Real-time monitoring of transcription with fluorescent light-up aptamers

Oskari Puro

Master's thesis

Subject: Molecular Biosciences (Biochemistry)
Author(s): Oskari Puro
Title: Real-time monitoring of transcription with fluorescent light-up aptamers
Supervisor(s): Georgi Belogurov, Janne Mäkinen
Number of pages: 62 pages
Date: 05.02.2023

All living species rely on transcription to express their genes and consequently their specific traits. Transcription is the synthesis of RNA from a DNA template, and it is carried out by an enzyme known as RNA polymerase (RNAP). Transcription can be divided into three main stages: initiation, elongation, and termination. So, far commonly used *in vitro* transcription methods are laborious and there aren't straightforward ways for monitoring the entire transcription cycle.

In this study we used a fluorescent light-up aptamer (FLAP) to develop assay for real-time monitoring of transcription. We prepared several plasmids encoding the FLAP positioned at varied distances from a strong bacterial promoter and evaluated their suitability for monitoring various stages of transcription cycle with *Escherichia coli* RNAP. We optimized reaction component concentrations including Mg^{2+}, NTPs, fluorophore, plasmid, and holoenzyme and additionally pH for further investigations with *Spirochaeta africana* transcription system.

Following the optimization assays, we investigated the mostly unknown transcription factors from *S. africana* with the optimized assay. The elongation factor, LoaP, apparently slows down transcription elongation, but steadily increases the transcription output in long-term. The elongation factor, NusA, and the termination factor, Rho, had strong negative effects on transcription similarly to their well-studied *E. coli* counterparts. We discovered that the unique spirochetal Gre factor apparently facilitates transcription by having a strong positive effect on transcription initiation. The effect of GreA was studied with truncated *S. africana* Gre proteins that did not exhibit the same effect than full-sized GreA. Studies with *S. africana* transcription factors will help understanding transcription in pathogenic spirochetes which cause diseases such as Lyme disease, syphilis and leptospirosis. Results of this study validate that FLAP assays are a novel, straightforward way of monitoring transcription in real-time.

Contents

Abbreviations

DFHBI-1T   (Z)-4-(3,5-difluoro-4-hydroxybenzylidene)-1,2-dimethyl-1H-
                    imidazol-5(4H)-one

DNA   deoxyribonucleic acid

dsDNA   double-stranded DNA

FLAP   fluorescence light-up aptamer

GFP   green fluorescent protein

rNTP   ribonucleoside triphosphate

RNA   ribonucleic acid

RNAP   RNA polymerase

TEC   transcription elongation complex

Introduction

All living species are dependent on cellular mechanisms known as transcription in process to express specific traits. Through transcription all living organisms make RNA according to DNA template (Furth et al. 1962). Transcription is the first step in gene expression. Transcribed RNA can be utilized, as messenger RNA for further processing into proteins or to construct structural complexes like ribosomal RNA or transfer RNA, which both are involved in translation stage of protein synthesis in living cell, to mention a few examples. Expression of genes is tightly regulated in cells, so the correct proteins are expressed in correct conditions in cell's life cycle such as normal growth or during infection. Regulation of gene expression can be achieved in multiple stages, but we will be returning to transcription regulation later in this study. (Watson 2014)

RNA polymerase (RNAP) is a processive enzyme that catalyzes RNA synthesis using DNA as a template (Hurwitz et al. 1960, Furth et al. 1962). RNAPs structure vary between prokaryotes and eukaryotes and even though their subunits and consequently structures are not the same in eukaryotes and prokaryotes, their main function has remained the same through evolution and the core structures of the multi-subunit RNAPs are homologous (Ebright 2000). Eukaryotes also have multiple types of RNAPs while bacteria and archaea have single type of RNAP. Bacterial RNAP consists of five subunits which are two α subunits, β and β' subunits and ω subunit. The two α subunits and ω subunit have structural and regulatory roles. β and β' subunit resemble a 'crab claw' structure creating the active site, which hosts the catalytic center of the enzyme. (Zhang et al. 1999)

RNAP is capable of synthesizing RNA without any additional factors, however this is not the reality in living organisms. Although, RNAP core enzyme could work independently (Burgess et al. 1969), for cells to adapt to their environment varying group of proteins known as transcription factors are needed to regulate transcription according to cell's needs (Perez-Rueda 2000). Transcription cycle includes three major stages: initiation, elongation, and termination. Transcription can be regulated at various stages but most of them work at the level of initiation (Browning and Busby 2004). Some transcription factors interact with DNA and help RNAP to bind promoters that are known as activators. Then, a group of factors that inhibit RNAP binding to promoters that are known as repressors (Perez-Rueda 2000).

Finally, there are some factors which interact with RNAP over various transcription stages. These factors are so called RNAP-associated transcription factors.

Transcription regulation can be investigated with various *in vitro* assays. However, common *in vitro* transcription methods are rather laborious and do not allow monitoring of entire transcription cycle. Here, we have utilized fluorescence light-up aptamers (FLAPs) to solve some of these problems that the common *in vitro* transcription assays have. In chapter 1, we will discover basic principles of transcription in bacteria, how transcription can be studied using FLAP based assays and how they can be utilized in the transcription research of rather unknown bacteria species, *Spirochaeta africana*.

1 Transcription in bacteria

Transcription is the synthesis of RNA. The synthesis uses ribonucleotides as substrate and DNA as a template (Furth et al. 1962) and consequently transcription is one of the key processes in cellular organisms throughout the domains of life (Crick 1970). It is essential in the process of expressing genetic information leading up to protein synthesis. Transcribed RNA in cells can be used as messenger RNA to make proteins, as ribosomal RNA which is structural part of ribosome or as transfer RNA that transfers amino acids in the cell for protein expression, to name a few examples. Transcription is facilitated by an enzyme known as RNAP. (Watson 2014)

1.1 RNA polymerase

RNAP is a processive DNA-dependent enzyme (Furth et al. 1962). Bacteria and archaea have only one type of RNAP whereas eukaryotes have multiple types of RNAPs (Chambon 1975, Cramer 2002) for different cellular activities. Bacteria have the simplest structure of multi-subunit RNAPs, which has five subunits (Zhang et al. 1999, Minakhin et al. 2001), whereas RNAP structure of archaea resembles more closely the structure of eukaryotic Pol II that has 12 different subunits (Cramer 2002). While the overall structure of RNAPs throughout the domains of life varies, the core structure of the enzyme remains highly conserved in all of them (Sweetser et al. 1987, Cramer 2002). The bacterial RNAP core enzyme consists of five subunits: two α-subunits and individual β-, β'-, and ω-subunits (Zhang et al. 1999, Minakhin et al. 2001).

The two α-subunits create a homodimer and consist of two domains which are connected by a linker domain. The two α-subunits have mainly a role in RNAP assembly by dimerization of the amino-terminal domains (Igarashi et al. 1991) and this dimer then forms a complex with the β-subunit (Murakami et al. 1997*a*). However, the α carboxy-terminal domain have been found to have a key role in transcription factor mediated transcription initiation while not having an effect in the transcription activity from transcription factor-independent promoters (Igarashi and Ishihama 1991). It interacts with DNA and various transcription factors during transcription initiation (Murakami et al. 1996, 1997*a*, *b*).

The ω-subunit is the smallest of bacterial RNAP subunits. Early studies with ω-subunit have shown that it is nonessential for the RNAP catalytic viability (Zalenskaya et al. 1990, Gentry et al. 1991). Structure studies with ω-subunit implicate that ω-subunit mainly interacts with β'-subunit consequently promoting RNAP assembly (Minakhin et al. 2001) and stability (Vassylyev et al. 2002). It also has been suggested to have a role in the response to ppGpp, which is a bacterial alarmone (Ross et al. 2013). Furthermore, in recent studies the ω-subunit has been suggested to have regulatory role in the selection of σ-factor. More research is required for understanding the role of ω-subunit for transcription. (Kurkela et al. 2021)

The β- and β'-subunits form together a structure that resembles a crab claw-shaped structure. This structure contains the active site of bacterial RNAP, the entry and exit channels for double-stranded DNA (dsDNA), the entry channel for ribonucleosidetriphosphates (rNTPs) and the exit channel for the transcribed RNA that are formed at the RNAP assembly (Zhang et al. 1999). The active site harbors two Mg^{2+} ions that are coordinated by three aspartic acid residues and are critical for the nucleotidyl transfer reaction (Vassylyev et al. 2002). They are located in two double-psi beta-barrel domains which form the active site (Basu et al. 2014). The β'-subunit also has two additional motifs that are individually important in two distinct functions: the trigger loop for the catalysis (Mishanina et al. 2017) and the bridge helix for translocation (Vassylyev et al. 2002). The assembly of the whole bacterial RNAP core enzyme is suggested to occur by the dimerization of the two α-subunits, which then recruit the β- and β'-subunits and finally the ω-subunit (Sutherland and Murakami 2018).

