core polymerase cleft's role in binding DNA. It would appear that the $\sigma_{1.1}$ in fact mimics the topology of DNA and is buried in the cleft

upon binding of the σ to the core polymerase. Thus, while the subunit as a whole supports specific recruitment to certain promoter sequences, the specific domain simultaneously inhibits nonspecific interactions.

2.1.2 The σ_2 and σ_4 domains recognise specific regions in DNA

All characterised members of the σ^{70} family have two elements: σ_2 and σ_4. Together, they are responsible for the recognition of two of the most important sequence determinants in the promoter DNA. Domain σ_2 is the most conserved domain in σ^{70} family proteins. Its structure has been further divided into several subregions. Of these, the widely conserved regions make contacts to the DNA -10 element, and, upon initial separation of DNA strands use their pockets to inhabit some of the bases of this region in a flipped conformation. While the domain as a whole is conserved, its finer structures differ between sigmas that represent different groups. Specifically, in group one and two proteins, an otherwise absent subregion exists: the $\sigma_{1.2}$, which is responsible for the recognition of a specific structural element in the DNA, the discriminator region. Another element, known as the non-conserved region (NCR) is found in group one proteins; different roles for it have been suggested.

The other broadly conserved domain is the σ_4 domain. It is composed of two subregions, $\sigma_{4.1}$ and $\sigma_{4.2}$. The domain is responsible for recognising another DNA element (-35 element). Additionally, there are extensive connections between the domain and β flap domain in the core polymerase. In addition to these, connections to transcriptional regulators have been described. Evidence comes from *E. coli*. In 2004, Wickstrum and Egan noticed that several of the seventeen alanine substitutions they introduced to the domain decreased the amount of transcription from a RhaR-controlled gene. RhaR was likewise subjected to mutagenesis; the results were similar. Finally, abnormal growth was observed for strains where either σ^{70} R599 or RhaR D276 were altered, providing definitive evidence for a contact.

2.1.3 The σ_3 domain offers flexibility to DNA recognition

The σ_3 domain is conserved in all but group four proteins. Structurally speaking, they consist of a bundle of three helices; this structure makes contacts to the major groove in the DNA immediately outside the -10 element. The interaction can be crucial for some biological functions, as it appears that the elements recognised by the σ_4 domain (that is, the -35 element) need not conform to the consensus sequence in case there is a strong interaction between the σ_3 and the extended promoter -10 element. It will later become clear that the group four proteins have stricter sequence determinants than do other sigmas (Koo *et al.*, 2009). One reason behind this can be traced back to the σ_3 domain. In non-group four proteins, the domain

is likely to offer some flexibility to the protein, allowing for slight variations to the distance between the recognition elements; lack of this makes the group four proteins very rigid. Also possibly related to the lack of the domain is that group four proteins have less characterised interacting partners (Rhodius *et al.*, 2013) than do other sigmas.

2.2 The pool of recruitable-to-RNAP σ subunits is tightly curated

The cytosol might contain, depending on the species, between one and a hundred species of σ factors — this group of potential holoenzyme members is referred to as the σ (factor) pool, from which the selection for RNAP is realised. In *E. coli*, both the affinity (Maeda *et al.*, 2000) and the copy number (Jishage *et al.*, 1996) of available sigmas contribute to the relative assembly ratios between different holoenzyme complexes. Recently, a theoretical model suggested that this type of passive regulation does, indeed, happen (Mauri and Klumpp, 2014). The authors prove that very strongly recognised promoters, so-called saturated promoters which are occupied by a holoenzyme most of the time, are relatively stably expressed even under passive regulation. However, suggest Mauri and Klumpp: "If promoters are recognized by two species of holoenzymes or promoters depending on different sigma factors overlap, even saturated promoters become affected by sigma factor competition." These examples clearly show that the study of transcription should always consider the transcriptional environment as a whole. If the individual phenomena are subjected to more detailed study in isolation, caution should be exercised in generalising their results to the goings-on of live cells.

While the mechanism is self-evidently driven by affinity constants, and its consequences are already understood, much remains to be learnt on how the cells regulate the availability of their different sigmas to these selection processes. Expansive toolkits have been characterised. Some group four proteins in *Bacillus subtilis* are translated as inactive prosigmas that can be selectively cleaved to release their active forms upon need (Hilbert and Piggot, 2004). In *S. coelicolor* oxidative compounds in the cytosol influence the selection of reading frame for σ^R, leading to the translation of two alternative isoforms with halflives of 10 and 70 minutes, respectively (Kim *et al.*, 2009); the stabler of these is the biochemically meaningful isoform. Also, the *E. coli* stress related σ^R is fast delivered for proteolysis in the exponential phase by the protein RssB in an ATP-dependent fashion (Becker *et al.*, 1999; Zhou *et al.*, 2001). Stress cues activate a host of adaptor proteins to interfere with RssB (Dorich *et al.*, 2019), savouring a rising proportion of σ^R for the σ pool. Select sigmas can also be sequestered from the active pool under or until specific conditions. Such anti-σ-dependent mechanisms are diverse and have been extensively reviewed, most recently by Paget (2015).

Complementing σ-specific regulatory mechanisms, the small molecular weight alarmone guanosine tetraphosphate (ppGpp; see chapter **4**) exerts a global effect on the entire σ pool (Jishage *et al.*, 2002). The molecule, which has effects that go beyond σ regulation, affects the binding properties of sigmas and the core polymerase — that is, the competitive edge of the sigmas rather than their stoichiometry. Thus, different ratios of holoenzymes can arise from stoichiometrically identical σ pools. To the best of my knowledge, ppGpp remains the sole small molecule for which this type of a functionality has been characterised, but what one molecule can do, others are also likely capable of. There is no reason to assume ppGpp be the only global modulator of σ competition.

There is an ongoing discourse relating to the (non)insularity of the σ-specific regulons. Four points offer an effective summary on the emerging consensus (Gruber and Gross, 2003; Österberg *et al.*, 2011; Feklístov *et al.*, 2014):

1) Group one and two proteins recognise divergent promoters and tolerate considerable deviations from the consensus sequences; environmental stimuli can reorder promoter preferences and lead the housekeeping sigmas to give rise to a variety of transcriptomes.
2) Group four proteins are frequently on-off regulated and strict in their promoter requirements. Divergence of a promoter from the consensus -10 and -35 elements quickly leads to a loss of specificity.
3) Group four-dependent stress programs can be redundant. That is, a gene in a group four σ-specific regulon can also belong to a regulon of another σ factor. Both clearly verified examples (Huang *et al.*, 1998; Dostálová *et al.*, 2019) and blurrier reports (Martinez-Bueno *et al*., 2002) attest to this line of thinking.
4) Some group four-dependent programs are truly insulated, or orthogonal. These types of regulons can be used to coordinate complicated multistep processes like sporulation. In the case of *Bacillus subtilis*, the induction of σ^F is mandatory before the activation of σ^E for spore formation to succeed (Narula *et al.*, 2012).

Work on regulon redundancy is in its infancy. At present, there is no consensus on whether the insulated promoter sigmas retain their orthogonality across the species boundaries. In the broadest sense, a dichotomy seems to be emerging: while some sigmas reconfigure transcriptional strength amongst already expressed genes, perhaps in response to some specific stress, others are used to distribute this strength to altogether novel transcriptional targets. In agreement with this, *Mycoplasma genitalium* (which inhabits only one environment, the human cell; Taylor-Robinson and Jensen, 2011) has only two sigmas (the view of *M. genitalium* as only having one σ is outdated; see Torres-Puig *et al.*, 2016); the gut commensal (and occasionally,

pathogen; Kaper *et al.*, 2004) *E. coli* has an extended battery of seven (Cook and Ussery, 2013). The species that experience variation in their environment harbour the broadest patterns of σ factors, frequently with insulated regulons — for the marine bacterium *Plesiocystis pacifica*, 118 group four proteins were reported in the Microbial Signal Transduction Database (MiST; for analysis, see Feklístov *et al.*, 2014; for the most recent version of the database, see Gumerov *et al.*, 2020).

Housekeeping genes are not always transcribed by the $E\sigma^{70}$ holoenzyme. In *E. coli*, it has been recognised, at least four different holoenzymes bind to their respective constitutive promoters (Shimada *et al*., 2017). It appears that perennially essential genes have, to an extent, been dispersed to separated regulons; why this is the case is curious but not immediately obvious, as constitutive genes are supposed to be expressed at a standard strength through the cell cycle and across different environmental conditions.

Promoters that are responsible for the transcription of ribosomal RNA are considered, to an extent, strictly insular to the housekeeping holoenzyme complex $E\sigma^{70}$ (see Section 4.3). This is not universally true, in fact. Transcription *in vitro* from the promoters by an alternative σ^{38} does happen, and even is promoted through some of the mechanisms that also drive forward $E\sigma^{70}$-mediated transcription, as was demonstrated by Newlands and colleagues in 1993. Most ribosomal transcription, it however seems safe to say, is initiated by the $E\sigma^{70}$ holoenzyme, and the promoter could then be characterised as “nearly insular”. Under energy-limited conditions, the many mechanisms of tilting the σ pool towards non-σ^{70} holoenzymes are activated; this in turn will reduce ribosome biogenesis and save energy for the cell.

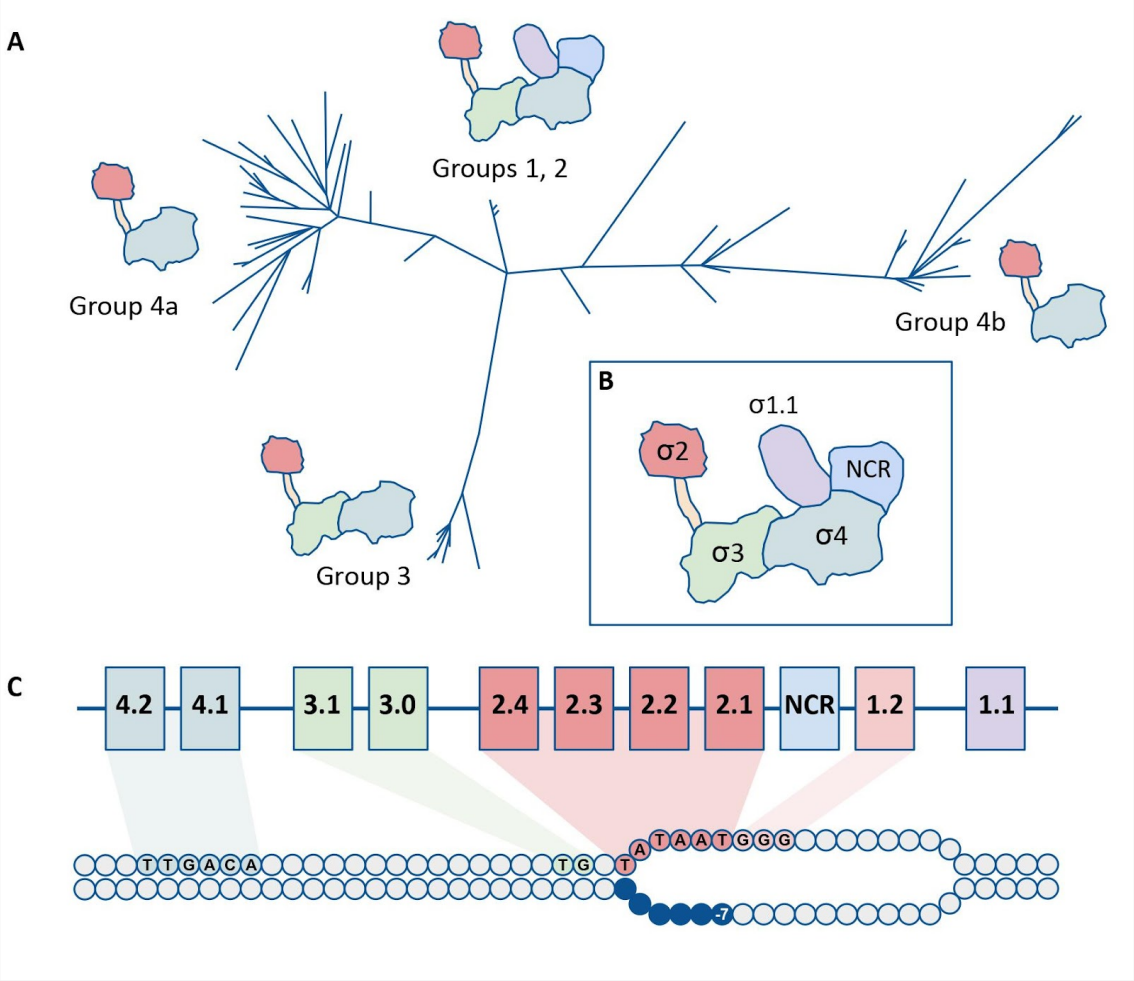


Figure 2.2 | **Architecture and diversity of σ^{70} family proteins.** In panel **A**, the approximate phylogenetic radiation of *Streptomyces coelicolor* σ^{70} family subunits is shown. The most diverse class of sigmas, the group four proteins, only have DNA-binding structural determinants. Their divergent evolution has produced a broad selection of nonredundant sigmas whose promoter sequences do not usually overlap. In *S. coelicolor*, most of the in-genome σ diversity is explained by the expansive radiation of group four sigmas. In panel **B**, the three dimensional structure of the protein is approximated. This is a reproduction from an old paper to mainly colour key to the figure; resolved whole-σ-alone structures have been elusive. In panel **C**, the polypeptide structure of a group one σ is shown. The four structural elements that make connections to the promoter DNA are connected with coloured bars. The $\sigma_{1.1}$ element, for which a cognate DNA sequence is not shown, rather makes its contact to the RNA polymerase core enzyme. See section 3 in the main text of this work for details regarding the DNA elements. Nucleotides from -12 to -7 are coloured blue. Figure has been inspired by Gruber and Gross (2003) and Paget (2015).

_____

The specificity subunits are required for the initiation of transcription but can be discarded post-initiation (Travers and Burgess, 1969) to return them to the σ pool in a process known as σ cycling. At least two lines of research support this. Firstly, RNA synthesis is coupled to an outward push of the σ subunit (Bowser and Hanna, 1991), which naturally supports dissociation. Secondly, the elongation factor NusA competes for the same binding surfaces on the polymerase (Yang and Lewis, 2010); most of the elongation complexes carry no σ but rather, NusA. Many transcription complexes in their early phases retain sigmas (Kapanidis *et al*., 2005), however, sometimes through the entire transcription cycle (Harden *et al.*, 2016). A succession of studies (Nickels *et al.*, 2004; Brodolin *et al.*, 2004) in the very same issue of *Nature Structural and Molecular Biology* indicated that retained sigmas might in some instances slow transcription down by transiently binding to -10 element-like sequences in the transcribed DNA as the polymerase moves along it, promoter proximal pausing. The σ^{70}-dependent pausing was reported in the 1990s as a phenomenon that limited phage DNA transcription (Ring *et al.*, 1996) but roles are currently being revealed for non-housekeeping proteins like *eco* σ^{38}, also (Petushkov *et al.*, 2016), and further regulatory components for promoter proximal pausing are actively studied (Sun *et al.*, 2021). These examples illustrate that transcription is affected by a multitude of factors.

3 The promoters and their complexes with Eσ: dynamics of transcription initiation

3.1 Promoters are a nonuniform family, if a family at all

Different promoters have radically different initiation rates *in vivo* even when transcribed by one type of a holoenzyme alone (e.g., $E\sigma^{70}$ alone, with no other σ subunit influencing the expression rates). This has generally been interpreted as a consequence of the architecture of promoters. They consist of discrete structural elements, for many of which consensus sequences have been determined. These elements are combined in a mix-and-match fashion, resulting in different promoters having a mosaic of determinants both enhancing and suppressing transcription. Promoters, then, are not a homogenic class of sequences; rather, being a promoter is something of a family resemblance. That is, a naturally occurring promoter is likely to have features in common with many other promoters, while no feature is likely to be common to all — or perhaps even the majority — of promoters. In the following, I introduce the most important sequence determinants of promoters and how their presence is thought to affect the dynamics of transcription. For a schematic representation of a promoter, see **Figure 2.2**.

The current technologies permit the evaluation of how a sequence feature affects the different phases of transcription. The mix-and-match architecture of the promoters does, however, collude the individual features' contribution to the general levels of transcription for the time

being. Data integration from different studies (combining the reaction constants correlated to different sequence determinants) remains blind to the additivity of some effects; the full significance of the promoter context remains unaccountable for. For two comprehensive expositions on promoter structure, refer to the works of Jensen and Galburt (2021) and Mejía-Almonte and colleagues (2020).

3.1.1 Upstream elements contact αCTDs

Located -60 to -40 bases from the transcription start site (TSS) in bacterial promoters, the upstream elements (UP elements) are the farthest removed promoter elements characterised (Jensen and Galburt, 2021). They are considered especially important for the ribosomal promoters and, following this, much of the work on them has been performed on ribosomal promoters. Studies have painted a picture of a beneficial, but, strictly speaking, not necessary promoter element. Estrem and colleagues (1998) sampled randomised UP sequences and selected those that performed better than the native *rrnaB* P1 promoter in *E. coli*. Similarities in 31 such promoters indicated a consensus sequence rich in AT. Indeed, the consensus promoter seemed to increase transcription from the *rrnaB* P1 promoter over 320-fold. Rao and colleagues (1994) showed that the *absence* of the native UP element in the *rrnaB* P1 resulted in a 30-fold decrease in transcription efficiency. These studies together illustrate the strong influence of the UP element.

The overall function of the element is well characterised, then. The reasons behind this are still a matter of some controversy. Specifically, at least two alternatives have been proposed. Working on the *lac* promoter, Ross and Gourse in 2005 attributed the 10-fold decrease in transcription that resulted from the deletion of the UP element to the decrease in the opening of the double stranded DNA (see 3.2), suggesting that the element's effects were mainly related to mechanism other than recruitment of the holoenzyme. On the contrary, a more recent preprint found that the presence of the UP element in the *rrnaB* increased holoenzyme recruitment around six-fold while all other effects were minimal (Mumm *et al.*, 2020). Considering that these studies used fundamentally different promoters, not much can be said with confidence on the mechanism; it might well be different in different promoter contexts.

