UNIVERSITY OF TURKU

Functional analysis of spirochaetal promoters using a reconstituted transcription system

Biochemistry/Department of Life Technologies

Master's thesis

Author:
Ville Levola

31.10.2025

Turku

Master's thesis

Subject: Biochemistry/Department of Life Technologies
Author: Ville Levola
Title: Functional analysis of spirochaetal promoters using a reconstituted transcription system
Supervisor(s): M.Sc. Vilma Trapp, Assoc. Prof. Georgy Belogurov
Number of pages: 59 pages
Date: 31.10.2025

Transcription is the first and most regulated step in gene expression. Transcription is catalysed by RNA polymerase (RNAP), a complex multi-subunit enzyme. RNAP can read regulatory signals encoded in the genomic DNA and respond to the changes in the concentration of substrate NTPs. However, most of the regulatory inputs are delivered to RNAP via accessory protein factors and regulatory RNAs. The transcription systems of model organisms such as *Escherichia coli* and *Bacillus subtilis* have been studied in great detail. In contrast, transcription systems of spirochetes are poorly understood. To study spirochaetal transcription, we selected *Spirochaeta africana* as a model organism. We expressed and purified components, reconstituted the core transcription system of *S. africana in vitro* and measured the activities of selected promoters and transcription factors.

We measured the *in vitro* activity of *S. africana* and *E. coli* RNAPs at nine selected spirochaetal promoters and a control *E. coli* promoter at pH 7.5, which is close to intracellular pH in *E. coli*. Our transcription templates encoded fluorogenic RNA aptamer called Broccoli downstream the studied promoters. We monitored the transcription output by following the fluorescence levels of the RNA aptamer bound to a fluorogen. The activity of spirochaetal RNAP was several folds lower than that of *E. coli* RNAP at most promoters. We then measured the transcription activities of spirochaetal RNAP at pH 9, which is close to the optimal growth pH for these bacteria. Increasing pH significantly increased the activity of *S. africana* RNAP at several promoters bringing the overall activity profile closer to that of *E. coli* RNAP. Prior studies in our laboratory indicate that activities of some *S. africana* promoters are modulated by transcription factor CarD. We reproduced previously observed effects of CarD at two promoters and discovered an additional CarD activatable promoter. Overall, the activity profile of spirochaetal RNAP matched the *E. coli* RNAP profile best when the former enzyme was assayed at pH 9 in the presence of CarD. As a part of the investigation, we also used primer extension technique to map the transcription start sites for a subset of promoters to confirm their identities.

Overall, our studies on spirochetes aim to enhance the understanding of their physiological mechanisms. This information can be used in the development of treatments for diseases propagated by pathogenic spirochetes, such as borreliosis.

Key words: fluorescence, primer extension, promoter, transcription start sites, transcription

Table of contents

Abbreviations

CAP	catabolite activator protein
CRE	core recognition element
DFHBI-1T	(Z)-4-(3,5-difluoro-4-hydroxybenzylidene)-1,2-dimethyl-1H-imidazol-5(4H)-one
I_1	open promoter complex intermediate 1
I_2	open promoter complex intermediate 2
OD600	optical density (λ=600 nm)
RID	regulatory interaction domain
RNAP	RNA polymerase
RNAPσ	RNA polymerase holoenzyme
RP_c	closed promoter complex
$RP_i1.5b$	open promoter complex intermediate 1.5b
RP_o	open promoter complex
UP element	upstream element of the promoter
αCTD	carboxy terminal domain of α subunit in RNA polymerase
αNTD	amino terminal domain of α subunit in RNA polymerase
σ1-4	the number after σ refers to the domain of the σ subunit
σ^A	*Thermus aquaticus* σ subunit, unless otherwise indicated
σ^{70}	*Escherichia coli* σ subunit, unless otherwise indicated

1 Introduction

1.1 Antimicrobial resistance

Antimicrobial resistance is a big threat globally, which leads to treatment failures and deaths (WHO, 2022). It has become a severe problem due to the overuse and misuse of the antimicrobial drugs in food production and medicine treatment (WHO, 2015). New antimicrobial drugs are being developed, but only two (Rifamycin and Fidaxomicin), which target RNA polymerase (RNAP), are in clinical use. More specifically Rifamycin blocks the extension of the nascent RNA during initiation of transcription and Fidaxomicin prevents the promoter open complex formation. (Calvori *et al.* 1965; Scott 2013). Therefore, the research targeting the qualities and the structure of bacterial RNAPs is essential.

1.2 Spirochaeta africana

Bacteria of *Spirochaeta* genus live in diverse environments from the hot springs of New Zealand to the hypersaline lakes. *Spirochaeta africana* was first found in alkali and salty Lake Magadi, Kenya in 1996. It is a gram negative and aerotolerant bacteria species, which lives in alkaline environment (Zhilina *et al.* 1996). It has a helical structure, as shown in Figure 1, and it moves with its periplasmic flagella. It moves like a worm and can penetrate through semi solid matter. The metabolism of *S. africana* includes glucose fermentation which products consist of acetate, hydrogen, ethanol, and lactate.

S. africana is potentially a good model spirochete because it is non-pathogenic and is easy to cultivate in the laboratory. Pathogenic spirochetes cause serious diseases for example Lyme's disease. Unsurprisingly the RNAPs of non-pathogenic spirochaetes are similar to pathogenic ones. RNAP of these species is poorly studied and needs more investigation.

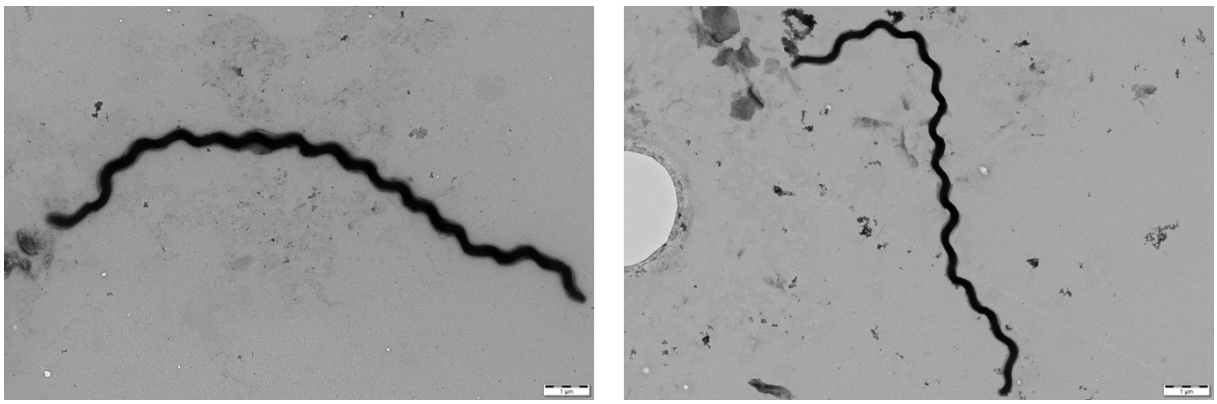

Figure 1. Electron microscopy images of S. africana. Images are taken by Georgy Belogurov lab members (Electron microscopy, Institute if Biomedicine, University of Turku).

1.3 A brief introduction to transcription

The recipe of every protein needed for creating the homeostasis of each living organism is encoded in their genome. For using this information encoded in the DNA, a living cell needs a working transcription system, which is usually highly regulated.

Bacterial transcription can be divided into three phases: initiation, elongation and termination. In the initiation of bacterial transcription RNAP binds to the promoter region using a separate sigma subunit. The double-stranded DNA is unwound and the template-directed synthesis of RNA begins when the first phosphodiester bond is made at the point of the transcription starting base. (Browning and Busby 2004).

RNAP escapes the promoter and the initiation phase after synthetising a short (13-15 bp) RNA strand and dissociating from the sigma subunit. It proceeds to the elongation phase where it adds nucleotides one at a time to the nascent RNA strand. The addition of nucleotides is a complex chain of events and occurs in the presence of two catalytic magnesium ions. (Belogurov and Artsimovitch 2015).

The bacterial transcription comes to an end with two ways. In the intrinsic mechanism the encoded features in DNA or nascent RNA can form a hairpin structure which leads to the dissociation of the RNAP from the DNA strand (Vassylyev *et al.* 2007a). The other mechanism requires a ring-shaped and homohexameric protein called Rho. Rho interacts with cytosine rich rut area in nascent RNA. Rho undergoes structure changes that are related to ATP hydrolysis and induce the dissociation of RNAP from the nucleotide strands leading to the termination of transcription (Richardson 2002).

A living organism needs a specific amount of each protein to maintain homeostasis in changing environmental conditions, which is necessary for metabolism, cell division and survival of the cell. Researchers propose that the initiation of transcription is the most regulated process during gene expression (Browning and Busby 2016). The main focus of my thesis is on this topic.

1.4 RNAP holoenzyme

RNAP is the ruling enzyme in the DNA dependent transcription across all three main domains of life (Werner and Grohmann 2011). In bacteria, the enzyme consists of five subunits: two identical α subunits, β subunit, β’ subunit and ω subunit (Murakami 2015). The subunits are assembled together with σ subunit (Figure 2), so that they can bind to the promoter region to initiate the transcription. The complex conformational changes in the holoenzyme lead to the formation of the open complex (described in section 1.6 below) before the first nucleotides of RNA are synthetized. Additionally, the addition of nucleotides to the nascent RNA and proof reading in the elongation step involve the

conformational changes in the subunits of RNAP (Vassylyev *et al.* 2007a). Also, the structure of RNAP is linked in the hairpin dependent termination of transcription. In conclusion, the structure of RNAP is crucial to the proper performance of the transcription process. Figure 2 shows the overall structure of the *Escherichia coli* RNAP holoenzyme. The key properties of RNAP subunit structures are presented below.

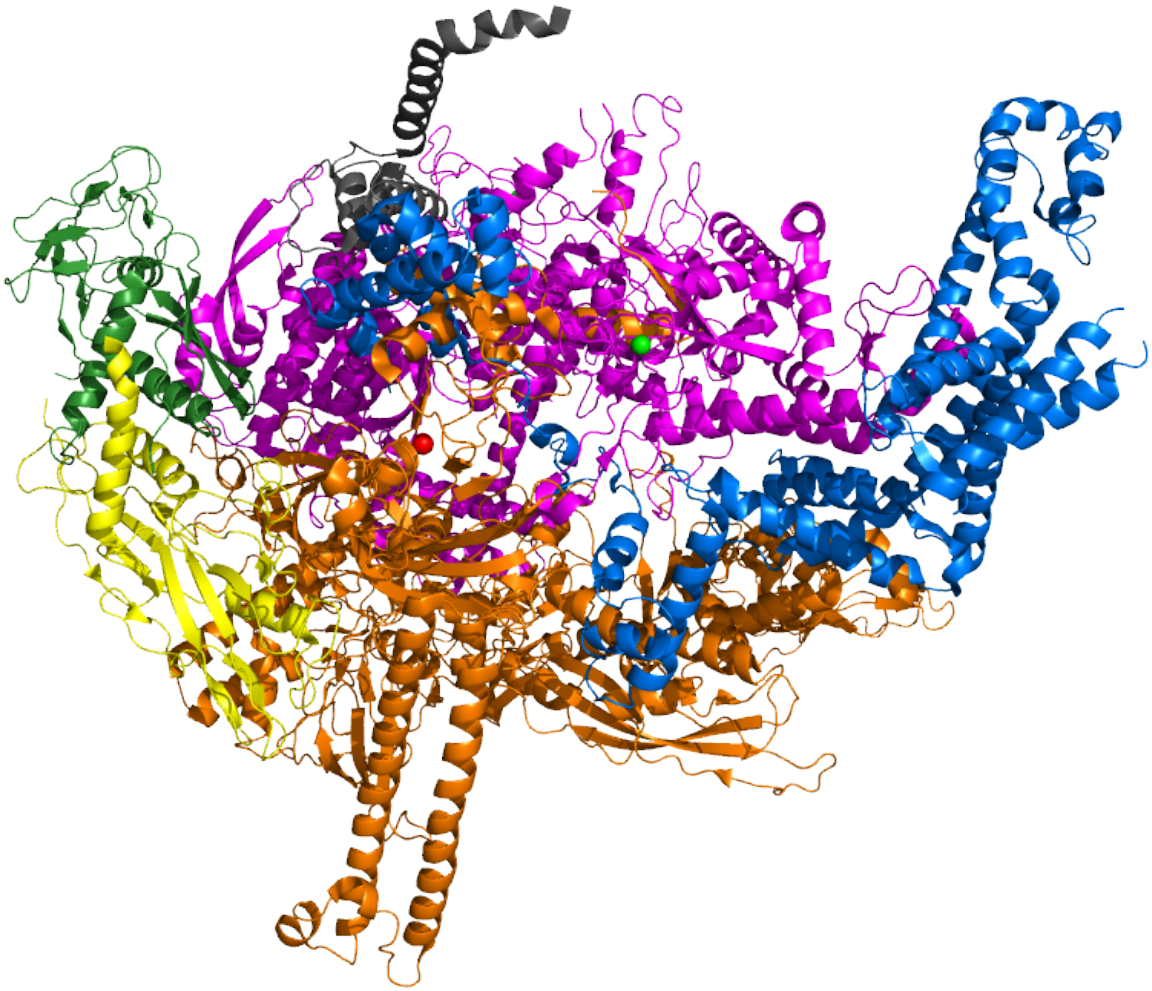

Figure 2. The overall structure of RNAP holoenzyme. E. coli RNAP holoenzyme (PDB code 4LJZ (Bae et al. 2013)) shows the crab claw-like structure of β (orange) and β' subunit (magenta) docked on the α subunits (green and yellow). ω subunit (dark gray) helps the β' subunit to fold correctly in the assembly of the core enzyme. σ subunit (blue) binds to the core enzyme to form the functional holoenzyme. The catalytically important Mg^{2+}-ion (red sphere) is located in the active cleft of the RNAP. Green spheres represent the Zn^{2+}-ions on the surface of the holoenzyme. Pymol (The PyMOL Molecular Graphics System, Version 3.0 Schrödinger) was used for visualization.

1.4.1 The crab claw-like structure is based on β and β' pinchers

Low resolution crystallographic and cryo-electron microscopic studies have revealed that the overall structure of RNAP resembles a crab claw (see Figure 2) and it is conserved across different bacterial species (Darst *et al.* 1998; Vassylyev *et al.* 2002; Zhang *et al.* 1999). The smaller top β subunit and the larger bottom β' subunit form the pinchers of the crab claw. β and β' subunits have major contacts between their structures. The largest interface between these two subunits takes place at the bottom of

the main channel in which the catalytically crucial Mg^{2+}-ion is chelated. Three aspartic acid residues in the widely conserved motif of -NADFDGD- chelate the Mg^{2+}-ion in the region D of the β' subunit. The main channel of the polymerase is formed between the β and β' subunits and it runs along the whole pincher structure (Zhang *et al.* 1999). When defining the overall structure of the *Thermus aquaticus* RNAP, Zhang *et al.* (1999) specify the dimensions of the whole core enzyme: it is 150 Å long, when measured from the back to the tips of the crab claws, 115 Å high and its width is 110 Å measured parallel with the main channel. The width of the main channel itself is 27 Å, but it has many internal structures.

Zhang *et al.* (1999) specify that the β subunit of *T. aquaticus* has two domains that don't interact with the β subunit core domains but extend away from it. One, which comprises C-terminal part of the I region of β subunit, makes extensive contacts with the β' subunit. The other domain, reaching from the region F to region H of the β subunit including the G region, forms a flap-like shape, which is shown to be flexibly attached with the rest of the β. The most notable sub-structure in the β' subunit is region F, which begins from the upper part of the β' and forms a helical structure and a loop, that cross the main channel. Finally, this domain anchors into the main body of the β' subunit. The catalytically active Mg^{2+}-ion lies on the base of the main channel directly opposite this helical structure. The β' F region helix interacts with the region G of the β' subunit forming a long loop extending to the main channel. This forms a wall-like structure that divides the main channel into two isolated channels. This secondary channel formed is reported to be roughly 10-12 Å in diameter in *T. aquaticus* RNAP crystal structure (Zhang *et al.* 1999). A structure that is also included in the internal topology of the main channel is a rudder-like structure which is formed by coiled-coil structure supporting another loop-like shape protruding upward in the main channel. This feature comprises the C region of the β' subunit. Additionally, so-called trigger loop structure occupies the main channel. Its refolding to two helix structures in the RNAP elongation complex has been proposed to be involved in the stabilization of the catalytically active intermediate in the process of nucleotide addition (Vassylyev *et al.* 2007b).

Among other features, the structural elements mentioned above establish the three channels in the RNAP: primary channel, secondary channel and RNA exit channel (Lee and Borukhov 2016). Primary channel is responsible for accommodation of the downstream duplex-DNA and RNA-DNA hybrid. Secondary channel serves as a site for incoming nucleotides and RNA exit channel acts as a way for the nascent RNA. The optimal folding of the structures mentioned earlier are the base of the proper functioning of these three crucial channels.

The structures, which aren't defined in the RNAP core structural studies, are only analysed in the crystallographic structure of the holoenzyme (Vassylyev *et al.* 2002). These are for example the large extension of the β' pincher formed by the non-conserved amino-terminal domain of β' subunit and the

zinc-finger domain in β', which lies in between of the amino- and carboxy-terminal domains of the σ subunit. Vassylyev *et al.* (2002) reveal that the zinc-finger does not itself have a secondary structure, but it needs zinc ions to interact with its four cysteines to maintain its integrity. The non-conserved amino-terminal domain of the β' subunit is also indicated to form an anchor when binding σ^{70} due to its close proximity to the amino terminal domain of σ subunit (region 1). In addition, researchers propose that Mg^{2+}-ions (approximately 250) coating the surface of the RNAP holoenzyme are involved in the σ subunit interactions, as well as DNA bending and binding.

1.4.2 α subunits help in the assembly of RNAP

Zhang *et al.* described the structure of RNAP core enzyme in 1999. The C terminal domain of the α subunit is not visible, but the N terminal domain is seen in the crystal structure. They propose that each αNTD monomer uses 24 % of its solvent-accessible surface for binding of other subunits of the RNAP. This is not a surprise, since Zhang and co-workers indicate in agreement with prevailing opinion that the role of the αNTD-dimer in the RNAP is to function as a docking surface of other subunits and to help them to assemble to a functional transcription machinery. Zhang *et al.* also reveal that the structures of αNTD dimer don't reach the internal channel of RNAP in which the catalytically active site of the RNAP is located. This indicates that this dimer structure doesn't play a role in the catalytical events of transcription.

The same study shows that the β and β' subunits interact each almost exclusively with one αNTD monomer. The α subunits are first proposed to interact with each other during the formation of the RNAP core enzyme (Ghosh *et al.* 2001). Secondly, Zhang and co-workers (1999) reveal that the next step would be the αI subunit interacting with the N terminal region of the segment I in the β subunit (residues 974-979 of *T. aquaticus)*. This triggers the formation of the $\alpha_2\beta$ intermediate. Finally the ω subunit helps the β' subunit to assemble to the final structure of the RNAP core enzyme (Ghosh *et al.* 2001).

Benoff and co-workers (2002) successfully determined the structure of αCTD domain, which is not seen in the core enzyme crystal structure, together with the catabolite activating protein (CAP) and DNA. This domain which is connected to the αNTD-end by flexible linkers plays a crucial role in the recognition of the UP element when RNAP binds to the promoter region (Estrem *et al.* 1999). The qualities of this interaction are described in more detail in the section 1.7.1 below.