While the RNAP core enzyme is sufficient to facilitate the RNA polymerization independently, a σ-factor is required for efficient transcription initiation from a promoter in a cellular environment (Burgess et al. 1969). The σ-factor is a protein cofactor that is

associated with RNAP core enzyme yielding an RNAP holoenzyme complex (Figure 1)(Vassylyev et al. 2002).
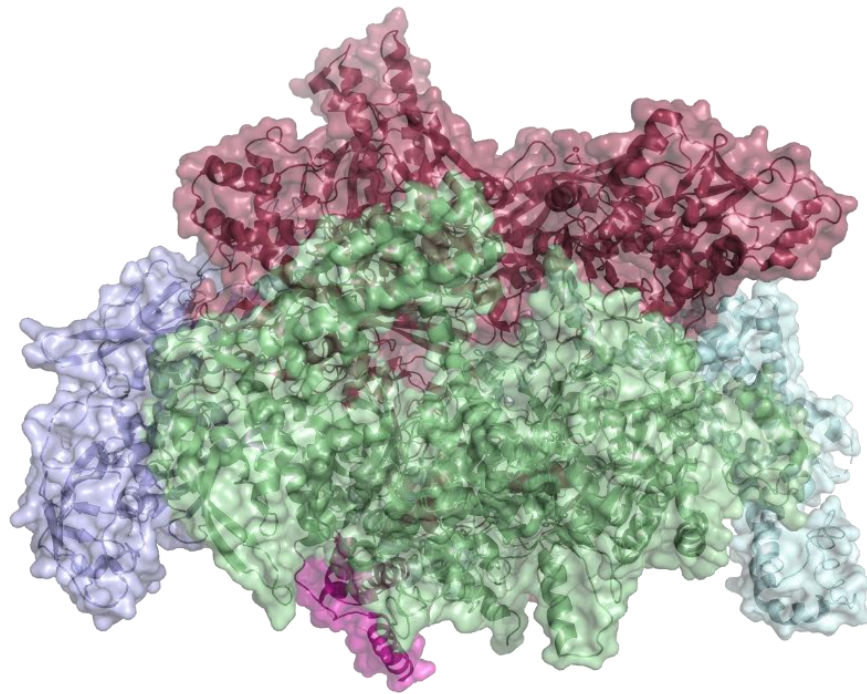
Figure 1 – Crystal structure of *E. coli* RNAP holoenzyme (PDB ID: 6C9Y). Bacterial holoenzyme consists of six subunits: two α-subunits (blue), β-subunit (red), β'-subunit (green), ω-subunit (magenta), and σ-factor (light blue).

In bacteria, most of the transcription initiation is established by the housekeeping σ-factor, which in *E. coli* is named σ^{70}-factor after the protein's molecular weight. Upon association to RNAP core enzyme, σ-factor undergoes a conformational change which then allows the RNAP holoenzyme bind to DNA (Callaci et al. 1998). When RNAP holoenzyme recognizes a promoter, the σ region 1.1 (σR1.1) domain of σ-factor, which mimics the negatively charged DNA, moves away from the active center (Mekler et al. 2002). Therefore, the σ-factor prevents unspecific association of RNAP to the DNA outside promoter regions (Dombroski et al. 1993). After the RNAP holoenzyme binds to promoter, transcription enters the first stage of transcription cycle, transcription initiation.

1.2   Transcription cycle

1.2.1 Transcription initiation

Transcription initiation is usually facilitated by σ^{70}-factor in *E. coli,* and we will be focusing on the housekeeping σ-dependent initiation. The σ-factor guides the RNAP to a promoter which in bacteria is generally composed of two different elements: -10 and -35 elements from transcription start site (Simpson 1979, Hawley and McClure 1983). The rate of initiation has been found to be promoter sequence dependent (Ruff et al. 2015). Promoter binding begins with two domains of the σ-factor binding to the -35 and -10 elements (Campbell et al. 2002, Feklistov and Darst 2011) and the α carboxy-terminal domain of RNAP recognizing the upstream element (Ross et al. 1993) forming a closed complex (Cook 2007) by loading DNA via loading gate in RNAP (Davis et al. 2007). The σR2 domain of σ-factor will facilitate the unwinding of DNA at the -10 element (Feklistov and Darst 2011). Formation of open complex is stabilized by DNA-protein contacts with the β-subunit and the σR1.2 domain of σ-factor (Haugen et al. 2006, Zhang et al. 2012).

Earlier studies suggested that the promoter melting is initiated by RNAP when it establishes specific interactions with the base at the -11 position of the non-template strand (Heyduk et al. 2006). This leads DNA to move towards the active center of RNAP (Feklistov and Darst 2011). Recent structural studies with single-particle cryo-electron microscopy have captured intermediate states between the closed complex and the open complex in the presence of transcription factor, TraR (Boyaci et al. 2019, Chen et al. 2020). This study validates the results from earlier studies, however it was achieved with additional factor present, so the results are not straightforward. Current understanding is that the σ-factor nucleates the -10 element melting and the initiation bubble expands by few nucleotides while the σR1.1 occupies the RNAP channel downstream in these structures. The other structure (Boyaci et al. 2019) suggests that the unwinding of the promoter occurs when the σR1.1 is ejected from the downstream channel. This leads to the final open complex formation where the template DNA is positioned in the active site of RNAP from which the transcription can then be initiated.

At the start of transcription, the RNAP will synthesize RNA in cycles that can lead to productive or abortive RNA synthesis (Gralla et al. 1980). In the abortive synthesis, the RNAP synthesizes short RNAs, which it releases and reverts to the open complex. The cycle

repeats while in the productive synthesis RNAP synthesizes 9- to 11-nucleotides long RNA, then RNAP can escape promoter into elongation (Gralla et al. 1980). The exact mechanism of promoter escape is still not perfectly understood but multiple studies demonstrate that it would take place through a mechanism known as scrunching where the RNAP remains fixed and pulls the downstream DNA into itself where the DNA strands create single-stranded bulges (Kapanidis et al. 2006, Revyakin et al. 2006, Winkelman et al. 2015). Following the scrunching, the RNAP enters the stage of transcription elongation.

1.2.2 Transcription elongation

Transcription proceeds to transcription elongation when it escapes the promoter and the RNAP will start synthesizing RNA in an efficient manner. In the core of the transcription elongation is the transcription elongation complex (TEC) that consists of the RNAP, the DNA, the transcribed RNA and possible additional transcription factors. Inside the primary channel of RNAP, DNA and RNA create an RNA:DNA hybrid through 9-10 base pairs distance (Nudler et al. 1997) and the non-template DNA strand is located away from the active site (Zhang et al. 1999). This 11-12 base pairs long melted region of DNA is known as the transcription bubble (Zaychikov et al. 1995, Turtola and Belogurov 2016). The TEC has multiple nucleic acid-protein interactions between the RNAP, RNA, and DNA, which are involved in maintaining the transcription bubble (Zhang et al. 1999, Cramer et al. 2001, Gnatt et al. 2001). These include the conserved D region loop 1 of the β-subunit (βD loop 1), the β'C rudder, and the βG flap. The βD loop 1 participates in the separation of downstream dsDNA, the β'C rudder separates the RNA:DNA hybrid and the βG flap forms the RNA exit channel with additional domains from the β'-subunit (Korzheva et al. 2000).

Essential steps of the rapid transcription elongation are the nucleotide addition reaction and the TEC translocation. The mobile trigger loop of RNAP participates in the nucleotide addition reaction by orientating the incoming rNTP in right orientation for catalysis (Vassylyev et al. 2007). The incoming rNTP is attached from the α-phosphate to the 3'-OH of the RNA through S_N^2 nucleophilic attack that is activated by the catalytic Mg^{2+} (Steitz 1998). This leads to pyrophosphate release and RNAP translocation. RNAP toggles between pre-translocated and post-translocated states during transcription. In the pre-translocated state, the 3'-OH of the RNA occupies the catalytic site and it is the target of nucleophilic

attack. Following the translocation, the active site is freed to bind the next substrate (Abbondanzieri et al. 2005, Bar-Nahum et al. 2005).

1.2.3 Transcription termination

The last stage of the transcription cycle is transcription termination. Transcription termination occurs if the TEC is dismantled after a transcriptional pause which is caused by an obstruction (Landick 2006). Transcription termination can be result of intrinsic or Rho-dependent pathways (Roberts 1969, Wilson and Von Hippel 1995). The Rho-dependent termination involves termination factor, Rho, and is assisted by elongation factors NusA and NusG. Rho is ATP-dependent hexameric helicase, which occurs to be associated with RNAP by traveling with it, rather than establishing contacts first with the transcribed RNA (Epshtein et al. 2010, Hao et al. 2021, Said et al. 2021) contrary to the classical model of Rho-dependent termination (Ray-Soni et al. 2016).

In the classical model, Rho binds with the nascent RNA that RNAP is synthesizing. In the nascent RNA, Rho will recognize cytosine-rich RNA segments known as Rho-utilization transcript (RUT) (Richardson and Richardson 1996). Binding of RUT and ATP will trigger helicase ring closure of Rho, which activates it for catalysis. The classical model of Rho termination has two suggested mechanisms. In the first one, Rho drives RNAP forward along the template DNA without nucleotide addition, which destabilizes the TEC and releases RNA. In the other one, Rho pulls RNA out of RNAP with force that shears the RNA-DNA hybrid, destabilizing TEC and releasing RNA. However, either of these mechanisms have not been directly observed. (Ray-Soni et al. 2016)

The novel, competing theory is based on biochemical evidence and recent structural results (Epshtein et al. 2010, Hao et al. 2021, Said et al. 2021). Rho occurs to travel with TEC in the open ring conformation. The transcribed RNA will first occupy the secondary binding site of Rho entering the central channel of hexameric Rho from its carboxy-terminal side. When Rho encounters RUT, the amino-terminal domains of Rho establish contacts with the RUT triggering the possible ring structure closure. This loading triggers the ATPase activity and starts pulling the transcribed RNA from RNAP resulting in transcription termination (Epshtein et al. 2010, Hao et al. 2021). However, it has been suggested that Rho could also trigger termination in RUT-independent manner with assistance of NusA and NusG by

inducing allosteric changes to the TEC that leads to transcription termination without Rho pulling RNA from RNAP with force (Said et al. 2021).

In the intrinsic termination pathway, TEC transcribes guanine-cytosine (GC)-rich repeat sequence followed by transcription pause induced by slowed transcription of thymine-rich section in downstream DNA. Transcription of uracil (U)-rich section leads to partial melting of the RNA:DNA hybrid few base pairs upstream from the 3' end of RNA. This melting induces the formation of GC-rich termination hairpin in RNA upstream from the U-rich section (Gusarov and Nudler 1999). Formation of the termination hairpin is suggested to be assisted by the flexible domains in the RNA exit channel (Epshtein et al. 2007). The growing termination hairpin invades the main channel of RNAP affecting the function of the trigger loop (Epshtein et al. 2007) and the RNAP clamp opening (Chakraborty et al. 2012) which results in a kinked bridge helix formation disfavoring the substrate entry to the active site (Weixlbaumer et al. 2013). Following changes in the TEC structure lead in the decrease of nucleic acid-protein interactions and finally to the dissociation of RNAP from the DNA and RNA, and to the termination in sequence-dependent manner.