3.1.2 The -35 hexamer promotes Eσ-promoter association

The -35 hexamer has the consensus sequence TTGACA (Jensen and Galburt, 2021). The consensus sequence is associated with the highest detected transcriptional outputs in experiments both *in vivo* and *in vitro* (Shimada *et al.*, 2014) from the constitutive promoters. At least two older studies (Kobayashi *et al.*, 1990; Moyle *et al.*, 1991) have found the -30-located

adenine the least important contributor to the activity of the element. Furthermore, authors of a recent paper used two bioinformatic softwares to study the composition of the -35 element across different phyla (Myers *et al.*, 2021): they found the initiating TTG triad almost universally preserved across different bacteria. The hexamer element is recognised by σ_4 (Campbell *et al.*, 2002) and, according to literature, the main role of the element is to increase Eσ recruitment to the promoter (Ross *et al*., 2003). Chen and others (2003) along with Ross and others (2003) gave the earliest demonstration of the avidity at the element: the -35 contacts seem likelier once the αCTD has already established a connection to the UP elements (see 3.1.1). Selectivity in the -35 region has been studied in greater detail, also. In *Streptomyces coelicolor*, the -31 consensus nucleotide was shown to differ from that of *E. coli* and the differential recognition of the promoters was traced down to a corresponding mutation in the σ subunit, where a methionine conferred specificity to -31 C, while an arginine mutant was able to recognise both variants (*Eco* -31 T and *Sc* -31 C) of the hexamer (Kim *et al.*, 2016). While important, the -35 hexamer either works or not in concert with the σ factor.

During the initiation pathway, double-stranded DNA is both bent and separated (see Section 3.4). Many studies have identified that the bending takes place in the immediate vicinity of the -35 hexamer, supporting the general hypothesis that the σ subunit is needed for meaningful promoter-RNAP contacts. However, while the bending of the DNA has been pinned down to the -35 hexamer, studies have demonstrated that the region is not necessary for the separation of the DNA strands (Niedziela-Majka and Heyduk, 2005).

3.1.3 AT-rich spacers promote unwinding

The region between the -35 and -10 elements is known as the spacer. Two factors contribute to its characteristics: length and make-up. Foremostly, the function of this sequence of DNA is to separate the -35 and -10 hexamers spatially. The ultimate logic behind this is readily explained by the polymerase holoenzyme structure: contact surfaces are flexible but still. Thus, it is not surprising that no consensus sequence for the region has been described (Hawley and McClure, 1983). In terms of length, 17 base pairs seem to provide optimal distance (Beutel and Record, 1990). Two streams of data support this conclusion. Firstly, spacers of 17 base pairs are the most commonly encountered spacers in the sequenced genomes (for an *E*. *coli* perspective, see Shimada *et al.*, 2014), indicating an evolutionarily optimised length; secondly, spacers with deviant lengths have been rendered more effective in many cases by correcting their spacers to 17 nucleotides (Aoyama *et al.*, 1983; Mulligan *et al.*, 1985).

While the spacer is a spatial element in the first sense, its sequence also affects the overall behaviour (Warne and deHaseth, 1993). Specifically, spacer sequences with a higher proportion of AT nucleotides — that is, nucleotides with less internucleotide hydrogen bonds — increase promoter activity (Hook-Barnard and Hinton, 2009) and the manipulation of AT-rich spacers into GC-rich diminishes transcription (Warne and deHaseth, 1993). Interestingly, AT-rich double helices are prone to helix deformation also for reasons that are not of immediate consequence of hydrogen bonding differences (Jensen and Galburt, 2021) but, irrespective of the fundamental reasons, AT-rich spacers seem to offer the best synergistic link between the -35 and -10 regions (Hook-Barnard and Hinton, 2009).

Finally, a more active role in the regulation of transcription has been proposed for select spacer subregions. In 2011, Yuzenkova and colleagues noted that the nucleotides at -17 and -18 positions could be directly responsible for polymerase recruitment. The -18 thymidine was suggested to be responsible for some of the conformational properties of the spacer region (Shin *et al.*, 2011) but the relevance of these reports, to my judgement, remains unsettled. Whatever their significance, it would appear that the noted sequence elements are relatively rare.

3.1.4 The -10 hexamer is essential in DNA melting

The -10 element is a multifunctional hexamer that runs from positions -12 to -7. Its consensus sequence is TATAAT (Jensen and Galburt, 2021). At least three distinctive roles have been characterised for this sequence: (1) It is directly involved in the promoter recognition events; (2) melting of the double-stranded DNA originates therefrom; and (3) its connections to the polymerase complex affect the stability of the open promoter complex. The consensus sequence, considering these, increases both the likelihood of reaching and the factual lifetime of the open promoter complex. (Whether this increases transcription would be reliant on the context; too stable complexes become unlikely to advance onwards from the open complex at the cost of transcriptional output. For the significance of this, see Section 3.2.)

The -12 thymine appears to be responsible for the direct contacts to the σ factor (Feklistov *et al.*, 2011), whereas the nucleotide at the position -9 is indicated to make a connection to the β subunit of the polymerase (Chen *et al.*, 2020). Nucleation of double strand melting is indicated to begin at -11 (Feklistov and Darst, 2011; Matlock and Heyduk, 2000) and structural data indicate that the entire length of the hexamer is flipped out (see references in Jensen and Galburt, 2021). The sequence of events through the promoter recognition and the accompanying conformational changes are not clear (Mazumder and Kapanidis, 2019). Heyduk and others in 2006 determined that adenine, especially, in -11 is important for the nucleation event (but

indications were there earlier; see Matlock and Heyduk, 2000). Part of the lacking characterisation might be explained by the fluid nature of the structural element. In 2013, Mekler and Severinov studied the importance of single nucleotides for the DNA-holoenzyme interaction. In brief, they tagged a polymerase fluorogenically and followed its association to different -10 sequences. They proposed a combined energy argument: that the holoenzyme either binds or does not bind to DNA as a function of the energetic properties of the entire -10 hexamer. This might be, according to the authors, a mechanism to maximise the differentiation between accidental -10 element-like sequences in the genome and actual promoters. Later experiments (Heyduk and Heyduk, 2014) supported these conclusions as the conservation of -12, -11 and -7 all proved nonessential; the combined energy argument was lent support as CG-rich elements proved less effective in the same experiments. In agreement with this, and in a notable contestment of the consensus sequence, Myers and others in 2021 found all bases to be equally commonly encountered at the -7 position, with *Alphaproteobacteria* rarely showing a thymine.

3.1.5 The discriminator sequence influences TSS selection

Immediately downstream of the -10 element is the discriminator sequence that displays differences between consensual and ribosomal promoters. The sequence is most commonly six bases long in general-use promoters (Vvedenskaya *et al.*, 2016) but many ribosomal promoters have longer consensus sequences, typically by two nucleotides (Winkelman *et al.*, 2016). In general, 5'-GGG(A)-3' motive in the immediate vicinity of the -10 element on the nontemplate strand is thought to increase the stability of the open promoter complex (Barinova *et al.*, 2008; Haugen *et al.*, 2006), possibly by modulating the behaviour of the holoenzyme complex but also through the autoregulation of the movements of the DNA strand itself; the lack of this motive and its replacement with a C-rich sequence in the template strand is a chief determinant of instability in the ribosomal promoters (Winkelman *et al.*, 2016).

Unsurprisingly, there is a clear correlation between the character of the discriminator sequence and its transcriptional behaviour: more stable open complexes generally produce longer abortive transcripts than the less table ones, and the extremely short-lived rRNA promoter open complexes seem to bypass abortive transcription altogether (Shin *et al.*, 2021; Henderson *et al.*, 2017); unstable ribosomal promoter open complexes are the natural targets of many regulatory mechanisms that destabilise the otherwise unstable open complex. The best characterised of these is, perhaps, the combination of two transcription factors DksA and guanosine tetraphosphate (see the next section).

3.1.6 The core recognition element extends downstream of the TSS

Not all sequence elements have been recognised because of their high homology. The arrival of high resolution structures made possible for example the recognition of the core recognition element (Zhang *et al.*, 2012). The structure showed that contacts are formed between nucleotides at positions -4 to +2 (excluding the -1 position); the +2 G is buried in the pocket in the β lobe. Targeted mutagenesis to the RNAP residues responsible for the contacts resulted in noticeable changes in transcription (Petushkov *et al.*, 2015). The inference is that while apparently meaningless in terms of promoter recognition, transcription from it is still affected; the interaction increases the stability of the open complex, for example (Jensen and Galburt, 2021).

3.2 Association of the holoenzyme to DNA and its unwinding

For transcription to initiate from any given double stranded DNA sequence, two things must be satisfied: firstly, the holoenzyme must recognise and stably bind to the promoter DNA, and, secondly, the two strands of DNA need to be separated from one another. As catalysed by the holoenzyme complex, these two conditions are realised in separate steps (Buc, 1985; Saecker, 2011; Chen *et al.*, 2020). Initially, the promoter DNA and the holoenzyme Eσ make stable contacts that effectively lock the transcription machinery in its place; a closed promoter complex (RP_c) is thus formed. The closed promoter complex is isomerised — that is, the chemical bonds within it are reorganised — to separate the strands to form the transcription bubble. This intermediate is referred to as the open promoter complex (RP_o). Following the formation of the open promoter complex, the coding DNA strand is loaded into the cleft of the core polymerase; the first phosphodiester bonds are catalysed by the initial transcribing complex (RP_{itc}). The start of processive transcription, coupled to promoter clearance, signifies the start of elongation. From this point, the complex is referred to as the elongation complex (EC). (The major events are represented in **Figure 3.2**.)

Modern structural biology methods are not constrained by the need to crystallise the studied complexes; consequently, it has become possible to investigate the molecular movements in greater detail than earlier. Most usually, however, only the most populated (stable) states are represented, since colliding the effects of short-lived intermediary steps into one step between the RP_c and RP_o is easy:

$$R + P \rightleftharpoons RP_c \rightleftharpoons RP_o \rightarrow R + P \qquad \textbf{Eqn 3.2.1 a}$$

Much is already known of the molecular movements behind these kinetic steps. One theme is overarching: the kinetic properties of each are determined by the interactions of the promoter

and the holoenzyme (Ruff, 2015). Because of the underlying differences, the kinetics of the different reactions are different, and, consequently, different promoter complexes are found to populate the same states to a different degree. In other words, for different promoter complexes, different steps in the initiation pathway become rate-limiting.

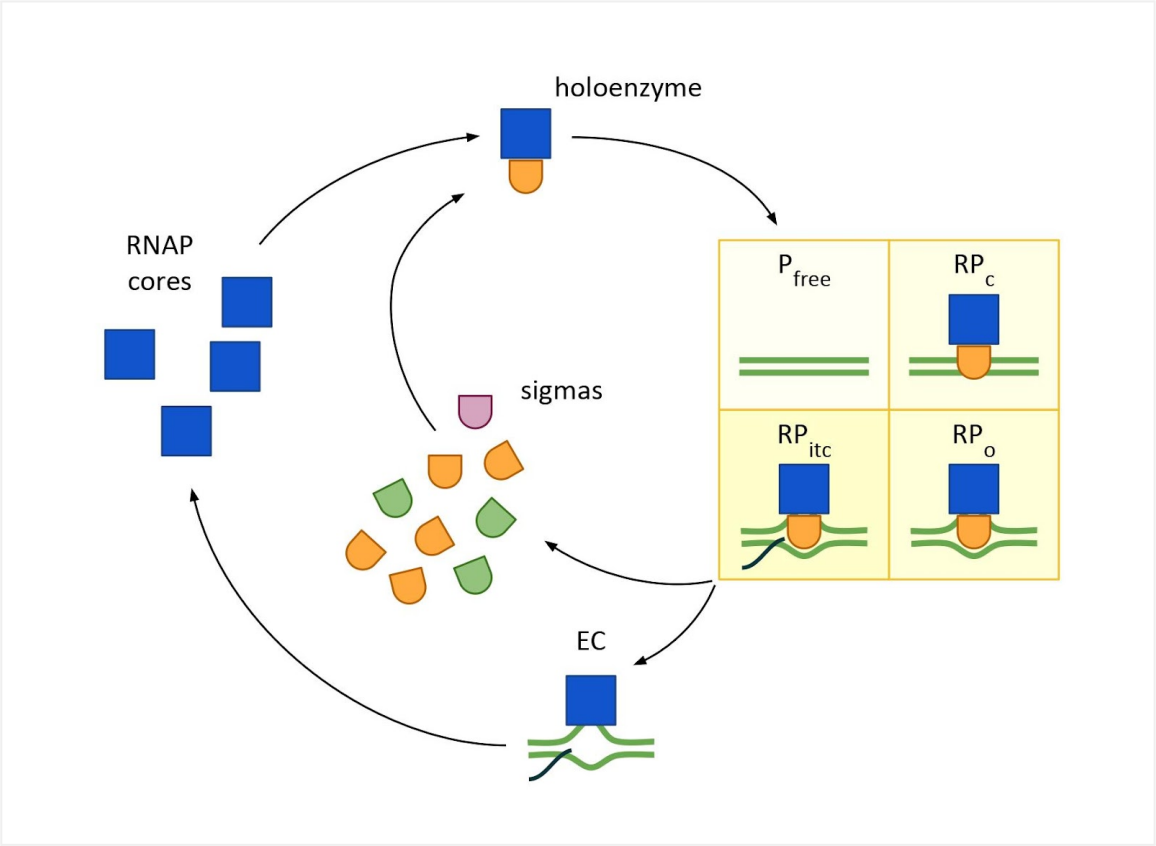


Figure 3.2 | **Some important events in bacterial transcription initiation.** The RNA polymerase core enzyme is catalytically competent but cannot recognise or melt promoter DNA on its own; to achieve these two goals, an accessory subunit (σ) is recruited to the polymerase from the σ pool. The so formed holoenzyme recognises its cognate promoter (represented as parallel green lines), forming a closed promoter complex (RP_c). The binding free energy is used to isomerise the RP_c into the open promoter complex (RP_o); the formation of this melted bubble is the prerequisite for the core polymerase to initiate the polymerisation reaction. Because the binding of the polymerase to the double stranded DNA is reversible, the open complex can collapse both before and after first polymerisation reactions, the pathway intermediates in the yellow boxes are effectively at an equilibrium. Entry of the species into the equilibrium is determined by the character of the holoenzymes (a holoenzyme only recognises some of the promoters in the genome); the escape of species is determined by the rate of promoter clearance (that is, to what extent the complex moves into processive elongation that no longer can revert back to a closed complex, for example). See text for details.

In *E. coli* and *M. tuberculosis*, several transient intermediates have been captured on the way from RP_c to RP_o. They differ to an extent (*eco* $E\sigma^{70}$ might undergo at least eight steps; Chen *et al.*, 2020) but five shared steps and their structural significance can be recognised (Chen *et al.*, 2021; Chakraborty *et al*., 2012; Duchi *et al*., 2018):

1. To initiate transcription, the holoenzyme Eσ binds to promoter DNA as previously indicated. Importantly for transcription initiation, an interaction between the -10 hexamer and its cognate σ_2 takes place. Strong interaction between the promoter and the σ subunit favours this intermediate.
2. The movement of β' clamp induces (nucleates) the DNA strands' separation. Subsequently, the nucleotide at -11 position, but possibly others, is destacked or removed from its preferential alignment in the DNA chain and flipped to its external, σ-facing side. The presence of nucleotide stabilising cavities in the σ subunit promotes this intermediate.
3. Starting from the destacked nucleotide and moving downstream, the energy difference between the native configuration of the promoter and its complex with the holoenzyme is used to distort DNA and induce its initial opening. The distortion pushes the template DNA to a closer proximity to the active site; some four to five bases are initially unwound.
4. The β lobe opens, and induces the ejection of $\sigma_{1.1}$. The ejection of $\sigma_{1.1}$ from the DNA cleft and the opening of the β lobe liberate space for the transcription bubble to propagate further. The bubble opens to its full extent in a promoter dependent way. To date, stabilised open complexes have been described that open up to positions between -4 and +2.
5. Finally, the open complex is stabilised by the contacts between the holoenzyme and DNA. In this intermediate the single stranded DNA in the template strand is proximally located to the catalytic site and ready for loading into it to initiate catalysis.

As has been mentioned, the transcription mechanism relies on the S_N2 reaction between two nucleotides, releasing in the process a pyrophosphate. Generally speaking, the two participatory nucleotides are coordinated by magnesium ions in two pockets: the *i* site (harbouring the 3' end of the RNA chain) and the *i*+1 site (for the incoming nucleotide). In initiation the *i*-located initiating NTP is stabilised by the *i*-1-located GTP until the S_N2 reaction has been completed between the incoming and initiating nucleotides; the GTP is released into the medium and the two first nucleotides joined by the newly-formed phosphodiester bond translocate to occupy positions *i*-1 and *i* (Kennedy *et al*., 2007).

The influence of the promoter extends beyond the formation of the open complex, and destine the initial transcribing complex to one of three reaction pathways: scrunching, abortion or escape (Xue *et al.*, 2008). In *E. coli*, a scrunching mechanism was described for the polymerase (Revyakin *et al.*, 2006). In the scrunching pathway, the polymerase scrunches DNA in itself to synthesise RNA but simultaneously maintains its contacts to the promoter region. As a result, nascent RNA is produced while the unwound DNA is accumulated inside the polymerase, increasing tension (Kapanidis *et al*., 2006). A productive and a nonproductive pathway compete for the resolving this tension: the polymerase can either leave the promoter (known as promoter escape or clearance) and advance to processive elongation (Vassylyev *et al.*, 2007), or it can divert to abortive cycling (Goldman *et al.*, 2009), thereby releasing the small fragments of nascent RNA and reverting back to its original position. Treating transcription as an equilibrium between the three alternatives, Xue and colleagues (2008) showed, in a Monte Carlo simulation, that the promoter structure is a determinant of the rates through the different routes. Confirmatory evidence has been accrued in the laboratory. For example, Tang and others in 2009 reported that abortive transcript length of a viral single subunit RNAP is correlated to the probability of the complex moving from initiation to elongation. Put differently, the strength of the Eσ-promoter contacts influences the escape rate. Importantly, a portion of promoters behave uniquely. Ribosomal promoters need free nucleotides to form open complexes and are reported to bypass the abortive cycling altogether (Winkelman *et al.*, 2016).

4 Auxiliary factors in transcription: variations to the theme

Transcription is initiated at different rates from different promoters; that is, firing rates are not alike. The structural biology behind these differences have been reviewed. In differently composed systems (containing sigmas, RNAP and promoters) different mechanistic steps become rate-limiting. In other words, while all promoters are fired more or less at different rates, the reasons behind this are *equally unique* for all promoters. Relying on this fact, bacterial cells have developed sophisticated mechanisms to selectively aid or hinder specific types of promoters. These transcription factors constitute a further regulatory level that is heavily utilised: *E. coli* genome, for example, was recently reported to code for 285 transcription factors (Ishihama *et al.*, 2016).