1.4.3 ω subunit; a chaperon for β' subunit

ω subunit, encoded by the gene *rpoZ*, is a small 7-11.5 kDa sized protein. It has been reported to be typically non-essential element of bacterial RNAP when estimating the cell viability (Kojima *et al.* 2002; Mathew *et al.* 2006; A. Weiss *et al.* 2017). At the same time, researchers have revealed its importance in the production of secondary metabolites, in the establishment of biofilms, and especially in acting as a chaperon for other subunits, particularly β', during the assembly of a functional RNAP core enzyme. These are just few examples of the potential roles of the ω subunit.

The ω subunit consists of three α-helices, that are conserved across the phylum of bacteria (Kurkela *et al.* 2021). Still, there are only five amino acids that are fully conserved in the sequence alignment: two in α3-helix and three in α2-helix. Four of these residues stabilise the three-dimensional structure of the RNAP. Glutamate in α3 and arginine in α2 contact each other through ionic interactions, while their hydrophobic regions pack against the two conserved alanine. The other alanine also makes essential contacts with the β' subunit via hydrophobic interactions. Kurkela *et al.* (2021) compare the N-terminus ends of the ω subunit of different bacterial species and show that mycobacteria have a 30 residues long extension in comparison with *T. aquaticus*, *E. coli* and *Bacillus subtilis*, whereas in cyanobacteria and *Streptomycetes* the extension is 10 residues long. On the other hand, the C terminal end of ω subunit and the loop structure between the α2 and α3 helices are extremely variable between different bacterial species.

Based on the crystal structures of *T. aquaticus*, *E. coli* and *Mycobacterium tuberculosis,* Kurkela *et al.* (2021) reveal that the ω subunit interacts almost exclusively with the β' subunit. One side of the α2-helix of the ω subunit contacts the conserved domain D of the β' subunit and the other side of the same structure interacts with the G domain of the β' subunit in *T. aquaticus* RNAP (Minakhin *et al.* 2001). Otherwise, the interactions between the ω subunit and the β' subunit differ across the bacterial species due to the non-conserved regions and other structural differences of RNAP core enzyme (Mao *et al.* 2018; Minakhin *et al.* 2001; Murakami 2013). Additionally, a recent study suggests that in *Porphyromonas gingivalis*, the ω subunit may interact with the σ4 domain in the holoenzyme form (Bu *et al.* 2024). The researchers indicate that this interaction leads to the assembly and stabilization of initiation complexes.

1.4.4 σ subunit is partially buried inside RNAP core

In the structural analysis of *T. aquaticus* and *Thermus thermophilus* RNAP holoenzymes, researchers show that the σ subunit interacts mainly with the surface of the core RNAP (Murakami *et al.* 2002; Vassylyev *et al.* 2002). Yet, a small segment of σ subunit connecting the amino-terminal and the carboxy-terminal domains (residues σ313-342 of *T. thermophilus*) is seen to be buried in the structures

of RNAP. The researchers show that the σ subunit divides to different domains (σ1, σ2, σ3 and σ4), which are connected with flexible linkers. The σ1 domain comprises regions 1.2-2.4 (residues σ74-254 of *T. thermophilus*) containing the non-conserved area between regions 1 and 2. It is composed of eight α helices that form four helix-turn-helix motifs. Formed by regions 2.4- 3.1, the σ2 domain (residues σ261-312 of *T. thermophilus*) consists of a bundle of three α-helices. The σ3 domain is considered to be a linker comprising a hairpin loop and consisting of roughly 30 residues (residues σ313-339 of *T. thermophilus*). As a linker, it is mainly unfolded and comprises region 3.2. Finally, the σ4 domain (residues σ340-423 of *T. thermophilus*) has four α-helices, which form two helix-turn-helix motifs comprising the conserved regions 4.1 and 4.2.

1.4.5 The conversion from the core RNAP to a functional holoenzyme

To establish proper promoter binding and to initiate the transcription RNAP core enzyme needs a σ subunit to form a holoenzyme structure. When comparing the crystal structures of *T. thermophilus* holoenzyme and *T. aquaticus* core RNAPs, the binding of the σ subunit is seen to change the conformation of the core enzyme in every domain from 2 Å to 12 Å (Borukhov and Nudler 2003; Vassylyev *et al.* 2002; Zhang *et al.* 1999). Most remarkable changes take place in the coiled-coil domain of β' (residues β'540-585 of *T. thermophilus*), αI subunit, large amino-terminal portion of β (residues β20-395 of *T. thermophilus*) and G flap domain of β (residues β702-828 of *T. thermophilus*). Inconsistently, these conformational changes upon σ subunit binding lead to the closing of the pincher, rather than opening of the cleft for the accommodation of double stranded DNA. The RNA exit channel is constricted by the shifting of the β G flap domain by 5-6 Å towards σ4 relative to the *T. aquaticus* core. The most outstanding change in the orientation occurs in the tip helix of the G flap, which conformation differs by 11 Å when comparing to the core enzyme. The αI domain and the large amino-terminal portion of β seem to change their conformation together with the β G flap. In conclusion, all these events seem to make the holoenzyme to enter a strained conformation, of which relaxation has been speculated to be involved in the σ subunit release when proceeding to the elongation part of the transcription (Borukhov and Nudler 2003).

The crystallographic studies of holoenzyme structures made by Vassylyev and co-workers (2002) or Murakami's team (2002) don't reveal all the regions of the holoenzyme. For example, the amino terminus of σ subunit including the conserved 1.1 region is not visible in these structures nor in the free σ subunit structures (Campbell *et al.* 2002). This mysterious domain is rather auto-inhibitory masking the σ subunit's DNA binding regions before attaching to the core enzyme (Bowers and Dombroski 1999 p. 1; Mekler *et al.* 2002; Wilson and Dombroski 1997). Additionally, Wilson and Dombroski (1997) propose that this domain is needed in a proper open promoter complex formation and transcription initiation at some promoters. It is also shown to be involved in the early steps of the

σ subunit binding to the core enzyme by interacting with the β G flap domain (Gruber *et al.* 2001; Sharp *et al.* 1999). In addition, carboxy terminal domains of both α subunits and the non-conserved region of β' (residues 252-363 of *T. thermophilus*) remain non-visible in the holoenzyme structures. However, the crystallographic structures of the αCTD-domains have been successfully defined when bound to catabolite activator protein (CAP) and DNA (Benoff *et al.* 2002).

1.4.6 The binding of σ subunit to the core enzyme results in multiple interface interactions

Borukhov and Nudler (2003) analysed the interface of the σ subunit and RNAP in the crystal structure of *T. thermophilus*. They point that the interface is very large containing altogether 115 potential contacts including 38 direct hydrogen bonds, 22 water mediated hydrogen bonds, 14 salt bridges and 41 van der Waal's contacts. Reviewers show that all of these interactions are focused on the β and the β' subunits. One would think that this many interactions would make the bond between the core enzyme and the σ subunit very strong, almost irreversible. However, the researchers reveal that these interactions besides the ones between σ1 and β' coiled-coil domain (residues 540-585 of *T. thermophilus*) are relatively weak and spread over a wide area. These statements match with the studied biochemical data (Gruber *et al.* 2001; Owens *et al.* 1998; Sharp *et al.* 1999). Additionally, Borukhov and Nudler (2003) state that despite the many interactions between σ subunit and non-conserved regions of β' (residues 163-452 of *T. thermophilus*) their role in binding may be limited. This is due to the observation, in which the deletion of this particular region in *E. coli* β' still results in higher σ-RNAP affinity than that of σ to RNAP core in *T. thermophilus*. However, this may also be due to the compensating effects of the large non-conservative region of σ132-385 interacting with the β domain in *E. coli*.

Many biophysical and biochemical evidence propose that binding of σ subunit to core RNAP is a cooperative process that occurs in many steps (Callaci *et al.* 1999; Mekler *et al.* 2002; Sharp *et al.* 1999; Zhang *et al.* 1999). In the beginning, σ1 binds with the β' coiled-coil structure and shifts σ2 and σ3 domains towards the β D region (residues 370-390 of *T. thermophilus*) and the β I region (residues 1010-1067 of *T. thermophilus*), respectively. At the same time 1.1 region of σ orientates towards the β G flap domain (residues 714-780 of *T. thermophilus*) and stimulates its opening. This movement helps σ3 and σ4 to slide under the β G flap into the RNA exit channel and finally towards the active site of the core enzyme. All these movements explain how the σ3 domain is eventually buried in the core enzyme, as mentioned earlier.

1.4.7 σ subunit release

Borukhov and Nudler (2003) ponder that the release of the σ subunit in the promoter escape would unlikely happen because of the many contacts between the core and σ and also the topological constraints mentioned above. On the other hand, they discuss the reasons that speak for the possibility for the σ subunit release despite these facts. First, in addition of those strong interactions between the σ subunit and the core there are many contacts that destabilize the bond. This is due to electrostatic repulsion, as well as energetically unfavourable located polar and hydrophobic residues. Secondly, the structural analysis of the β G flap domain indicates its flexible properties. Borukhov and Nudler (2003) propose that this makes the binding of the σ subunit to the core enzyme not that strict. Thirdly, studies that investigated single-point mutations in the σ subunit or the core enzyme show that these mutations make the binding affinity of the σ subunit weaker (Arthur *et al.* 2000; Fenton *et al.* 2000; Shuler *et al.* 1995). These results indicate that the σ subunit is in a metastable and strained conformation when bound to the core enzyme. Finally, it has been proved that relatively low concentrations of leucine and phenylalanine in solution trigger the σ subunit release with *E. coli* holoenzyme. (Maitra *et al.* 2002).

1.5 The structure of the promoter elements

As mentioned earlier, the initiation of transcription is the most regulated process in gene expression. The regulation occurs mostly in the promoter recognition, when RNAP binds to the specific sequence via σ subunit (Browning and Busby 2004). The most important regions for promoter binding are -10 and -35 elements, but generally there are up to five elements that are involved in RNAP holoenzyme binding and transcription initiation: UP-element, -35 element, extended -10 element, -10 element and discrimination region. Figure 3 shows a simplified view of these elements and their binding to the RNAP holoenzyme.

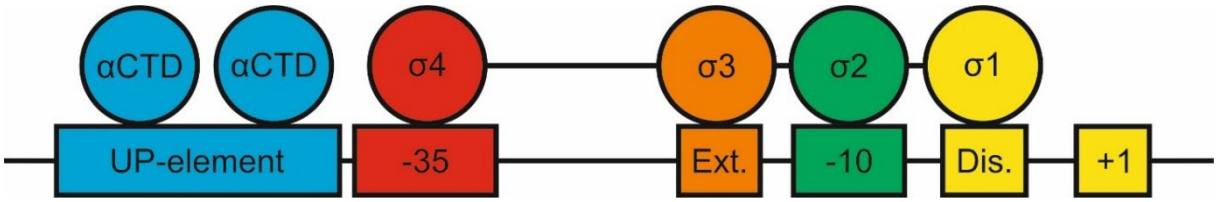

Figure 3. The five promoter regions involved in RNAP holoenzyme binding. Rectangular shapes represent the promoter sequences (UP element: blue; -35 element: red; extended -10 element: orange; -10 element: green and both discrimination and transcription start site (+1): yellow) and the circles above them show the RNAP holoenzyme regions that are involved in the promoter binding (αCTD: blue; σ4: red; σ3 orange; σ2: green and σ1: yellow.

1.5.1 UP element

The UP element, which is usually an adenosine- and thymine-rich sequence, is located upstream of the -35 element. A study shows that the *rrnB* P1 promoters interact directly with the two α subunits of the RNAP via their C terminal domains, which bind to two distinct sites within the UP element (Estrem *et al.* 1999). Another study suggests that the binding sites in the UP element are recognized by the αCTD tails of *E. coli* RNAP through the minor-groove electrostatic potential of the DNA sequence (Lara-Gonzalez *et al.* 2020). The UP element stimulates the transcription at least 30-fold for *E. coli* ribosomal RNA promoter *in vivo* (Ross *et al.* 1993). A separate study revealed that the transcription rate increased by up to 326-fold when these UP element sequences were modified (Estrem *et al.* 1998). In addition, the UP element of *E. coli guaB* promoter has been shown to be involved in the growth rate-dependent control of the cell (Husnain and Thomas 2008). Also, a study indicated that the UP element of the promoter encoding the flagellin gene *hag* stimulates the transcription from the SigA and SigD dependent promoters in *B. subtilis* (Caramori and Galizzi 1998). These milestones show that the UP element has an important role as the structural component when analysing the transcription rate of a studied promoter.

1.5.2 -35 element

The -35 element is a six-nucleotide long sequence located between the UP element and the extended -10 element roughly 35 nucleotides upstream from the transcription start site (+1). It has been proposed to bind the 4.2 region of the sigma factor in the structural studies with the *T. aquaticus* RNAP and its σ^A (Campbell *et al.* 2002). The same study shows that the 4.2 region of the sigma factor consists of three α helices and has a helix-turn-helix motif, other of which recognizes the conserved sequence of the -35 element (TTGACA).

1.5.3 Extended -10 element

The two nucleotides located straight upstream of the -10 element are identified as extension of -10 region. A study proposes that one of the three helixes in the 3.0 region of σ^A bound to the *T. aquaticus* RNAP recognizes this region. (Campbell *et al.* 2002). In *Borrelia burgdorferi,* the thymine-rich spacer, including the extended -10 element, was found to be essential for optimal transcription, when the -35 element was disrupted by deletion (Gautam *et al.* 2009). This suggests that, for some promoters, the -35 element is not essential for proper transcription as long as the extended -10 element is present. On the other hand, a promoter that had a strong -35 and extended -10 element, but a very weak -10 element was studied using the *E. coli* RNAP-σ^{70} holoenzyme and was reported to be active both *in vivo* and *in vitro* (Hook-Barnard *et al.* 2006). In this study, the mutations in the sequence of the -35 element or the extended -10 element made the transcription from this promoter inactive. This

emphasizes the importance of the extended -10 element, when either the -35 or the -10 element is poorly structured. Furthermore, a study showed that other properties of the spacer region, such as its sequence composition and length, also influence promoter strength, not just the extended −10 element (Klein *et al.* 2021).

1.5.4 -10 element

The -10 element (or the Pribnow Box) of the promoter consists of conserved bases in the sequence (TATAAT), which reach from the -12 base to the -7 base respectively to the start site of the transcription (+1) (Feklistov and Darst 2011). A study reveals that the melting of the promoter DNA in the initiation of the transcription starts at the -10 element of the promoter approximately from the distance of 12 base pairs from the transcription start site when using the *E. coli* RNAP (deHaseth and Helmann 1995). Researchers propose that there are highly conserved basic and aromatic residues in the region 2.3 of the σ^A which are involved in the process of promoter melting in the *T. aquaticus* RNAP holoenzyme (Campbell *et al.* 2002). At the promoter, the melting of double-stranded DNA by the *E. coli* RNAP holoenzyme is indicated to involve the non-template strand of the -10 element (Young *et al.* 2004).

1.5.5 Discriminator element

In some promoters, a GC-rich region is located downstream of the -10 element, separating it from the transcription start site. This region is called the discriminator element. A study shows that the 1.2 region of the *E. coli* σ^{70} subunit interacts with the discriminator element with many or even all the promoters in the RNAP holoenzyme binding (Haugen *et al.* 2008). However, the strength of the interaction depends on the discriminator sequence. This study indicates that the optimal sequence in the non-template strand for proper binding is GGG, when starting next to the -10 hexamer. In addition, Haugen and co-workers (2008) proposed that the strength of this interaction affects to the lifetime of the promoter complexes. The strong interaction at the discrimination element activates transcription with the promoter complexes that have intrinsically short lifetime.

1.5.6 Transcription start site

The first nucleotide transcribed in the promoter area marks the transcription start site (+1). Most commonly the starting base of the transcription is purine (G or A) (Mendoza-Vargas *et al.* 2009). The optimal spacing between the -10 element and the transcription start site is ~6 bp, but the spacer length isn't the only thing that determines the start site. For example, RNAP holoenzyme makes sequence-specific interactions with the non-template strand at the downstream part of the transcription bubble called core recognition element (CRE; bases from -4 to +2) (Vvedenskaya *et al.* 2016). These interactions between RNAP holoenzyme and CRE are proposed to affect the expansion and

contraction of the transcription bubble, which could influence the determination of transcription start site.

1.6 From promoter recognition to a stable open promoter complex

The formation of the transcription bubble and a stable open promoter complex begins when RNAP holoenzyme binds specifically to a promoter resulting a closed promoter complex (Mazumder and Kapanidis 2019). This is followed by the conformation changes of both the RNAP holoenzyme and the DNA promoter area. As a result, the transcription bubble is formed in the length of ~13 base pairs. There have been intensive studies concerning the structural transformation from the closed promoter complex to the open promoter complex. The formation of the stable open promoter complex is a crucial step of transcription initiation. Next, a few insights into this area of research are presented.

1.6.1 Promoter recognition forms a closed promoter complex

The binding of RNAP holoenzyme to a promoter forms a closed complex (RP_c). Studies using the *E. coli* RNAP and the *T7A1* promoter suggest that the promoter binding begins when the αCTD of the polymerase recognizes the UP element. (Sclavi *et al.* 2005). This is followed by the interaction between the -35 element and the σ4-region. Next the σ2- and σ3-regions interact with the -10 element and the spacer area (Rogozina *et al.* 2009). The RNAP holoenzyme, when fully interacting with the promoter region, forms a complete closed promoter complex RP_c. A team investigated the structural details of open promoter complex formation using *E. coli* RNAP and the transcription factor TraR on the *rpsT*P2 promoter, employing single particle cryo-electron microscopy (Chen *et al.* 2020). Figure 4 shows the closed promoter complex, in which the RNAP holoenzyme binds to the promoter region. I will be presenting figures of the potential intermediate and the open promoter complex using the PDB structures defined in this study by visualizing them with PyMOL (Schrodinger). These figures demonstrate the structural changes in the polymerase and especially the bending and unwinding of the duplex DNA. The binding of the RNAP holoenzyme to the promoter region is discussed in more detail in section 1.5.