1.3   Transcription regulation

Transcription, being the first step in gene expression in cellular organisms leading up to the production of proteins, is necessary to be regulated for the organisms to react to changes in their environment. Generally, transcription regulation occurs in the stage of transcription initiation by regulatory proteins known as transcription factors. Common transcription factors which participate in the regulation of initiation step are known as activators or repressors (Perez-Rueda 2000). These transcription factors regulate transcription by binding to the upstream DNA of transcription start sites. Briefly, the activators assist the RNAP holoenzyme binding to the promoter while the repressors inhibit the RNAP holoenzyme binding (Perez-Rueda 2000). However, there are multiple other transcription factors which participate in the regulation of various transcription stages through association to the RNAP and allosteric regulation of the function of RNAP. Later, we will explore some common RNAP-associated transcription factors and their function in *Spirochaeta africana*.

Traditional *in vitro* transcription assays are laborious and do not allow monitoring of the entire transcription cycle. Here, we will give a few examples of these assays (Figure 2). There are commercial rolling circle-based assays that utilize single-stranded circular DNA that is transcribed (Daubendiek and Kool 1997). Something such as intercalating fluorophores can be then used to detect the accumulation of RNA. Then, there are assays that utilize chemically produced non-template and template DNAs and 5' fluorophore labeled RNA. This assay is used to study processive transcription elongation step-by-step by assembling TEC using DNAs and RNA and then initiating transcription by addition of NTPs. Accumulated RNAs are then separated on denaturing PAGE gel and detected by the 5' fluorophore label.
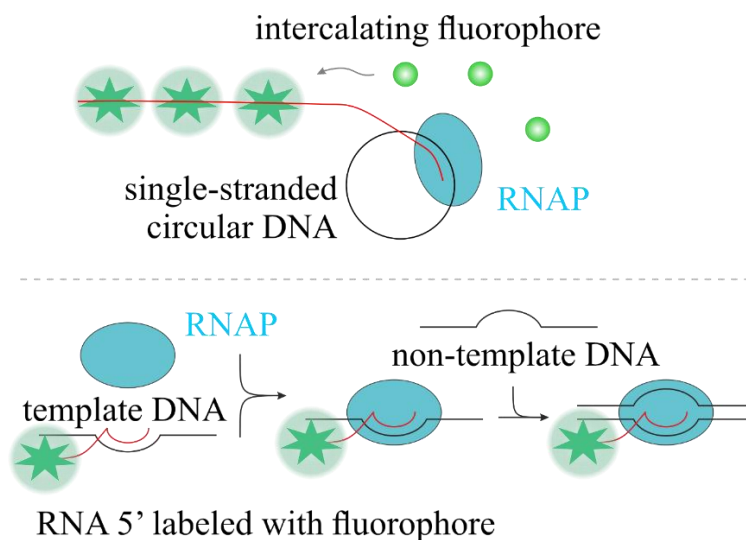
Figure 2 – Commonly used *in vitro* transcription assays. Top: Rolling-circle based assay that utilizes intercalating fluorophores for monitoring transcription. Bottom: Processive transcription elongation assay that utilizes chemically produces non-template and template DNAs and 5' fluorophore labeled RNA. Transcription elongation complex is assembled with the chemically produced DNAs and RNA and transcription is initiated from the elongation phase with the addition of the rNTPs.

While the later assay is useful in studying transcription process step-by-step since it can separate singular nucleotide differences, they are both problematic for studying the function of some transcription factors that influence at multiple stages of transcription as well as being laborious.

Aptamers are usually oligonucleotides that bind a specific target molecule. In the recent decade, there has been considerable progress in methods to image RNAs with fluorescence. A new method of monitoring RNAs derived from the use of green fluorescent proteins (GFP) in the analysis of proteins. GFPs produce fluorescence when they fold. GFP folding induces conformation change in three residues that create the fluorophore, 4-hydroxybenzlidene (HBI). HBI is encased in the protein which enables the fluorescence of the complex. GFPs can be utilized in the imaging of RNAs and DNAs but this results in high background signal which is produced by unbound GFPs (Chudakov et al. 2010).

To overcome this challenge, there has been development of RNA sequences that would exhibit GFP-like properties. RNA mimics GFP bind fluorophore specifically and binding of the fluorophore increases fluorescence of the fluorophore greatly from unbound state (50- to 100-fold). First, different HBI derivatives were experimented with, so that they are not activated by other cellular constituents. Second, a large number of distinct RNA molecules were identified by their ability of binding and activating fluorescence of HBI variant. These experiments were done utilizing systematic evolution of ligands by exponential enrichment (SELEX) and after 10 rounds of SELEX, the fluorescence induced by individual RNAs was determined. Third, new HBI derivatives, which were developed based on properties of enhanced GFP, were tested against RNAs that exhibited the highest aptamer-induced fluorescence. From these experiments, there was identified RNA sequence 24-2 named 'Spinach' and fluorophore, 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI). This complex produces green fluorescence light and is considerably resistant to photobleaching. These qualities make Spinach and other FLAP based assays viable for RNA imaging and specifically for transcription factor research. (Paige et al. 2011)

2.1 Broccoli fluorescence light-up aptamer

Spinach was one of the brightest aptamers found utilizing SELEX with DFHBI. However, multiple FLAPs have been identified that have even better properties than Spinach. This has been achieved by utilizing SELEX and fluorescence-activated cell sorting (FACS). While SELEX is an effective method to sort out RNAs that bind to DFHBI, it does not measure the efficiency of RNA folding in cellular environment or the ability of turning on the fluorescence of DFHBI. FACS allows the screening of aptamers based on properties. Following SELEX and FACS screening, potential RNA aptamer fluorescence properties

were tested in the presence of new higher fluorescence output fluorophore, (Z)-4-(3,5-difluoro-4-hydroxybenzylidene)-1,2-dimethyl-1H-imidazol-5(4H)-one (DFHBI-1T). Directed evolution was applied to improve the properties of the potential aptamer. The desired properties were improved folding in cellular environment and improved *in vivo* fluorescence output. This led to discovery of aptamer named Broccoli. Broccoli works without transfer RNA, which is used to improve folding, *in vivo*, has higher thermostability and requires lower Mg^{2+} concentrations for folding than Spinach. Broccoli is also 49 nucleotides long while Spinach is significantly longer. All the mentioned properties establish Broccoli advantageous to use as a tag for imaging RNA both *in vitro* and *in vivo* over Spinach. (Filonov et al. 2014)

2.2   Fluorescence light-up aptamer assays

Fluorescence increase can be monitored using FLAPs to study transcription in real-time. For monitoring the entire transcription cycle, a DNA template is required which has a promoter region, the sequence encoding a FLAP and an intrinsic terminator region. The basic principle of *in vitro* assays that utilize fluorescence increase created by fluorophore binding to FLAPs is to assemble RNAP holoenzyme complex and initiate transcription from the template with the addition of rNTPs (Figure 3). Step-by-step, the process involves the sigma-dependent initiation from the promoter, transcription of the template which has the sequence encoding a FLAP and finally the termination. After RNAP is dissociated from the template, the transcription cycle repeats. When the sequence encoding the FLAP is transcribed, there is a delay before the FLAP adopts a three-dimensional fold and binds a fluorophore present in the mixture. Binding the fluorophore to the FLAP causes the fluorophore to be 'switched on'. As mentioned, transcription cycle repeats in these assays, which causes increasing amounts of FLAPs to be transcribed and consequently increase in fluorescence signal. The increase in fluorescence signal is measured with a fluorometer and the data is analyzed.
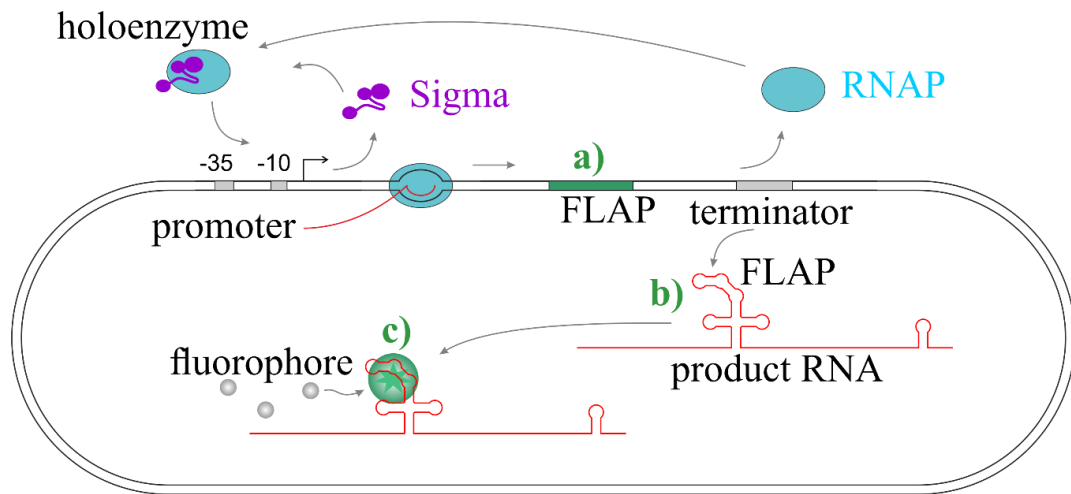
Figure 3 – Multi round FLAP assay for real-time monitoring of transcription. A) Transcription elongation complex transcribes the sequence encoding FLAP. B) Transcribed FLAP adopts a complex 3D fold. C) Unbound fluorophore is virtually non-fluorescent. Folded FLAP binds fluorophore enhancing the fluorescence emitted by the fluorophore greatly.