Two types of transcription factors exist: those that make contact to the polymerase and those that do so to the DNA. In the following, I review two transcription factors (CarD and DksA) that are present in the genome of *Spirochaeta africana*.

4.1 CarD increases the lifetime of the open complex

Spirochaeta africana genome contains the gene for CarD, a transcriptional regulator whose kinetic properties have been studied in great detail. Interestingly, CarD is not present in the genome of *E. coli*; consequently most studies have been conducted on the transcription systems of its most prominent host organism, mycobacteria. Using *Mycobacterium tuberculosis* as the model organism, Stallings and colleagues (2009) showed that CarD is especially important in the regulation of ribosomal RNA genes in the species. Specifically, they knocked out the gene and observed the mutant strain in laboratory culture, and in a mouse infection model. In both experimental setups, loss of CarD was detrimental and death was accompanied by constantly high levels of rRNA; for this reason they suggested that a CarD-dependent mechanism regulates rRNA synthesis in mycobacteria.

Structural studies have elucidated the mechanism of CarD-mediated regulation of gene expression to a high precision. In the main, CarD stabilises the open complex RP_o and increases the fraction of the species in this isomerisation state at the cost of others in the equilibrium (Zhu and Stallings, 2023; see **Equation 3.2.1a** and **Figure 3.2**). The picture has been built piece by piece: In 2013, structural studies indicated that the protein docks to the RNA polymerase, and forms connections to the junction of the upstream end of the transcription bubble, at the interphase of double and single stranded DNA (Srivastava *et al.*, 2013). The authors noted the reaction mechanism then has to do with the open promoter complex; soon afterwards the mycobacterial polymerase was compared to that of *E. coli in vitro* to verify the polymerase-dependent difference in baseline stability of the RP_o (Davis *et al*., 2014), and that CarD functions to stabilise the unstable open complexes in mycobacteria. At saturating conditions, the collapse of the open complex back to the closed complex became slower while the strand separation leading to the open complex was accelerated (Rammohan *et al*., 2015). Finally, studied by Jensen and coworkers in 2019, CarD was shown to slow down promoter escape in mycobacterial transcription system. These results together show why and how CarD can both increase the levels of transcription from some promoters and inhibit it elsewhere.

In the case of CarD research, one extrascientific consideration requires attention. While the mechanism behind the CarD function has been revealed in an exemplary manner and nothing in the data offers a reason to suggest any mistakes, all research has still been done by a very small circle of people.

4.2 The DksA-guanosine tetraphosphate axis controls stress responses

DksA is a broadly conserved regulatory protein of transcription. While constitutively present in cells, DksA's effects are potentiated by guanosine tetraphosphate (ppGpp) that is mainly synthesised under resource limitation and other stresses. Consequently, guanosine tetraphosphate concentrations vary greatly throughout the lifecycle of the bacterium (Varik *et al*., 2017), suggesting that a change in concentration is responsible for eliciting biochemical consequences. The mechanism of DksA must be different, however, since its concentration is not fluctuating over the life histories of most cells. One possibility is that DksA is responsive to the changes in the physico-chemical environment: *eco* protein at least is sensitive to pH (Furman *et al.*, 2015) and might realign itself into an active conformation upon physical cues (in this case, hydrogen ions). There is a conserved zinc finger in the protein (Perederina *et al.*, 2004). As to the mechanism of action, cryo-EM structures were obtained that indicated contact surfaces to the β' rim helices, active site, bridge helix, the trigger loop and β subunit (Shin *et al.*, 2021). As such, the DksA protein body forms a steric (physical) hindrance for the incoming free nucleotides, thus lowering their availability for RNA biosynthesis. Alternative mechanisms have been proposed. As for the ppGpp, different mechanisms have been described for the different species studied. Barker and others in 2001 noted that the molecule destabilises the open complex in ribosomal RNA promoters, especially. In some species, the mechanism might advance through the regulation of nucleotide pools (Hayruliuk *et al.*, 2015).

Åberg and colleagues (2009) noted that depletion of ppGpp and DksA had resemblant but not equivalent effects. This is a quite significant result; the factors' individual effects had not been studied earlier. However, the importance of understanding the effects separately is clear as DksA is almost constantly present in the cell, while ppGpp is synthesised especially under stress conditions. In their work, Åberg and colleagues quantified transcriptomes of both wildtype and manipulated strains. They found that transcriptomes from ppGpp^0 and *ΔdksA* strains exhibited at least two-fold changes in 6% and 7% of the transcripts; in the double negative ppGpp^0-*ΔdksA* strain such effect was noted in 13% of the genes. These results were obtained using *E*. *coli* as the model organism. Results to the same effect were later reported for another species, *Haemophilus ducreyi*, where 28% and 17% of the transcripts were significantly affected in the ppGpp^0 and *ΔdksA* strains (Holley *et al.*, 2015). Furthermore, some promoters for ribosomal proteins display biassed responses to either ppGpp or DksA (Lemke *et al.*, 2011).

Some generalisations can be made on the target promoters of the axis. The promoters that are activated by the axis usually have AT-rich discriminator regions that are easily caught in the abortive cycling; DksA-ppGpp axis here destabilises the open complex, helping the polymerase

escape from the promoter (Travis and Schumacher, 2022). Conversely, GC-rich discriminator confers susceptibility to DksA-ppGpp-mediated inhibition (Lemke *et al.*, 2011), and a cytidine in the nontemplate strand is especially beneficial. Gummesson and colleagues (2013) reported results in general agreement with these claims. Specifically, they showed that manipulating the discriminator sequence of a strongly activated promoter (in a stress response gene, AAGGA) to the corresponding sequence in a ribosomal promoter (GCGCCACC) was enough to turn the entire gene DksA-inhibited.

The DksA-ppGpp axis functions in the allocation of cellular energy resources. Its chief target are the ribosomal RNAs (whose promoters confer them a tendency to be suppressed), and, according to the current understanding, also the ribosomal proteins. Lemke and others (2011) proved this by selecting seventeen promoters from *E. coli* — including ribosomal RNA, ribosomal protein and metabolic promoters — and transcribing from them *in vitro* with and without DksA and ppGpp. Changes between 4.0-fold and 1.8-fold were observed in the activity of r-protein promoters. Importantly, these studies did not address the question of DksA and ppGpp's differing targets. The structural basis for the additivity of the activities was, however, recently correlated to the binding of ppGpp to the interphase of the core polymerase and DksA (Ross *et al.*, 2016).

A variety of targets beyond ribosomal RNA and r-proteins have been reported. In *Acitenobacter baumannii*, for instance, DksA appears truly pleiotropic (Maharjan *et al.*, 2023). Two points deserve attention: Firstly, Maharjan and colleagues showed that, in addition to just ribosomal or stress response related targets, central carbon metabolism was also greatly affected. Some secondary metabolism targets, like the pathways for aliphatic compound degradation, were affected in addition. Secondly, a curious possibility that more than one response might be coded in the protein was raised. Specifically, experiments showed that the DksA-mediated responses to zinc stress were different from those elicited by copper stress; also interestingly, ppGpp appeared not necessary to combat copper stress, while still being important for zinc stress, on the other hand.

Another example of DksA diversity concerns *Pseudomonas aeruginosa*, a notable human pathogen. In an uncommon fashion, two DksA paralogues have been characterised in its genome. They seem to be mostly, but again not entirely, interchangeable, with small subsets of paralogue-specific regulatory targets (Fortuna *et al*., 2021). (The other paralogue is zinc independent, and mainly expressed in such conditions; Furman *et al.*, 2013; Blaby-Haas *et al.*, 2011.) In *Pseudomonas*, the regulation through DksA overrode other well characterised

regulatory mechanisms like quorum sensing showcasing how these factors do not necessarily operate in isolation but as an additional layer over other mechanisms. It might be possible that quorum sensing mechanisms recruit polymerases to QS-regulated promoters, only some of which environmental conditions would permit transcription from.

The diversity of these biological processes do well to remind the reader to never take the function of a factor at face value; context dependent differences, not immediately obvious in sequence alignments alone, are likely to be found in all studied systems. In a final note of diversity, guanosine tetraphosphate and the related molecules appear to bind to several other targets, too, in addition to RNAP alone (Kanjee *et al.*, 2012). While the present Master's thesis will deploy an *in vitro* methodology, working on purified components, all live cell data will need to be evaluated against the possibility that whatever effects are seen, might stem from a variety of sources. In part, this convolution was recognised early on (Dennis and Nomura, 1974; Dennis and Nomura, 1975), but work disentangling this web is still in its infancy.

II AIMS OF THE STUDY

5 The aims of the present study

Spirochaetes are a unique group of helical, motile Gram-negative bacteria that mostly inhabit anaerobic environments (Paster, 2010). Three factors make these bacteria interesting:

1. Purely intellectually speaking, they are evolutionarily and physiologically distant from *E. coli*, the most common model organism in biochemistry (Hug *et al.*, 2016).
2. They cause disease in both man (e.g., syphilis, aetiology reviewed in Tampa *et al*., 2014; borreliosis, Steere *et al.*, 1983; periodontitis, Simonson *et al*., 1988) and cattle (digital dermatitis, Evans *et al.*, 2008), raising the need to characterise the possible differences of these bacteria from other pathogens.
3. They play a major part in some ecosystems (van de Water *et al.*, 2016), suggesting that understanding their general biology can help to better understand the ecology of these environments as a whole.

Motivated by factors like these, this study set out to explore the functioning of the transcriptional apparatus of *Spirochaeta africana*, a nonpathogenic spirochaete (Zhilina *et al*., 1996). Using purified proteins from this species, our lab previously optimised a fluorescence based method to study biosynthesis of transcription in real time. Here, this methodology was to be applied to understand how two transcription factors known to inhibit rRNA synthesis in different species (CarD and DksA) affected transcription in *S. africana*.

III MATERIALS AND METHODS

6 Production of transcription templates and other DNA

6.1 General remarks on plasmid construction and quality control

1.5 µg of plasmid backbones were linearised overnight (30-µL 10x FastDigest, digesting enzymes, +37 °C) and separated on an agarose gel (loading dye, catalogue number R0611, Thermo Fisher as instructed; 1% agarose in TAE; 80V for 60 min). Backbone fragments were purified from the gel (catalogue number 740609.50, Macheroy-Nagel as instructed) and ligated with likewise ligated fragments (see below) in a Gibson assembly (catalogue number #E2621S, New England Biolabs as instructed). Constructs were transformed to *Escherichia coli* XL-1 with a heat shock and plated; antimicrobial selected cultures (5 mL) were grown and purified (catalogue number 740588.50, Macheroy-Nagel as instructed). Constructs were sequenced to verify their quality (Mix-2-Seq, Eurofins). DNA concentrations were obtained spectroscopically according to the manufacturer's instructions (NanoDrop, Thermo Fisher). Aliquots of actively growing cultures were stored in -80 °C (15% glycerol). For backbone, fragment, primer and construct identities, refer to subsections below.

6.2 *sfc* rRNA promoter construct with a distant Broccoli sequence

The backbone plasmid pOP004 was restricted with XbaI and BamH. In a Gibson assembly, the rrna_prom_native fragment was used. The construct was sequenced using the T7 primer. The construct was named pVN003.

6.3 *sfc* consensual promoter construct with a distant Broccoli sequence

The backbone plasmid pOP004 was restricted with XbaI and BamHI. In a Gibson assembly, the Gre_prom_native fragment was used. The construct was sequenced using the T7 primer. The construct was named pVN001.

7 Production of transcription and other proteins

7.1 Induction of gene products

7.1.1 Overnight induction

A glycerol preparation was plated on an LB plate and grown overnight +37 °C. A colony was used to inoculate a 50-mL preculture in LB and was grown (+37 °C, 250 rpm) until turbid. Turbid preculture was diluted 1000-fold in 4000 mL overnight expression medium (1 mM $MgSO_4$, 100 mM PO_4, 25 mM SO_4, 50 mM NH_4, 100 mM Na, 50 mM K, 0.5% glycerol, 0.05% glucose, 0.2% α-lactose, [0.1% arabinose for Xjb *E. coli* strain] in LB medium) and grown until OD600 stabilised for three hours. Cells were harvested by centrifuging at 7000 × *g* +4 °C for 15 minutes. Cells

were weighted and stored in -80 °C until further use. Unless otherwise specified, 100 µg mL^{-1} ampicillin selection was always used when cells were grown.

7.1.2 IPTG induction

A 50-mL preculture in LB was inoculated from a plate or from a glycerol stock and grown overnight (+37 °C, 250 rpm) or until turbid. Preculture was diluted 1000-fold in 4000 mL of expression culture medium and grown until OD600 reached 0.8. The culture was then inducted using 1 mM IPTG and grown for six hours. Cells were harvested by centrifuging at 7000 × *g* +4 °C for 15 minutes. Cells were weighted and stored in -80 °C until further use. Unless otherwise specified, 100 µg mL^{-1} ampicillin selection was always used when cells were grown.

7.2 Separation of proteins

7.2.1 General remarks on protein chromatography

Proteins were purified using either one step (proteases) or many steps (transcription proteins). Below, the five protein purification methods are specified. In all purification steps, four standard buffers (A, B, L and S) were used. Buffer A contained 50 mM Tris-HCl pH 7.9, 5% glycerol, 0.1 mM Na-EDTA and 10 mM β-mercaptoethanol; buffer B contained 1500 mM NaCl in buffer A. Buffers A and B were filtered and sonicated before use. Lysis (L) buffer contained 50 mM Tris-HCl pH 7.9, 500 mM NaCl, 5% glycerol, 1 mM β-mercaptoethanol and 1 tablet of protease inhibitor (catalogue number A32955, Thermo Scientific) per 15 mL of solution.

7.2.2 Nickel chromatography

Cells were dissolved in buffer L and disrupted by sonication (in cycles of 20s of sonication followed by 40s of recovery until the cells were lysed). Solution was cleared by centrifugation (45,000-50,000 × *g* for 45-60 min) and 0.002% Tween-20 was added. Lysate was then applied to 700-1500 µL nickel sepharose resin (catalogue number 7-5268-01, Cytiva or 17-5318-01, Cytiva) using either gravity flow or incubating the resin in lysate for 60 minutes and collecting it by centrifugation (“batch binding”). The loaded resin was washed with 0 and 20 mM imidazole in buffer L (10 mL), and the protein was eluted using 50, 100, and 200/250 mM imidazole. Resin was prepared and regenerated as instructed by the manufacturer.

7.2.3 Heparin (DNA) affinity chromatography

A 5-mL heparin column (catalogue number 17040701, Cytiva) was used with ÄKTApurifier (GE Healthcare). The fractions from the previous chromatographic step were combined and diluted to <200 mM NaCl in buffer A. The column was equilibrated with 10% buffer B in buffer A. Two column volumes (CV) of flow without a change in conductivity and UV absorption was

considered indicative of a completed equilibration. The sample was loaded to the column using a syringe pump (catalogue number 14-831-200, Thermo Fisher), and eluted with 100 ml gradient of buffer B (10-100%) at a rate of 1 mL min^{-1}. The eluate was collected in 1-mL fractions. The column was used and stored as instructed by the manufacturer. Fractions with highest UV peaks were assumed to contain protein of interest.

7.2.4 Anion exchange chromatography

A 6-mL quaternary ammonium column (catalogue number 17117901, Cytiva) was used with ÄKTApurifier (GE Healthcare). The fractions from the previous chromatographic step were combined and diluted to <150 mM NaCl in buffer A. The column was equilibrated with 10% buffer B in buffer A. Two column volumes (CV) of flow without a change in conductivity and UV absorption was considered indicative of a completed equilibration. The pooled fractions were loaded to the column using a syringe pump (catalogue number 14-831-200, Thermo Fisher), and eluted with 100 ml gradient of buffer B (10-100%) at a rate of 1 mL min^{-1}. The eluate was collected in 1-mL fractions. The column was used and stored as instructed by the manufacturer. Fractions with highest UV peaks were assumed to contain protein of interest.

7.2.5 Concentration and dialysis

Purified protein was centrifuged 4,000 × *g* through a regenerated cellulose membrane with a nominal molecular weight cut-off value of 3,500 Daltons (catalogue number UFC800324, Merck) until retentate volume was <400 µL. Retentate was dialysed through a filter unit with a molecular weight cut-off limit of 3,500 Daltons (catalogue number 69550, Thermo Scientific) for two to three days at +4 °C in excess S buffer to obtain <200 µL protein concentrate.

7.2.6 Quality control and storage of proteins

Aliquots of 20 µL of protein and 5 µL of a molecular weight marker (catalogue number 26619, Thermo Scientific) were separated on a denaturing agarose gel (catalogue number NP0321BOX, Thermo Fisher or NW04120BOX, Invitrogen) for 24 minutes at 200 volts as instructed by the manufacturer. The gels were stained overnight as instructed by the manufacturer (catalogue number 24596, Thermo Scientific) and the purest fractions were used further. Poorly expressed or purified proteins could be discarded. Final concentrations were determined spectroscopically according to the manufacturer's instructions (catalogue number ND2000CLAPTOP, Thermo Scientific). Purified proteins were stored in -20 °C in the S buffer and handled on ice.

7.2.7 *sfc* RNA polymerase

Xjb cells were transformed with pOP009 plasmid and inducted as described in section **7.1**. Then, nickel chromatography (see **7.2.2**), heparin chromatography (**7.2.3**) and anion exchange chromatography (**7.2.4**) were performed on consecutive days. Proteins were concentrated and dialysed as indicated in section **7.2.5**.

7.2.8 *sfc* σ^{70} transcription factor

T7 Express lysY/Iq *E. coli* cells were transformed with pGB217 plasmid and the protein expression was induced with IPTG at room temperature overnight (see **7.1.2**). Then, nickel chromatography (**7.2.2**) and heparin chromatography (**7.2.3**) were performed, and the protein solution was digested overnight with Ulp1 at +4 °C; purification was continued with anion exchange chromatography (**7.2.4**) and concentration and dialysis (**7.2.5**).

7.2.9 *sfc* transcription factors CarD and DksA

The transcription factors *sfc* CarD and DksA were previously purified. Their purification and quality control is indicated in Puro, 2023. Generally, they were expressed in *E. coli* and purified using several chromatographic steps.

8 Following transcription *in situ*

8.1 General remarks on fluorescent light-up aptamer assays

Transcription was followed in real time from purified reaction components following the previously reported protocol (Huang *et al.*, 2022). Briefly, three separate mixtures of excess reaction components were prepared; these were, via one intermediary mixture, combined into the active transcription mixture that was observed spectrophotometrically.