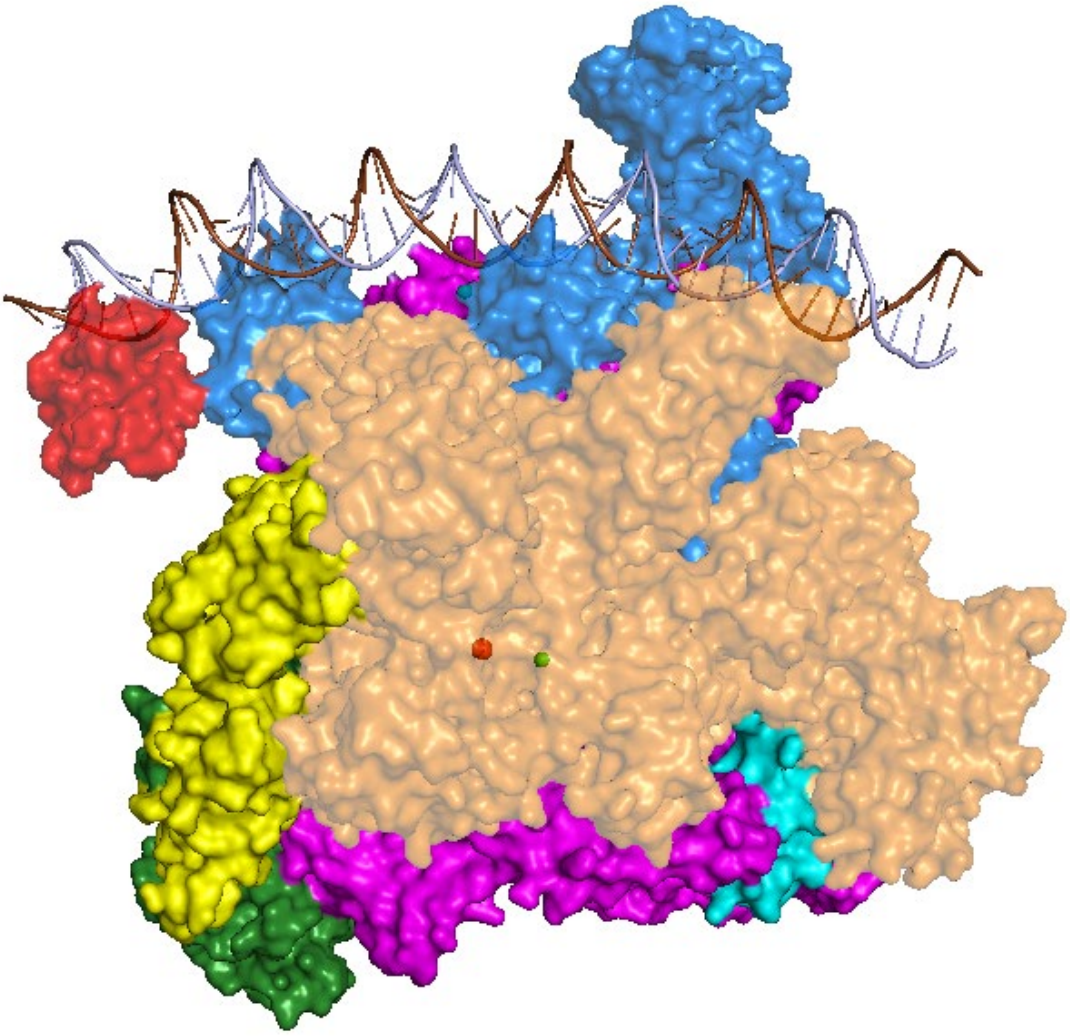

Figure 4. The structure of E. coli RNAP holoenzyme bound to the rpsTP2 promoter with the transcription factor TraR (PDB code 6PSQ (Chen et al. 2020)). In the figure, the double stranded DNA is bound by the holoenzyme structure. The subunits of the holoenzyme are coloured as follows: β subunit: orange, β' subunit: magenta, α subunits: green and yellow, αCTDs: red, ω subunit: dark gray, σ subunit: blue and transcription factor TraR: cyan. The catalytically important Mg^{2+}-ion is shown as red sphere, Zn^{2+}-ions are shown as green spheres, and the DNA is coloured in brown (template strand) and light blue (non-template strand). The direction of the DNA is from left (downstream) to right (upstream). Pymol (The PyMOL Molecular Graphics System, Version 3.0 Schrödinger) was used for visualization.

In some studies, the isomerisation steps leading to the formation of the open promoter complex are described by the simplified equation R + P → RP_c →RP_o (Saecker *et al.* 2011). Still, two kinetically important intermediates between the closed (RP_c) and the stable open promoter complexes (RP_o), I_1 and I_2, have been identified (see equation 1), although characterizing the properties of these complexes has proven challenging. However, studies employing real-time hydroxyl radical and potassium permanganate footprinting, as well as structural modelling, propose successfully features of these cryptic intermediates.

$$RNAP\sigma + P(promoter\ DNA) \leftrightarrow RP_c \leftrightarrow I_1 \leftrightarrow I_2 \leftrightarrow RP_o$$

Equation 1. The steps from promoter recognition to a stable open complex. RNAPσ: RNAP holoenzyme; P: promoter DNA; RP_c: closed promoter complex; I_1: intermediate 1; I_2: intermediate 2; RP_o: stable open promoter complex.

1.6.2 Organisation of the DNA strands in the intermediate I_1

Conformational changes occur in both RNAP holoenzyme and the double stranded DNA during the transition from RP_c to the first intermediate I_1. These steps lead, for example, to interactions between RNAP holoenzyme and upstream DNA, as well as to the movement of DNA closer to the active cleft of the polymerase. Evidence supporting these interactions are presented below.

A study using potassium permanganate and real-time hydroxyl radical footprints, when investigating *E. coli* RNAP-σ^{70} holoenzyme and λP_R promoter, proposes the first intermediate (I_1) to protect DNA upstream from position -81 to -71 respect to the start site (Davis *et al.* 2007). Additionally, this study shows that the interactions between closed complex and upstream DNA above position -47 rearrange the conformation of the cleft between RNAP pinchers so that the downstream DNA duplex can enter it. After entering the cleft, the downstream DNA bases from -11 to +20 are shown to be buried in the active site of the polymerase and the DNA is shown to be double stranded in this conformation. The process of DNA entering the active cleft is described in more detail below.

A study focusing on kinetics and structural properties of the promoter complex intermediates with the same RNAP/promoter combination as above proposes that the protection from position -11 to +15 is due to a sharp 90 °C kink in the DNA that leads the strands between the cleft between β and β' subunits (Saecker *et al.* 2002). Additionally, the study shows that these interactions with the DNA downstream from position -5 and RNAP active site result in the closure of the RNAP jaws and finally opening the double stranded DNA leading to the second intermediate.

1.6.3 The separation of the strands takes place when moving from I_1 to I_2

Studies investigating the first two intermediates in the λP_R promoter using the real-time footprinting method found that, in I_1, the bases from −11 to ~20 are unreactive to permanganate and protected from hydroxyl radical and DNase I on both strands (Craig *et al.* 1998; Davis *et al.* 2007). On the contrary, in the studies with the *E. coli* RNAP–λP_R promoter complex the bases from -11 to +2 were shown to be permanganate-reactive in complex I_2 (Gries *et al.* 2010). In conclusion, the results of these studies indicate that the opening of the duplex DNA happens when moving from I_1 to I_2.

In a study covering the kinetics of the open complex formation using *E. coli* RNAP and *2-AP* substituted promoter, the increase in the concentration of RNAP holoenzyme in the beginning of the

reaction leads to the accumulation of the closed complex intermediates (I_1). This indicates that isomerisation step from the first intermediate to the second, including the strand separation, is rate determining and has been shown to be a slow process compared to the other steps of open complex formation. (Schroeder *et al.* 2009).

The suggested conformation changes that lead to the subsequent intermediate (I_2) are not unanimous. One hypothesis based on the structural studies of *T. aquaticus* and *T. thermophilus* RNAPs proposes a model in which the separation of the double stranded DNA takes place above the active site cleft of the RNAP, and this is then followed by the entry of the template strand into the active site (Murakami *et al.* 2003; Vassylyev *et al.* 2002). This occurs due to the narrow width of cleft (~25 Å), which indicates that only a single stranded DNA could enter the active site. Narrowness of the cleft is not the only obstacle to overcome. Studies report that, before the formation of the open promoter complex by *E. coli* RNAP, the σ^{70} region 1.1, which mimics single stranded DNA, must be repositioned from the active cleft (Mekler *et al.* 2002; Nagai and Shimamoto 1997). Studies using Brownian dynamic simulations, as well as time-resolved potassium permanganate and hydroxyl-radical X-ray footprinting support this hypothesis (Chen *et al.* 2010; Rogozina *et al.* 2009).

The second hypothesis describes a situation, in which the double stranded DNA of the λP_R promoter is first packed into the cleft between the β and β' subunits of *E. coli* RNAP, and then the RNAP elements carry out the separation of the strands (Gries *et al.* 2010; Kontur *et al.* 2010). Again, it has been hypothesized that in the intermediate I_1, the 1.1 region of σ^{70} remains bound in the active site, blocking the entry of the duplex DNA as it progresses toward the cleft for strand separation (Kontur *et al.* 2006). Researchers propose that the exposure of the active site occurs during the transition toward intermediate I_2, when region 1.1 is relocated outside of the cleft.

Supporting the second hypothesis, a study investigating *E. coli* RNAP and the λP_R promoter demonstrates that the I_1 complex has only small effects on DNA melting in the presence of different sodium salts (Cl-, Glu- and F-), whereas this process was shown to be salt dependent in solution (Kontur *et al.* 2010). This may be due to the burial of duplex DNA in the sequestered environment of the active cleft. This study indicates that DNA duplex binds first in the active site of the RNAP, and this is then followed by the conformational changes that open the transcription bubble.

A study demonstrates that the second intermediate, I_2, in open complex formation is heparin insensitive unlike the first intermediate (Craig *et al.* 1998). Again, this supports the theory that the opening of the strands occurs in the transition from I_1 to I_2. This study investigated I_2 by decreasing the temperature in the reaction to accumulate the intermediate. In addition, Craig and co-workers (1998)

propose that clamp closure by *E. coli* RNAP at the downstream region of the λP_R promoter initiates DNA unwinding.

In addition of these mentioned above, other types of intermediates have been proposed to exist in the formation of the open promoter complex (Chen *et al.* 2020; Hubin *et al.* 2017). Some studies also suggest that the intermediates are species and promoter specific (Chen *et al.* 2020). Figure 5 shows the open promoter complex intermediate RP_i1.5b formed by *E. coli* RNAP holoenzyme in the presence of the transcription factor TraR on the *rpsT*P2 promoter. Chen *et al.* (2020) report it to be a crucial intermediate on the way toward the open promoter complex in *E. coli* even without the transcription factor TraR. In this intermediate, the transcription bubble has opened in the length of 5 base pairs (-11 to -7). The non-template adenine in the -11 position is shown to interact with σ2-pocket, while the non-template strand from -10 to -8 obtains the conformation in which the bases contact the σ2 in the similar way as in the RP_o. The unwinding of the duplex DNA is a complex process, which leads closer to the stable open promoter complex RP_o and transcription initiation. The key players of the strand separation are discussed below.

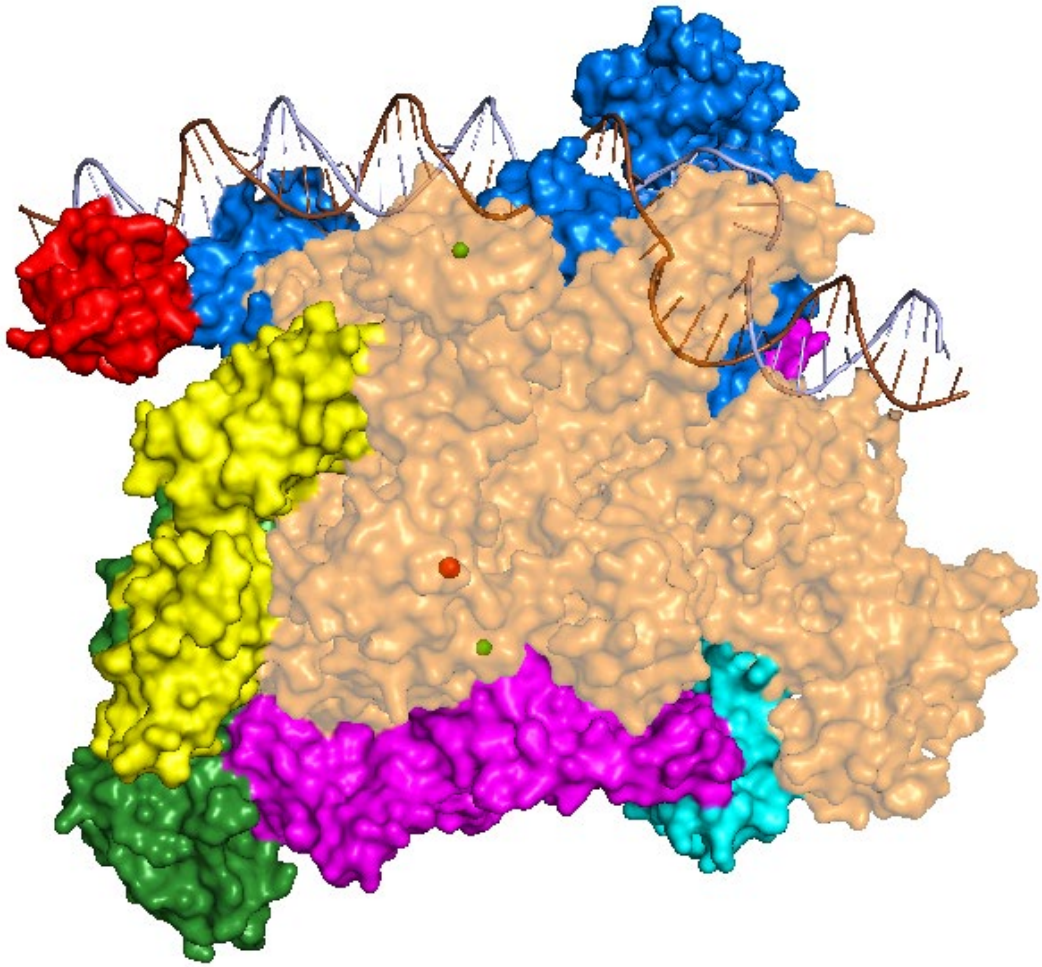

Figure 5. The structure of RNAP holoenzyme open promoter complex intermediate $RP_i1.5b$ with the rpsTP2 promoter and the transcription factor TraR (PDB code 6PST (Chen et al. 2020)). The figure shows the initiation of the strand separation. The transcription bubble has opened in the length of 5 base pairs (-11 to -7) and the non-template adenine in the -11 position interacts with σ2-pocket, while the non-template strand from -10 to -8 obtains the conformation in which the bases contact the σ2 in the similar way as in the RP_o. Figure shows that the bending of the DNA is established, and the next conformational changes will lead the template strand even closer to the catalytically active Mg^{2+}-ion. The subunits of the holoenzyme are coloured as follows: β subunit: orange, β' subunit: magenta, α subunits: green and yellow, αCTDs: red, ω subunit: dark gray, σ subunit: blue and transcription factor TraR: cyan. The catalytically important Mg^{2+}-ion is shown as a red sphere, Zn^{2+}-ions are shown as green spheres, and the DNA is coloured in brown (template strand) and light blue (non-template strand). The direction of the DNA is from left (downstream) to right (upstream). Pymol (The PyMOL Molecular Graphics System, Version 3.0 Schrödinger) was used for visualization.

1.6.4 The structural elements involved in the initiation of strand separation

The melting of the duplex DNA is proposed to occur by base flipping. Researchers suggest that the conserved non-template base -11A is the first base to initiate duplex DNA separation and to interact favourably with the RNAP holoenzyme (Helmann and deHaseth 1999; Heyduk *et al.* 2006; Shultzaberger *et al.* 2007). Alternatively, the researchers propose that the base at the position of -7 flips first or at least just after the opening of the strands. Studies supporting the first proposal have shown that substitutions and depurinations at -11A had a negative effect on the open complex formation (Li and McClure 1998; Schroeder *et al.* 2007).

Substitutes in amino acids T429, Y430 and W433, which all locate in the region 2.3 of *E. coli* σ^{70} and interact with promoter area in which DNA opening takes place, are shown to have a negative effect in the open complex formation (Fenton *et al.* 2000; Panaghie *et al.* 2000). This indicates that these amino acids are crucial for favourable RNAP holoenzyme/promoter interactions and the strand separation. A study proposes that especially tyrosine at position 430 in region 2.3 of σ^{70} interacts directly with the non-template -11 adenine (Schroeder *et al.* 2009). This may also indicate the importance of this amino acid in base flipping and in the initiation of strand separation. Additionally, it has been suggested that tryptophan at position 433 is involved in base flipping in the early steps of DNA unwinding (Fenton *et al.* 2000; Tomsic *et al.* 2001). Finally, flipping of the bases continues downstream from the -11 to +2 and results in a complete transcription bubble.

The separation of duplex DNA over the distance of 12-14 base pairs at 25-37 °C without RNAP being present is a highly enthalpic process (~70-84 kcal) (Holbrook *et al.* 1999). When investigating the kinetics of the strand separation process ($I_1 \rightarrow I_2$) with *E. coli* RNAP-σ^{70} holoenzyme and λP_R promoter, researchers report a kinetic signature of approximately ~34 kcal (Saecker *et al.* 2002). This indicates that when RNAP holoenzyme is involved in the promoter melting there must be some energetically favourable changes that decrease the enthalpy of the reaction. Researchers suggest that this may be due to the interactions with the non-template strand and the aromatic residues including the ones mentioned above (Saecker *et al.* 2011). In conclusion, the interactions between the RNAP holoenzyme and the promoter region appear to play an active role, first in bending the duplex DNA within the active site cleft, and secondly in strand separation during the transition from the first intermediate (I_1) to the second (I_2), through these favourable interactions between the aromatic residues and non-template DNA.

1.6.5 The final step: large conformational changes lead to a stable open promoter complex

The transition from the unstable intermediate, I_2, to the final open promoter complex, RP_o, involves large conformational changes. A study proposes that over 100 residues in RNAP holoenzyme fold and interact with downstream DNA backbone (Kontur *et al.* 2010). Researchers demonstrate that this folding occurs in the regions of β'- clamp and -jaw. This closure of the clamp is indicated to take place after the DNA is successfully unwound and loaded in the active site of the cleft. These interactions are also shown to stabilise the open conformation of the strands in RP_o. Another study proposes that the starting base of transcription (+1) is at the same place in both I_2 and RP_o, since its similar reactivity with permanganate in both complexes (Gries *et al.* 2010). This indicates that the opening of the

transcription bubble in $I_1 \rightarrow I_2$ transition already moves the +1 base of the template strand to the active site.

On the other hand, Gries and co-workers show that the non-template thymines -4/-3 and +2 have 2-fold reactivity when comparing the I_2 and RP_o complexes. This suggests that the non-template strand is relocated when moving from the second intermediate to the final stable open complex. This relocation of the non-template strand may be linked with reorganization of the switch regions and the region 1.1 of σ^{70} (Belogurov *et al.* 2009; Gries *et al.* 2010; Mekler *et al.* 2002). The researchers hypothesize that these events and the closure of the RNAP clamps together lead to the stableness of RP_o compared to the second intermediate I_2. Additionally, a recent study demonstrates that the final progression to RP_o is partially driven by clamp closure, that is fueled by the favourable free energy generated through in-cleft DNA-RNAP interactions (Saecker *et al.* 2024).

Finally, Figure 6 shows the open promoter complex of *E. coli* RNAP on the *rpsT*P2 promoter with the transcription factor TraR, as detemined by cryo-electron microscopy (Chen *et al.* 2020). The transcription bubble has fully formed, and the start site of transcription locates right next to the catalytically active Mg^{2+}-ion (red sphere). From this point on, the actual transcription process and addition of the first nucleotides are ready to begin.