Different templates, where the position of the FLAP varies relative to the promoter and the terminator, for the assays can be designed for investigating transcription and various transcription factors, making it versatile tool for transcription research. The fluorescence curves produced by the assay utilize the increase of fluorescence induced by FLAP-fluorophore binding (Figure 6). Fluorescence curves can be divided into two individual phases: a lag phase and a linear phase. The lag phase includes all the steps before the generation of fluorescence such as aptamer folding and transcription elongation in case of long templates, while the linear phase reports the rate of transcription initiation (Huang et al. 2022).

3 Spirochetes

Spirochetes are free-living or host-associated bacteria that have some distinct features compared to other bacteria. They exist as free-living in distinct water-based environments such as lakes, ponds and as extreme as deep-sea hydrothermal vents. They also exist as associated to eukaryotic hosts but can also be pathogenic to animals and humans causing various diseases. Spirochetes have flagella inside the long, helical cell body contrary to many

other bacteria that have helical flagella which extends outside the cell body to enable the movement of these bacteria (Figure 4). Instead, spirochetes motility is dictated by movement of the periplasmic flagella and consequently inducing movement of the whole cell body (Nakamura 2020).
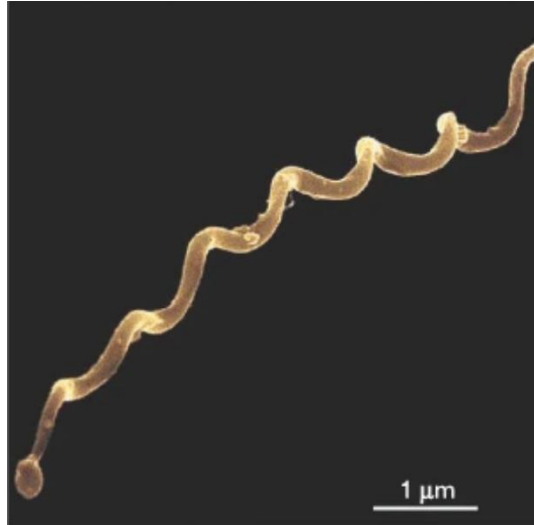
Figure 4 – Scanning electron micrograph of *Borrelia burgdorferi* (Rosa et al. 2005). Spirochetes have long, helical cell body.

This unique motility could also explain why pathogenic spirochetes are able to move around in viscous environments and penetrate tissues rather efficiently. Some pathogenic spirochetes have also been reported to evade the complement immune response in humans (Skare and Garcia 2020). Pathogenic spirochetes include some of *Leptospira* species, few *Borrelia* species, *Treponema pallidum*, *Brachyspira pilosicoli* and *Brachyspira aalborgi*. These species cause diseases such as leptospirosis, Lyme disease, syphilis and intestinal spirochaetosis. The pathogenicity of some spirochetes and the shortage of transcription research of them creates interest to study how spirochetal transcription functions. As it is easier to study organisms with non-pathogenic species, we utilized *Spirochaeta africana* for our transcription research. (Harwood and Canale-Parola 1984)

3.1   *Spirochaeta africana*

S. africana is halophilic, aerotolerant bacterium found from basic Lake Magadi in Africa (Zhilina et al. 1996). They are non-pathogenic bacteria and consequently easier to handle than pathogenic spirochetes. Genome of *S. africana* was sequenced in 2013 (Liolos et al.

2013) and from the genome, various transcription factors were identified based on bioinformatics analysis.

3.2 RNAP-associated transcription factors in *S. africana*

RNAP-associated transcription factors can be found across all domains of life. While some factors are universally conserved amongst all domains of life, some function differently between species. Following RNAP-associated factors were identified from *S. africana* genome for this study: few σ-factors, CarD, DksA, GreA, LoaP, NusA, NusG, Rho, Tex, and TPR. Elongation Factor Thermo Unstable (EF-Tu) was also studied since it was observed to form a complex with *S. africana* RNAP. Here, we will explore RNAP-associated transcription factors that are found in *S. africana* and their function in other bacteria.

3.2.1 Transcription initiation factors

DksA is a transcription initiation factor that associates with RNAP by binding to the secondary channel rim of the β'-subunit and altering the DNA binding channel region allosterically in proteobacteria. It has coiled-coil domain that reaches into the RNAP active site interacting with the trigger loop, C-terminal helix domain and a zinc-binding globular domain. DksA has been found to be tightly related to the function of bacterial alarmone ppGpp which is induced in cells as a response to stress. DksA inhibits transcription from rRNA promoters and promotes transcription from metabolic promoters as well as contributes to the change from exponential to stationary cell cycle phase. The inhibition from rRNA promoters with both DksA and ppGpp present is greatly increased relative to ppGpp independently. DksA association to RNAP creates a second binding site for ppGpp, the other being at the β'-ω subunits interface of RNAP. The activation of transcription has only been observed from the site which requires DksA association to RNAP and seems to affect the open complex formation. DksA has been found to have some effect on elongation as well, but these effects are unclear. Function of DksA in *S. africana* has not been studied and consequently the effects to transcription are unknown. (Gourse et al. 2018)

CarD acts as a transcription initiation factor in mycobacteria. It associates with the β-subunit of RNAP and minor groove of DNA upstream of the -10 element of the promoter (Srivastava et al. 2013). CarD appears to promote the closed complex DNA melting and stabilize the open complex during transcription initiation through the interactions with RNAP and DNA in mycobacteria. Increased affinity of CarD for RNAP increases the growth rate of the cells

but decreases the virulence, suggesting that CarD might be essential for global transcription regulation in mycobacteria (Garner et al. 2017). However, in *Rhodobacter sphaeroides* CarD has been found to inhibit the activity of its own promoter suggesting that it acts as an activator or an inhibitor depending on the promoter sequence. The mechanism of inhibition has not been determined. The autoregulation of CarD promoter is suggested to be important in maintaining homeostasis during exponential cell cycle phase (Henry et al. 2021). CarD has not been studied in global transcription regulation and consequently its role is unclear (Garner et al. 2017).

3.2.2 Transcription elongation factors

NusA is a transcription elongation factor that associates with the RNA exit channel of RNAP and has interactions with various domains in the channel. NusA is bound to all RNAPs during elongation. It is suggested to have a role as mediating transcription pausing by making RNAP more sensitive to pause signals, facilitating Rho-dependent, and intrinsic termination as well as participating in anti-termination activities, and having a role as a chaperone in the folding of structural RNAs. In RNA hairpin-dependent transcription pausing, NusA alters the RNA exit channel domains to promote the RNA hairpin stability or to enhance allosteric effect to RNAP by the hairpin. (Belogurov and Artsimovitch 2015)

Another Nus factor, NusG is universally conserved through all domains of life. It is also associated with most RNAPs together with NusA in transcription elongation. It is suggested to reduce transcription elongation pausing and to facilitate promoter escape during transcription initiation. NusG has an essential role in silencing aberrant RNA transcription together with Rho-factor in *E. coli*. It is part of multi-component anti-termination complexes that promote assembly of rRNAs by ensuring transcription of rRNA genes. The effect of NusG as reducing transcription pauses during elongation is established by stabilizing the upstream edge of the transcription bubble leading to decreased ability for RNAP backtracking. The role of NusG as promoting and inhibiting termination is explained by the interactions established between NusG and Rho/ribosome, both of which are mediated by the Kyrpines-Ouzonis-Woese motif in the C-terminal domain of NusG. While NusG exhibits aforementioned functions in *E. coli*, there are reports of NusG in other bacteria species having alternative functions. (Wang and Artsimovitch 2021)

Multiple NusG paralogs have been discovered in bacteria, one of which is LoaP that can be found also in *S. africana*. While NusG associates with the most of RNAPs and the association mechanism with RNAP occurs to be quite similar between NusG and its paralogs, LoaP occurs to be involved only in regulation of few operons and is encoded by a sequence adjacent to its target operons. It has been found to facilitate anti-termination by interactions with intrinsic termination RNA hairpins and consequently promote expression of various antibiotic synthesis gene clusters in various *Bacillus* species (Goodson et al. 2017, Elghondakly et al. 2021).

GreA is a transcription elongation factor that has similar coiled-coil domain, which reaches into the active site of RNAP through the NTP entry channel, as DksA. In most bacteria, Gre factors have only this coiled-coil domain and β-barrel domain that facilitates the association with RNAP. Gre factors reportedly have anti-pause activity by triggering RNAP activity from polymerase to nuclease and they are also important for proof-reading during transcription. RNAP's nuclease activity is triggered when the acidic residues in the tip of the coiled-coil domain of Gre stabilize the catalytic Mg^{2+}-ions of RNAP, which allows the Gre-dependent cleavage of the nascent RNA in backtracked RNAP complexes (Nickels and Hochschild 2004) rescuing them from backtracked state. It has been reported that the trigger loop of RNAP plays important role in Gre-dependent cleavage and especially explains the differences in RNA cleavage activity between different bacterial RNAPs (Miropolskaya et al. 2017). In a recent study (Fernández-Coll et al. 2020), overexpression of GreA was found to be lethal in *E. coli* independently of its anti-pause activity. It was proposed that overexpressed Gre factors would interrupt reactivation of transcription after RNA cleavage by blocking the NTP enter channel from NTPs or other transcription factors that interact with the channel. The toxicity was decreased when DksA was present. Spirochetes are interesting bacteria regarding Gre factors since they have Gre factors with additional domains. Especially, *S. africana* has GreA (Figure 5A) with two additional domains consisting mostly of α-helices connected with a domain consisting of β-sheets. These additional domains are connected to the N-terminus of the coiled-coil domain of the usual Gre factors. The function of GreA of *S. africana* is unknown and the effect of additional domains has not been determined.
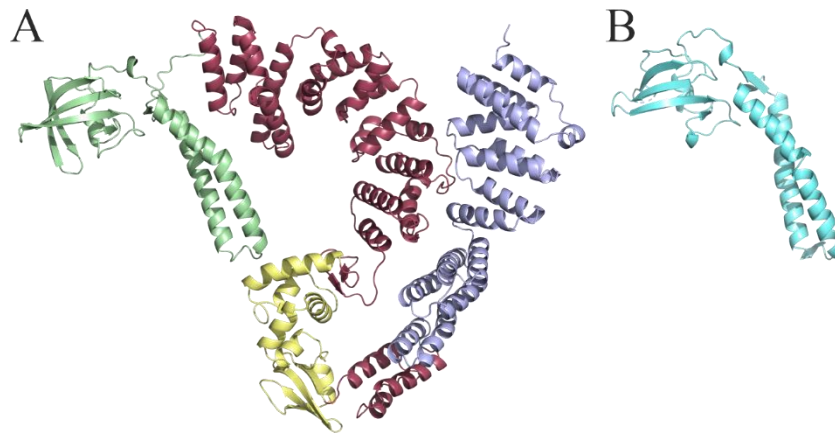
Figure 5 – Comparison of transcription elongation factor, GreA, structure from two different bacteria. A) Structure prediction of *S. africana* GreA from AlphaFold (Jumper et al. 2021, Varadi et al. 2022). Usual structure of Gre factors (coiled-coil and globular domains) illustrated as green, additional α-helical domains of *S. africana* GreA illustrated as light red and light blue. CarD-like domain illustrated in light yellow. B) Crystal structure of *E. coli* GreA (PDB ID: 1GRJ).