First, the holoenzyme solution was prepared in storage buffer by incubating purified *sfc* RNAP (10 µM) and *sfc* σ^{70} (40 µM) for a minimum of twenty minutes at +30 °C (20-30 minutes were used). In case a transcription factor was used (see **8.3**), it was co-incubated in this mixture to allow for its binding to the holoenzyme. The mixture was transferred to -20 °C after incubation, as previous experiments indicated the holoenzyme maintains its viability better this way (Vilma Trapp, personal communication). In a second mixture, the transcription template (125 nM; see **8.2**) was combined with the fluorophore DFHBI-1T (25 µM) and pyrophosphatase (0.25 µM) in TB10. In a third mixture, equimolar amounts (2 mM) of the ribonucleotides ATP, CTP, GTP and UTP were combined in TB10. The latter mixtures were kept cold and dark.

Second, the first and second mixtures were combined to an intermediary mixture (2 µM *sfc* RNAP, 8 µM *sfc* σ^{70}, 100 nM template DNA and 20 µM DFHBI-1T); this and the third mixture were simultaneously warmed up to +37 °C. Equal volumes of the warmed mixtures were combined into the active transcription mixture. The active transcription mixtures thus contained 1 µM *sfc* RNAP and 4 µM *sfc* σ^{70}; 0.1 µM pyrophosphatase, 10 µM DFHBI-1T and 50 nM linearised template DNA (see **8.2**); 1 mM ATP, CTP, GTP and UTP each; and 40 mM TAPS-KOH pH 9.0, 80 mM KCl, 10 mM $MgCl_2$, 0.1 mM Na-EDTA, 0.1 mM DTT and 5% glycerol from the TB10 buffer. A 55-µL aliquot of this mixture was immediately transferred into a cuvette (catalogue number 105-251-15-40, Hellma Analytics; temperature maintained at +37 °C by water bath).

Finally, aptamer-ligand complexes (complexes of Broccoli RNA and the small fluorogen DFHBI-1T) in the active transcription mixture were excited at the wavelength of 472 nm and fluorescence was observed at the emission wavelength of 507 nm. A measurement was obtained at each second between 20-620 seconds post-initiation (slit widths, 10 nm; LS-55 Fluorescence Spectrometer, Perkin Elmer).

8.2 Transcription from *sfc* rRNA and consensual promoters

Reactions were assembled using either pVN003 (distant Broccoli sequence under the *sfc* rRNA promoter) or pVN001 (distant Broccoli sequence under the *sfc* Gre promoter) as a template. Template DNA was linearised overnight at +37 °C using XhoI as before in 1x Red buffer (catalogue number BR5, Thermo Scientific). Experiments were conducted in duplicate and repeated at least three times. The template that showed less variation (pVN003) in the readings was chosen as the reference reaction condition and was measured in each reaction series.

8.3 Transcription from *sfc* rRNA and consensual promoters with ppGpp, *sfc* CarD and *sfc* DksA

Reactions were supplemented with the final concentration of either 5 µM *sfc* CarD; 10 µM *sfc* CarD; 5 µM *sfc* DksA; 10 µM *sfc* DksA; 5 µM *sfc* CarD and 5 µM *sfc* DksA; 5 µM *sfc* DksA and 1 mM ppGpp; or 1 mM ppGpp alone in the active transcription mixture. Experiments were conducted in duplicate and experiments with noticeable effects were replicated for statistics. Where possible, all independent repeats were performed using proteins from the same production batch. At least one reaction with pVN003 without additional compounds was used as a reference in every reaction series.

9 Statistical procedures and data

9.1 Data normalisation

All experiments were normalised against a reference reaction. Transcription factor-independent transcription from rRNA promoter was chosen as the reference reaction and remeasured alongside all other experimental conditions. The raw data from each measurement was normalised by dividing it with the end fluorescence of the corresponding reference reaction. Data was *not* manipulated to start from the same RFU value.

9.2 Statistical analyses

A minimum of three independent repeats were obtained for experiments whose preliminary data indicated significant effects. Data was assumed normally distributed and analysis of variance (ANOVA) was used to compare means of many groups simultaneously; a post-hoc Tukey test was used in pairwise comparisons between the groups to identify the statistically significantly different populations. In case of just two groups compared, one-way ANOVA or Student's *t* test was used. In all statistical tests, the level of significance was set at α=0.05. Data was excluded from the analyses in the case of clear sources of error in the experiments and consequently flawed data. Reanalysis of data was performed for some experimental setups and the rationale of doing so is indicated in the Results section. Analyses were performed on SPSS and Excel.

9.3 Representation of data as graphs

All graphs were created using the Origin software package. Briefly, data was mean and standard deviation of experiments were represented as a function of time. Google Slides was used to create schematics.

IV RESULTS

10 The results from transcription experiments

10.1 Transcription from *sfc* rRNA and consensual promoters

Transcription from the ribosomal and consensual promoters was measured using fluorescent light-up aptamers (FLAPs). Briefly, purified transcription components in these experiments were assembled *in vitro*. Transcription templates coded for a Broccoli aptamer sequence (Huang *et al*., 2022) under either the ribosomal RNA promoter of *Spirochaeta africana* or the spirochaetal consensual promoter. Their transcripts specifically recognise a small fluorogen in the solution (DFHBI-1T), forming a fluorescent complex. The accumulation of fluorescence in the reactions was followed for approximately eleven minutes after the start of the reactions *in situ* to understand the kinetics of transcription. All fluorescence data is reported as relative fluorescence as compared to the fluorescence readings of the ribosomal RNA promoter (the mean mathematically manipulated to be 100). The data here is reported as the mean of a minimum of three independent replicates (line), surrounded by the standard deviation (coloured area surrounding the line). At least three independent repeats were obtained for all experiments.

Transcription from the consensual promoter construct was 140% higher than from the rRNA promoter (average relative fluorescences 240 RFU and 100 RFU, respectively; **Figure 10.1**), a difference which one-way ANOVA confirmed as statistically significant (p=0.0497). Lower scattering was observed in the rRNA promoter (98-102 RFU against 175-345 RFU in the consensual promoter). rRNA promoter was thus chosen as the reference reaction for normalisation.

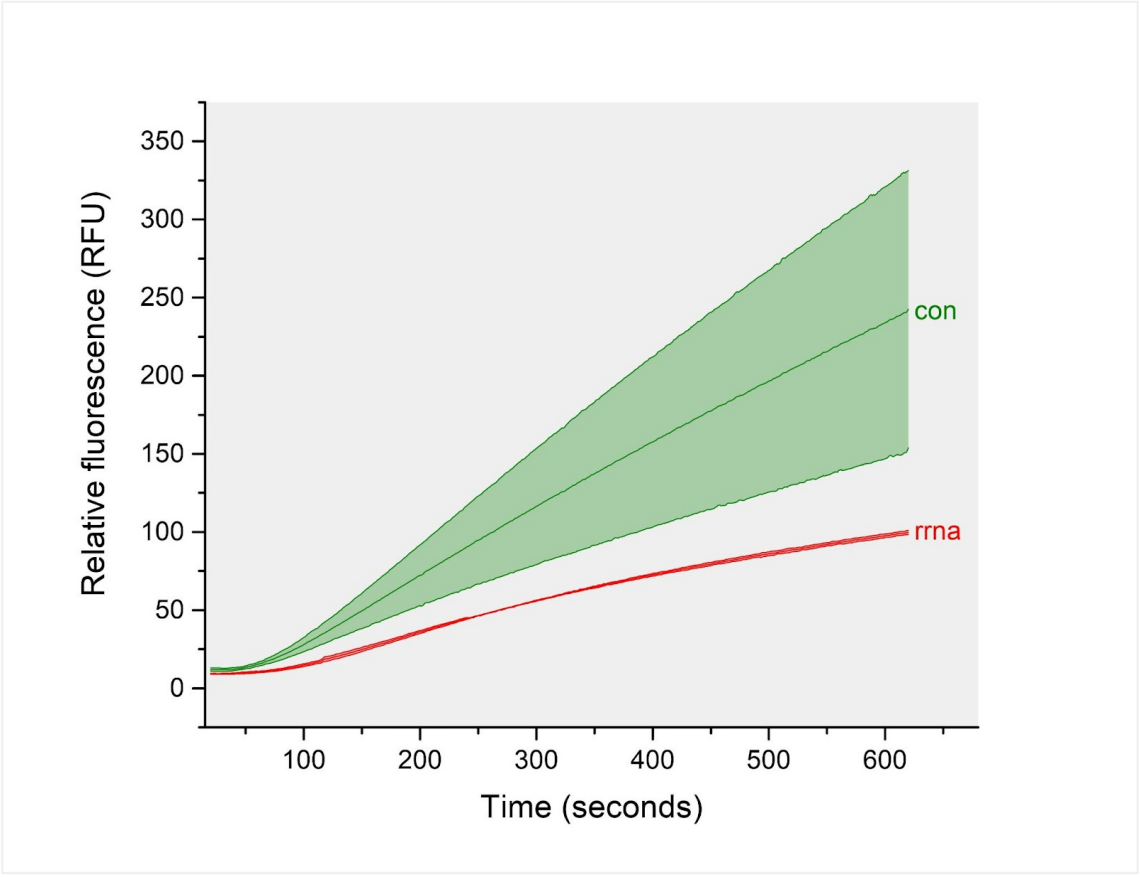


Figure 10.1 | **Transcription from the two promoters without any additional transcription factors.** Transcription reactions (1 µM *sfc* RNAP, 4 µM *sfc* σ^{70}, 0.1 µM pyrophosphatase, 10 µM DFHBI-1T, 50 nM linearised template DNA, 1 mM ATP, 1 mM CTP, 1 mM GTP, 1 mM UTP, 40 mM TAPS-KOH pH 9.0, 80 mM KCl, 10 mM $MgCl_2$, 0.1 mM Na-EDTA, 0.1 mM DTT, 5 % glycerol) were warmed up to +37 °C and combined *in vitro*. Fluorescence was quantified from 55-µL aliquots between 20-620 seconds post-mixing (LS-55 Fluorescence Spectrometer, Perkin Elmer) in a cuvette with a 3 x 3 mm light path (catalogue number 105-251-15-40, Hellma Analytics) as indicated in the methods. The mean and standard deviation (*n*=3) is represented.

_____

10.2 Influence of CarD and DksA on transcription from *sfc* promoters

The influence of two transcription factors, CarD and DksA, were studied by assembling and quantifying transcription reactions promoter-wise as before and supplementing them with purified CarD or DksA. 5 µM and 10 µM transcription factor was used, were used for all other experiments except for DksA in the case of the consensual promoter, where only 5 µM DksA was used.

In the case of the rRNA promoter, opposing effects were observed for supplementation with CarD or DksA. Specifically, the addition of CarD (**Figure 10.2a**) increases transcription by approximately 190% from 100 RFU to 291 RFU (234-342 RFU, 5 µM CarD) and 290 RFU (233-372 RFU, 10 µM CarD). One-way ANOVA indicated a significant difference between at least one pair of experiments (p=0.0055), and a post-hoc Tukey test was used to determine the groups between which these differences existed. Briefly, both CarD-supplemented reactions were statistically different from the CarD-independent experimental setup (p=0.0090670 for 5 µM CarD; p=0.0086854 for 10 µM CarD), while no difference was found between the two concentrations of CarD (p=0.8999947). These results indicate that while CarD significantly and strongly increases transcription from the rRNA promoter, the effect saturates at or below 5 µM CarD, at least *in vitro*.

As for DksA, the biological effect is reversed. The averaged end fluorescences for DksA-supplemented reactions were 85 RFU (79-97 RFU, 5 µM DksA) and 81 RFU (74-91 RFU, 10 µM DksA), respectively, representing a 15 to 19% decrease (**Figure 10.2b**). Statistically, one-way ANOVA indicated no significant difference between these groups and the DksA-independent reference group (p=0.0512); however, in the experiments themselves, DksA-supplemented reactions constantly showed lower fluorescences than DksA-independent reactions when assembled from the same components, still suggesting a possible biological relevance of the results. Under the assumption that one value (91) in the 10 µM series is not normally distributed but rather is an outlier and can be excluded from the analysis, one-way ANOVA would, conversely, indicate statistically significant differences (p=0.0224), and a post-hoc Tukey test would reveal this to be the case for the difference between the means of DksA-independent and 10 µM DksA groups (p=0.0214259).

The effects were similarly analysed for the consensual promoter. CarD-supplementation indicated an opposing, transcription-inhibiting effect from what is seen with the rRNA promoter (**Figure 10.2c**). Specifically, the two CarD-supplemented reactions had lower end fluorescences of 145 RFU (122-178 RFU, 5 µM CarD; 40% lower) and 161 RFU (160-162, 10 µM CarD; 33% lowering) compared to the CarD-independent reaction (243 RFU, 176-343 RFU). One-way ANOVA indicated no statistically significant differences between the means of these groups (p=0.1346). This statistical insignificance is due to large scatter in the CarD-independent series. The biological effect, however, was replicated in all experiments: when CarD-supplemented and CarD-independent reactions that were assembled on the same day from the same biological reagents, lower readings were always observed in the supplemented reactions.

DksA has no effect on transcription from the consensual promoter (**Figure 10.2d**). Reaction supplemented with 5 μM DksA produces, on average, end fluorescence of 252 RFU (124-327 RFU; 4% higher). One-way ANOVA indicated no statistically significant differences between the two groups (p=0.9088), and no further tests were conducted with 10 μM supplementation.

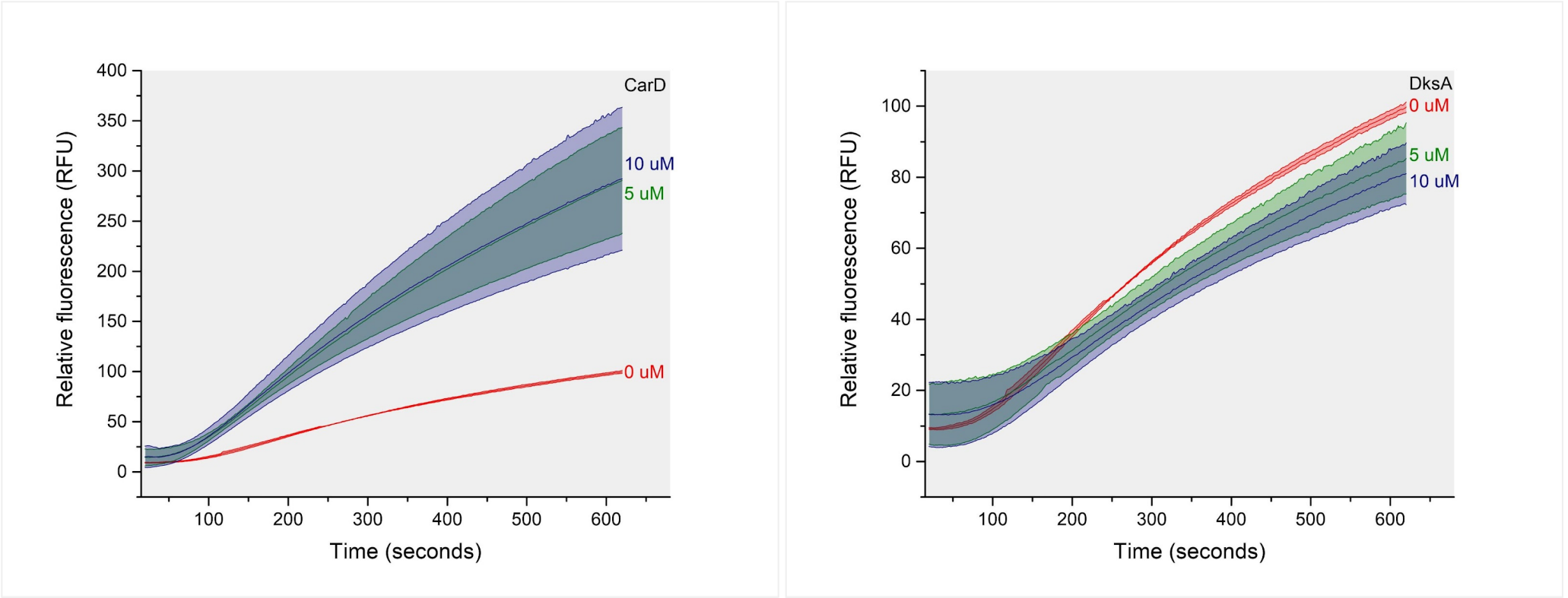


Figure 10.2 a, b | **Effects of CarD** (left) **and DksA** (right) **on transcription from the *sfc* rRNA promoter.** Transcription reactions (1 µM *sfc* RNAP, 4 µM *sfc* σ^{70}, *sfc* transcription factor, 0.1 µM pyrophosphatase, 10 µM DFHBI-1T, 50 nM linearised template DNA, 1 mM ATP, 1 mM CTP, 1 mM GTP, 1 mM UTP, 40 mM TAPS-KOH pH 9.0, 80 mM KCl, 10 mM $MgCl_2$, 0.1 mM Na-EDTA, 0.1 mM DTT, 5 % glycerol) were warmed up to +37 °C and combined *in vitro*. Fluorescence was quantified from 55-µL aliquots between 20-620 seconds post-mixing (LS-55 Fluorescence Spectrometer, Perkin Elmer) in a cuvette with a 3 x 3 mm light path (catalogue number 105-251-15-40, Hellma Analytics) as indicated in the methods. The mean and standard deviation (*n*=3) is represented.