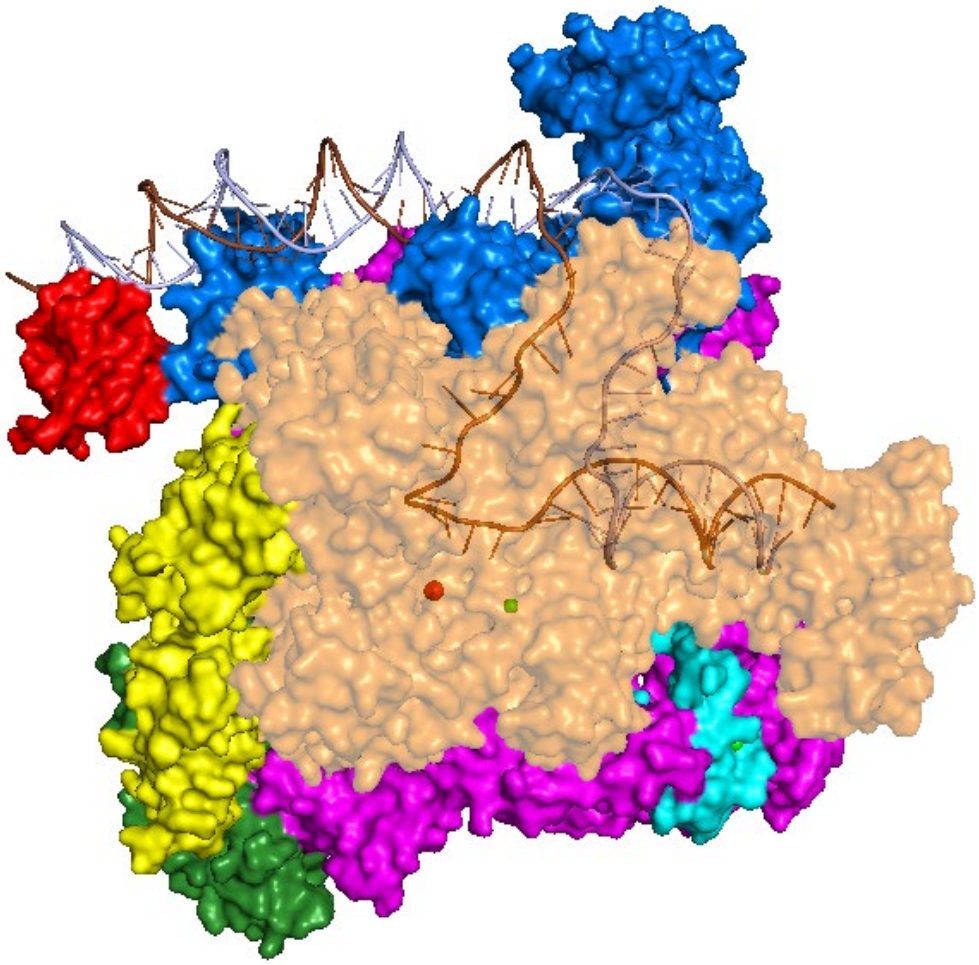

Figure 6. The structure of E. coli RNAP holoenzyme open promoter complex with the rpsTP2 promoter and transcription factor TraR (PDB code 6PSW (Chen et al. 2020)). The figure shows the fully formed transcription bubble. The kink in the DNA leads the template strand into the active site of the RNAP holoenzyme next to the catalytically active Mg^{2+}-ion. The subunits of the holoenzyme are coloured as follows: β subunit: orange, β' subunit: magenta, α subunits: green and yellow, αCTDs: red, ω subunit: dark gray, σ subunit: blue and transcription factor TraR: cyan. The catalytically important Mg^{2+}-ion is shown as red sphere, Zn^{2+}-ions are shown as green spheres, and the DNA is coloured in brown (template strand) and light blue (non-template strand). The direction of the DNA is from left (downstream) to right (upstream). Pymol (The PyMOL Molecular Graphics System, Version 3.0 Schrödinger) was used for visualization.

1.7 CarD

CarD is a transcription factor found in several bacterial species including for example *M. tuberculosis, Mycobacterium smegmatis* and *Myxococcus xanthus* (García-Moreno *et al.* 2010; Srivastava *et al.* 2013). A study shows that it interacts directly with the mycobacterial RNAP β1 region (L. A. Weiss *et al.* 2012). The same study presents the importance of this interaction for viability, rifampicin resistance, and stress resistance of *M. tuberculosis.* This indicates that the regulatory effects of CarD are mediated via its contacts with RNAP. Srivastava *et al.* (2013) also propose that the effects of CarD particularly take place in the transcription initiation at the promoter complexes in *M. smegmatis*. Therefore, it can be considered as a global regulator of transcription initiation in these bacteria.

The crystal structure of *T. thermophilus* CarD shows that it contains two domains (Srivastava *et al.* 2013). The first one is a Tudor-like folded structure in the N terminal end of the protein. The second one, located at the C terminus, is a compact structure comprising α helical folds that do not resemble any previously described structures. The orientation of these two elements is most likely rigid due to their interactions with each other in a small, but significant surface, which comprises conserved amino acids across CarD protein family. Additionally, Srivastava and co-workers (2013) performed structural modelling that suggests that CarD-RID binds to RNAP β1 region such that the C terminal domain of CarD is orientated towards -10 element of the promoter DNA on the opposite side of the σ subunit. The structural modelling had no indications of interactions between CarD and σ subunit.

Researchers demonstrate that, in mycobacteria, CarD forms favourable protein-protein (CarD-RID/RNAP β1) and protein-DNA interactions at the upstream edge of the transcription bubble (Srivastava *et al.* 2013). This has been shown to help in promoter binding, to activate the transcription initiation, and finally to activate the transcription globally in mycobacteria. Conversely, a group studying *Rhodobacter sphaeroides* suggests that CarD has an inhibitory rather than activating effect when regulating transcription from its own promoter. (Henry *et al.* 2021). This supports the idea that CarD can regulate transcription in different ways. Figure 7 shows the binding of CarD to the RNAP holoenzyme.

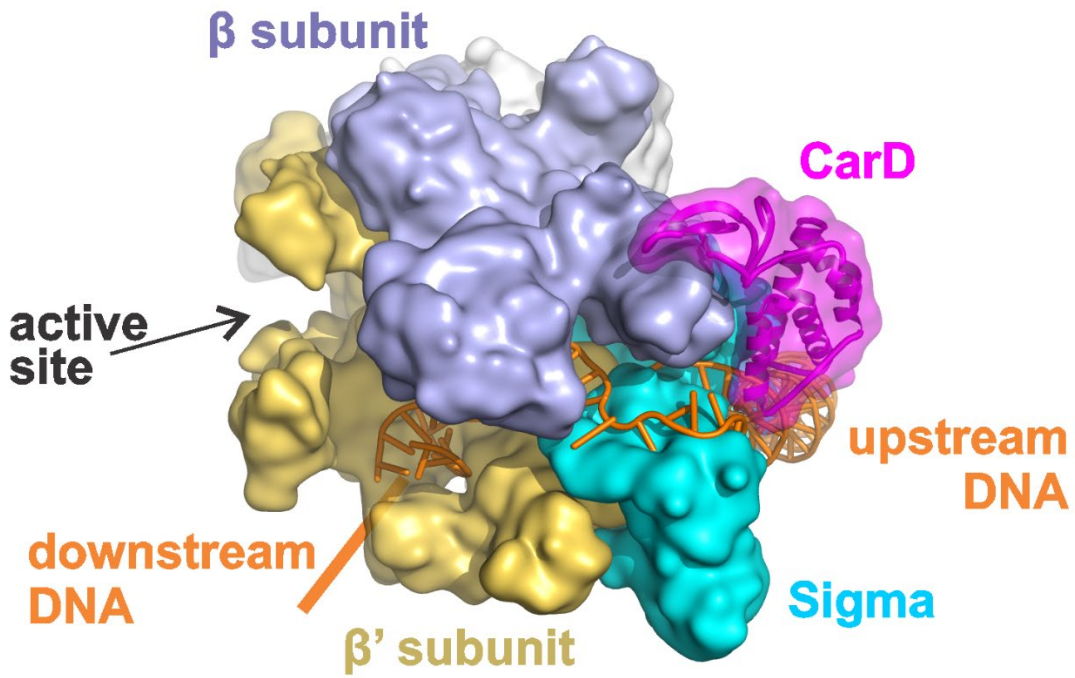

Figure 7. RNAP holoenzyme with the transcription factor CarD bound to the unwound DNA. The figure shows that CarD is bound in the side of β subunit and orientated towards the DNA at the opposite side of σ subunit. The colour codes of different subunits are presented in the figure. The figure is visualised by my supervisor Georgy Belogurov using PDB ID 4XLR.

1.8 The goal of the project

The goal of this study is to investigate spirochaetal transcription initiation by using a reconstituted transcription system. The study focuses on transcription activity of *S. africana* RNAP compared to that of *E. coli* RNAP using ten transcription templates encoding nine spirochaetal promoters of different strengths and one *E. coli* control promoter. Transcription activities of both *E. coli* and *S. africana* RNAPs were measured at pH 7.5, and additionally, the transcription activity of *S. africana* RNAP was measured at pH 9. The change of the pH was due to the fact that *S. africana* is an alkalophilic bacterial species and possibly doesn't control its intracellular pH. Moreover, we studied the effects of CarD on spirochaetal transcription activities with selected promoters. As a part of the project, we constructed eight of the transcription templates used in the transcription activity assays. Finally, we defined the actual start sites of the selected transcription templates to confirm their identities.

2 Materials and methods

2.1 Protein expression

2.1.1 Overnight expression

Plasmids were first transformed to *E. coli* cells (strains Xjb or T7) using $CaCl_2$ mediated DNA uptake method or a ready cell stock with wanted plasmid was used. The cells were plated on LB (Luria-Bertani) agar with antibiotics and incubated overnight at 37 °C. Next day a colony was inoculated to a 50 ml preculture LB medium with antibiotics and let it grow turbid at 37 °C at 250 rpm. The preculture was inoculated to the main culture in a ratio of one to one thousand. Altogether the main culture volume was four litres in protein expression. The overnight expression reagents were added to the final concentrations of 1 mM $MgSO_4$; 5 mg/ml glycerol; 500 µg/ml glucose; 2 mg/ml alpha-lactose; 100 mM PO_4; 25 mM SO_4; 50 mM NH_4; 100 mM Na; 50 mM K; 200 µg/ml arabinose and antibiotics (ampicillin 100 µg/ml or kanamycin 30 µg/ml). The main culture was incubated overnight at 37 °C at 250 rpm in flasks or in a fermenter (aeration 6 l/min). Next day the OD600 of the culture was measured with a spectrophotometer to confirm that the cell growth had stopped. The cells were harvested by centrifuging 7000 x g for ten minutes at +4 °C and the pellet was stored at -80 °C for further use.

2.1.2 IPTG induction

E. coli cells (strains Xjb or T7) from transformation or a glycerol stock were plated on LB agar with antibiotics. A 50 ml LB preculture with antibiotics was inoculated from a colony and the culture was incubated at 37 °C at 250 rpm until it was turbid. The preculture was inoculated to the main culture with antibiotics in a ratio of one to one hundred and the OD600 of the main culture was followed until it reached 0.6 - 0.8. Then the protein expression was induced with 1 mM IPTG and also arabinose was added if culturing Xjb cells. The culture was incubated at 37 °C at 250 rpm for three hours in bottles or in a fermenter (aeration 6 l/min). Cells were harvested by centrifuging at 7000 x g at +4 °C for ten minutes. The pellet was weighted and stored at -80 °C for further use.

2.2 Protein purification

Proteins were separated from the cells using different chromatography techniques that are listed in the sections below. Four main buffers were used in those purification steps (A buffer, B buffer, lysis buffer and dialysis/storage buffer). Buffer A contained 50 mM Tris-HCl pH 7.9; 1 mM β-mercaptoethanol; 5 % (v/v) glycerol; 100 µM Na-EDTA and Buffer B 50 mM Tris-HCl pH 7.9; 1 mM β-mercaptoethanol; 5 % (v/v) glycerol; 100 µM Na-EDTA; 1.5 M NaCl. These two buffers were filtered and sonicated before use. Lysis buffer contained 50 mM Tris-HCl pH 7.9; 1 mM β-

mercaptoethanol; 5 % (v/v) glycerol; 500 mM NaCl and dialysis/storage buffer 20 mM Tris-HCl pH 7.9; 100 µM Na-EDTA; 50 % (v/v) glycerol; 150 mM NaCl; 0.1 mM DTT.

2.2.1 Nickel affinity chromatography

The frozen cells were suspended in lysis buffer, and one protease tablet (Pierce Protease Inhibitor Tablets, Thermo Scientific) was added per 50 ml of cell suspension. When T7 cells were used, lysozyme was added to the cell solution at the concentration of 1 mg/ml, and the mixture was incubated on ice for one hour. The final cell lysis was performed by sonicating the cell solution in cycles where there were a 20-second sonication period and a 40-second rest period. The lysate was separated by centrifuging at 50 000 x g at 4 °C for one hour. Tween-20 (0.2 % final) was added to the lysate and a nickel sepharose column was equilibrated with lysis buffer. The lysate was loaded to the nickel gravity column containing 1.5 ml nickel sepharose resin and let it flow through by gravity. In some cases, the lysate was incubated with the resin for 60 minutes and the loaded resin was collected by centrifugation (4000 x g, 4 °C, 10 min). The loaded resin was washed with lysis buffer (10 ml) and then eluted with 20-, 50-, 100- and 250-mM imidazole in lysis buffer. Samples from each elution were run on an SDS-PAGE gel (BoltTM 4 to 12% Bis-Tris Plus Gel, Invitrogen) in MES buffer (200 V, 22 min) with a ladder (PageRulerTM Prestained Protein Ladder, Catalog number: 26616, Thermo Scientific) to select the correct fractions based on protein quantity and purity for further purification steps. The gels were stained overnight according to the manufacturer's instructions (GelCodeTM Blue Safe Protein Stain, Thermo Scientific). The nickel columns were stored with MQ-water and regenerated before next use.

2.2.2 Heparin 5 ml chromatography

The eluted fractions with the protein of interest were combined and diluted to <200 mM NaCl with buffer A. The 5 ml heparin column (HiTrapTM Heparin HP 5ml, Cytiva Life Science) was equilibrated to 10 % buffer B in buffer A with chromatography system ÄKTAprime plus (Cytiva Life Sciences) running the buffers through the column two column volumes with stabilised conductivity and UV absorption. The sample containing the protein of interest was loaded to the column with syringe pump (Catalogue number: 9034914, Fischer Scientific) through 0.45 µM filter (Filtropur S 0.45). The column inserted to the chromatography system was washed with 10 % buffer B in buffer A for two column volumes with stabilised UV absorbance spectrum. The protein was eluted by running a 100 ml gradient from 10 % to 100 % buffer B and collected in 1 ml fractions while following the UV absorbance spectrum. A peak in the UV absorbance spectrum indicated an eluted protein. The samples from the fractions with the highest UV peaks were run on an SDS-PAGE gel as described above to decide which ones to use in next purification steps. The column was stored according to manufacturer's instructions.

2.2.3 Anion exchange chromatography

The fractions with the protein of interest were combined and diluted to <200 mM NaCl with buffer A. A quaternary ammonium ion exchange column (Resource^{TM} Q 6 ml, Cytiva Life Sciences) was equilibrated with 10 % buffer B in buffer A in ÄKTAprime plus the same way as in the case of the heparin column. The diluted fractions were loaded to the column with syringe pump through a 0.45 µM filter. The column inserted to the chromatography system was washed with 10 % buffer B in buffer A for two column volumes with stabilised UV absorbance spectrum. The protein was eluted from the column with a 100 ml gradient which was run from 10 % to 100 % buffer B and collected in 1 ml fractions following the UV absorbance spectrum for peaks indicating eluted proteins. The samples from the fractions containing the protein were run on an SDS-PAGE gel as described in section 2.2.1. The column was stored according to manufacturer's instructions.

2.2.4 Heparin 20 ml chromatography

In some cases, a 20 ml heparin column (HiPrepTM Heparin FF 16/10, Cytiva Life Sciences) was used in ÄKTAprime plus as a first purification step. The frozen cells were first suspended in 10 % buffer B in buffer A with one protease tablet per 50 ml of solution and sonicated as described in section 2.2.1. The lysate was separated by centrifuging at 50 000 x g at 4 °C for one hour and Tween-20 (0.2 % final) was added to the clarified lysate. The heparin column was equilibrated to 10 % buffer B in buffer A the same way as described in the section 2.2.2. The clarified lysate was loaded to the column using a 0.45 µM filter with syringe pump. The column inserted in the chromatography system was washed with 10 % buffer B in buffer A for three column volumes so that the UV absorbance spectrum had stabilised. The protein was eluted with a 20 ml gradient from 10 % to 100 % buffer B and collected in 1 ml fractions following the UV absorbance spectrum for peaks indicating eluted proteins. In some cases, the potential fractions with protein were run in a gel to confirm the purification result, but in other cases all the fractions at the peak were collected and used in the next step of the purification. This was due to the broadness of the peak and also the specificity of the column. The column was stored according to the manufacturer's instructions.

2.2.5 Strep-tag affinity chromatography

This purification step was performed fully with syringe pump. The 5 ml strep-tactin column (StrepTrapTM XT 5 ml, Cytiva Life Sciences) was equilibrated with 50 % buffer B in buffer A. The combined fractions from the previous chromatography step were loaded into the column with syringe pump through a 0.45 µM filter and the column was washed with 50 % of buffer B in buffer A for three column volumes. The protein was eluted with biotin buffer (50 mM biotin, 100 mM Tris, 500mM NaCl, 1mM β-mercaptoethanol) and collected in 1 ml fractions manually. The selected fractions were

run on an SDS-PGE gel as described earlier. The column was stored according to the manufacturer's instructions.

2.2.6 Concentration and dialysis of the purified proteins

After the gel run the fractions containing the highest amount and purest protein were selected for the concentration step. The concentration was performed using regenerated cellulose membrane filters with 3 kDa or 10 kDa cutoff (Amicon® Ultra-4, Merck Millipore). The purified proteins were concentrated by centrifuging 4000 x g at 4 °C until the concentration of the protein solution was reasonable for dialysis. The retentate was dialysed using a 3.5 kDa cutoff membrane (Slide-A-Lyzer™ MINI Dialysis Devices, 3.5K MWCO, Thermo Fischer Scientific) in dialysis/storage buffer overnight. Final concentration of the protein stock was measured spectroscopically (NanoDrop, Thermo Fischer Scientific). The protein stocks were stored at -20 °C in dialysis/storage buffer and handled with cold block or ice.

2.2.7 *S. africana* RNAP expression and purification

pVN005 plasmid was transformed into *E. coli* Xjb cells and the overnight expression protocol was performed for protein expression as described in section 2.1.1. The protein was purified using the heparin 20 ml column (2.2.4) and the strep-tactin column (2.2.5). This protein was also purified with another combination of the chromatography columns first using the nickel chromatography column (section 2.2.1), second the heparin chromatography 5 ml column (section 2.2.2) and finally the anion exchange chromatography column (section 2.2.3). After the purification steps the protein was concentrated and dialysed as described in section 2.2.6.

2.2.8 MMLV Reverse transcriptase expression and purification

pGB225 plasmid was transformed into *E. coli* Xjb cells and the IPTG induction protocol (2.1.2) was performed for protein expression. The protein was purified starting with nickel chromatography. Then a modified strep-tag chromatography was performed. The protocol was same as described in section 2.2.5 except that the Resource Q column was assembled in tandem with the strep-tactin column to eliminate all nucleic acid contaminations. The purified protein was concentrated and dialysed as described in section 2.2.6.

2.2.9 *S. africana* CarD, *E. coli* RNAP, *E. coli* σ-factor and *S. africana* σ-factor expressions and purifications

S. africana CarD, *E. coli* RNAP, *E. coli* σ-factor and *S. africana* σ-factor expressions and purifications were performed previously by the members of RNAP laboratory. In general, the described protein

expression protocols and purification methods were used. Let it be mentioned that the expression of σ-factors was usually performed at room temperature o/n. I want to give special thanks to Janne Mäkinen, Sari Paavilainen and Vilma Trapp who have been expressing and purifying these proteins.

2.3 Construction of the transcription templates

The backbone of the construct including the Broccoli sequence (Huang *et al.* 2022) was made by digesting pOP004 plasmid with FastDigest restriction enzymes (Thermo Scientific). The reaction was incubated in FastDigest Green buffer for one hour at 37 °C and the temperature was set to 70 °C for ten minutes to inactivate the enzymes at the end of the reaction. The restriction enzymes and fragments used for cloning are presented in Table 1.