TPR is a protein that is suggested to consist of multiple tetratricopeptide repeat (TPR) domains. It is expressed by a gene that is located upstream of the GreA gene in all spirochetes and the additional domains of the GreA are TPR domains. Consequently, this connection between these factors suggests that they could function together, which is why TPR was included in FLAP assays.

EF-Tu is a protein which primary function is to transport aminoacylated tRNAs to the ribosome for translation of mRNAs and consequently is the most abundant protein in bacterial cells (Weijland et al. 1992). EF-Tu's function in *S. africana* transcription is unknown and due to observations of it forming a complex with RNAP, it is included in this study.

Tex is a transcription factor that has a role in toxin expression and was first found in *Bordetella pertussis* (Fuchs et al. 1996). It has been found to have a negative effect on transcription of toxin genes and is found to bind to DNA and thus regulating transcription (He et al. 2006). Its function has not been studied broadly, which is why Tex was included to the screening of *S. africana* transcription factors with the FLAP assay.

4 Aims

Broccoli-FLAP assays are novel and relatively straightforward way to study multiple stages of the transcription cycle. For *in vitro* transcription research, Broccoli-FLAP assays create solutions to drawbacks that some other common transcription assays have. The aims of my study are to explore how Broccoli-FLAP assays can be utilized in *in vitro* transcription research, optimize the assay to work for the two organisms that are mostly used in our research group and then study the RNAP-associated transcription factors of *S. africana* with the optimized assay.

Broccoli-FLAP assay utilizes transcription of plasmids that include Broccoli-RNA aptamer encoding sequence to produce fluorescence that can be monitored in real-time. In this study, plasmids constructs will be made that encode Broccoli-FLAP at varied distances from strong bacterial promoter for the Broccoli-FLAP assays. The assay will be optimized for *E. coli* and *S. africana* since they are distinctly different bacteria and live in totally different environments. Various reactant concentration changes will be tested with *E. coli* to produce sufficient fluorescence output with minimal reagent usage Regarding optimizations with *S. africana* especially higher buffer pH level will be experimented with due to the contrast in the environment between *S. africana* and *E. coli.* Following optimizations, various RNAP-associated transcription factors of *S. africana* will be experimented with using the optimized Broccoli-FLAP assay. Spirochetes have few transcription factors which are found in other bacteria, but their properties are unique compared to other bacteria such as GreA. Even though the studied transcription factors are found widely in other bacteria, the functions of these factors in spirochetes are mostly unknown.

Methods

5 Constructing Broccoli-FLAP encoding plasmids with Gibson assembly

5.1 Plasmid digestion

Plasmid, pJM020, was used for Broccoli-FLAP constructs. For all the constructs, plasmid was digested with designated restriction enzymes (1: MunI-Bsp1407I, 2: NheI-BamHI, 3: SacI-XhoI). Digestions were done overnight (O/N) at 37 °C in reaction mixes of 30 µl, which contained 1.5 µg of plasmid, FastDigest Buffer (10X, Thermo Scientific), and corresponding restriction enzymes (FastDigest, Thermo Scientific) for each construct according to manufacturer's guide.

5.2 Purification of digestion reactions

Agarose gel electrophoresis samples in DNA loading dye (6X, Fermentas) were done from the digestion products and run in 1 % agarose TAE gel (80 V, 1 h). Digestion products were analyzed against a ladder (GeneRuler DNA Ladder Mix, Thermo Scientific) and dissected from the gel for digested plasmid purification with a gel purification kit (NucleoSpin Gel and PCR Clean-up, Macherey-Nagel). Digested plasmid concentration was measured with NanoDrop2000 (Thermo Scientific).

5.3 Gibson assembly and transformation of Gibson assembly products

Gibson assembly reactions were done according NEBuilder HiFi DNA Assembly Reaction protocol, but the reaction volume was reduced to 10 µl with 50 ng of digested plasmid and 100 ng of insert fragment (Table 1). After the reaction, Gibson assembly products were transformed to competent *E. coli* XL-1 cells by incubating 5 µl of Gibson assembly products with 100 µl of competent cells on ice for 30 min, then inducing heat shock by incubation at 42 °C for 2 min. Fresh LB medium was added to the cells and cells were recovered at 37 °C shaker (250 rpm) for 1 h. Cells were plated on LA plates with carbenicillin (100 µg/ml) and grown O/N at 37 °C.
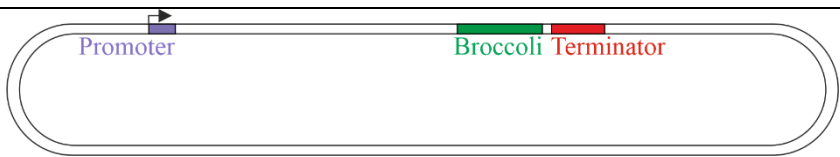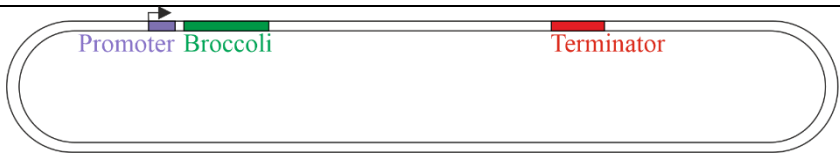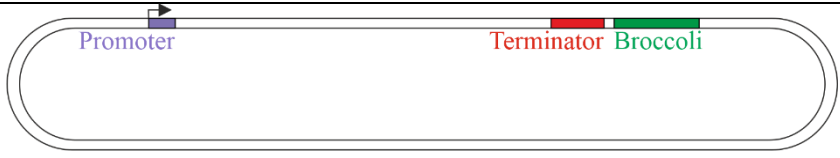
Table 1 – Gibson assembly of the digested plasmids needed for Broccoli-FLAP plasmid constructs.

Digested plasmid	Insert
pJM020 – MunI-Bsp1407I	broccoli_distant
pJM020 – NheI-BamHI	broccoli_proximal
pJM020 – SacI-XhoI	broccoli_postterm

5.4 Purification of Broccoli-FLAP encoding plasmid constructs and sequencing

From the plate, further cultivations were done from singular colonies to 5 ml of LB medium (100 µg/ml carbenicillin) and grown at 37 °C until turbid. Plasmids were purified from these cultivations according to a plasmid purification kit (NucleoSpin Plasmid, Macherey-Nagel) and the final plasmid concentrations were measured with NanoDrop2000 (Thermo Scientific). Plasmid constructs were sequenced with Mix2Seq kit (Eurofins) and sequences were analyzed with sequence alignment software. Plasmids that had correct sequence were named based on the position of the Broccoli-FLAP to the promoter or terminator (Table 2).

Table 2 – Plasmid constructs made for FLAP assays.

Construct name	Plasmid name	Broccoli-FLAP sequence position
Broccoli Distant	pOP004	
Broccoli Proximal	pOP005	
Broccoli Post termination	pOP007	

5.5 Large-scale purification of template plasmids

Constructed plasmids were cultivated and purified in large-scale according to the kit instructions (NucleoBond Xtra Maxi, Macherey-Nagel) for FLAP assays. Plasmids, pOP004 (#185985) and pOP005 (#186030), were deposited to Addgene repository (Huang et al. 2022).

6 Protein purifications

6.1 Cell cultivation and protein expression

All proteins contained histidine tag (6x) and were heterologously expressed in *E. coli* T7 express lysY/Iq (New England Biolabs) or Xjb cells and were purified with the following methods after cell disruption. Pre-cultures of the *E. coli* were cultivated in LB-medium (100 µg/ml ampicillin) with a shaker (37 °C, 250 rpm) overnight. Pre-culture was inoculated to LB-medium (100 µg/ml ampicillin), and culture was induced with 1 mM IPTG when OD_{600} reached 0.8. Culture was grown for 4 h after inducing at 30 °C and gathered by centrifuging in 7000 x g 4 °C for 12 min. Cells were transferred to -80 °C until further procedures.

6.1.1 Overnight expression

S. africana RNAP was expressed with overnight expression. Pre-culture was inoculated in LB-medium (100 µg/ml ampicillin) in the morning and was grown in a shaker (37 °C, 250 rpm) for 6-8 h. In the evening, overnight expression culture (LB with 1 mM $MgSO_4$, 100 mM PO_4, 25 mM SO_4, 50 mM NH_4, 100 mM Na, 50 mM K, 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose, (0.1 % arabinose for Xjb cells), 100 µg/ml ampicillin) was inoculated with the preculture (0.1 % volume of the overnight expression culture). Culture was grown overnight in a shaker (37 °C, 250 rpm) and cells were harvested in the morning after OD_{600} stayed the same for 3 h. Cells were harvested by centrifuging at 7000 x g 4 °C for 12 min. Cells were transferred to -80 °C until further procedures.