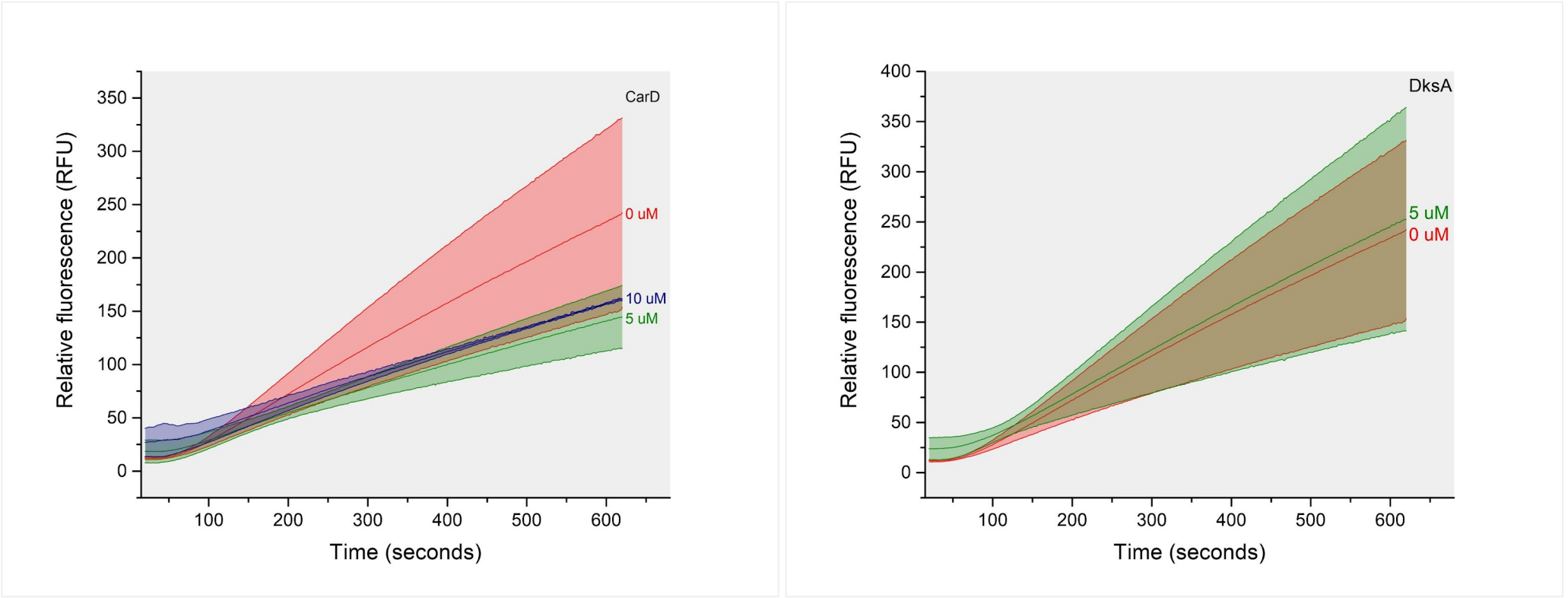
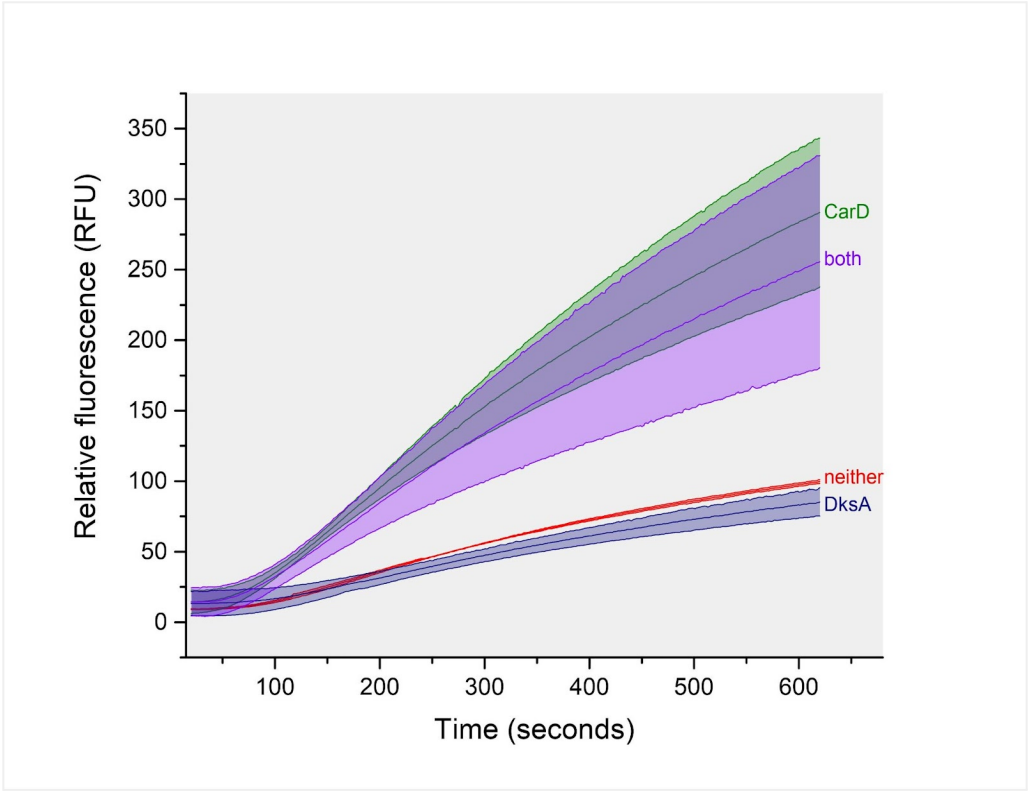


Figure 10.2 c, d | **Effects of CarD** (left) **and DksA** (right) **on transcription from the *sfc* consensual promoter.** Transcription reactions (1 µM *sfc* RNAP, 4 µM *sfc* σ^{70}, *sfc* transcription factor, 0.1 µM pyrophosphatase, 10 µM DFHBI-1T, 50 nM linearised template DNA, 1 mM ATP, 1 mM CTP, 1 mM GTP, 1 mM UTP, 40 mM TAPS-KOH pH 9.0, 80 mM KCl, 10 mM $MgCl_2$, 0.1 mM Na-EDTA, 0.1 mM DTT, 5 % glycerol) were warmed up to +37 °C and combined *in vitro*. Fluorescence was quantified from 55-µL aliquots between 20-620 seconds post-mixing (LS-55 Fluorescence Spectrometer, Perkin Elmer) in a cuvette with a 3 x 3 mm light path (catalogue number 105-251-15-40, Hellma Analytics) as indicated in the methods. The mean and standard deviation (*n*=3) is represented.

10.3 The combined effects of CarD and DksA on transcription from *sfc* promoters

Having established the effects of the two transcription factors independently of one another, their combined effects were evaluated. In these experiments equimolar CarD and DksA (5 µM) were added to the reaction and the reactions were quantified as before. The combined supplementation of CarD and DksA increased normalised average fluorescence by 155% from 100 to 255 RFU (202-341 RFU) from the rRNA promoter (**Figure 4.3a**); there was considerable scatter in the upper end of the measurements in the supplemented reaction. This value was analysed together with the previously obtained results that described the non-supplemented reaction and single factor-supplemented reactions; one-way ANOVA indicated a significant difference existed between a pair of experiments (p=0.0009). A post-hoc Tukey test indicated significant differences between Independent vs CarD-and-DksA groups (p=0.0065966), Independent vs CarD-only groups (p=0.0058894), CarD-only vs DksA-only groups (p=0.0034333), and DksA-only vs CarD-and-DksA groups (p=0.0038269). Statistically significant differences were not shown in Independent vs DksA-only groups (p=0.8999947), but also not in CarD-only vs CarD-and-DksA groups (p=0.8999947). No experiment was conducted using other concentrations of CarD and DksA. These results suggest DksA's transcription-inhibiting effect, itself not statistically significant but constantly reproduced in experiments, is subordinate to the effects of CarD.

Similar experimental setup lowered transcription from the consensual promoter by 57% from 234 RFU to 99 RFU (77-124 RFU; **Figure 4.3b**). Statistically significant differences were not identified between the four groups (one-way ANOVA, p=0.0893). Reanalysis of data was performed by excluding one possibly outlying data point from the DksA-only group (124) and comparing CarD-only, DksA-only and CarD-and-DksA groups (p=0.0005); a post-hoc Tukey test showed statistically significant differences between CarD-only vs DksA-only groups (p=0.0014040) and DksA-only vs CarD-and-DksA groups (p=0.0010053) and no difference between CarD-only vs CarD-and-DksA groups (p=0.1528825). This suggests that, in the case of consensual promoter and the presence of both of the transcription factors, CarD has a dominating effect over DksA.



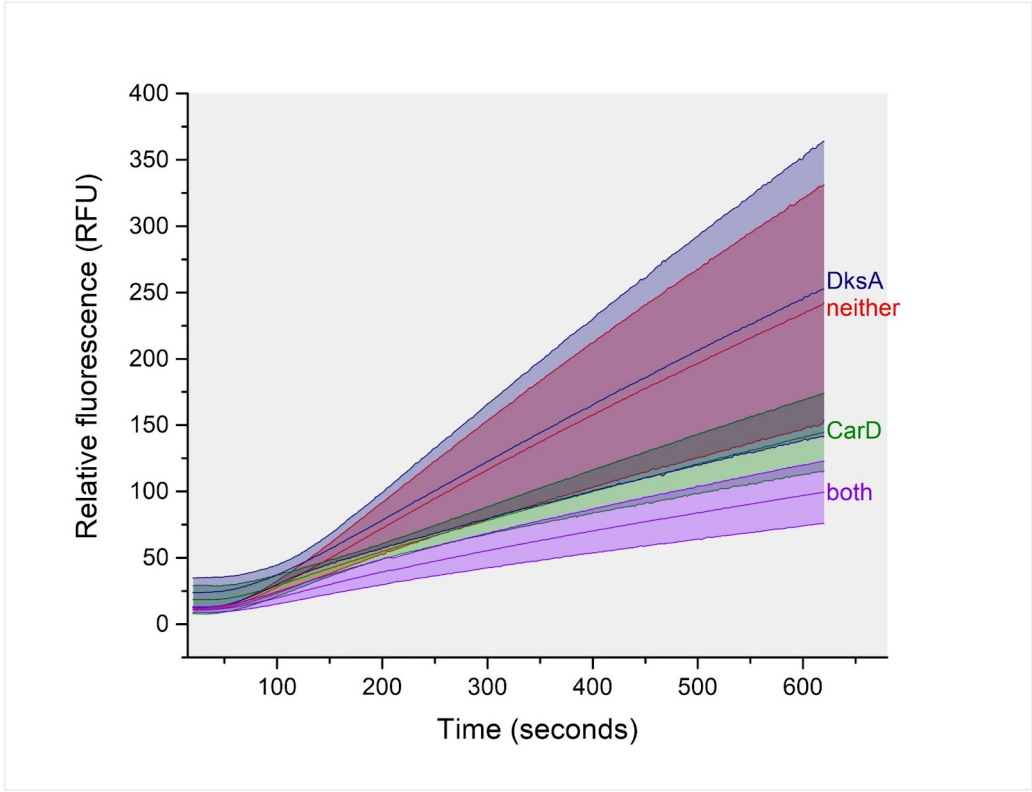


Figure 10.3 a, b | **Combined effect of CarD and DksA on transcription from the *sfc* rRNA promoter** (left) **and *sfc* consensual promoter** (right)**.** Transcription reactions (1 µM *sfc* RNAP, 4 µM *sfc* σ^{70}, *sfc* transcription factors, 0.1 µM pyrophosphatase, 10 µM DFHBI-1T, 50 nM linearised template DNA, 1 mM ATP, 1 mM CTP, 1 mM GTP, 1 mM UTP, 40 mM TAPS-KOH pH 9.0, 80 mM KCl, 10 mM $MgCl_2$, 0.1 mM Na-EDTA, 0.1 mM DTT, 5 % glycerol) were warmed up to +37 °C and combined *in vitro*. Fluorescence was quantified from 55-µL aliquots between 20-620 seconds post-mixing (LS-55 Fluorescence Spectrometer, Perkin Elmer) in a cuvette with a 3 x 3 mm light path (catalogue number 105-251-15-40, Hellma Analytics) as indicated in the methods. The mean and standard deviation (*n*=3) is represented.

10.4 Guanosine tetraphosphate's effect on transcription from the *sfc* consensual promoter

Having established the role of the two proteins in regulating the two promoters, I wished to evaluate the effect, if any, introduced by guanosine tetraphosphate (ppGpp). In these experiments, the *in vitro* reactions were assembled as before and supplemented with 1 mM ppGpp. The effect of ppGpp-supplementation was investigated in three experimental setups: non-supplemented (**Figure 10.4a**), CarD-supplemented (**Figure 10.4b**) and DksA-supplemented (**Figure 10.4c**). All experiments used the same concentration of transcription factors (5 µM); material was not sufficient to probe the effect of ppGpp on CarD-and-DksA-supplemented reactions. Numerically, supplementing ppGpp to reactions with no transcription factors decreased transcription by 11% from 243 RFU (176-343 RFU) to 216 RFU (158-274 RFU). The supplementation of CarD-only reaction with ppGpp further lowers transcription by 19% from 144 RFU (122-178 RFU) to 116 RFU (114-118 RFU). As per DksA, the corresponding change is 253 RFU (124-327 RFU) to 252 RFU (240-276 RFU), respectively, signifying a decrease by 0%. (Again excluding the possibly outlying result of 124, the corresponding numbers would be 317 RFU to 253 RFU, or a 20-percent decrease.) Finally, CarD-only, DksA-only, CarD-and-ppGpp, and DksA-and-ppGpp groups were compared statistically; one-way ANOVA returned a *p*-value of p=0.000029337. A Tukey post-hoc test indicated that statistically significant differences between CarD-only vs DksA-only groups (p=0.0010053), CarD-only vs DksA-and-ppGpp groups (p=0.0012391), DksA-only vs DksA-and-ppGpp groups (p=0.0359847), DksA-only vs CarD-and-ppGpp groups (p=0.0010053), and DksA-and-ppGpp vs CarD-and-ppGpp groups (p=0.0010053) exist. CarD-only vs CarD-and-ppGpp groups were not likely to differ in a statistically significant manner (p=0.3617754).

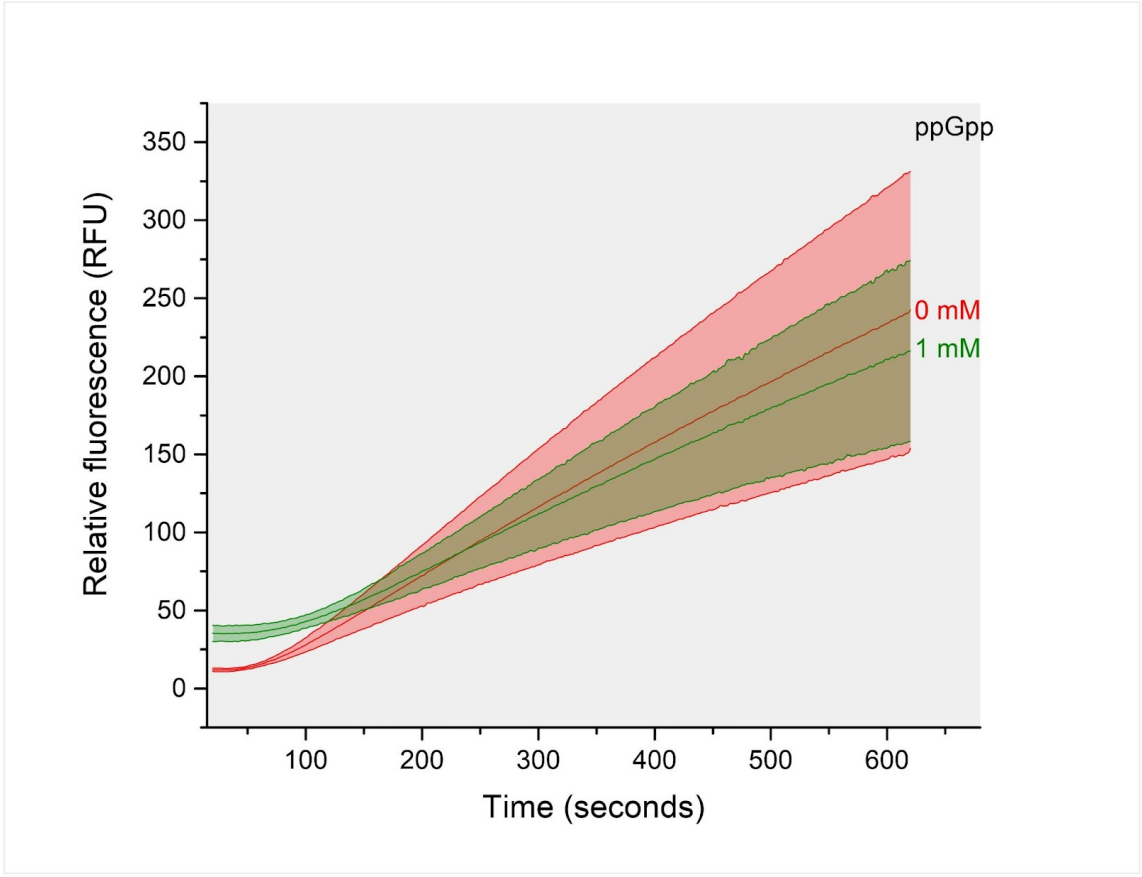


Figure 10.4 a | **Effect of guanosine tetraphosphate (ppGpp) on transcription from *sfc* consensual promoter independently of other transcription factors.** Transcription reactions (1 µM *sfc* RNAP, 4 µM *sfc* σ^{70}, 1 mM ppGpp, 0.1 µM pyrophosphatase, 10 µM DFHBI-1T, 50 nM linearised template DNA, 1 mM ATP, 1 mM CTP, 1 mM GTP, 1 mM UTP, 40 mM TAPS-KOH pH 9.0, 80 mM KCl, 10 mM $MgCl_2$, 0.1 mM Na-EDTA, 0.1 mM DTT, 5 % glycerol) were warmed up to +37 °C and combined *in vitro*. Fluorescence was quantified from 55-µL aliquots between 20-620 seconds post-mixing (LS-55 Fluorescence Spectrometer, Perkin Elmer) in a cuvette with a 3 x 3 mm light path (catalogue number 105-251-15-40, Hellma Analytics) as indicated in the methods. The mean and standard deviation (*n*=3) is represented.