The digested plasmid backbones were separated on an 1 % agarose gel for isolating a pure vector for the cloning. The vector band was cut from the gel and purified using a gel extraction kit (NucleoSpin® Gel and PCR Clean up kit, Macherey-Nagel) according to manufacturer's instructions. The amount of extracted backbone was confirmed with spectrophotometer (Nanodrop, Thermo Scientific).

The cloning was performed with Gibson Assembly technique using the New England Biolabs Gibson Assembly Master Mix. The reaction was performed incubating the reaction mixture at 50 °C for 20 minutes. The reaction mixtures were stored at -20 °C until the transformation was made.

The cloned plasmids were transformed to *E. coli* XL-1 cells using $CaCl_2$ mediated DNA uptake method, and the cells were plated on LB plates with ampicillin for selection. The colonies were inoculated to a 5 ml LB medium with ampicillin and grown overnight. Next day the constructs were isolated using a MiniPrep kit (NucleoSpin® Plasmid, Macherey-Nagel) according to the manufacturer's instructions. The concentrations of the isolated constructs were measured spectroscopically (NanoDrop, Thermo Scientific) and stored at -20 °C until amplifying. The sequences of the constructs were confirmed with whole plasmid sequencing (Whole Plasmid Sequencing, Eurofins Genomics) and primer sequencing (Mix2Seq Kit, Eurofins Genomics).

The amplifying of the constructs was done by first transforming the plasmids to *E. coli* XL-1 cells using $CaCl_2$ mediated DNA uptake method and plating them on ampicillin plates. Then a 5 ml LB pre-culture with antibiotics was established and finally the preculture was inoculated to two 500 ml LB main cultures with antibiotics. After an overnight incubation at 37 °C the cells were harvested by centrifuging (7000 x g, 10 min, 4 °C). The plasmids were isolated with a Maxiprep kit (NucleoBond® Xtra Maxi Plus, Macherey-Nagel) according to the manufacturer's instructions. The concentration of

amplified plasmids was measured spectroscopically (NanoDrop, Thermo Scientific) and they were stored at -20 °C.

Before the assays the constructs were linearised by digesting with XhoI restriction enzyme (Thermo Scientific) in a Red buffer (Thermo Scientific) overnight at 37 °C aiming to a 250 nM stock.

Table 1. The backbones, restriction enzymes, and cloned fragments used in cloning. In addition, the names given to the cloned transcription templates are presented in the last column.

Back bone	**Restriction enzymes**	**Cloned fragment**	**Name of the transcription template**
OP4	XbaI and BamHI	*eftu*P	VL8
OP4	XbaI and BamHI	*rrf2*P	JK2
OP4	XbaI and BamHI	*rppk*P	JK3
OP4	XbaI and BamHI	*fre*P	VL3
OP4	XbaI and BamHI	*dksA*P	VL6
OP4	XbaI and BamHI	*nusG*P	JK6
OP4	XbaI and BamHI	*if-1*P	VL2
OP4	XbaI and BamHI	*rrnb*P	JK4

Plasmids used in the protein expression (pVN005/*S. africana* RNAP and pGB225/reverse transcriptase) and two additional templates used in the transcription activity assays (VN1/*greA*P and VN3/*rrna*P) were cloned previously by the members of the Georgy Belogurov lab.

2.4 *In vitro* transcription activity assay

The templates used in transcription activity assays encoded a fluorogenic RNA aptamer called Broccoli downstream the studied promoters. In the transcription activity assays the transcription output was monitored for 10 minutes by following the fluorescence levels of the transcribed RNA aptamer bound to a fluorogen (Figure 8). This method is described more specifically by Huang *et al.* (2022).

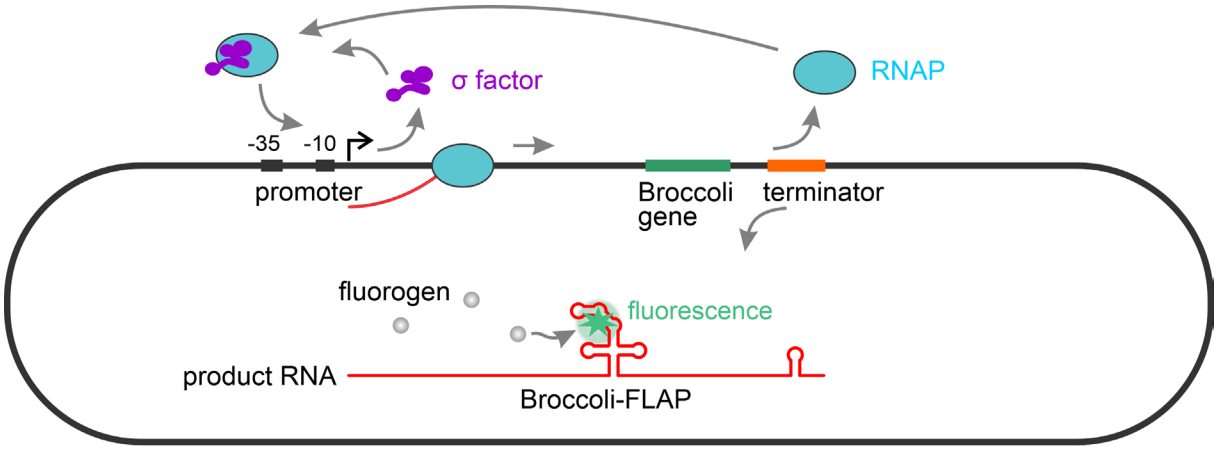

Figure 8. The principle of the transcription activity assay experimental setup (Huang et al. 2022). Transcription templates used in the transcription activity assays encoded fluorogenic RNA aptamer called Broccoli downstream the studied promoters. The transcription output was monitored by following the fluorescence levels of the transcribed RNA aptamer bound to a fluorogen during a 10-minute time frame.

For this assay, three subsets of mixtures were prepared to form the final reaction mixture. In the first one (H mix) RNAP holoenzyme was assembled by mixing *S. africana* or *E. coli* RNAP (10 µM) and their σ-factor (40 µM) in storage buffer. This mixture was incubated at 30 °C for 20 minutes and then stored at -20 °C. The second mixture (F mix) contained the studied template (125 nM), fluorogen DFHBI-1T (25 µM) and pyrophosphatase (0.25 µM) in TB10 buffer (50 mM HEPES-KOH pH 7.5/9; 80 mM KCl; 0.2 mM DTT; 0.1 mM EDTA; 10 mM $MgCl_2$). The third mixture (N mix) was made by mixing ribonucleotides ATP, GTP, UTP and CTP (2 mM each, Jena Bioscience) in TB10 buffer. The H mix was mixed with F mix resulting to a mixture which contained 2 µM *E. coli*/*S. africana* RNAP; 8 µM σ-factor; 100 nM transcription template; 20 µM DFHBI-1T and 0.20 µM pyrophosphate. Both H+F mix and N mix were prewarmed on a heat block at 37 °C. Finally, the reaction was started by mixing equal volumes of H+F mix and N mix and transferring 55 µl of the final reaction mixture to a fluorometer (LS 55 Fluorescence Spectrometer, PerkinElmer) in a cuvette (Catalog number: 105-251-15-40, Hellma Analytics) which was also prewarmed at 37 °C on water bath. The active reaction mix contained 1 µM *E. coli*/*S. africana* RNAP; 4 µM of σ-factor; 0.1 µM pyrophosphatase; 10 µM DFHBI-1T; 50 nM linearised studied template; 1mM of each NTP; 50 mM HEPES-KOH pH 7.5/9; 80 mM KCl; 10 mM $MgCl_2$; 0.1 mM EDTA and 0.2 mM DTT. The pH of HEPES-KOH was chosen depending on which conditions were used in the experiment. In measurements where CarD was present, it was added to the H mix so that the final concentration of CarD was 5 µM in the final reaction mixture. The mixtures where the fluorogen was present were covered to prevent exposure to light.

The transcription output was monitored by using 472 nm excitation wavelength and 507 nm emission wavelength in the fluorometer. The activity of transcription was measured every second between 20 and 620 seconds from the beginning of the reaction. The temperature was kept at 37 °C during the

whole measurement with water bath. In the assays the transcription activities were measured in duplicates or more where possible. The bar diagrams in the results section were constructed by subtracting the background fluorescence level measured at 25 seconds from the fluorescence level at 600 seconds (counting from the start of the reaction).

2.5 Primer extension technique

First, RNA was transcribed in the same way as described in the transcription activity assay. RNA was isolated from the reaction mixture with Qiagen RNeasy Mini Kit according to manufacturer's instructions and the concentration of RNA was measured with a spectrophotometer (NanoDrop, Thermo Fischer Scientific). Then, a fluorescent DNA primer (1 µM) (Sigma-Aldrich) was annealed to the isolated RNA (1.5 µM) in 10 mM HEPES-KOH pH 7.5 buffer with 100 µM EDTA. The annealing was performed by incubating the mixture 3 minutes at 95 °C on a heat block and then setting the heat block to cool down to 20 °C. Reverse transcriptase (SuperScrpit™, Thermo Scientific/self-made) was used to extend the fluorescent DNA primer until it reached the 5'-end of RNA. The reaction mixture contained 0.2 µM primer-RNA scaffold; 0.1 mM dNTPs each (Bioline, Meridian Bioscinece); 400 U/0.2 µM reverse transcriptase; 50 mM HEPES-KOH pH 8; 80 mM KCl; 3 mM $MgCl_2$; 0.1 mM EDTA and 10 mM DTT. The reaction mixture was incubated at room temperature for 5 minutes with self-made reverse transcriptase and at 40 °C with SuperScrpit™-reverse transcriptase. The reaction was stopped after 5 minutes with formamide stop buffer (94.6 % (v/v) formamide; 13 mM Li-EDTA and Orange G dye 0.21 % (m/v)).

The ladder was made by using a DNA oligonucleotide corresponding to the 5' region (20-30 nt) of the expected RNA transcript (Eurofins Genomics) and a fluorescent DNA primer (Sigma Aldrich). DNA polymerase (Klenow, Thermo Scientific) was used to extend the DNA primer by adding nucleotides one by one for each reaction. So, the shortest band of the ladder was made by adding only for example cytosine and the next band by adding cytosine and thymine etc.

When making the ladder the template DNA oligonucleotide (14 µM) and the fluorescent primer (10 µM) were annealed in 10 mM HEPES-KOH pH 7.5 buffer with 100 µM EDTA. This mixture was incubated for 3 minutes at 95 °C on a pre-warmed heat block and the annealing was performed by setting the heat block to cool down to 20 °C. When extending the DNA primer, the reaction mixture contained 1 µM primer-DNA scaffold; 20 µM dNTPs each (Bioline, Meridian Bioscience); 20 U Klenow DNA polymerase; 1x Klenow DNA polymerase buffer (Thermo Scientific). The reaction was incubated at room temperature for 5 minutes and stopped with formamide stop buffer.

The extended DNA primers were then run on a 16 % denaturing PAGE gel with the ladder to determine their length. The length of the extended DNA primer marked the transcription start site.

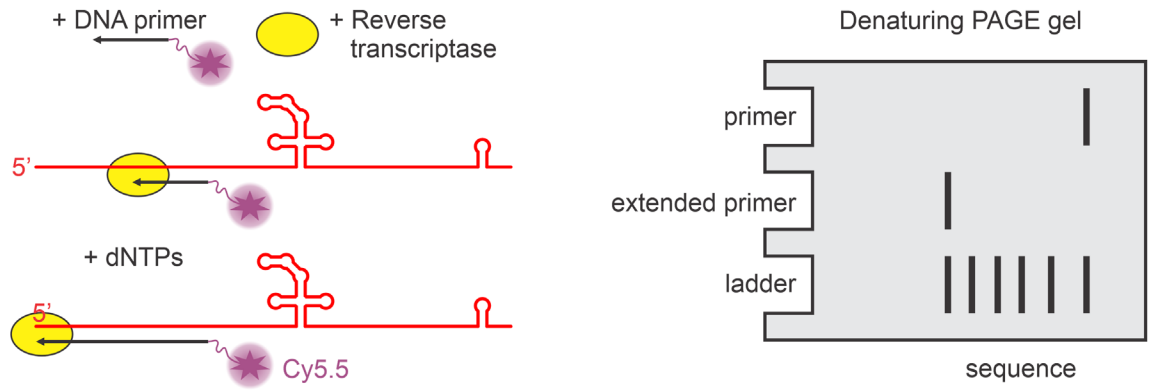

Figure 9. The principle of the experimental setup in the primer extension assay. In these assays, the isolated RNA (red) was annealed with the fluorescent DNA primer (black) and the reverse transcriptase (yellow) was used to extend the DNA primer. The extended DNA primers were run on a denaturing PAGE gel with a self-made ladder to define the length of the extended DNA primer. The length of the extended DNA primer marked the start site of transcription. The self-made ladder was made by using a template DNA oligonucleotide corresponding to the expected 5' region (20-30 nt) of the nascent RNA and a fluorescent DNA primer. DNA polymerase was used to extend the DNA primer by adding nucleotides one by one for each reaction to make the extended DNA primers of certain lengths.

2.6 Use of AI-Assisted Language Tools

ChatGPT was used to assist with language editing in this Master's thesis.

3 Results

3.1 Transcription activity assays

We constructed VL (made by Ville Levola) and JK (made by Juho Kotikoski) series of transcription templates that were used in the transcription activity assays. These templates include various promoters (Table 2). VL8 contains the promoter for the *S. africana* gene encoding EF-Tu and ribosomal protein genes. EF-Tu is a translation elongation factor, which functions as a GTPase. JK2 contains the *S. africana rrf2* promoter for gene encoding *S. africana* translation termination factor. JK3 carries the promoter for *S. africana* ribose-phosphate pyrophosphokinase gene. Ribose-phosphate pyrophosphokinase is involved in several pathways in the cell for example pentose phosphate pathway, purine metabolism, metabolic pathways, biosynthesis of secondary metabolites and carbon metabolism. The template VL3 includes the *fre* promoter for *S. africana* fructose operon transcriptional repressor gene, whereas VL6 contains the promoter for *S. africana dksA*. DksA, an RNAP binding transcription factor, is involved, for example, in the stringent response in several bacteria. The promoter present in the template JK6 is involved in transcription initiation for *S. africana nusG* and ribosomal protein genes. NusG is a transcription factor that plays a role, for example, in the regulation of transcription elongation. The transcription template VL2 carries the promoter for *S. africana* translation initiation factor *if-1* gene whereas VN1 (made by Verneri Nissilä) contains the promoter for *S. africana greA.* GreA is known as a transcription elongation factor. The transcription template VN3 (made by Verneri Nissilä) includes the promoter for *S. africana* ribosomal RNA genes and finally serving as a control the template JK4 carries the promoter for *E. coli* ribosomal RNA genes. The information regarding the encoded proteins is based on the KEGG (Kyoto Encyclopedia of Genes and Genomes) database.

Table 1. The abbreviations and definitions of the promoters encoded in the transcription templates that were used in the transcription activity assays.

Template name	Promoter abbreviation	Protein synthetized from the promoter
VL8	*eftu*P	*S. africana* EF-Tu, ribosomal proteins
JK2	*rrf2*P	*S. africana* RF2 (translation termination factor)
JK3	*rppk*P	*S. africana* ribose-phosphate pyrophosphokinase
VL3	*fre*P	*S. africana* fructose repressor
VL6	*dksA*P	*S. africana* DksA
JK6	*nusG*P	*S. africana* NusG, ribosomal proteins
VL2	*if-1*P	*S. africana* IF-1 (translation initiation factor)
VN1	*greA*P	*S. africana* GreA
VN3	*rrna*P	*S. africana* rRNA
JK4	*rrnb*P	*E. coli* rRNA

3.1.1 Analysis of the promoter sequences in the studied templates

We tested the transcription activities of *E. coli* and *S. africana* RNAPs with nine selected templates (VL8/*eftu*P, JK2/*rrf2*P, JK3/*rppk*P, VL3/*fre*P, VL6/*dksA*P, JK6/*nusG*P, VL2/*if-1*P, VN1/*greA*P and VN3/*rrna*P) encoding nine different *S. africana* promoters and a template encoding a control *E. coli* promoter (JK4/*rrnb*P). The sequences of the promoters are presented in Figure 10. The consensus sequence for -10 element is TATAAT and for -35 element TTGACA (Feklistov and Darst 2011). The optimal length of the spacer between -35 and -10 elements is shown to be 17±1 base pairs including the extended -10 element (CG), if present (Feklistov and Darst 2011; Harley and Reynolds 1987). In addition, the optimal spacing between the -10 element and the starting base of transcription, which is optimally purine (G or A), is six base pairs (Mendoza-Vargas *et al.* 2009).

When analysing the -10 elements of the studied promoters it is seen that only VN1/*greA*P and JK4/*rrnb*P have the optimal sequence. VL8/*eftu*P, JK2/*rrf2*P, JK3/*rppk*P, VL3/*fre*P, JK6/*nusG*P and VN3/*rrna*P have one base difference to the consensus sequence, while VL2/*if-1*P has two and VL6/*dksA*P has three base differences. In -35 element analysis VL8/*eftu*P, VL6/*dksA*P, JK6/*nusG*P and VL2/*if-1*P have the identical match with consensus sequence. JK3/*rppk*P, VL3/*fre*P, VN1/*greA*P, VN3/*rrna*P and JK4/*rrnb*P have one base difference and JK2/*rrf2*P has two different bases, when comparing to the consensus sequence of the -35 element. The optimal spacer length (17±1 bp) is found in each studied promoter while it is 17 base pairs in VL8/*eftu*P, VL3/*fre*P, VL6/*dksA*P, JK6/*nusG*P, VL2/*if-1*P and VN3/*rrna*P, 16 base pairs in JK2/*rrf2*P, VN1/*greA*P and JK4/*rrnb*P, and 18 base pairs in JK3/*rppk*P. JK6/*nusG*P and VN1/*greA*P have the optimal extended -10 element motif TG, while

almost identical motif CG is found in JK2/*rrf2*P, JK3/*rppk*P and VL6/*dksA*P. Templates JK2/*rrf2*P, JK3/*rppk*P, VL6/*dksA*P, JK6/*nusG*P, VL2/*if-1*P and VN1/*greA*P have the optimal spacing between -10 element and the start site (6 bp), while it is 5 base pairs in VL8/*eftu*P and VL3/*fre*P, and 7 base pairs in VN3/*rrna*P. The predicted start site has 8 base pairs in spacing in JK4/*rrnb*P, but there is also another adenosine which has spacing of 5 base pairs. The predicted starting base of transcription marked on green background in Figure 10 is adenosine in every promoter except in VN1/*greA*P, in which it is guanine and VL2/*if-1P*, in which it is thymine. All these similarities and differences in the promoters affect to the transcription activity of a certain promoter as seen below.

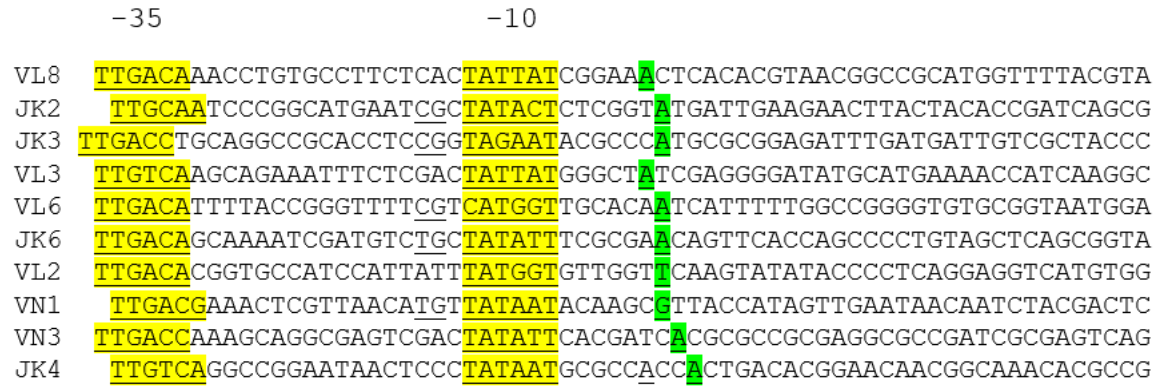

Figure 10. The promoter sequences of the studied templates in transcription activity assays. -35 and -10 elements are in yellow, and the predicted start site is on green background. In addition, the motif of extended -10 element is underlined, if present.