6.2 Cell disruption

Cells were suspended in 45 ml of lysis buffer (50 mM Tris-HCl pH 7.9, 500 mM NaCl, 5 % glycerol, 10 mM β-mercaptoethanol) with 1 mg/ml lysozyme (only for T7 cells), 1 protease inhibitor tablet (Pierce Protease Inhibitor Mini Tablets (ThermoFisher)) and incubating the

suspension on ice for 1 h (only for T7 cells). Cells were sonicated for 15 min (10/15 s sonication, 45 s break) on ice and the sonicated cells were centrifuged at 48000 x g 4 °C for 32 min. After centrifugation supernatant was gathered for further purification steps (Table 3) and Tween-20 was added (0.2 %). All the following purification steps were done at +8 °C if not reported otherwise.

Table 3 – Purification information of the proteins used in the study.

Protein	Cells used/Expression plasmid	Purification procedures
E. coli (eco) RNAP	T7/Xjb / pVS10	6.3, 6.6, 6.7, 6.6, 6.8
eco σ^{70}	T7 / pGB172	6.3, 6.6, 6.7, 6.4, 6.6, 6.8
S. africana (sfc) RNAP	T7/Xjb / pJM019	6.3, 6.6, 6.7, 6.6, 6.8
sfc σ^{70}	T7 / pGB172	6.3, 6.6, 6.7, 6.4, 6.6, 6.8
sfc CarD	T7 / pGB179	6.3, 6.4, 6.6, 6.8
sfc DksA	T7 / pGB175	6.3, 6.4, 6.10
sfc NusA	Xjb / pGB174/177	6.3, 6.10, 6.6, 6.8
sfc NusG	T7 / pGB171	6.3, 6.7, 6.4, 6.6, 6.8
sfc GreA	T7 / pGB170	6.3, 6.10
sfc EF-Tu	T7 / pGB204	6.3, 6.5, 6.6, 6.8
sfc TPR	T7 / pGB173	6.3, 6.4, 6.10, 6.6, 6.8
sfc Tex	T7 / pGB203	6.3, 6.4, 6.6, 6.7, 6.6, 6.8
sfc LoaP	T7 / pGB178	6.3, 6.4, 6.6, 6.9
sfc Rho	T7/Xjb / pGB209	6.3, 6.6, 6.7, 6.5, 6.6, 6.8
sfc GreMidEnd	T7 / pGB192	6.3, 6.4, 6.6, 6.8 (ResourceQ), 6.6, 6.8 (MonoQ)
sfc GreEnd	T7 / pGB194	6.3, 6.4, 6.6, 6.8

6.3 Purification of histidine-tagged proteins

The supernatant was purified by running it through NiSepharose column that has been washed with few column volumes (CV) of Milli-Q (MQ) water and balanced with few CVs of lysis buffer. After adding lysate to column, it was washed with one CV of lysis buffer and protein was eluted in 10 ml fractions using lysis buffer with 20-, 50-, 100-, and 200-mM imidazole. SDS-PAGE was run from the resulting fractions to determine which fractions

contained desired protein. Fractions that contained the desired protein were pooled together for further purification steps.

6.4 TEV digestion

Some proteins were digested overnight with TEV at +4 °C (Table 3). TEV concentration used for digestion was one-hundredth of the NiSepharose fraction's protein concentration.

6.5 Ulp1 digestion

Some proteins were digested with Ulp1 protease overnight at +4 °C. 100 µl of Ulp1 with 1 mM DTT was used for purified protein from 4 l of cultivation.

6.6 Fraction dilution

Fractions from NiSepharose column were diluted 3-4 times with ÄktaPurifier buffer A (50 mM Tris-HCl pH 7.9, 5 % glycerol, 0.1 mM Na-EDTA, 10 mM β-mercaptoethanol) to reduce the salt concentration for ion exchange chromatography steps.

6.7 Purification with affinity chromatography

NiSepharose fractions were purified by affinity chromatography on HiTrap Heparin HP (Cytiva) column with ÄKTApurifier (GE Healthcare). ÄKTApurifier and the column were washed with MQ water for multiple CVs, then the column was balanced with 100 % buffer B (50 mM Tris-HCl pH 7.9, 1.5 M NaCl, 5 % glycerol, 0.1 mM Na-EDTA, 1 mM β-mercaptoethanol). After balancing, the concentration of buffer B was lowered to 2.5 % with buffer A and the column was loaded with the diluted NiSepharose fractions. Finally, gradient was run from 2.5 % buffer B to 100 % buffer B for 90 ml and fractions of 1 ml were gathered. Fractions were analyzed with SDS-PAGE and fractions that contained desired protein were identified for further purification steps.

6.8 Purification with anion exchange chromatography

ÄKTApurifier was used with ResourceQ (GE Healthcare) or MonoQ (Cytiva) anion exchange column for further purification of NiSepharose fractions or affinity chromatography fractions. They were diluted for anion exchange chromatography with buffer A. After dilution, the balancing of the column, diluted fraction loading to column and gradient run proceeded similar to chapter 6.7. Resulted fractions were analyzed with SDS-

PAGE and fractions that contained desired protein were identified for further purification steps.

6.9 Purification with cation exchange chromatography

ÄKTApurifier was used with ResourceS (GE Healthcare) anion exchange column for further purification of NiSepharose fractions or affinity chromatography fractions. They were diluted for cation exchange chromatography with buffer A. After dilution, the balancing of the column, diluted fraction loading to column and gradient run proceeded similar to chapter 6.7. Resulted fractions were analyzed with SDS-PAGE and fractions that contained desired protein were identified for further purification steps.

6.10 Purification with size exclusion chromatography

ÄKTApurifier was used with Superdex200hr 10/30 (GE Healthcare) or HiPrep 16/60 Sephacryl S-200 HR (Cytiva) gel filtration column for further purification of NiSepharose fractions or affinity chromatography fractions. Fractions were concentrated with concentration columns (Amicon Ultracel (Merck Millipore Ltd)) for loading to the column. Washing of the column was done similar to chapter 6.7. The column was loaded with loading loop, sample was run in gel filtration buffer (50 mM Tris-HCl pH 6.9, 5% glycerol, 150 mM - 1.0 M NaCl, 1 mM β-mercaptoethanol, 0.1 mM EDTA), and fractions were gathered. Resulted fractions were analyzed with SDS-PAGE and fractions that contained desired protein were identified for further purification steps.

6.11 Concentration and dialysis

In all purifications, fractions containing desired protein were concentrated with concentration columns (Amicon Ultracel (Merck Millipore Ltd)) to appropriate volume with suitable molecular weight cut-off relative to the purified protein. Resulting concentrate were dialyzed in storage buffer (20 mM Tris-HCl pH 7.9, 150 mM NaCl, 0.1 mM Na-EDTA, 50 % glycerol, 0.1 mM DTT) with dialysis device (Slide-A-Lyzer Dialysis device or cassette (ThermoScientific)) with appropriate molecular weight cut-off relative to the purified protein at +8 °C overnight. Dialyzed protein's concentration was measured with NanoDrop2000.

6.12 SDS-PAGE

SDS-PAGE was run on different stages of purification. SDS-PAGE samples of 40 µl were made with loading buffer (NuPAGE™ LDS Sample Buffer, 4X (ThermoFisher)) and they

were incubated at 85 °C for 2 min. SDS-PAGE (NuPAGE™ 4 to 12%, Bis-Tris, 1.0 mm, Mini Protein Gel (ThermoFisher)) was run in MES buffer (Bolt™ MES SDS Running Buffer, 20X (ThermoScientific)) for 22 min (200 V, 100 mA). Gels were stained (GelCode Blue Safe Protein Stain (ThermoFisher)) and imaged with a scanner.

7 Fluorescence light-up aptamer assay

7.1 Proteins and reagents

All plasmids and proteins used in the assays were produced and purified according to chapters 5 and 6. NTPs and DFHBI-1T for assays were ordered from Jena Bioscience GmbH.

7.2 Initial assay

In the initial assays, all constructed plasmids were experimented with. First, 20 µl of RNAP holoenzyme was assembled (20 µM *E. coli* RNAP, 80 µM *E. coli* σ^{70}) in storage buffer by incubating for 20 min at 30 °C. Assembled RNAP holoenzyme was stored on ice for immediate use or at -20 °C if not used immediately.

Two different mixtures were prepared: 100 µl plasmid-fluorophore mixture (96 nM plasmid, 20 µM DFHBI-1T), including the assembled RNAP holoenzyme, in transcription buffer (40 mM HEPES-KOH pH 7.5, 80 mM KCl, 10 mM MgCl$_2$, 0.1 mM Na-EDTA, 0.1 mM DTT, 5 % glycerol) and 100 µl rNTP mixture (2 mM ATP, 2 mM CTP, 2 mM GTP, 2 mM UTP) in the transcription buffer. Plasmid-fluorophore mixture needed to be protected from light. Both mixtures were pre-heated to 37 °C prior to transcription reaction.

Transcription reaction was initiated by combining the plasmid-fluorophore and NTP reaction mixtures together in a cuvette (2 µM holoenzyme, 48 nM plasmid, 10 µM DFHBI-1T, 1 mM rNTPs). The mixing time was recorded, and fluorescence was measured using TimeDrive (472 nm excitation wavelength, 507 nm emission wavelength, width of excitation and emission slits 10 nm) for 1800 s at one second intervals at 37 °C with a fluorometer (LS-55 (PerkinElmer)). Fluorescence data was processed (Origin, Version 2016. (OriginLab Corporation, Northampton, MA, USA)) and plotted to a curve. Cuvette (16.160F-Q-10/Z15 (Starna GmbH)) was used for the initial and optimization assays.

7.3 Optimization of the assay

Various changes to holoenzyme, plasmid and NTP concentrations were experimented with during the optimization of the assay with *E. coli.* Broccoli Distant plasmid was used for *E. coli* optimization assays. For experiments with *S. africana* RNAP, optimizations included various Mg^{2+} and NTP concentrations and two different pH levels. Broccoli Proximal plasmid was used for *S. africana* optimization assays.