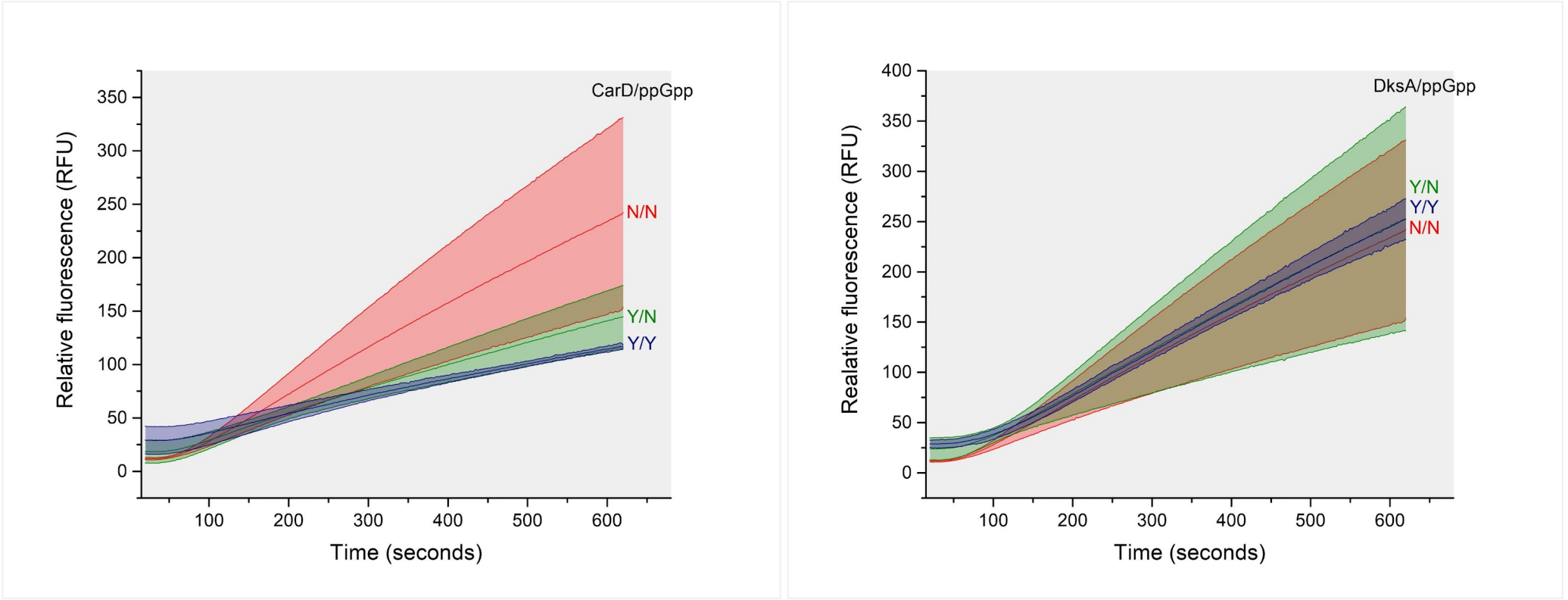


Figure 10.4 b, c | **Effects of guanosine tetraphosphate (ppGpp) on transcription from sfc consensual promoter together with CarD** (left) **and DksA** (right)**.** Transcription reactions (1 µM *sfc* RNAP, 4 µM *sfc* σ^{70}, *sfc* transcription factor, ppGpp, 0.1 µM pyrophosphatase, 10 µM DFHBI-1T, 50 nM linearised template DNA, 1 mM ATP, 1 mM CTP, 1 mM GTP, 1 mM UTP, 40 mM TAPS-KOH pH 9.0, 80 mM KCl, 10 mM $MgCl_2$, 0.1 mM Na-EDTA, 0.1 mM DTT, 5 % glycerol) were warmed up to +37 °C and combined *in vitro*. Fluorescence was quantified from 55-µL aliquots between 20-620 seconds post-mixing (LS-55 Fluorescence Spectrometer, Perkin Elmer) in a cuvette with a 3 x 3 mm light path (catalogue number 105-251-15-40, Hellma Analytics) as indicated in the methods. The mean and standard deviation (*n*=3) is represented.

_____

V DISCUSSION

11 Discussion

From this work, several conclusions can be drawn to a reasonable degree of certainty. They are as follows:

1. In *Spirochaeta africana*, transcription from the native rRNA promoter is rate-limited by the formation of the open complex; the metabolic promoter's open complex is more stable.
2. In *Spirochaeta africana*, the regulation of ribosomal biogenesis differs from that of *E. coli*, while its similarity or difference to that of *M*. *tuberculosis* remains unsettled.
3. In the case of the *sfc* metabolic promoter, the effects of CarD and ppGpp might be additive, supporting the existence of a CarD-ppGpp axis in spirochaetes.

11.1 Differences in RP_o stability confer differential regulatory properties for metabolic and ribosomal promoters in *S. africana*

To understand *Spirochaeta africana* molecular biology, I in this study explored transcription from two different promoters. To establish a baseline against which it is possible to analyse other results, I first compared transcription by the $E\sigma^{70}$ holoenzyme from the promoters alone, without any additional transcription factors. A statistically significant difference in transcription rates was observed in the favour of the consensual promoter (140% stronger than the rRNA promoter). Results to this effect were obtained by transcribing polymerases from multiple purification batches, eliminating the possibility that the effect was limited to a single batch. To gain a deeper understanding of the spirochaetal transcription system, and of its possible differences to that of other bacteria, however, I wanted to understand the biophysical basis of this difference in transcription rates. For this reason, experiments with different transcription factors were performed.

The key to these experiments is transcription's nature as an equilibrium reaction (**Equation 3.2.1a**). There is a simple consequence from this nature: since there is one reaction alone in the equilibrium that can be approximated as an irreversible reaction (that is, the promoter clearance; Hsu, 2002), the commitment of a holoenzyme promoter complex to that reaction will remove it from the left side of the equilibrium between the other intermediates. In other words, increasing the relative population of the open complex will likely be seen as an increase in transcriptional output. If what separates the two studied promoters is the stability of the open complex, increasing the stability of this intermediate will increase transcription; if the bottleneck of the reaction is earlier with promoter recognition, stabilising the open complex will not help the reaction notably as the complexes fail to reach that stage either way.

Considering then that *Spirochaeta africana* genome encodes CarD, a well-characterised transcriptional regulator that increases the formation of the open complex (and simultaneously decreases its collapse back to the closed complex), I supplemented the transcription reaction from the rRNA promoter with CarD. These experiments produced, using two concentrations, an increase in transcription that averaged 190%. It can, then, be concluded that the promoter's transcription was limited by the formation and the stability of the open promoter complex. This implies, in turn, the following: there are no problems in the recruitment of the polymerase to the rRNA promoter. Putting this in context of the higher transcription from the metabolic promoter, two further conclusions can be drawn: as the metabolic promoter is more strongly transcribed, *it cannot experience massive problems in the recruitment of the holoenzyme, either*. Therefore, the difference in transcription strengths must come down to the stability of the open complex. Following this logic, a larger fraction of the not-escaped metabolic promoter complexes can be found at any given moment in the open complex intermediate than those of the rRNA promoter.

Supposing the previous, it seems logical to assume that DksA, which in *E. coli* destabilises the open complexes of the rRNA promoters (Barker *et al*., 2001), would decrease transcription from the promoter. Indeed, the DksA-supplemented transcription from the rRNA promoter is lower than transcription without DksA, but only by 15-19%. (See Results for whether this is a statistically significant difference.) In case of the metabolic promoter, an incremental increase in transcription was observed. Considering how stabilising effect from CarD was enough to severely impair transcription from the metabolic promoter, possibly by overstabilising its open complex, the transcription-promoting effect of DksA might suggest that the Gre promoter is very close to the limit of being rate-limited by the transitioning from the open promoter complex to the processive elongation; slight destabilisation of the open complex could then tilt the energetic profile in a way that would promote promoter clearance and processive elongation.

The conclusion from the experiments is, undoubtedly, that the two promoters are differently regulated. My experiments with RP_o-stabilising and RP_o-destabilising transcription factors also indicated that the *sfc* rRNA promoter is rate-limited by the formation of the open complex. The metabolic promoter, on the other hand, is stabler than the rRNA promoter to begin with, explaining its decrease of function when stabilised further. The results here are in broad agreement with those published in the literature, but, in some instances, differ slightly. As per the agreement, it has been described previously that both CarD and DksA can function in both

promotion and inhibition of transcription from different genes (Kaur *et al*., 2018; Huang *et al*., 2022). Two possible discrepancies are worthy of mention: In *E. coli*, purified DksA caused approximately a 20-fold decrease in rRNA transcription (Paul *et al.*, 2004), while here a decrease of only 15-19% was described. Also, in the case of CarD, some indication exists that the factor's main function would be to overstabilise open complexes of the promoters that are responsible for ribosome biogenesis, thus preventing transcription from them; this can be inferred from an early report that *ΔcarD* strains are nonviable due to their inability to stop ribosome biogenesis (Stallings *et al*., 2009). On the contrary, I noticed here that CarD increases transcription from at least one ribosomal promoter; in fact, a similar result was described by Zhu and Stallings (2023).

11.2 Regulatory mechanism of the ribosomal biogenesis might differ in *S. africana* from other model organisms

The experiments I used to infer the relative stabilities of the two promoters implied that CarD and DksA have opposing effects on the rRNA promoter (and possibly on the metabolic promoter also). As has been mentioned earlier, in many species, it is the DksA axis that negatively regulates ribosomal biogenesis. However, the relatively small effect of DksA on the *sfc* rRNA promoter raises the question of whether a transcription factor with such a limited effect could be functional as a metabolic switch with accentuated effects. Specifically, as DksA's influence on rRNA transcription is very small, and at least separately tested, that of CarD is more pronounced but in a different direction, DksA's potential as a chief regulator of ribosome biogenesis can be rightfully questioned. Admittedly, as the mechanisms of action of the two transcription factors are different, their combined effect need not be the exact sum of the two individual effects; that is, when both of the factors are present, the other one's biological effect might dominate disproportionately from what could be assumed from just calculating the combined effect algebraically. To better understand the relative roles of CarD and DksA, then, and to explore the possibility of DksA still, I performed experiments where both of the transcription factors were supplemented to the transcription reactions.

Such experiments were performed on both of the promoters. In the case of the rRNA promoter, the observed effect was only slightly (but not statistically significantly) different from that of CarD's effect alone (**Figure 10.3a**). Context will clarify the meaning of these results. Firstly, experiments that were performed using one transcription factor only (CarD or DksA) cause significant changes in transcription from the rRNA promoter, for example; however, the different concentrations (5 μM and 10 μM) do not seem to induce significantly different effects. Thus it can be assumed that the saturating condition of the transcription factors are reached at

five micromoles or less in the *in vitro* system. Secondly, such biologically saturating effects are quite possibly present in living cells. The logic here is as follows. One, the copy number of the rRNA promoter is, in living cells, considerably smaller than in this experimental setup (see Methods). This is clear since the experimental setup uses nanomolar template, whereas the sequenced *sfc* genome only contains 53 RNA genes and <2800 protein coding genes (Liolos *et al*., 2010). Two, in *E. coli* the most studied transcription factors are present in the copy numbers of hundreds or low thousands (Ishihama *et al*., 2014) and the RNA polymerase, respectively, in some 5000 copies (Bakshi *et al*., 2012), we can infer, assuming similar levels, that upon *sfc* rRNA RP_c formation, biologically meaningful concentrations of the factors are likely to be present. Taken together, biologically meaningful concentrations of the molecules are likely to be found in a living cell should the aforementioned assumptions hold; this effectively suggests that the CarD-dominated behaviour in CarD-and-DksA supplemented *in vitro* reaction is, to an extent at least, reflective of the biological environment.

This points to a discrepancy between the biochemical logic of *E. coli* and *S. africana*. If at the saturating conditions the effect of CarD dominates, the regulation of ribosome biogenesis through the adjustment of DksA levels becomes futile. An interesting similarity, however, can be noted between these results and what was previously described for *M*. *tuberculosis*.
As noted above, the *rrnA* P3 promoter is significantly upregulated by CarD in this species.
I suggest, in light of this, a possible regulatory mechanism for ribosomal biogenesis in *S. africana*: namely that it might operate through adjusting the levels or activity of CarD rather than DksA. In this proposed paradigm, the baseline of ribosome biogenesis would be set by the expression of CarD under normal circumstances and the stringent response-like down regulation would be realised through the cessation of CarD synthesis or other mechanisms that shrink its activity on the promoter. Clearly, more research is urgently needed to set these preliminary findings in their appropriate proportions.

11.3 *sfc* transcriptional apparatus is targeted by the guanosine tetraphosphate but the finer details require more research

Having established that in the case of rRNA promoter, CarD's effect seems to be biologically dominant, the same type of experiments were performed on the metabolic promoter. In essence, the results here agree with those obtained using the rRNA promoter: the dominating effect is that of CarD. In the case of the metabolic promoter, however, it is DksA that increases transcriptional output and CarD that inhibits it, bringing the saturated combined clearly on the negative side (**Figure 10.3b**), or put otherwise, inhibiting transcription. In the experiments where one transcription factor alone was used, CarD seemed to decrease transcription by a

third but any statistically meaningful differences were lost to the high scatter in the non-supplemented reactions. (**Figure 10.2d**); in an interesting discontinuity, DksA was not found to increase transcription from this promoter. In principle what this shows is that the two studied promoters are not polar opposites; their regulation through CarD is to a different effect, but responses to at least DksA are not each other's mirror images.

In principle, the stringent response can either decrease or increase transcription of a gene (Irving *et al*., 2021), presumably with the goal of reassigning the cellular energy resources to more reasonable use. Considering that a SpoT-family protein, responsible for the synthesis of ppGpp, is encoded in the S. africana genome, I became interested in the possible modulating effect of the DksA's effect by ppGpp. In short, the effect of ppGpp on transcription from the promoter in the absence of DksA was measured to deduce the independent effect of the molecule. Then, DksA-supplied reactions were spiked with ppGpp; ambiguous results were obtained (**Figure 10.4c**). For one, alone the molecule seems to slightly decrease transcription. This is in line with the previously published reports according to which the main mechanism of action of ppGpp is through the destabilisation of the open promoter complexes; thus, a decrease in transcription is a logical outcome when a promoter with presumably stable open complex is studied. Less clear is the result of the experiment with DksA and ppGpp, where two "opposing forces" are, presumably, at play: a small but statistically insignificant decrease in transcription was observed. The result is inconclusive, and it is possible that larger sample sizes could allow for the finding of a statistically significant difference; in the absence of them, however, one can already concur that even if such statistical significance could be shown, the biological significance of the effect would be minimal.

Finally, as ppGpp seemed to decrease transcription from the promoter, rather than work in unison with DksA, the effects of CarD and ppGpp were explored together. This line of research was purely explorative but yielded interesting results. Namely, in comparison to the ppGpp or CarD alone, a statistically significant combined effect was noted in the experiment that combined CarD to ppGpp (**Figure 10.4b**). Indeed, this is an indication that while CarD alone can decrease transcription from the promoter, the combined effect with ppGpp is even more accentuated. In other words, there can possibly be another regulatory axis at play in spirochaetes. It must be admitted that the difference between CarD and CarD-and-ppGpp is not particularly large (decreases of 11% vs 19%, respectively) but its biophysical basis is interesting. CarD's presumed mechanism of action is through the stabilisation of the open complex; as has now been mentioned many times, ppGpp's is that of destabilising it (Barker *et al*., 2001). This of course seems quite counterintuitive at first, and the data I have accrued

cannot be used to say anything definitive about the mechanism that is in the works here. Three facts or assumptions allow one, however, to accept this conundrum, even if they don't offer an obvious resolution. Firstly, Rammohan and colleagues in 2015 suggested for CarD a two-tiered mechanism of action. In this model, CarD both increases the formation of the open complex and decreases its collapse back to the closed complex (these two need not necessarily be coupled). Secondly, ppGpp is thought to modify the overall conformation of the polymerase; in *E. coli* this leads to the diminished efficiency of the RP_{itc}, leading to less probable stabilisation of the open complex by RNA:DNA hybrid. Combining the many mechanisms involved and remembering that in use here is a new species of polymerase, lineage-specific effects cannot be ruled out. Thirdly, the existence of a CarD-ppGpp axis, while counterintuitive at first, is an established phenomenon (Stallings *et al.*, 2009). Specifically, their synergy has been previously described, raising confidence that more research will both corroborate this result and, hopefully, explain its underlying biophysics.

VI CONCLUDING REMARKS

12 Perspective and path forward

In this work, transcription was framed as a multifaceted process that is driven by the biophysics between promoters and holoenzymes transcribing them. Using a previously unexplored model organism, *Spirochaeta africana*, important characteristics were shown. Firstly, the ribosomal and metabolic promoters' relative strengths were characterised and the metabolic promoter was shown to be stronger; secondly, a possible mechanism relating to the regulation of ribosome biosynthesis was explored; and finally, cooperativity in CarD-ppGpp axis was shown.

This work offers preliminary insight into the transcriptional apparatus of *Spirochaeta africana*. It has been constrained by the lack of previous research; if anything the results highlight a need to further explore the topics as the promoter behaviour was noted to be in many ways different from the comparable networks in *E. coli*. It is possible that these avenues of research might lead one day to more effective medications or, possibly, conservation efforts in the ecosystems that depend on spirochetes. To get there, however, work on these promoters must be expanded to other promoters and, eventually, a transition to studies on live cells must be made. In the meantime, transcriptome sequencing and profiling of the RNAP localisation in the genome should offer valuable new data.

Lastly, and to myself most importantly, this work shows yet again that what is true for *E. coli* is not necessarily true for an elephant. In fact, the advancing high throughput methodologies and possibility to obtain structural data of even transient complexes seem to continuously revolutionise our understanding of even the most fundamental processes of life. Ultimately, then, and not surprisingly, life is much more complicated, in ways we cannot always see beforehand, than could be inferred from the undergraduate textbook; the task remains to make sense out of it.

VII LITERATURE CITED

Aoyama T, Takanami M, Ohtsuka E, et al. Essential structure of E. coli promoter: effect of spacer length between the two consensus sequences on promoter function. Nucleic Acids Res. 1983;11(17):5855-5864. doi:10.1093/nar/11.17.5855.

Artsimovitch I, Svetlov V, Murakami KS, Landick R. Co-overexpression of Escherichia coli RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. J Biol Chem. 2003;278(14):12344-12355. doi:10.1074/jbc.M211214200.

Bae B, Davis E, Brown D, Campbell EA, Wigneshweraraj S, Darst SA. Phage T7 Gp2 inhibition of Escherichia coli RNA polymerase involves misappropriation of σ70 domain 1.1. Proc Natl Acad Sci U S A. 2013;110(49):19772-19777. doi:10.1073/pnas.1314576110.

Bakshi S, Siryaporn A, Goulian M, Weisshaar JC. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. Mol Microbiol. 2012;85(1):21-38. doi:10.1111/j.1365-2958.2012.08081.x.

Ball AS, van Kessel JC. The master quorum-sensing regulators LuxR/HapR directly interact with the alpha subunit of RNA polymerase to drive transcription activation in Vibrio harveyi and Vibrio cholerae. Mol Microbiol. 2019;111(5):1317-1334. doi:10.1111/mmi.14223.

Barinova N, Kuznedelov K, Severinov K, Kulbachinskiy A. Structural modules of RNA polymerase required for transcription from promoters containing downstream basal promoter element GGGA. J Biol Chem. 2008;283(33):22482-22489. doi:10.1074/jbc.M802445200.

Barker MM, Gaal T, Josaitis CA, Gourse RL. Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. J Mol Biol. 2001;305(4):673-688. doi:10.1006/jmbi.2000.4327.