The transcription templates encoded fluorogenic RNA-aptamer called broccoli downstream the studied promoters. We monitored the transcription output by following the fluorescence emitted by transcribed fluorogenic RNA aptamer bound to a fluorogen during a 10-minute time frame. We present the transcription activity values here on bars that are constructed by subtracting the background fluorescence level measured at 25 seconds from the fluorescence level at 600 seconds (counting from the start of the reaction). The units of the transcription activities are arbitrary units of fluorescence measured at 507 nm wavelength and those are proportional to the amount of transcribed RNA as described above. We tested the transcription activities of *E. coli* RNAP at pH 7.5, *S. africana* RNAP at pH 9 and also those with transcription factor CarD with at least two replicates. My supervisor Vilma Trapp kindly assayed and analysed the transcription activity of *S. africana* at pH 7.5. The average of the replicates with their standard deviations are presented in the bar diagrams shown below.

3.1.2 Transcription activities of *E. coli* and *S. africana* RNAPs measured at pH 7.5

First, we measured the transcription activities of both *E. coli* (red) and *S. africana* (blue) at pH 7.5 (Figure 11). We selected the studied promoters including intentionally both weak and medium strength ones. As expected, the transcription activities varied by manyfold on different promoters as seen in Figure 11. Most of the spirochaetal RNAP transcription activities were significantly lower than those

of *E. coli* RNAP. For example, for the template VL8/*eftu*P the spirochaetal transcription activity was 54 % of that of *E. coli*. Same ratios for other templates were JK2/*rrf2*P: 30 %, JK3/*rppk*P: 16 %, JK6/*nusG*P: 8 %, VL2/*if-1*P: 37 %, VN3/*rrna*P: 17 %, JK4/*rrnb*P: 4 %. The spirochaetal transcription activity was higher at few templates (VL3/*fre*P: 139 %, VL6/*dksA*P: 123 % and VN1/*greA*P: 135 %) when compared to that of *E. coli*. The differences between transcription activities of *E. coli* and *S. africana* RNAPs are statistically significant, since the standard deviations don't overlap except in the templates VL3/*fre*P and VL6/*dksA*P, in which the results can be considered marginally significant.

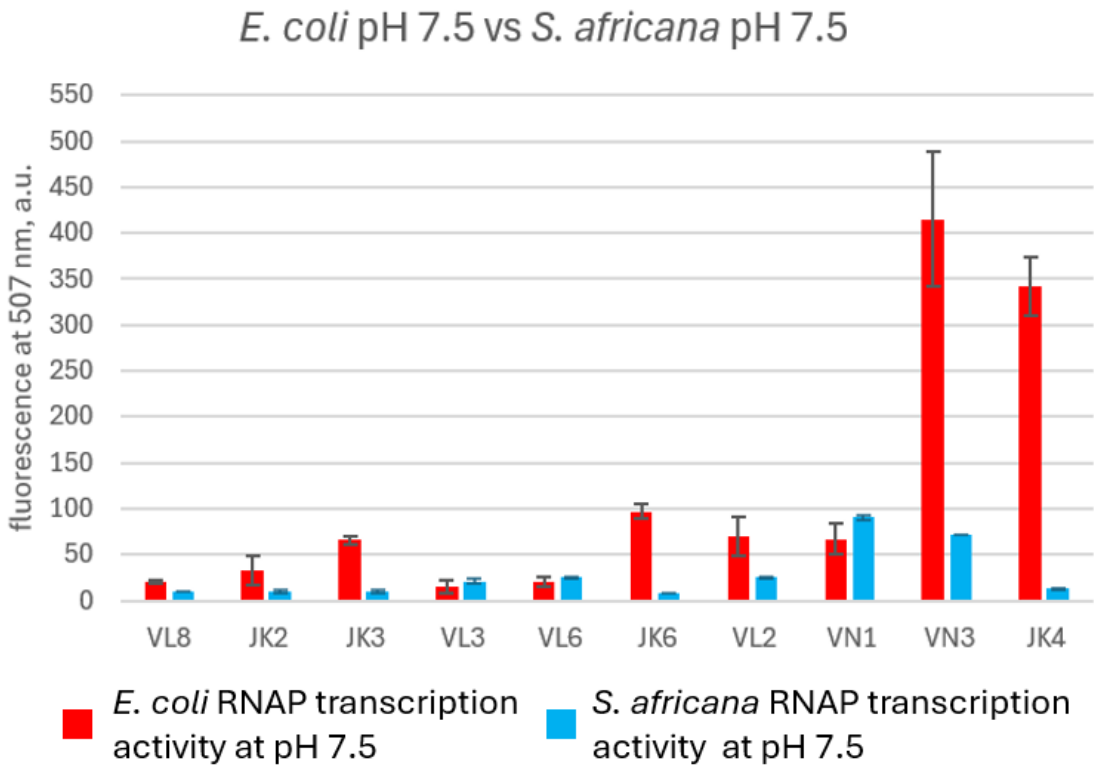

Figure 11. Transcription activities of E. coli (red) and S. africana (blue) RNAPs for studied templates measured at pH 7.5. The error bars represent the standard deviation. Intentionally, we included both weak and medium strength promoters in the studies and as expected the transcription activities vary between different promoters. Overall, it is seen that the spirochaetal transcription activity is significantly lower than that of E. coli for several promoters. The values are presented in arbitrary units of fluorescence measured at 507 nm. Connections between the templates and the promoters are seen in Table 2.

3.1.3 Transcription activity of *S. africana* RNAP measured at pH 9

Given that *S. africana* is an alkalophilic bacteria and possibly doesn't control its intracellular pH we also measured the transcription activities of spirochaetal RNAP at pH 9, which is close to the optimal growth pH of *S. africana*. Figure 12 shows that the transcription activity of *S. africana* RNAP increased at each promoter when the pH was raised from 7.5 (magenta) to 9 (cyan). For example, the increase was 134 % for the template VL8/*eftu*P. The increases to transcription activities for other templates were JK2/*rrf2*P: 145 %, JK3/*rppk*P: 191 %, VL3/*fre*P: 80 %, VL6/*dksA*P: 50 %, JK6/*nusG*P: 661 %, VL2/*if-1*P: 279 %, VN1/*greA*P: 144 %, VN3/*rrna*P: 41 % and JK4/*rrnb*P: 24 %. The

differences are statistically significant in all except in the case of the template JK4, in which the standard deviations overlap making the result marginally significant.

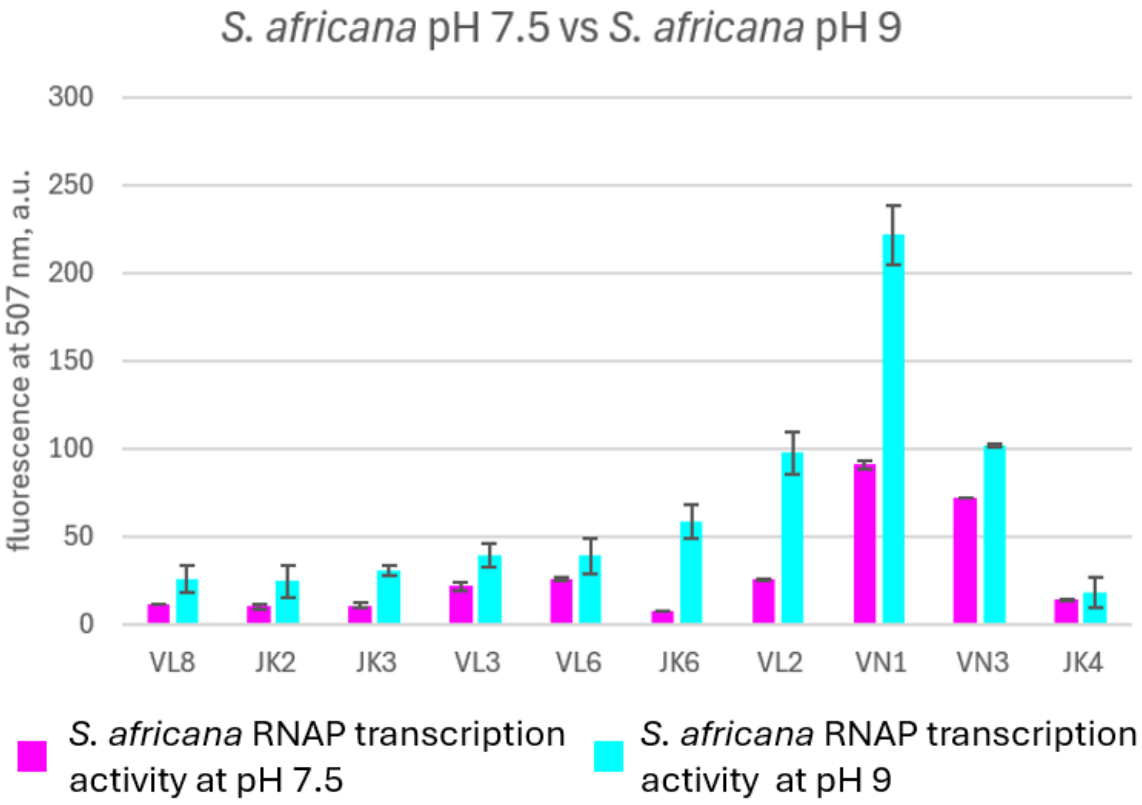

Figure 12. The transcription activities of S. africana RNAP measured at pH 7.5 (magenta) and at pH 9 (cyan) for studied templates. The error bars represent the standard deviation. The transcription activities increased for each template when the pH was raised from 7.5 to 9. The values are presented in arbitrary units of fluorescence measured at 507 nm. Connections between the templates and the promoters are seen in Table 2.

3.1.4 Transcription activities of *E. coli* RNAP at pH 7.5 and *S. africana* RNAP at pH 9

The transcription activity of *S. africana* RNAP at pH 9 (blue) and that of *E. coli* RNAP at pH 7.5 (red) are compared in Figure 13. The figure shows that the spirochaetal transcription activity profile matches better to that of *E. coli* when measured at pH 9 instead of pH 7.5. We compared the ratio of transcription activity between *S. africana* and *E. coli* at pH 7.5 to the ratio obtained when *S. africana* was tested at pH 9 and *E. coli* at pH 7.5. The ratio for the template VL8/*eftu*P increased from 54 % to 127 %. The ratios increased for other templates as follows: JK2/*rrf2*P: 30 % → 74 %, JK3/*rppk*P: 16 % → 46 %, VL3/*fre*P: 139 % → 251 %, VL6/*dksA*P: 123 % →185 %, JK6/*nusG*P: 8 % →60 %, VL2/*if-1*P: 37 % → 140 %, VN1/*greA*P: 135 % → 329 %, VN3/*rrna*P: 17 % → 24 % and JK4/*rrnb*P: 4 % → 5 %.

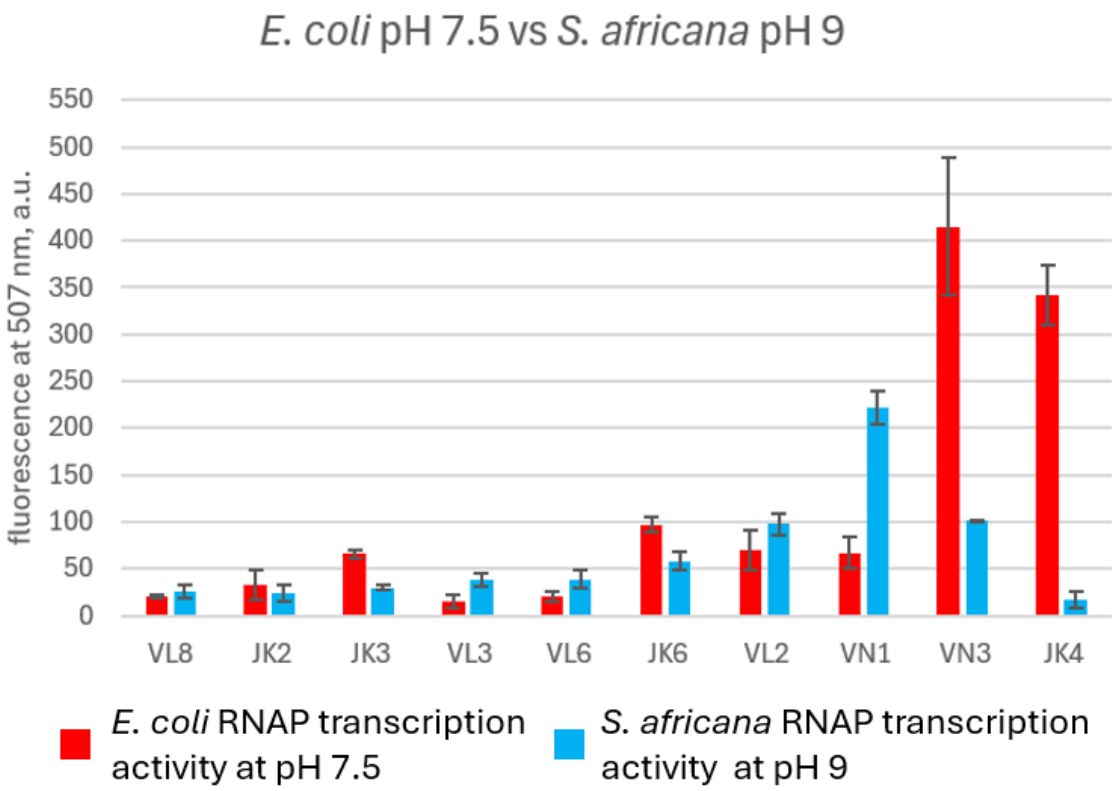

Figure 13. The transcription activity of S. africana RNAP at pH 9 (blue) compared to that of E. coli measured at pH 7.5 (red) for studied templates. The error bars represent the standard deviation. It is seen that spirochaetal transcription activity profile matches better to that of E. coli when measured at pH 9. The values are presented in arbitrary units of fluorescence measured at 507 nm. Connections between the templates and the promoters are seen in Table 2.

3.1.5 Effects of CarD on the transcription activity of *S. africana* RNAP at pH 9

Prior studies in RNAP group indicate that CarD has an effect on spirochaetal RNAP and its transcription activity. We also tested the effect of CarD in this study with selected templates (JK6/*nusG*P, VL2/*if-1*P, VN1/*greA*P and VN3/*rrna*P) by adding it to the final concentration of 5 µM in the reaction mixture. Figure 14 shows that the effect of CarD varied on different templates. For example, CarD increased the transcription activity of VN3/*rrna*P and an opposite effect was found on the template VN1/*greA*P. CarD enhanced the transcription activity by 90 % for JK6/*nusG*P, 23 % for VL2/*if-1*P and 160 % for VN3/*rrna*P when added to the experimental setup. On the contrary, transcription activity was decreased by 40 % in the case of VN1/*greA*P by its effect. We reproduced the results that were observed earlier in RNAP group for the templates VN1/*greA*P and VN3/*rrna*P carrying spirochaetal promoters. In addition, we discovered two new promoters that are regulated by the transcription factor CarD. CarD has a strong effect for the template JK6/*nusG*P containing spirochaetal *nusG* promoter and a weaker effect for the template VL2/*if-1*P carrying the spirochaetal translation initiation factor 1 promoter. The results are statistically significant, since the standard deviations of transcription activities do not overlap in any cases.

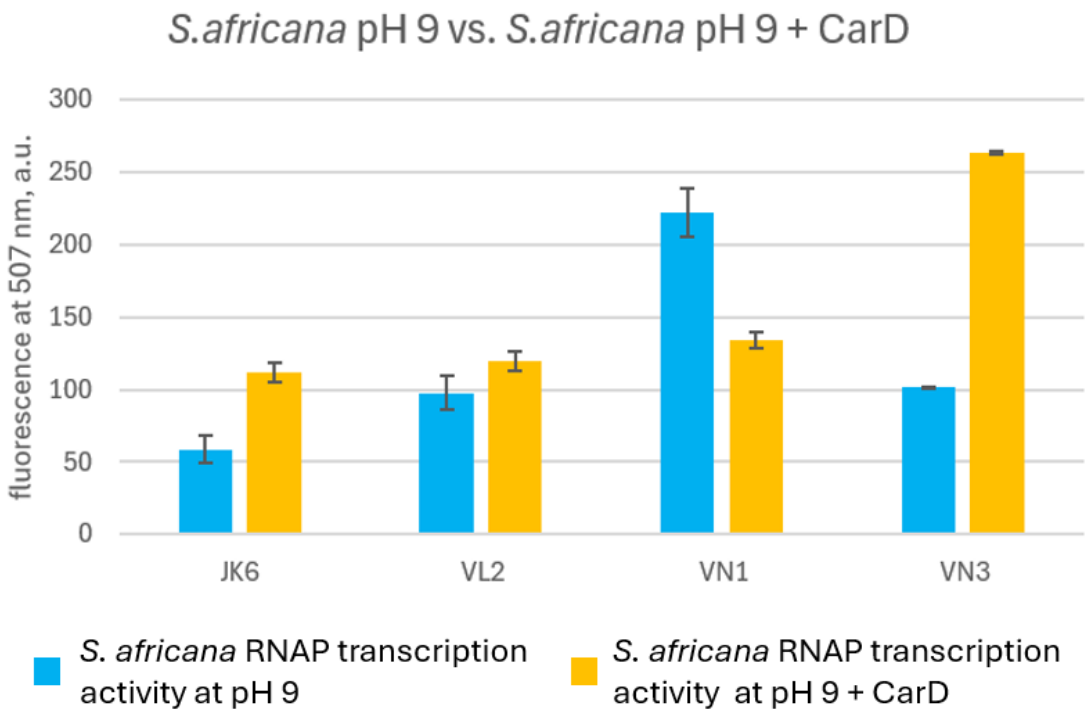

Figure 14. Transcription activities of S. africana RNAP at pH 9 measured with (yellow) and without (blue) the transcription factor CarD. The error bars represent the standard deviation. As seen, the effect of CarD varies on different templates: it activates the transcription for JK6, VL2 and VN3 and deactivates it for VN1. The values are presented in arbitrary units of fluorescence measured at 507 nm. Connections between the templates and the promoters are seen in Table 2.

3.1.6 Transcription activities of *E. coli* RNAP at pH 7.5 and *S. africana* RNAP at pH 9 with CarD

We compared the transcription activity of *S. africana* measured at pH 9 (blue) including those measurements, in which CarD was present in the reaction (yellow) to that of *E. coli* measured at pH 7.5 (red) (Figure 15). We compared the ratio of transcription activity between *S. africana* at pH 9 and *E. coli* at pH 7.5 to the ratio obtained when *S. africana* was tested at pH 9 in the presence of CarD, and *E. coli* at pH 7.5. Upon the addition of CarD the spirochaetal transcription activity profile settled even closer to that of *E. coli* for templates JK6/*nusG*P (60 % → 114 %), VN1/*greA*P (329 % → 199%) and VN3/*rrna*P (24 % → 64 %). For the template VL2/*if-1*P the difference increased from 140 % → 171 % when using the ratios described above. Overall, it can be said that spirochaetal transcription activity profile matches better to that of *E. coli* when measured at pH 9 and in the presence of transcription factor CarD.