7.4 Investigation of the RNAP-associated transcription factors of *S. africana* with optimized assay

Assays with transcription factors were done with 45 µl cuvette (3x3 mm, QS High Precision Cell, Article number: 105-251-15-40 (Hellma Analytics)). First, the 5 µl RNAP holoenzyme assembly mixture (10 µM RNAP and 40 µM σ^{70}) was prepared in storage buffer and incubated for 20 min at 37 °C. Assembled RNAP holoenzyme was stored on ice for immediate use or at -20 °C if not used immediately.

Two different mixtures were prepared: 25 µl plasmid-fluorophore mixture (96 nM plasmid, 20 µM DFHBI-1T, 0.2 µM pyrophosphatase), including the assembled RNAP holoenzyme, in transcription buffer (40 mM TAPS-KOH pH 9.0, 80 mM KCl, 10 mM $MgCl_2$, 0.1 mM Na-EDTA, 0.1 mM DTT, 5 % glycerol) and 100 µl rNTP mixture (2 mM ATP, 2 mM CTP, 2 mM GTP, 2 mM UTP, 50 µM transcription factor) in the transcription buffer. Plasmid-fluorophore mixture needed to be protected from light. Both mixtures were pre-heated to 37 °C prior to transcription reaction.

Transcription reaction was initiated by combining the plasmid-fluorophore and NTP reaction mixtures together in a cuvette (1 µM holoenzyme, 48 nM plasmid, 10 µM DFHBI-1T, 0.1 µM pyrophosphatase, 1 mM rNTPs, 10 µM transcription factor). The mixing time was recorded, and fluorescence was measured using TimeDrive (472 nm excitation wavelength, 507 nm emission wavelength, width of excitation and emission slits 10 nm) for 600 s at one second intervals at 37 °C with a fluorometer. All measurements with transcription factors were duplicated. Individual fluorescence dataset was adjusted by the difference relative to the average value of the duplicates during the first 40 seconds of measurement. Normalized data was then averaged in duplicates and plotted with standard deviation in Origin.

7.5 Validation of the fluorescent data from the assay with gel electrophoresis

Fluorescent data from the initial assays was validated by running samples of the assay reactions in 1.5 % (m:v) agarose TBE (UltraPure TBE buffer, 10X (Invitrogen)) gels containing ethidium bromide (EtBr)(0.58 µg/ml) or gels stained afterwards with DFHBI-1T without EtBr. After an assay, transcription reactions were terminated with two-fold volume of STOP buffer (94 % formamide, 0.2 % Orange G, 13 mM Li-EDTA) for EtBr or with 6 X loading buffer (100 mM EDTA, 30 % glycerol, 0.2 % Orange G) for DFHBI-1T and samples were loaded on the gel. Gel electrophoresis was run for 1 h (80 V, 100 mA, 30 W).

After the run, the EtBr containing gel was imaged. When stained with DFHBI-1T, the gel was washed three times with MQ water for 5 min after the electrophoresis run and then stained with 10 µM DFHBI-1T (40 mM HEPES-KOH pH 7.4, 100 mM KCl, 1 mM $MgCl_2$) for 30 min. DFHBI-1T-stained gel was imaged using fluorescence (Alexa 488nm, 50 µm pixel size) with Sapphire Biomolecular Imager (Azure Biosystems).

Results

8 Validation and optimization of the fluorescence light-up aptamer assays

8.1 Initial assays

From the initial assays, we wanted to identify the difference in the fluorescence curves between the three plasmids that encode Broccoli-FLAP in varying positions relative to the promoter (Figure 6A). Whereas fluorescent curve of the Broccoli Proximal was expected to possess steep, close to linear slope straight from the initiation of the assay, fluorescent curve of the Broccoli Distant was expected to possess a lag phase after initiation of the assay. The position of the Broccoli-FLAP relative to the promoter or terminator, in the case of Broccoli Post termination, was expected to be the primary cause to the varying shape of the fluorescent curves monitored between different plasmid constructs.

From the initial fluorescent curves with all template plasmid (Figure 6B), Broccoli Proximal had a steeper slope from the beginning of the measurement whereas the fluorescent signal from Broccoli Distant begun to increase much later compared to Broccoli Proximal. Broccoli Post termination had significantly lower fluorescent signal output compared to the other plasmids and long delay period before the slope began rising. Both Broccoli Distant and Broccoli Proximal had almost linear curve, but the measurement delay with Broccoli Distant made the observations from the early stages of the assay unavailable for any comparisons between these two plasmids.
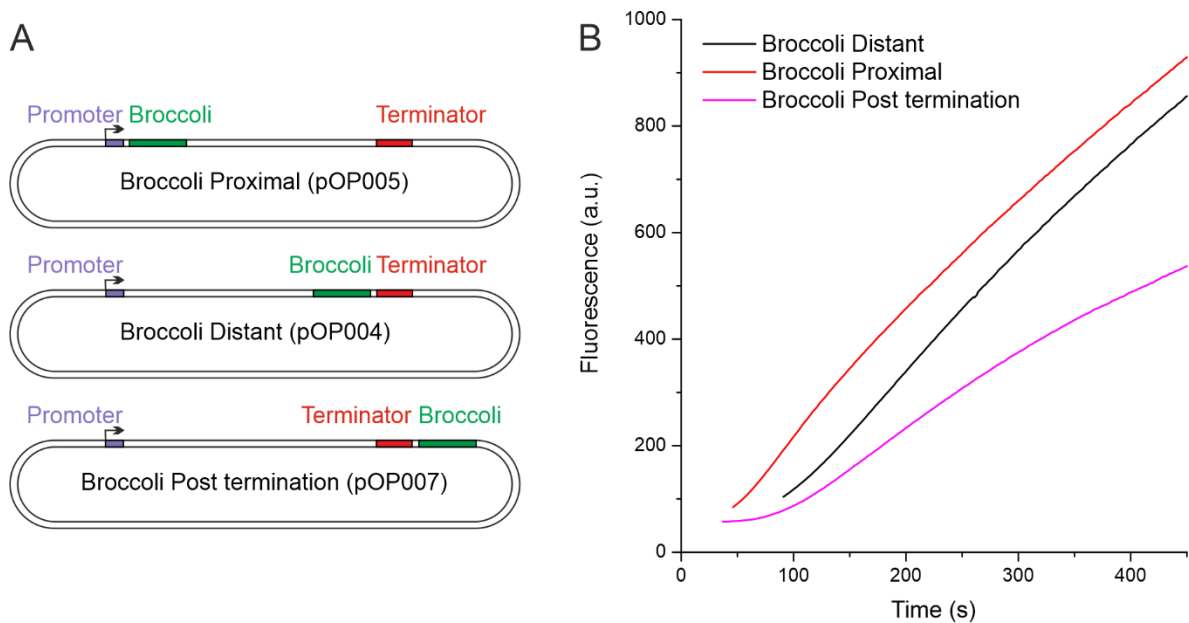
Figure 6 – Initial FLAP assays with all plasmid constructs. A) Plasmid template schematics of all plasmid constructs. B) Effects of the varying positions of the sequence encoding Broccoli-FLAP were discovered in the initial assays using the different plasmid constructs. [rNTPs] = 1 µM, transcription buffer (40 mM HEPES-KOH pH 7.5, 80 mM KCl, 10 mM MgCl$_2$, 0.1 mM Na-EDTA, 0.1 mM DTT, 5 % glycerol), cuvette (16.160F-Q-10/Z15 (Starna GmbH)).

8.2 Validation of the fluorescent data from the assays

Following the initial assay, we wanted to validate our data from the fluorescence measurements by running two agarose gels with different stains (Figure 7). Both gels have a reaction without rNTPs as a negative control and other the reaction was done regularly with rNTPs present. In both gels, there were two bands in the reaction with rNTPs present compared to the negative control where there is only one band. The first band on both gels was the plasmid DNA template that was observed from both reactions. The second band, observed only in the reactions with NTPs, was the RNA transcript.

33
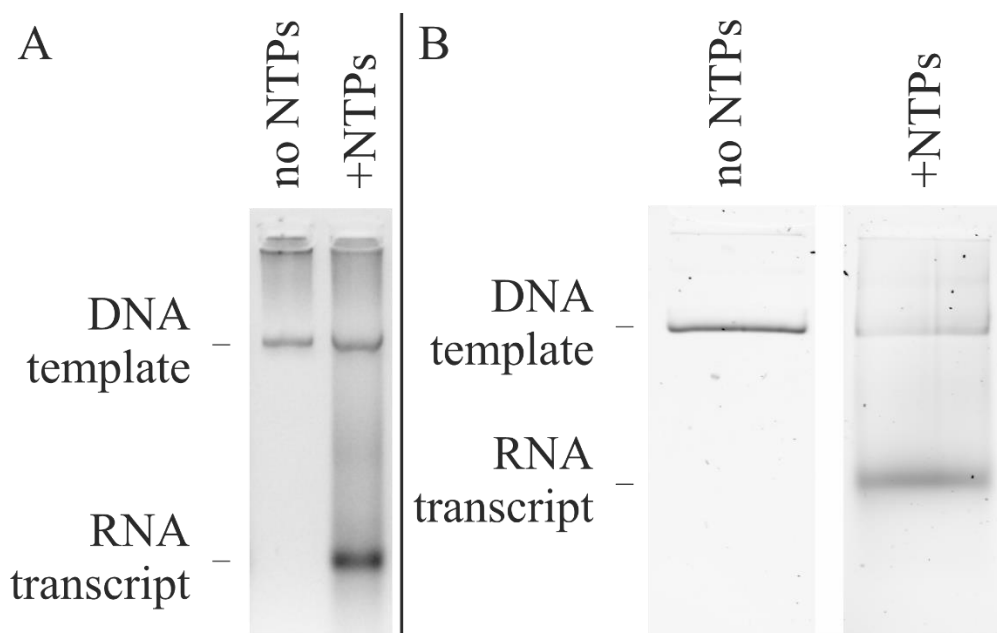
Figure 7 – Agarose gel electrophoresis run of FLAP assays. Samples of FLAP assays with and without rNTPs ran to validate the fluorescent data. Reactions without rNTPs were used as controls. A) Gel stained with EtBr. B) Gel stained with DFHBI-1T.