Becker G, Klauck E, Hengge-Aronis R. Regulation of RpoS proteolysis in Escherichia coli: the response regulator RssB is a recognition factor that interacts with the turnover element in RpoS. Proc Natl Acad Sci U S A. 1999;96(11):6439-6444. doi:10.1073/pnas.96.11.6439.

Beutel BA, Record MT Jr. E. coli promoter spacer regions contain nonrandom sequences which correlate to spacer length. Nucleic Acids Res. 1990;18(12):3597-3603. doi:10.1093/nar/18.12.3597.

Binder SC, Eckweiler D, Schulz S, et al. Functional modules of sigma factor regulons guarantee adaptability and evolvability. Sci Rep. 2016;6:22212. Published 2016 Feb 26. doi:10.1038/srep22212.

Blaby-Haas CE, Furman R, Rodionov DA, Artsimovitch I, de Crécy-Lagard V. Role of a Zn-independent DksA in Zn homeostasis and stringent response. Mol Microbiol. 2011;79(3):700-715. doi:10.1111/j.1365-2958.2010.07475.x.

Blatter EE, Ross W, Tang H, Gourse RL, Ebright RH. Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. Cell. 1994;78(5):889-896. doi:10.1016/s0092-8674(94)90682-3.

Borukhov S, Nudler E. RNA polymerase holoenzyme: structure, function and biological implications. Curr Opin Microbiol. 2003;6(2):93-100. doi:10.1016/s1369-5274(03)00036-5.

Brodolin K, Zenkin N, Mustaev A, Mamaeva D, Heumann H. The sigma 70 subunit of RNA polymerase induces lacUV5 promoter-proximal pausing of transcription. Nat Struct Mol Biol. 2004;11(6):551-557. doi:10.1038/nsmb768.

Buc H, McClure WR. Kinetics of open complex formation between Escherichia coli RNA polymerase and the lac UV5 promoter. Evidence for a sequential mechanism involving three steps. Biochemistry. 1985;24(11):2712-2723. doi:10.1021/bi00332a018.

Burgess RR. Separation and characterization of the subunits of ribonucleic acid polymerase. J Biol Chem. 1969;244(22):6168-6176.

Burgess RR, Travers AA, Dunn JJ, Bautz EK. Factor stimulating transcription by RNA polymerase. Nature. 1969;221(5175):43-46. doi:10.1038/221043a0.

Campbell EA, Muzzin O, Chlenov M, et al. Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol Cell. 2002;9(3):527-539. doi:10.1016/s1097-2765(02)00470-7.

Chakraborty A, Wang D, Ebright YW, et al. Opening and closing of the bacterial RNA polymerase clamp. Science. 2012;337(6094):591-595. doi:10.1126/science.1218716.

Chen J, Boyaci H, Campbell EA. Diverse and unified mechanisms of transcription initiation in bacteria. Nat Rev Microbiol. 2021;19(2):95-109. doi:10.1038/s41579-020-00450-2.

Chen J, Chiu C, Gopalkrishnan S, et al. Stepwise Promoter Melting by Bacterial RNA Polymerase. Mol Cell. 2020;78(2):275-288.e6. doi:10.1016/j.molcel.2020.02.017.

Chen J, Gopalkrishnan S, Chiu C, et al. E. coli TraR allosterically regulates transcription initiation by altering RNA polymerase conformation. Elife. 2019;8:e49375. Published 2019 Dec 16. doi:10.7554/eLife.49375.

Chen H, Tang H, Ebright RH. Functional interaction between RNA polymerase alpha subunit C-terminal domain and sigma70 in UP-element- and activator-dependent transcription. Mol Cell. 2003;11(6):1621-1633. doi:10.1016/s1097-2765(03)00201-6.

Cheng CY, Yu YJ, Yang MT. Coexpression of omega subunit in E. coli is required for the maintenance of enzymatic activity of Xanthomonas campestris pv. campestris RNA polymerase. Protein Expr Purif. 2010;69(1):91-98. doi:10.1016/j.pep.2009.07.002.

Chlenov M, Masuda S, Murakami KS, Nikiforov V, Darst SA, Mustaev A. Structure and function of lineage-specific sequence insertions in the bacterial RNA polymerase beta' subunit. J Mol Biol. 2005;353(1):138-154. doi:10.1016/j.jmb.2005.07.073.

Condon C, Philips J, Fu ZY, Squires C, Squires CL. Comparison of the expression of the seven ribosomal RNA operons in Escherichia coli. EMBO J. 1992;11(11):4175-4185. doi:10.1002/j.1460-2075.1992.tb05511.x.

Cook H, Ussery DW. Sigma factors in a thousand E. coli genomes. Environ Microbiol. 2013;15(12):3121-3129. doi:10.1111/1462-2920.12236.

Cramer P. Multisubunit RNA polymerases. Curr Opin Struct Biol. 2002;12(1):89-97. doi:10.1016/s0959-440x(02)00294-4.

Crick F. Central dogma of molecular biology. Nature. 1970;227(5258):561-563. doi:10.1038/227561a0.

Dennis PP, Nomura M. Stringent control of ribosomal protein gene expression in Escherichia coli. Proc Natl Acad Sci U S A. 1974;71(10):3819-3823. doi:10.1073/pnas.71.10.3819.

Dennis PP, Nomura M. Stringent control of the transcriptional activities of ribosomal protein genes in E. coli. Nature. 1975;255(5508):460-465. doi:10.1038/255460a0.

Dorich V, Brugger C, Tripathi A, et al. Structural basis for inhibition of a response regulator of σS stability by a ClpXP antiadaptor. Genes Dev. 2019;33(11-12):718-732. doi:10.1101/gad.320168.118.

Duchi D, Gryte K, Robb NC, et al. Conformational heterogeneity and bubble dynamics in single bacterial transcription initiation complexes. Nucleic Acids Res. 2018;46(2):677-688. doi:10.1093/nar/gkx1146.

Dostálová H, Busche T, Holátko J, et al. Overlap of Promoter Recognition Specificity of Stress Response Sigma Factors SigD and SigH in Corynebacterium glutamicum ATCC 13032. Front Microbiol. 2019;9:3287. Published 2019 Jan 9. doi:10.3389/fmicb.2018.03287.

Ebright RH, Busby S. The Escherichia coli RNA polymerase alpha subunit: structure and function. Curr Opin Genet Dev. 1995;5(2):197-203. doi:10.1016/0959-437x(95)80008-5.

Egan SM, Pease AJ, Lang J, et al. Transcription activation by a variety of AraC/XylS family activators does not depend on the class II-specific activation determinant in the N-terminal domain of the RNA polymerase alpha subunit. J Bacteriol. 2000;182(24):7075-7077. doi:10.1128/JB.182.24.7075-7077.2000.

Estrem ST, Gaal T, Ross W, Gourse RL. Identification of an UP element consensus sequence for bacterial promoters. Proc Natl Acad Sci U S A. 1998;95(17):9761-9766. doi:10.1073/pnas.95.17.9761.

Evans NJ, Brown JM, Demirkan I, et al. Three unique groups of spirochetes isolated from digital dermatitis lesions in UK cattle. Vet Microbiol. 2008;130(1-2):141-150. doi:10.1016/j.vetmic.2007.12.019.

Feklistov A, Darst SA. Structural basis for promoter-10 element recognition by the bacterial RNA polymerase σ subunit. Cell. 2011;147(6):1257-1269. doi:10.1016/j.cell.2011.10.041.

Feklístov A, Sharon BD, Darst SA, Gross CA. Bacterial sigma factors: a historical, structural, and genomic perspective. Annu Rev Microbiol. 2014;68:357-376. doi:10.1146/annurev-micro-092412-155737.

Fortuna A, Bähre H, Visca P, Rampioni G, Leoni L. The two Pseudomonas aeruginosa DksA stringent response proteins are largely interchangeable at the whole transcriptome level and in the control of virulence-related traits. Environ Microbiol. 2021;23(9):5487-5504. doi:10.1111/1462-2920.15693.

Furman R, Biswas T, Danhart EM, Foster MP, Tsodikov OV, Artsimovitch I. DksA2, a zinc-independent structural analog of the transcription factor DksA. FEBS Lett. 2013;587(6):614-619. doi:10.1016/j.febslet.2013.01.073.

Furman R, Danhart EM, NandyMazumdar M, Yuan C, Foster MP, Artsimovitch I. pH dependence of the stress regulator DksA. PLoS One. 2015;10(3):e0120746. Published 2015 Mar 23. doi:10.1371/journal.pone.0120746.

Goldman SR, Ebright RH, Nickels BE. Direct detection of abortive RNA transcripts in vivo. Science. 2009;324(5929):927-928. doi:10.1126/science.1169237.

Gruber TM, Gross CA. Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol. 2003;57:441-466. doi:10.1146/annurev.micro.57.030502.090913.

Gumerov VM, Ortega DR, Adebali O, Ulrich LE, Zhulin IB. MiST 3.0: an updated microbial signal transduction database with an emphasis on chemosensory systems. Nucleic Acids Res. 2020;48(D1):D459-D464. doi:10.1093/nar/gkz988.

Gummesson B, Lovmar M, Nyström T. A proximal promoter element required for positive transcriptional control by guanosine tetraphosphate and DksA protein during the stringent response. J Biol Chem. 2013;288(29):21055-21064. doi:10.1074/jbc.M113.479998.

Harden TT, Wells CD, Friedman LJ, et al. Bacterial RNA polymerase can retain σ70 throughout transcription. Proc Natl Acad Sci U S A. 2016;113(3):602-607. doi:10.1073/pnas.1513899113.

Haugen SP, Berkmen MB, Ross W, Gaal T, Ward C, Gourse RL. rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. Cell. 2006;125(6):1069-1082. doi:10.1016/j.cell.2006.04.034.

Hawley DK, McClure WR. Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic Acids Res. 1983;11(8):2237-2255. doi:10.1093/nar/11.8.2237.

Hauryliuk V, Atkinson GC, Murakami KS, Tenson T, Gerdes K. Recent functional insights into the role of (p)ppGpp in bacterial physiology. Nat Rev Microbiol. 2015;13(5):298-309. doi:10.1038/nrmicro3448.

Henderson KL, Felth LC, Molzahn CM, et al. Mechanism of transcription initiation and promoter escape by E. coli RNA polymerase. Proc Natl Acad Sci U S A. 2017;114(15):E3032-E3040. doi:10.1073/pnas.1618675114.

Heyduk E, Heyduk T. Next generation sequencing-based parallel analysis of melting kinetics of 4096 variants of a bacterial promoter. Biochemistry. 2014;53(2):282-292. doi:10.1021/bi401277w.

Heyduk E, Kuznedelov K, Severinov K, Heyduk T. A consensus adenine at position -11 of the nontemplate strand of bacterial promoter is important for nucleation of promoter melting. J Biol Chem. 2006;281(18):12362-12369. doi:10.1074/jbc.M601364200.

Holley CL, Zhang X, Fortney KR, et al. DksA and (p)ppGpp have unique and overlapping contributions to Haemophilus ducreyi pathogenesis in humans. Infect Immun. 2015;83(8):3281-3292. doi:10.1128/IAI.00692-15.

Hook-Barnard IG, Hinton DM. The promoter spacer influences transcription initiation via sigma70 region 1.1 of Escherichia coli RNA polymerase. Proc Natl Acad Sci U S A. 2009;106(3):737-742. doi:10.1073/pnas.0808133106.

Huang X, Fredrick KL, Helmann JD. Promoter recognition by Bacillus subtilis sigmaW: autoregulation and partial overlap with the sigmaX regulon. J Bacteriol. 1998;180(15):3765-3770. doi:10.1128/JB.180.15.3765-3770.1998.

Huang YH, Trapp V, Puro O, et al. Fluorogenic RNA aptamers to probe transcription initiation and co-transcriptional RNA folding by multi-subunit RNA polymerases. Methods Enzymol. 2022;675:207-233. doi:10.1016/bs.mie.2022.07.010.

Hug LA, Baker BJ, Anantharaman K, et al. A new view of the tree of life. Nat Microbiol. 2016;1:16048. Published 2016 Apr 11. doi:10.1038/nmicrobiol.2016.48.

Hsu LM. Promoter clearance and escape in prokaryotes. Biochim Biophys Acta. 2002;1577(2):191-207. doi:10.1016/s0167-4781(02)00452-9.

Irving SE, Choudhury NR, Corrigan RM. The stringent response and physiological roles of (pp)pGpp in bacteria. Nat Rev Microbiol. 2021;19(4):256-271. doi:10.1038/s41579-020-00470-y.

Ishihama A, Kori A, Koshio E, et al. Intracellular concentrations of 65 species of transcription factors with known regulatory functions in Escherichia coli. J Bacteriol. 2014;196(15):2718-2727. doi:10.1128/JB.01579-14.

Ishihama A, Shimada T, Yamazaki Y. Transcription profile of Escherichia coli: genomic SELEX search for regulatory targets of transcription factors. Nucleic Acids Res. 2016;44(5):2058-2074. doi:10.1093/nar/gkw051.

Jensen D, Galburt EA. The Context-Dependent Influence of Promoter Sequence Motifs on Transcription Initiation Kinetics and Regulation. J Bacteriol. 2021;203(8):e00512-20. Published 2021 Mar 23. doi:10.1128/JB.00512-20.

Jishage M, Iwata A, Ueda S, Ishihama A. Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli: intracellular levels of four species of sigma subunit under various growth conditions. J Bacteriol. 1996;178(18):5447-5451. doi:10.1128/jb.178.18.5447-5451.1996.

Jishage M, Kvint K, Shingler V, Nyström T. Regulation of sigma factor competition by the alarmone ppGpp. Genes Dev. 2002;16(10):1260-1270. doi:10.1101/gad.227902.

Kanjee U, Ogata K, Houry WA. Direct binding targets of the stringent response alarmone (p)ppGpp. Mol Microbiol. 2012;85(6):1029-1043. doi:10.1111/j.1365-2958.2012.08177.x.

Kannan N, Chander P, Ghosh P, Vishveshwara S, Chatterji D. Stabilizing interactions in the dimer interface of alpha-subunit in Escherichia coli RNA polymerase: a graph spectral and point mutation study. Protein Sci. 2001;10(1):46-54. doi:10.1110/ps.26201.

Kapanidis AN, Margeat E, Laurence TA, et al. Retention of transcription initiation factor sigma70 in transcription elongation: single-molecule analysis. Mol Cell. 2005;20(3):347-356. doi:10.1016/j.molcel.2005.10.012.

Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, Ebright RH. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. Science. 2006;314(5802):1144-1147. doi:10.1126/science.1131399.

Kaper JB, Nataro JP, Mobley HL. Pathogenic Escherichia coli. Nat Rev Microbiol. 2004;2(2):123-140. doi:10.1038/nrmicro818.

Kaur G, Kaundal S, Kapoor S, Grimes JM, Huiskonen JT, Thakur KG. Mycobacterium tuberculosis CarD, an essential global transcriptional regulator forms amyloid-like fibrils. Sci Rep. 2018;8(1):10124. Published 2018 Jul 4. doi:10.1038/s41598-018-28290-4.

Kazmierczak MJ, Wiedmann M, Boor KJ. Alternative sigma factors and their roles in bacterial virulence. Microbiol Mol Biol Rev. 2005;69(4):527-543. doi:10.1128/MMBR.69.4.527-543.2005.

Keller AN, Yang X, Wiedermannová J, Delumeau O, Krásný L, Lewis PJ. ε, a new subunit of RNA polymerase found in gram-positive bacteria. J Bacteriol. 2014;196(20):3622-3632. doi:10.1128/JB.02020-14.

Kennedy WP, Momand JR, Yin YW. Mechanism for de novo RNA synthesis and initiating nucleotide specificity by t7 RNA polymerase. J Mol Biol. 2007;370(2):256-268. doi:10.1016/j.jmb.2007.03.041.

Kim KY, Park JK, Park S. In Streptomyces coelicolor SigR, methionine at the -35 element interacting region 4 confers the -31'-adenine base selectivity. Biochem Biophys Res Commun. 2016;470(2):257-262. doi:10.1016/j.bbrc.2016.01.075.

Kim MS, Hahn MY, Cho Y, Cho SN, Roe JH. Positive and negative feedback regulatory loops of thiol-oxidative stress response mediated by an unstable isoform of sigmaR in actinomycetes. Mol Microbiol. 2009;73(5):815-825. doi:10.1111/j.1365-2958.2009.06824.x.

Kimura M, Ishihama A. Subunit assembly in vivo of Escherichia coli RNA polymerase: role of the amino-terminal assembly domain of alpha subunit. Genes Cells. 1996;1(6):517-528. doi:10.1046/j.1365-2443.1996.d01-258.x.

Kobayashi M, Nagata K, Ishihama A. Promoter selectivity of Escherichia coli RNA polymerase: effect of base substitutions in the promoter -35 region on promoter strength. Nucleic Acids Res. 1990;18(24):7367-7372. doi:10.1093/nar/18.24.7367.

Koo BM, Rhodius VA, Nonaka G, deHaseth PL, Gross CA. Reduced capacity of alternative sigmas to melt promoters ensures stringent promoter recognition. Genes Dev. 2009;23(20):2426-2436. doi:10.1101/gad.1843709.

Kornberg RD. The molecular basis of eukaryotic transcription. Proc Natl Acad Sci U S A. 2007;104(32):12955-12961. doi:10.1073/pnas.0704138104.

Kurkela J, Fredman J, Salminen TA, Tyystjärvi T. Revealing secrets of the enigmatic omega subunit of bacterial RNA polymerase. Mol Microbiol. 2021;115(1):1-11. doi:10.1111/mmi.14603.

Lane WJ, Darst SA. (*a*) Molecular evolution of multisubunit RNA polymerases: sequence analysis. J Mol Biol. 2010;395(4):671-685. doi:10.1016/j.jmb.2009.10.062.

Lane WJ, Darst SA. (*b*) Molecular evolution of multisubunit RNA polymerases: structural analysis. J Mol Biol. 2010;395(4):686-704. doi:10.1016/j.jmb.2009.10.063.

Lemke JJ, Sanchez-Vazquez P, Burgos HL, Hedberg G, Ross W, Gourse RL. Direct regulation of Escherichia coli ribosomal protein promoters by the transcription factors ppGpp and DksA. Proc Natl Acad Sci U S A. 2011;108(14):5712-5717. doi:10.1073/pnas.1019383108.

Liolos K, Abt B, Scheuner C, et al. Complete genome sequence of the halophilic bacterium Spirochaeta africana type strain (Z-7692(T)) from the alkaline Lake Magadi in the East African Rift. Stand Genomic Sci. 2013;8(2):165-176. Published 2013 May 20. doi:10.4056/sigs.3607108.