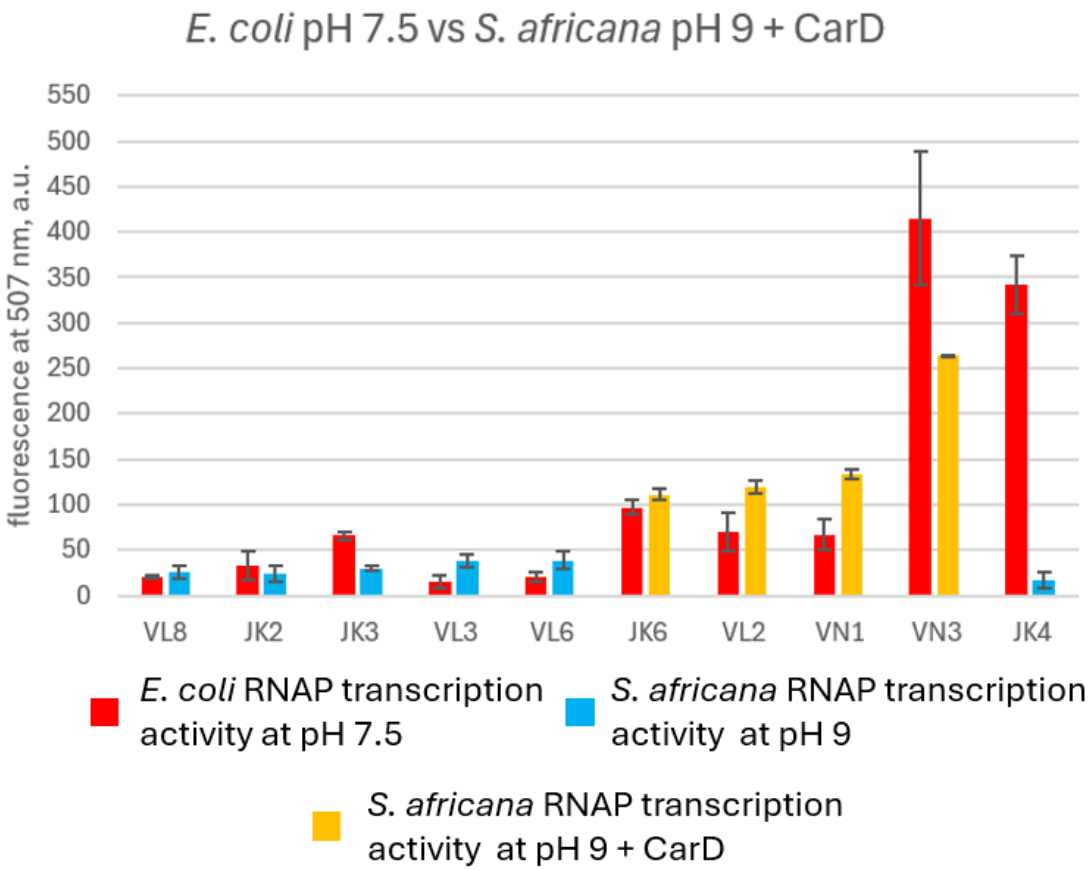

Figure 15. The transcription activities of E. coli at pH 7.5 (red) and S. africana at pH 9 (blue) including the transcription activities in the presence of transcription factor CarD (yellow) compared for the studied templates. The error bars represent the standard deviation. It can be said that the transcription activity profile of S. africana RNAP matches better with that of E. coli RNAP when measured at pH 9 and in the presence of CarD. The values are presented in arbitrary units of fluorescence measured at 507 nm. Connections between the templates and the promoters are seen in Table 2.

3.2 Confirmation of the start sites of the promoters used in transcription assays

We mapped the start sites of the promoters used in transcription assays to identify or confirm the real transcription starting point. For this assay, RNA was first transcribed from the studied promoters with the same method that we used in the transcription activity assays. We isolated the RNA from the reaction mixture after the reaction had proceeded for 10 minutes and annealed it to a fluorescent DNA primer. We used reverse transcriptase to extend the fluorescent DNA primer towards the RNA 5' end. The extended DNA primers were run on a denaturing PAGE gel together with a self-made ladder. We identified the transcription start site by calculating the length of the extended DNA primer base by base using the self-made ladder.

3.2.1 Start sites were determined using both *E. coli* and *S. africana* RNAPs

Figure 16 shows the sequences of the studied promoters. We investigated the start sites of IP4/synthetic derivative of VN1, VN1/*greA*P and VN3/*rrna*P using *E. coli* RNAP and those of JK6/*nusG*P and VL2/*if-1*P using *S. africana* RNAP. The -35 and -10 elements of the promoters are shown in yellow, and the predicted transcription start site according to sequence analysis on green background in Figure 16.

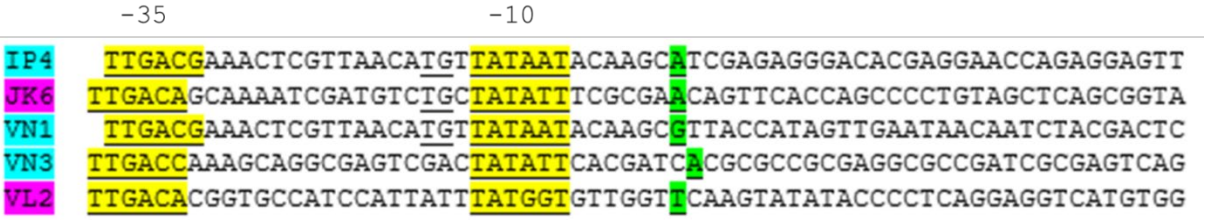

Figure 16. The sequences of the promoters of which start sites were studied. The promoters studied with E. coli RNAP are marked on cyan background and those studied with S. africana RNAP are marked on magenta background. The -35 and -10 regions of the promoters are shown in yellow, and the predicted transcription start site is marked on green background.

3.2.2 Start sites were defined using self-made ladders

We show the mapping of the start site of the template IP4 containing the synthetic derivative of *S. africana greA* promoter (the initially transcribed sequence is different from the native *greA* promoter) in Figure 17. The schematic of the experimental setup is shown above the gel image: the promoter sequence appears in black on the top row, and the transcribed RNA sequence is shown below it in red. Finally, the sequence of the extended DNA primer that was synthesized by reverse transcriptase is marked below the transcribed RNA sequence in black. The predicted transcription start site according to the sequence analysis is marked on green background. The top bands on the gel are the extended DNA primer and the DNA primer that did not get extended in the reaction. The four bands below are the self-made ladder, which we used to map the real transcription start site of the promoter. Additionally, we explain the principle of making the ladder below the gel pic (the shortest band of the ladder was made by adding only cytosine (+C) in the reaction and the second shortest by adding only cytosine and thymine (+C, +T) etc.) The length of the extended DNA primer marks the actual start site of the transcription and as Figure 17 shows, it is the same as the predicted one. In this case we confirmed the start site.

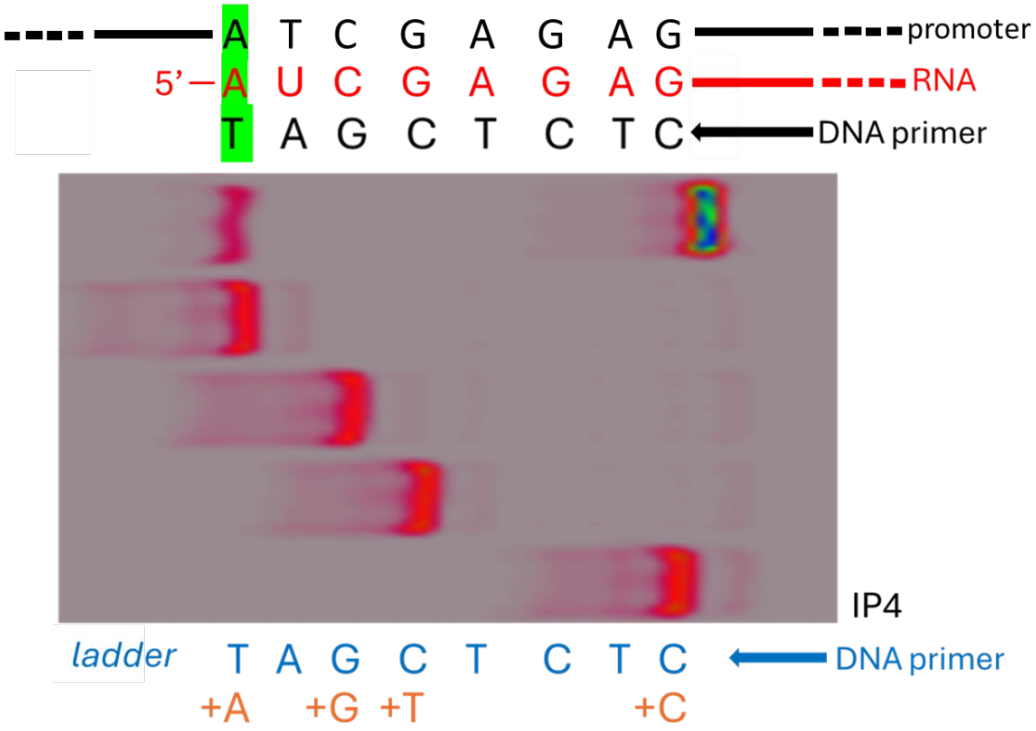

Figure 17. The gel image that was used to determine the actual start site of transcription for template IP4. The sequences of the promoter, transcribed RNA and extended DNA primer are shown above the gel image. The predicted transcription start site according to sequence analysis is marked on green background. The top bands on the gel are the extended DNA primer and the DNA primer that didn't get extended in the reaction. The four bands below are the self-made ladder, which we used to map the transcription start site of the promoter. The principle of making the ladder is explained below the gel pic (the shortest band of the ladder was made by adding only cytosine (+C) in the reaction and the second shortest by adding cytosine and thymine (+C, +T) etc.) We confirmed the start site for this template, since the length of the extended DNA matches with the predicted start site.

We demonstrate the mapping of the start sites of the promoters for templates JK6 carrying the *S. africana nusG* promoter, VN1 containing the *S. africana greA* promoter and VN3 carrying the *S. africana* ribosomal RNA promoter in Figure 18. The principle is the same as for template IP4: the sequences of the promoter, transcribed RNA and extended DNA primer appear above the gel image. This time, we present only the extended DNA primer bands without the ladder in the gel images for simplicity. However, in all cases, ladders were used to determine the length of each extended DNA primer. Again, the result was the same for each template (JK6/*nusG*P, VN1/*greA*P, and VN3/*rrna*P) as for template IP4: the band marking the length of the extended DNA primer leads to the real transcription start site, which is the same as the predicted one marked on green background. Thus, we also confirmed the start site for these promoters.

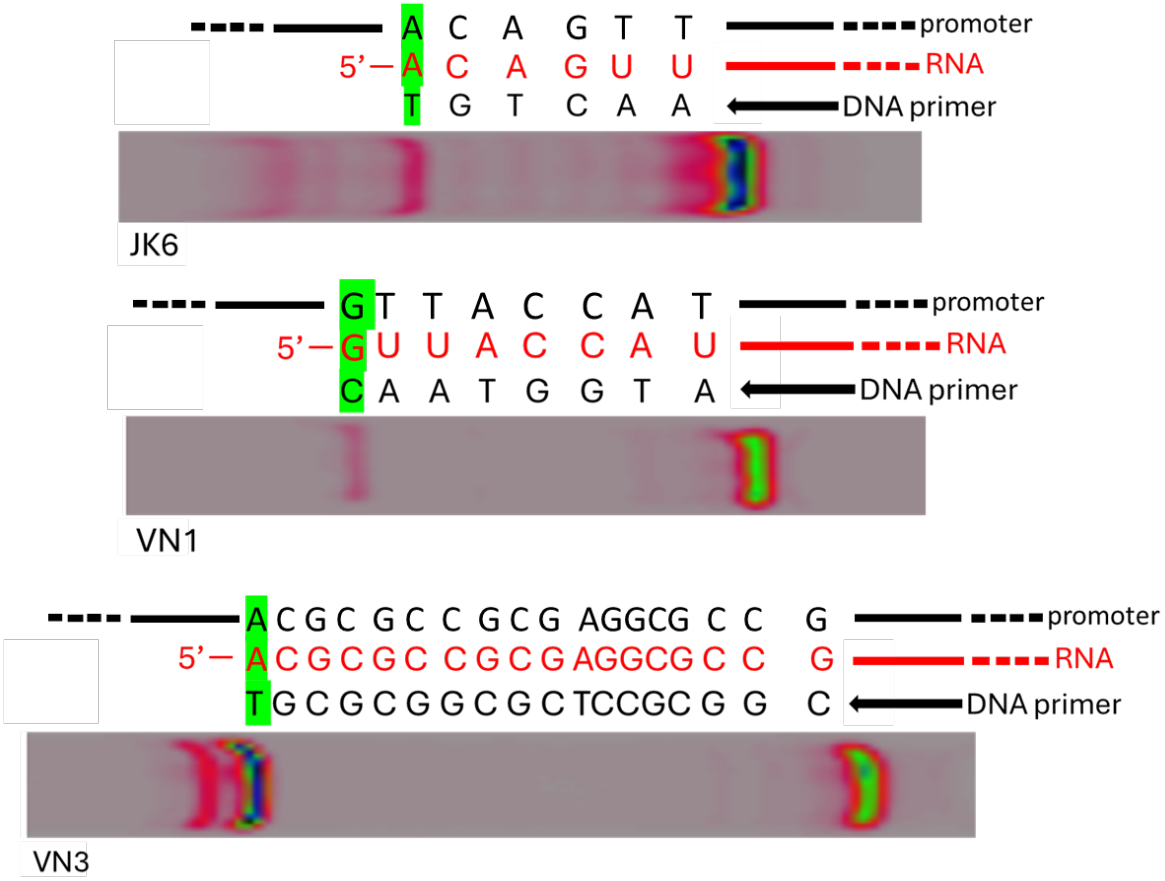

Figure 18. The gel images that were used to map the transcription start site for templates JK6, VN1 and VN3. The sequences of the promoter, transcribed RNA and extended DNA primer appear above the gel images. The predicted transcription start site according to sequence analysis is shown on green background. This time, only the extended DNA primer bands are presented in the gel images for simplicity. However, ladders were used to determine the length of each extended DNA primer. Again, in all cases, the length of the extended DNA primer leads to the real transcription start site, which matches with the predicted one. Thus, we also confirmed the start site for these promoters.

3.2.3 The cryptic start site of the template VL2

When mapping the real transcription start site of the template VL2 containing the *S. africana* translation initiation factor 1 promoter the outcome was not that simple. Again, the sequences of the promoter, transcribed RNA and extended DNA primer are presented above the gel image in Figure 19, and only the extended DNA primers are shown for simplicity. Similarly, we used a ladder to determine the length of the extended DNA primers. The figure shows a band, which length matches with the predicted start site, but a stronger and brighter band is seen on the gel at the black arrow. This band marking the length of another extended DNA primer indicates that there is another, more active start site in this template upstream the predicted one. When analysing the sequence of the VL2/*if-1*P template we indeed found potential -35 and -10 regions that could fit this cryptic start site. This is a good example of the importance of transcription start site confirmation.

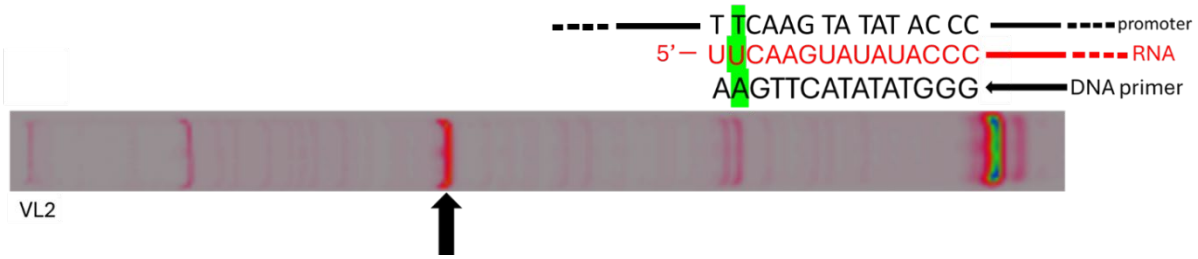

Figure 19. The gel image that was used to determine the real transcription start site for template VL2. The sequences of the promoter, transcribed RNA and extended DNA primer are shown above the gel image. The predicted transcription start site according to sequence analysis appears on green background. Only the extended DNA primer bands are presented in the gel image for simplicity. However, a ladder was used to determine the length of the extended DNA primers. A band, which length matches with the predicted start site is seen on the gel, but a stronger and brighter band is seen on the gel at the black arrow. This band marking the length of another extended DNA primer indicates that there is another, more active start site in this template upstream the predicted one.

4 Discussion

We studied the transcription activity with selected *S. africana* promoter group and a control *E. coli* promoter using both *E. coli* RNAP and *S. africana* RNAP. The transcription activities of *E. coli* RNAP were measured at pH 7.5. Furthermore, the transcription activities of *S. africana* RNAP were measured at pH 7.5 and pH 9. Additionally, the transcription activities of *S. africana* RNAP were measured at pH 9 on selected promoters in the presence of transcription factor CarD.

The results indicate that the transcription activity profile of *S. africana* RNAP show more similarities to that of *E. coli* when measured at pH 9. Could this tell something about the intracellular environment of *S. africana*? *S. africana* was first found in Lake Magadi, Kenya in 1996 (Zhilina *et al.* 1996). The Lake Magadi is an alkali and salty environment, so it would not be a surprise if the environmental conditions would affect also the intracellular pH of *S. africana*. This would suggest that *S. africana* does not control its intracellular pH. Surely this hypothesis needs more investigation. The ongoing research in Georgy Belogurov laboratory group investigates the properties and features of spirochaetal transcription and one of the future aspects could be mapping the conditions of the spirochaetal transcription environment.

The prior unpublished studies in Georgy Belogurov lab suggest that CarD has a role in the regulation of transcription initiation in spirochetes. We proved that CarD has an effect on the transcription activity of four promoters (JK6/*nusG*P, VN1/*greA*P, VN3/*rrna*P and VL2/*if-1*P). However, the effect of CarD varies with different promoters: with JK6/*nusG*P, VL2/*if-1*P and VN3/*rrna*P CarD boosted the transcription activity, while with VN1/*greA*P it restrains the transcription rate. The researchers studying CarD have previously reported similar, different types of regulatory mechanisms of this transcription factor (Henry *et al.* 2021; Srivastava *et al.* 2013).

Considering that CarD regulates spirochaetal ribosomal RNA promoter and brings the spirochaetal transcription activity profile closer to that of *E. coli* on the same set of promoters, we indicate that in *S. africana* CarD acts as a global regulator of transcription initiation. Serving as a transcription factor, CarD is found in some bacteria species, but not in others (García-Moreno *et al.* 2010; Henry *et al.* 2021; Srivastava *et al.* 2013). How has the transcription factor CarD evolved to be an essential transcription regulator in certain bacteria and not in others? Can the distribution of CarD across different bacteria species be explained through taxonomy and evolution of the transcription machinery or are there other factors involved? Future studies in the investigation of CarD will perhaps answer these questions and deepen our understanding with this broadly but sporadically distributed transcription factor.

5 References

Arthur, T. M., Anthony, L. C. & Burgess, R. R. (2000) Mutational analysis of beta '260-309, a sigma 70 binding site located on Escherichia coli core RNA polymerase. *J Biol Chem* **275**:23113–23119.

Bae, B., Davis, E., Brown, D., Campbell, E. A., Wigneshweraraj, S. & Darst, S. A. (2013) Phage T7 Gp2 inhibition of Escherichia coli RNA polymerase involves misappropriation of σ70 domain 1.1. *Proc Natl Acad Sci U S A* **110**:19772–19777.

Belogurov, G. A. & Artsimovitch, I. (2015) Regulation of Transcript Elongation. *Annu Rev Microbiol* **69**:49–69.

Belogurov, G. A., Vassylyeva, M. N., Sevostyanova, A., Appleman, J. R., Xiang, A. X., Lira, R., … Vassylyev, D. G. (2009) Transcription inactivation through local refolding of the RNA polymerase structure. *Nature* **457**:332–335.

Benoff, B., Yang, H., Lawson, C. L., Parkinson, G., Liu, J., Blatter, E., … Ebright, R. H. (2002) Structural basis of transcription activation: The CAP-alpha CTD-DNA complex. *Science* **297**:1562–1566.

Borukhov, S. & Nudler, E. (2003) RNA polymerase holoenzyme: Structure, function and biological implications. *Curr Opin Microbiol* **6**:93–100.

Bowers, C. W. & Dombroski, A. J. (1999) A mutation in region 1.1 of sigma70 affects promoter DNA binding by Escherichia coli RNA polymerase holoenzyme. *EMBO J* **18**:709–716.

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

Browning, D. F. & Busby, S. J. W. (2016) Local and global regulation of transcription initiation in bacteria. *Nat Rev Microbiol* **14**:638–650.

Bu, F., Wang, X., Li, M., Ma, L., Wang, C., Hu, Y., … Liu, B. (2024) Cryo-EM Structure of Porphyromonas gingivalis RNA Polymerase. *J Mol Biol* **436**:168568.

Callaci, S., Heyduk, E. & Heyduk, T. (1999) Core RNA polymerase from E. coli induces a major change in the domain arrangement of the sigma 70 subunit. *Mol Cell* **3**:229–238.

Calvori, C., Frontali, L., Leoni, L. & Tecce, G. (1965) Effect of rifamycin on protein synthesis. *Nature* **207**:417–418.

Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., … Darst, S. A. (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol Cell* **9**:527–539.

Caramori, T. & Galizzi, A. (1998) The UP element of the promoter for the flagellin gene, hag, stimulates transcription from both SigD- and SigA-dependent promoters in Bacillus subtilis. *Mol Gen Genet MGG* **258**:385–388.

Chen, James, Chiu, C., Gopalkrishnan, S., Chen, A. Y., Olinares, P. D. B., Saecker, R. M., … Darst, S. A. (2020) Stepwise Promoter Melting by Bacterial RNA Polymerase. *Mol Cell* **78**:275-288.e6.

Chen, Jie, Darst, S. A. & Thirumalai, D. (2010) Promoter melting triggered by bacterial RNA polymerase occurs in three steps. *Proc Natl Acad Sci U S A* **107**:12523–12528.

Craig, M. L., Tsodikov, O. V., McQuade, K. L., Schlax, P. E., Capp, M. W., Saecker, R. M. & Record, M. T. (1998) DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase-promoter open complex: Evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA. *J Mol Biol* **283**:741–756.

Darst, S. A., Polyakov, A., Richter, C. & Zhang, G. (1998) Structural studies of Escherichia coli RNA polymerase. *Cold Spring Harb Symp Quant Biol* **63**:269–276.

Davis, C. A., Bingman, C. A., Landick, R., Record, M. T. & Saecker, R. M. (2007) Real-time footprinting of DNA in the first kinetically significant intermediate in open complex formation by Escherichia coli RNA polymerase. *Proc Natl Acad Sci U S A* **104**:7833–7838.

deHaseth, P. L. & Helmann, J. D. (1995) Open complex formation by Escherichia coli RNA polymerase: The mechanism of polymerase-induced strand separation of double helical DNA. *Mol Microbiol* **16**:817–824.

Estrem, S. T., Gaal, T., Ross, W. & Gourse, R. L. (1998) Identification of an UP element consensus sequence for bacterial promoters. *Proc Natl Acad Sci U S A* **95**:9761–9766.

Estrem, S. T., Ross, W., Gaal, T., Chen, Z. W., Niu, W., Ebright, R. H. & Gourse, R. L. (1999) Bacterial promoter architecture: Subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes Dev* **13**:2134–2147.

Feklistov, A. & Darst, S. A. (2011) Structural basis for promoter-10 element recognition by the bacterial RNA polymerase σ subunit. *Cell* **147**:1257–1269.

Fenton, M. S., Lee, S. J. & Gralla, J. D. (2000) Escherichia coli promoter opening and -10 recognition: Mutational analysis of sigma70. *EMBO J* **19**:1130–1137.

García-Moreno, D., Abellón-Ruiz, J., García-Heras, F., Murillo, F. J., Padmanabhan, S. & Elías-Arnanz, M. (2010) CdnL, a member of the large CarD-like family of bacterial proteins, is vital for Myxococcus xanthus and differs functionally from the global transcriptional regulator CarD. *Nucleic Acids Res* **38**:4586–4598.

Gautam, A., Hathaway, M. & Ramamoorthy, R. (2009) The Borrelia burgdorferi flaB promoter has an extended -10 element and includes a T-rich -35/-10 spacer sequence that is essential for optimal activity. *FEMS Microbiol Lett* **293**:278–284.

Ghosh, P., Ishihama, A. & Chatterji, D. (2001) Escherichia coli RNA polymerase subunit omega and its N-terminal domain bind full-length beta' to facilitate incorporation into the alpha2beta subassembly. *Eur J Biochem* **268**:4621–4627.

Gries, T. J., Kontur, W. S., Capp, M. W., Saecker, R. M. & Record, M. T. (2010) One-step DNA melting in the RNA polymerase cleft opens the initiation bubble to form an unstable open complex. *Proc Natl Acad Sci U S A* **107**:10418–10423.

Gruber, T. M., Markov, D., Sharp, M. M., Young, B. A., Lu, C. Z., Zhong, H. J., … Gross, C. A. (2001) Binding of the initiation factor sigma(70) to core RNA polymerase is a multistep process. *Mol Cell* **8**:21–31.

Harley, C. B. & Reynolds, R. P. (1987) Analysis of E. coli promoter sequences. *Nucleic Acids Res* **15**:2343–2361.

Haugen, S. P., Ross, W., Manrique, M. & Gourse, R. L. (2008) Fine structure of the promoter-sigma region 1.2 interaction. *Proc Natl Acad Sci U S A* **105**:3292–3297.

Helmann, J. D. & deHaseth, P. L. (1999) Protein-nucleic acid interactions during open complex formation investigated by systematic alteration of the protein and DNA binding partners. *Biochemistry* **38**:5959–5967.

Henry, K. K., Ross, W. & Gourse, R. L. (2021) Rhodobacter sphaeroides CarD Negatively Regulates Its Own Promoter. *J Bacteriol* **203**:e0021021.

Heyduk, E., Kuznedelov, K., Severinov, K. & Heyduk, T. (2006) A consensus adenine at position -11 of the nontemplate strand of bacterial promoter is important for nucleation of promoter melting. *J Biol Chem* **281**:12362–12369.

Holbrook, J. A., Capp, M. W., Saecker, R. M. & Record, M. T. (1999) Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: Interpretation in terms of coupled processes of formation and association of single-stranded helices. *Biochemistry* **38**:8409–8422.

Hook-Barnard, I., Johnson, X. B. & Hinton, D. M. (2006) Escherichia coli RNA polymerase recognition of a sigma70-dependent promoter requiring a -35 DNA element and an extended -10 TGn motif. *J Bacteriol* **188**:8352–8359.

Huang, Y.-H., Trapp, V., Puro, O., Mäkinen, J. J., Metsä-Ketelä, M., Wahl, M. C. & Belogurov, G. A. (2022) Fluorogenic RNA aptamers to probe transcription initiation and co-transcriptional RNA folding by multi-subunit RNA polymerases. *Methods Enzymol* **675**:207–233.

Hubin, E. A., Fay, A., Xu, C., Bean, J. M., Saecker, R. M., Glickman, M. S., … Campbell, E. A. (2017) Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA. *eLife* **6**:e22520.

Husnain, S. I. & Thomas, M. S. (2008) The UP element is necessary but not sufficient for growth rate-dependent control of the Escherichia coli guaB promoter. *J Bacteriol* **190**:2450–2457.

Klein, C. A., Teufel, M., Weile, C. J. & Sobetzko, P. (2021) The bacterial promoter spacer modulates promoter strength and timing by length, TG-motifs and DNA supercoiling sensitivity. *Sci Rep* **11**:24399.

Kojima, I., Kasuga, K., Kobayashi, M., Fukasawa, A., Mizuno, S., Arisawa, A. & Akagawa, H. (2002) The rpoZ gene, encoding the RNA polymerase omega subunit, is required for antibiotic

production and morphological differentiation in Streptomyces kasugaensis. *J Bacteriol* **184**:6417–6423.

Kontur, W. S., Capp, M. W., Gries, T. J., Saecker, R. M. & Record, M. T. (2010) Probing DNA binding, DNA opening, and assembly of a downstream clamp/jaw in Escherichia coli RNA polymerase-lambdaP(R) promoter complexes using salt and the physiological anion glutamate. *Biochemistry* **49**:4361–4373.

Kontur, W. S., Saecker, R. M., Davis, C. A., Capp, M. W. & Record, M. T. (2006) Solute probes of conformational changes in open complex (RPo) formation by Escherichia coli RNA polymerase at the lambdaPR promoter: Evidence for unmasking of the active site in the isomerization step and for large-scale coupled folding in the subsequent conversion to RPo. *Biochemistry* **45**:2161–2177.

Kurkela, J., Fredman, J., Salminen, T. A. & Tyystjärvi, T. (2021) Revealing secrets of the enigmatic omega subunit of bacterial RNA polymerase. *Mol Microbiol* **115**:1–11.

Lara-Gonzalez, S., Dantas Machado, A. C., Rao, S., Napoli, A. A., Birktoft, J., Di Felice, R., … Lawson, C. L. (2020) The RNA Polymerase α Subunit Recognizes the DNA Shape of the Upstream Promoter Element. *Biochemistry* **59**:4523–4532.

Lee, J. & Borukhov, S. (2016) Bacterial RNA Polymerase-DNA Interaction-The Driving Force of Gene Expression and the Target for Drug Action. *Front Mol Biosci* **3**:73.

Li, X. Y. & McClure, W. R. (1998) Stimulation of open complex formation by nicks and apurinic sites suggests a role for nucleation of DNA melting in Escherichia coli promoter function. *J Biol Chem* **273**:23558–23566.

Maitra, A., Moreno, J. & Hernandez, V. J. (2002) Low concentrations of free hydrophobic amino acids disrupt the Escherichia coli RNA polymerase core-sigma(70) protein-protein interaction. *Protein Expr Purif* **24**:163–170.

Mao, C., Zhu, Y., Lu, P., Feng, L., Chen, S. & Hu, Y. (2018) Association of ω with the C-Terminal Region of the β’ Subunit Is Essential for Assembly of RNA Polymerase in Mycobacterium tuberculosis. *J Bacteriol* **200**:e00159-18.

Mathew, R., Mukherjee, R., Balachandar, R. & Chatterji, D. (2006) Deletion of the rpoZ gene, encoding the omega subunit of RNA polymerase, results in pleiotropic surface-related phenotypes in Mycobacterium smegmatis. *Microbiol Read Engl* **152**:1741–1750.

Mazumder, A. & Kapanidis, A. N. (2019) Recent Advances in Understanding σ70-Dependent Transcription Initiation Mechanisms. *J Mol Biol* **431**:3947–3959.

Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A. N., … Ebright, R. H. (2002) Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **108**:599–614.

Mendoza-Vargas, A., Olvera, L., Olvera, M., Grande, R., Vega-Alvarado, L., Taboada, B., … Morett, E. (2009) Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in E. coli. *PloS One* **4**:e7526.

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E. A., Darst, S. A., Ebright, R. H. & Severinov, K. (2001) Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci U S A* **98**:892–897.

Murakami, K. S. (2013) X-ray crystal structure of Escherichia coli RNA polymerase σ70 holoenzyme. *J Biol Chem* **288**:9126–9134.

Murakami, K. S. (2015) Structural biology of bacterial RNA polymerase. *Biomolecules* **5**:848–864.

Murakami, K. S., Masuda, S. & Darst, S. A. (2002) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. *Science* **296**:1280–1284.

Murakami, K. S., Masuda, S. & Darst, S. A. (2003) Crystallographic analysis of Thermus aquaticus RNA polymerase holoenzyme and a holoenzyme/promoter DNA complex. *Methods Enzymol* **370**:42–53.

Nagai, H. & Shimamoto, N. (1997) Regions of the Escherichia coli primary sigma factor sigma70 that are involved in interaction with RNA polymerase core enzyme. *Genes Cells Devoted Mol Cell Mech* **2**:725–734.

Owens, J. T., Miyake, R., Murakami, K., Chmura, A. J., Fujita, N., Ishihama, A. & Meares, C. F. (1998) Mapping the sigma70 subunit contact sites on Escherichia coli RNA polymerase with a sigma70-conjugated chemical protease. *Proc Natl Acad Sci U S A* **95**:6021–6026.

Panaghie, G., Aiyar, S. E., Bobb, K. L., Hayward, R. S. & de Haseth, P. L. (2000) Aromatic amino acids in region 2.3 of Escherichia coli sigma 70 participate collectively in the formation of an RNA polymerase-promoter open complex. *J Mol Biol* **299**:1217–1230.

Richardson, J. P. (2002) Rho-dependent termination and ATPases in transcript termination. *Biochim Biophys Acta* **1577**:251–260.

Rogozina, A., Zaychikov, E., Buckle, M., Heumann, H. & Sclavi, B. (2009) DNA melting by RNA polymerase at the T7A1 promoter precedes the rate-limiting step at 37 degrees C and results in the accumulation of an off-pathway intermediate. *Nucleic Acids Res* **37**:5390–5404.

Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., … Gourse, R. L. (1993) A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **262**:1407–1413.

Saecker, R. M., Mueller, A. U., Malone, B., Chen, J., Budell, W. C., Dandey, V. P., … Darst, S. A. (2024) Early intermediates in bacterial RNA polymerase promoter melting visualized by time-resolved cryo-electron microscopy. *Nat Struct Mol Biol* **31**:1778–1788.

Saecker, R. M., Record, M. T. & Dehaseth, P. L. (2011) Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J Mol Biol* **412**:754–771.

Saecker, R. M., Tsodikov, O. V., McQuade, K. L., Schlax, P. E., Capp, M. W. & Record, M. T. (2002) Kinetic studies and structural models of the association of E. coli sigma(70) RNA polymerase with the lambdaP(R) promoter: Large scale conformational changes in forming the kinetically significant intermediates. *J Mol Biol* **319**:649–671.

Schroeder, L. A., Choi, A.-J. & DeHaseth, P. L. (2007) The -11A of promoter DNA and two conserved amino acids in the melting region of sigma70 both directly affect the rate limiting step in formation of the stable RNA polymerase-promoter complex, but they do not necessarily interact. *Nucleic Acids Res* **35**:4141–4153.

Schroeder, L. A., Gries, T. J., Saecker, R. M., Record, M. T., Harris, M. E. & DeHaseth, P. L. (2009) Evidence for a tyrosine-adenine stacking interaction and for a short-lived open intermediate subsequent to initial binding of Escherichia coli RNA polymerase to promoter DNA. *J Mol Biol* **385**:339–349.

Sclavi, B., Zaychikov, E., Rogozina, A., Walther, F., Buckle, M. & Heumann, H. (2005) Real-time characterization of intermediates in the pathway to open complex formation by Escherichia coli RNA polymerase at the T7A1 promoter. *Proc Natl Acad Sci U S A* **102**:4706–4711.

Scott, L. J. (2013) Fidaxomicin: A review of its use in patients with Clostridium difficile infection. *Drugs* **73**:1733–1747.

Sharp, M. M., Chan, C. L., Lu, C. Z., Marr, M. T., Nechaev, S., Merritt, E. W., … Gross, C. A. (1999) The interface of sigma with core RNA polymerase is extensive, conserved, and functionally specialized. *Genes Dev* **13**:3015–3026.

Shuler, M. F., Tatti, K. M., Wade, K. H. & Moran, C. P. (1995) A single amino acid substitution in sigma E affects its ability to bind core RNA polymerase. *J Bacteriol* **177**:3687–3694.

Shultzaberger, R. K., Chen, Z., Lewis, K. A. & Schneider, T. D. (2007) Anatomy of Escherichia coli sigma70 promoters. *Nucleic Acids Res* **35**:771–788.

Srivastava, D. B., Leon, K., Osmundson, J., Garner, A. L., Weiss, L. A., Westblade, L. F., … Campbell, E. A. (2013) Structure and function of CarD, an essential mycobacterial transcription factor. *Proc Natl Acad Sci U S A* **110**:12619–12624.

Tomsic, M., Tsujikawa, L., Panaghie, G., Wang, Y., Azok, J. & deHaseth, P. L. (2001) Different roles for basic and aromatic amino acids in conserved region 2 of Escherichia coli sigma(70) in the nucleation and maintenance of the single-stranded DNA bubble in open RNA polymerase-promoter complexes. *J Biol Chem* **276**:31891–31896.

Vassylyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S. & Yokoyama, S. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature* **417**:712–719.

Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H. & Artsimovitch, I. (2007a) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448**:157–162.

Vassylyev, D. G., Vassylyeva, M. N., Zhang, J., Palangat, M., Artsimovitch, I. & Landick, R. (2007b) Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **448**:163–168.

Vvedenskaya, I. O., Vahedian-Movahed, H., Zhang, Y., Taylor, D. M., Ebright, R. H. & Nickels, B. E. (2016) Interactions between RNA polymerase and the core recognition element are a determinant of transcription start site selection. *Proc Natl Acad Sci U S A* **113**:E2899-2905.

Weiss, A., Moore, B. D., Tremblay, M. H. J., Chaput, D., Kremer, A. & Shaw, L. N. (2017) The ω Subunit Governs RNA Polymerase Stability and Transcriptional Specificity in Staphylococcus aureus. *J Bacteriol* **199**:e00459-16.

Weiss, L. A., Harrison, P. G., Nickels, B. E., Glickman, M. S., Campbell, E. A., Darst, S. A. & Stallings, C. L. (2012) Interaction of CarD with RNA polymerase mediates Mycobacterium tuberculosis viability, rifampin resistance, and pathogenesis. *J Bacteriol* **194**:5621–5631.

Werner, F. & Grohmann, D. (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* **9**:85–98.

Wilson, C. & Dombroski, A. J. (1997) Region 1 of sigma70 is required for efficient isomerization and initiation of transcription by Escherichia coli RNA polymerase. *J Mol Biol* **267**:60–74.

World Health Organization (2022) Antimicrobial resistance surveillance in Europe 2022 – 2020 data ISBN 978-92-890-5668-7

World Health Organization (2015) Global action plan on antimicrobial resistance ISBN 978 92 4 150976 3

Young, B. A., Gruber, T. M. & Gross, C. A. (2004) Minimal machinery of RNA polymerase holoenzyme sufficient for promoter melting. *Science* **303**:1382–1384.

Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. & Darst, S. A. (1999) Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. *Cell* **98**:811–824.

Zhilina, T. N., Zavarzin, G. A., Rainey, F., Kevbrin, V. V., Kostrikina, N. A. & Lysenko, A. M. (1996) 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* **46**:305–312.