Lonetto M, Gribskov M, Gross CA. The sigma 70 family: sequence conservation and evolutionary relationships. J Bacteriol. 1992;174(12):3843-3849. doi:10.1128/jb.174.12.3843-3849.1992.

Maeda H, Jishage M, Nomura T, Fujita N, Ishihama A. Two extracytoplasmic function sigma subunits, sigma(E) and sigma(FecI), of Escherichia coli: promoter selectivity and intracellular levels. J Bacteriol. 2000;182(4):1181-1184. doi:10.1128/JB.182.4.1181-1184.2000.

Maharjan RP, Sullivan GJ, Adams FG, et al. DksA is a conserved master regulator of stress response in Acinetobacter baumannii. Nucleic Acids Res. 2023;51(12):6101-6119. doi:10.1093/nar/gkad341.

Martínez-Bueno MA, Tobes R, Rey M, Ramos JL. Detection of multiple extracytoplasmic function (ECF) sigma factors in the genome of Pseudomonas putida KT2440 and their counterparts in Pseudomonas aeruginosa PA01. Environ Microbiol. 2002;4(12):842-855. doi:10.1046/j.1462-2920.2002.00371.x.

Matlock DL, Heyduk T. Sequence determinants for the recognition of the fork junction DNA containing the -10 region of promoter DNA by E. coli RNA polymerase. Biochemistry. 2000;39(40):12274-12283. doi:10.1021/bi001433h.

Mazumder A, Kapanidis AN. Recent Advances in Understanding σ70-Dependent Transcription Initiation Mechanisms. J Mol Biol. 2019;431(20):3947-3959. doi:10.1016/j.jmb.2019.04.046.

Mauri M, Klumpp S. A model for sigma factor competition in bacterial cells. PLoS Comput Biol. 2014;10(10):e1003845. Published 2014 Oct 9. doi:10.1371/journal.pcbi.1003845.

Mejía-Almonte C, Busby SJW, Wade JT, et al. Redefining fundamental concepts of transcription initiation in bacteria. Nat Rev Genet. 2020;21(11):699-714. doi:10.1038/s41576-020-0254-8.

Mekler V, Kortkhonjia E, Mukhopadhyay J, et al. Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. Cell. 2002;108(5):599-614. doi:10.1016/s0092-8674(02)00667-0.

Mekler V, Severinov K. Cooperativity and interaction energy threshold effects in recognition of the -10 promoter element by bacterial RNA polymerase. Nucleic Acids Res. 2013;41(15):7276-7285. doi:10.1093/nar/gkt541.

Mooney RA, Darst SA, Landick R. Sigma and RNA polymerase: an on-again, off-again relationship?. Mol Cell. 2005;20(3):335-345. doi:10.1016/j.molcel.2005.10.015.

Moyle H, Waldburger C, Susskind MM. Hierarchies of base pair preferences in the P22 ant promoter. J Bacteriol. 1991;173(6):1944-1950. doi:10.1128/jb.173.6.1944-1950.1991.

Mulligan ME, Brosius J, McClure WR. Characterization in vitro of the effect of spacer length on the activity of Escherichia coli RNA polymerase at the TAC promoter. J Biol Chem. 1985;260(6):3529-3538.

Mumm JP, Friedman LJ, Gelles J. Mechanism of upstream promoter element stimulation of transcription at a ribosomal RNA promoter determined by single-molecule imaging. bioRxiv 2020.02.17.953182; doi: https://doi.org/10.1101/2020.02.17.953182 [preprint]

Murakami KS. Structural biology of bacterial RNA polymerase. Biomolecules. 2015;5(2):848-864. Published 2015 May 11. doi:10.3390/biom5020848.

Myers KS, Noguera DR, Donohue TJ. Promoter Architecture Differences among Alphaproteobacteria and Other Bacterial Taxa. mSystems. 2021;6(4):e0052621. doi:10.1128/mSystems.00526-21.

Niedziela-Majka A, Heyduk T. Escherichia coli RNA polymerase contacts outside the -10 promoter element are not essential for promoter melting. J Biol Chem. 2005;280(46):38219-38227. doi:10.1074/jbc.M507984200.

Narula J, Devi SN, Fujita M, Igoshin OA. Ultrasensitivity of the Bacillus subtilis sporulation decision. Proc Natl Acad Sci U S A. 2012;109(50):E3513-E3522. doi:10.1073/pnas.1213974109.

Newlands JT, Gaal T, Mecsas J, Gourse RL. Transcription of the Escherichia coli rrnB P1 promoter by the heat shock RNA polymerase (E sigma 32) in vitro. J Bacteriol. 1993;175(3):661-668. doi:10.1128/jb.175.3.661-668.1993.

Nickels BE, Mukhopadhyay J, Garrity SJ, Ebright RH, Hochschild A. The sigma 70 subunit of RNA polymerase mediates a promoter-proximal pause at the lac promoter. Nat Struct Mol Biol. 2004;11(6):544-550. doi:10.1038/nsmb757.

Nudler E. RNA polymerase active center: the molecular engine of transcription. Annu Rev Biochem. 2009;78:335-361. doi:10.1146/annurev.biochem.76.052705.164655.

Paget MS. Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. Biomolecules. 2015;5(3):1245-1265. Published 2015 Jun 26. doi:10.3390/biom5031245.

Paster, B.J. Phylum XV. Spirochaetes. In: Krieg, N.R., et al. Bergey's Manual® of Systematic Bacteriology. 2010. Springer, New York, NY. https://doi.org/10.1007/978-0-387-68572-4_4.

Pei HH, Hilal T, Chen ZA, et al. The δ subunit and NTPase HelD institute a two-pronged mechanism for RNA polymerase recycling. Nat Commun. 2020;11(1):6418. Published 2020 Dec 18. doi:10.1038/s41467-020-20159-3.

Perederina, A., Svetlov, V., Vassylyeva, M., Tahirov, T., Yokoyama, S., Artsimovitch, I., and Vassylyev, D. (2004) Regulation through the secondary channel – structural framework for ppGpp-DksA synergism during transcription. Cell 118: 297–309.

Petushkov I, Esyunina D, Kulbachinskiy A. σ38-dependent promoter-proximal pausing by bacterial RNA polymerase. Nucleic Acids Res. 2017;45(6):3006-3016. doi:10.1093/nar/gkw1213.

Petushkov I, Pupov D, Bass I, Kulbachinskiy A. Mutations in the CRE pocket of bacterial RNA polymerase affect multiple steps of transcription. Nucleic Acids Res. 2015;43(12):5798-5809. doi:10.1093/nar/gkv504.

Puro, O. Real-time monitoring of transcription with fluorescent light-up aptamers. Master's thesis. 2023. https://urn.fi/URN:NBN:fi-fe2023030930927.

Rammohan J, Ruiz Manzano A, Garner AL, Stallings CL, Galburt EA. CarD stabilizes mycobacterial open complexes via a two-tiered kinetic mechanism. Nucleic Acids Res. 2015;43(6):3272-3285. doi:10.1093/nar/gkv078.

Revyakin A, Liu C, Ebright RH, Strick TR. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. Science. 2006;314(5802):1139-1143. doi:10.1126/science.1131398.

Rhodius VA, Segall-Shapiro TH, Sharon BD, et al. Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. Mol Syst Biol. 2013;9:702. Published 2013 Oct 29. doi:10.1038/msb.2013.58.

Ring BZ, Yarnell WS, Roberts JW. Function of E. coli RNA polymerase sigma factor sigma 70 in promoter-proximal pausing. Cell. 1996;86(3):485-493. doi:10.1016/s0092-8674(00)80121-x.

Ross W, Gourse RL. Sequence-independent upstream DNA-alphaCTD interactions strongly stimulate Escherichia coli RNA polymerase-lacUV5 promoter association. Proc Natl Acad Sci U S A. 2005;102(2):291-296. doi:10.1073/pnas.0405814102.

Ross W, Sanchez-Vazquez P, Chen AY, Lee JH, Burgos HL, Gourse RL. ppGpp Binding to a Site at the RNAP-DksA Interface Accounts for Its Dramatic Effects on Transcription Initiation during the Stringent Response. Mol Cell. 2016;62(6):811-823. doi:10.1016/j.molcel.2016.04.029.

Ross W, Schneider DA, Paul BJ, Mertens A, Gourse RL. An intersubunit contact stimulating transcription initiation by E coli RNA polymerase: interaction of the alpha C-terminal domain and sigma region 4. Genes Dev. 2003;17(10):1293-1307. doi:10.1101/gad.1079403.

Ruff EF, Record MT Jr, Artsimovitch I. Initial events in bacterial transcription initiation. Biomolecules. 2015;5(2):1035-1062. Published 2015 May 27. doi:10.3390/biom5021035.

Saecker RM, Chen J, Chiu CE, et al. Structural origins of Escherichia coli RNA polymerase open promoter complex stability. Proc Natl Acad Sci U S A. 2021;118(40):e2112877118. doi:10.1073/pnas.2112877118.

Saecker RM, Record MT Jr, Dehaseth PL. Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. J Mol Biol. 2011;412(5):754-771. doi:10.1016/j.jmb.2011.01.018.

Savery NJ, Lloyd GS, Kainz M, et al. Transcription activation at Class II CRP-dependent promoters: identification of determinants in the C-terminal domain of the RNA polymerase alpha subunit. EMBO J. 1998;17(12):3439-3447. doi:10.1093/emboj/17.12.3439.

Shimada T, Yamazaki Y, Tanaka K, Ishihama A. The whole set of constitutive promoters recognized by RNA polymerase RpoD holoenzyme of Escherichia coli [published correction appears in PLoS One. 2014;9(6):e100908]. PLoS One. 2014;9(3):e90447. Published 2014 Mar 6. doi:10.1371/journal.pone.0090447.

Shimada T, Tanaka K, Ishihama A. The whole set of the constitutive promoters recognized by four minor sigma subunits of Escherichia coli RNA polymerase. PLoS One. 2017;12(6):e0179181. Published 2017 Jun 30. doi:10.1371/journal.pone.0179181.

Shin Y, Qayyum MZ, Pupov D, Esyunina D, Kulbachinskiy A, Murakami KS. Structural basis of ribosomal RNA transcription regulation. Nat Commun. 2021;12(1):528. Published 2021 Jan 22. doi:10.1038/s41467-020-20776-y.

Simonson LG, Goodman CH, Bial JJ, Morton HE. Quantitative relationship of Treponema denticola to severity of periodontal disease. Infect Immun. 1988;56(4):726-728. doi:10.1128/iai.56.4.726-728.1988.

Skinner GM, Baumann CG, Quinn DM, Molloy JE, Hoggett JG. Promoter binding, initiation, and elongation by bacteriophage T7 RNA polymerase. A single-molecule view of the transcription cycle. J Biol Chem. 2004;279(5):3239-3244. doi:10.1074/jbc.M310471200.

Stallings CL, Stephanou NC, Chu L, Hochschild A, Nickels BE, Glickman MS. CarD is an essential regulator of rRNA transcription required for Mycobacterium tuberculosis persistence. Cell. 2009;138(1):146-159. doi:10.1016/j.cell.2009.04.041.

Sun Z, Yakhnin AV, FitzGerald PC, Mclntosh CE, Kashlev M. Nascent RNA sequencing identifies a widespread sigma70-dependent pausing regulated by Gre factors in bacteria. Nat Commun. 2021;12(1):906. Published 2021 Feb 10. doi:10.1038/s41467-021-21150-2.

Sutherland C, Murakami KS. An Introduction to the Structure and Function of the Catalytic Core Enzyme of Escherichia coli RNA Polymerase. EcoSal Plus. 2018;8(1):10.1128/ecosalplus.ESP-0004-2018. doi:10.1128/ecosalplus.ESP-0004-2018.

Tampa M, Sarbu I, Matei C, Benea V, Georgescu SR. Brief history of syphilis. J Med Life. 2014;7(1):4-10.

Tang GQ, Roy R, Bandwar RP, Ha T, Patel SS. Real-time observation of the transition from transcription initiation to elongation of the RNA polymerase. Proc Natl Acad Sci U S A. 2009;106(52):22175-22180. doi:10.1073/pnas.0906979106.

Taylor-Robinson D, Jensen JS. Mycoplasma genitalium: from Chrysalis to multicolored butterfly. Clin Microbiol Rev. 2011;24(3):498-514. doi:10.1128/CMR.00006-11.

Travers AA, Burgess RR. Cyclic re-use of the RNA polymerase sigma factor. Nature. 1969;222(5193):537-540. doi:10.1038/222537a0.

Travis BA, Schumacher MA. Diverse molecular mechanisms of transcription regulation by the bacterial alarmone ppGpp. Mol Microbiol. 2022;117(2):252-260. doi:10.1111/mmi.14860.

Torres-Puig S, Broto A, Querol E, Piñol J, Pich OQ. A novel sigma factor reveals a unique regulon controlling cell-specific recombination in Mycoplasma genitalium. Nucleic Acids Res. 2015;43(10):4923-4936. doi:10.1093/nar/gkv422.

Turowski TW, Petfalski E, Goddard BD, French SL, Helwak A, Tollervey D. Nascent Transcript Folding Plays a Major Role in Determining RNA Polymerase Elongation Rates. Mol Cell. 2020;79(3):488-503.e11. doi:10.1016/j.molcel.2020.06.002.

van de Water JA, Melkonian R, Junca H, et al. Spirochaetes dominate the microbial community associated with the red coral Corallium rubrum on a broad geographic scale. Sci Rep. 2016;6:27277. Published 2016 Jun 6. doi:10.1038/srep27277.

Varik V, Oliveira SRA, Hauryliuk V, Tenson T. HPLC-based quantification of bacterial housekeeping nucleotides and alarmone messengers ppGpp and pppGpp. Sci Rep. 2017;7(1):11022. Published 2017 Sep 8. doi:10.1038/s41598-017-10988-6.

Vassylyev DG, Sekine S, Laptenko O, et al. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. Nature. 2002;417(6890):712-719. doi:10.1038/nature752.

Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH, Artsimovitch I. Structural basis for transcription elongation by bacterial RNA polymerase. Nature. 2007;448(7150):157-162. doi:10.1038/nature05932.

Vvedenskaya IO, Vahedian-Movahed H, Zhang Y, Taylor DM, Ebright RH, Nickels BE. Interactions between RNA polymerase and the core recognition element are a determinant of transcription start site selection. Proc Natl Acad Sci U S A. 2016;113(21):E2899-E2905. doi:10.1073/pnas.1603271113.

Warne SE, deHaseth PL. Promoter recognition by Escherichia coli RNA polymerase. Effects of single base pair deletions and insertions in the spacer DNA separating the -10 and -35 regions are dependent on spacer DNA sequence. Biochemistry. 1993;32(24):6134-6140. doi:10.1021/bi00075a003.

Wickstrum JR, Egan SM. Amino acid contacts between sigma 70 domain 4 and the transcription activators RhaS and RhaR. J Bacteriol. 2004;186(18):6277-6285. doi:10.1128/JB.186.18.6277-6285.2004.

Wiedermannová J, Sudzinová P, Kovaľ T, et al. Characterization of HelD, an interacting partner of RNA polymerase from Bacillus subtilis. Nucleic Acids Res. 2014;42(8):5151-5163. doi:10.1093/nar/gku113.

Windgassen TA, Mooney RA, Nayak D, Palangat M, Zhang J, Landick R. Trigger-helix folding pathway and SI3 mediate catalysis and hairpin-stabilized pausing by Escherichia coli RNA polymerase. Nucleic Acids Res. 2014;42(20):12707-12721. doi:10.1093/nar/gku997.

Winkelman JT, Vvedenskaya IO, Zhang Y, et al. Multiplexed protein-DNA cross-linking: Scrunching in transcription start site selection. Science. 2016;351(6277):1090-1093. doi:10.1126/science.aad6881.

Xie WQ, Jäger K, Potts M. Cyanobacterial RNA polymerase genes rpoC1 and rpoC2 correspond to rpoC of Escherichia coli. J Bacteriol. 1989;171(4):1967-1973. doi:10.1128/jb.171.4.1967-1973.1989.

Xue XC, Liu F, Ou-Yang ZC. A kinetic model of transcription initiation by RNA polymerase. J Mol Biol. 2008;378(3):520-529. doi:10.1016/j.jmb.2008.03.008.

Yang X, Lewis PJ. Overproduction and purification of recombinant Bacillus subtilis RNA polymerase. Protein Expr Purif. 2008;59(1):86-93. doi:10.1016/j.pep.2008.01.006.

Yang X, Lewis PJ. The interaction between bacterial transcription factors and RNA polymerase during the transition from initiation to elongation. Transcription. 2010;1(2):66-69. doi:10.4161/trns.1.2.12791.

Yuzenkova Y, Tadigotla VR, Severinov K, Zenkin N. A new basal promoter element recognized by RNA polymerase core enzyme. EMBO J. 2011;30(18):3766-3775. Published 2011 Jul 26. doi:10.1038/emboj.2011.252.

Zalenskaya K, Lee J, Gujuluva CN, Shin YK, Slutsky M, Goldfarb A. Recombinant RNA polymerase: inducible overexpression, purification and assembly of Escherichia coli rpo gene products. Gene. 1990;89(1):7-12. doi:10.1016/0378-1119(90)90199-2.

Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA. Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. Cell. 1999;98(6):811-824. doi:10.1016/s0092-8674(00)81515-9.

Zhilina TN, Zavarzin GA, Rainey F, Kevbrin VV, Kostrikina NA, Lysenko AM. Spirochaeta alkalica sp. nov., Spirochaeta africana sp. nov., and Spirochaeta asiatica sp. nov., alkaliphilic anaerobes from the Continental Soda Lakes in Central Asia and the East African Rift. Int J Syst Bacteriol. 1996;46(1):305-312. doi:10.1099/00207713-46-1-305.

Zhou Y, Gottesman S, Hoskins JR, Maurizi MR, Wickner S. The RssB response regulator directly targets sigma(S) for degradation by ClpXP. Genes Dev. 2001;15(5):627-637. doi:10.1101/gad.864401.

Åberg A, Fernández-Vázquez J, Cabrer-Panes JD, Sánchez A, Balsalobre C. Similar and divergent effects of ppGpp and DksA deficiencies on transcription in Escherichia coli. J Bacteriol. 2009;191(10):3226-3236. doi:10.1128/JB.01410-08.

Österberg S, del Peso-Santos T, Shingler V. Regulation of alternative sigma factor use. Annu Rev Microbiol. 2011;65:37-55. doi:10.1146/annurev.micro.112408.134219.