8.3 Optimization of the assay with *E. coli* RNA polymerase

Varying concentrations were tested during optimizations of the assay to assess the optimal reactant usage for the assay. Final holoenzyme, plasmid and rNTP concentrations were decreased to half (1 μM holoenzyme, 24 nM plasmid, 0.5 mM NTP) one by one from the initial assays. From these optimizations (Figure 8), holoenzyme decrease had hardly an effect on the fluorescent signal while compared to the control reaction, the slope with lower holoenzyme concentration appeared to be less linear as the reaction advanced. Decreasing plasmid concentration had the most significant negative effect on the signal output and decreasing rNTPs concentration had a negative effect on the signal compared to the control. White precipitate in the cuvette was noticed during optimizations and it is likely to be consequence of magnesium pyrophosphate forming as a byproduct from transcription (Akama et al. 2012). 0.1 μM of pyrophosphatase was added to all reactions to counter the formation of precipitate. Since decreasing holoenzyme concentration had hardly any effect on the fluorescent signal output, subsequent assays were done with 1 μM holoenzyme, 48 nM plasmid, and 1 mM rNTP.
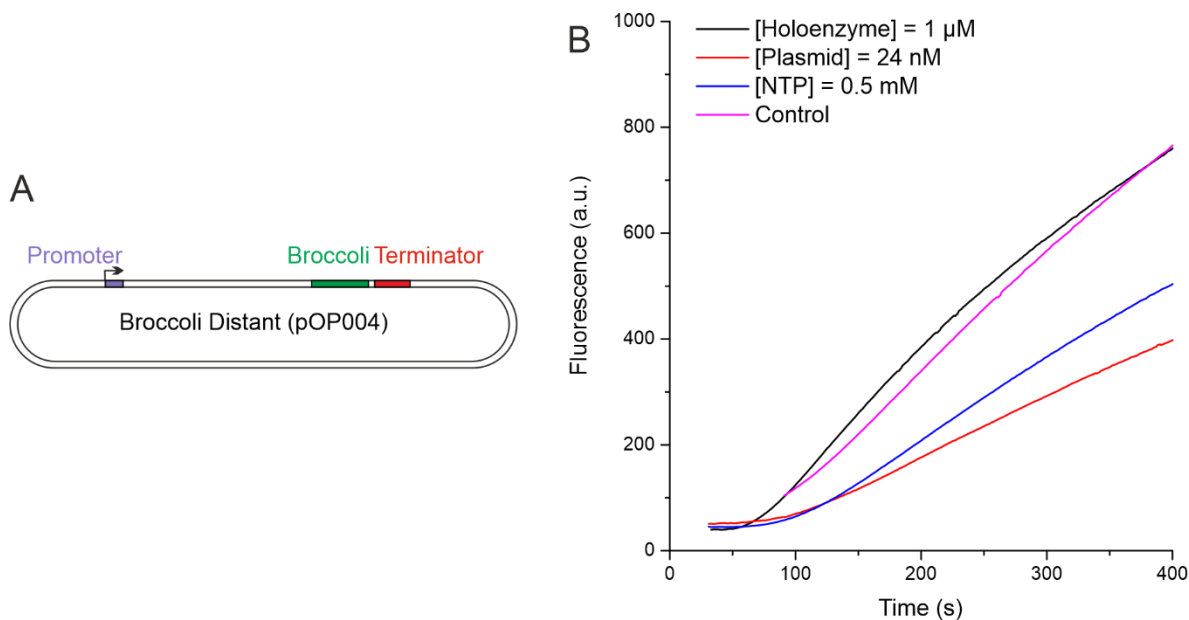
Figure 8 – Optimization assays with *E. coli* RNAP. A) Broccoli Distant plasmid template. B) Effect of reducing the reactant concentrations to half from the initial assays. Control reaction: [Holoenzyme] = 2 µM, [Plasmid] = 48 nM, [DFHBI-1T] = 10 µM, [rNTPs] = 1 mM.

8.4 Optimization of the assay with *S. africana* RNA polymerase

Since *S. africana*'s physiological environment differs from the conditions of *E. coli*, the conditions for assay had to be optimized for *S. africana* RNAP. One of the most notable features observed in these optimizations was the difference in the shape of the curve between *E. coli* RNAP (Figure 6, red curve) and *S. africana* RNAPs (Figure 9, green curve) with Broccoli Proximal plasmid. With *E. coli* RNAP the curve was close to linear the whole measurement time, but with *S. africana* RNAP the curve had a very steep almost linear slope at the start which then declined towards horizontal linear section of the curve. This effect was mostly distinguishable with the highest fluorescence output conditions, but it appeared with the second highest fluorescence output conditions as well.

Varying Mg^{2+} and rNTP concentrations and pH were experimented with (Figure 9). Difference in signal output between pH of 8.0 and 9.0 at 5 mM Mg^{2+} and 1 mM rNTPs was very minor. In contrast, increasing $[Mg^{2+}]$ from 5 to 10 mM dramatically (~3-fold) increased the transcription output during the first 200 s. Similarly, increasing [rNTPs] from 0.5 mM to 1 mM approximately doubled the signal output during the first 200 s. However, pH, $[Mg^{2+}]$ and [NTPs] had little effect on the slope of the fluorescence curves in the time interval 300-

600 s. Final conditions for the assays with *S. africana* transcription factors were chosen to be pH of 9, 10 mM Mg^{2+} and 1 mM rNTPs concentrations.
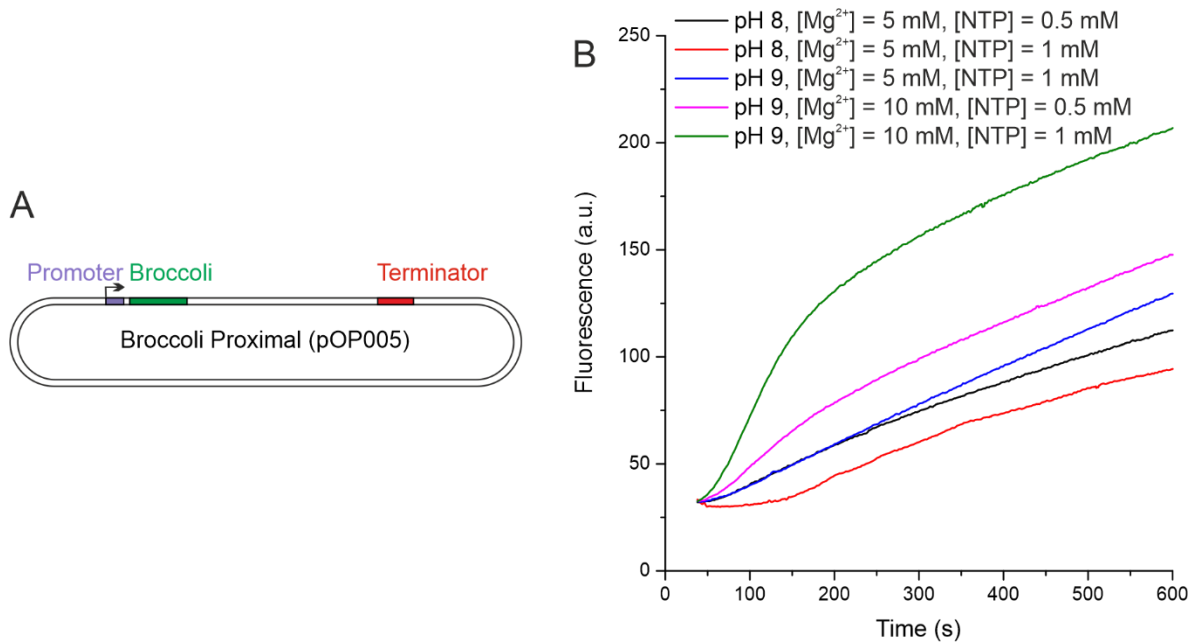
Figure 9 – Optimization assays with *S. africana* RNAP. A) Broccoli Proximal plasmid template. B) Effect of pH change and various Mg^{2+} and rNTP concentrations. Transcription buffer (40 mM TAPS-KOH pH 8.0/9.0, 80 mM KCl, 10 mM $MgCl_2$, 0.1 mM Na-EDTA, 0.1 mM DTT, 5 % glycerol), cuvette (16.160F-Q-10/Z15 (Starna GmbH)).

9 Investigating the RNAP-associated transcription factors of *S. africana* with the optimized fluorescent light-up aptamer assay

FLAP assays have not been widely used for *in vitro* transcription factor studies. We wanted to experiment with our optimized assay to identify how it could be used for investigating various stages of transcription and RNAP-associated transcription factors, some of which influence transcription at multiple stages. We investigated various transcription factors of which functions have been reported in other organisms and we wanted to identify what FLAP assays could reveal about these transcription factors of *S. africana*. RNAP-associated transcription factors can roughly be divided into three groups based on their main function in an individual stage of transcription: initiation, elongation and termination factors.

9.1 Initiation factors

FLAP assays with initiation factors were done using Broccoli Proximal plasmid since it has the Broccoli-FLAP positioned immediately after the promoter and we hypothesize that it would reflect initiation better than the other plasmid constructs. Following transcription initiation factors were investigated: CarD, DksA, and DksA with ppGpp.

CarD decreased the signal output ~1.5 fold during the first 200 s but had little effect on the slope of the signal increase after 300 s (Figure 10B). Both control and CarD, had upward slope right after the initiation of the assay, but the curve turned more linear after approximately 150 seconds. The purification batch 2 of *S. africana* RNAP was considerably less active than batch 1, but effects of CarD were consistent between the different batches of *S. africana*.

Addition of DksA did not alter the fluorescence output comapring to the control reaction beyond the margins of experimental uncertainty (Figure 10C). Addition of DksA together with ppGpp sligtly reduced the fluorescence output (Figure 10D). The effect was observed with two batches of RNAP, but the marked difference in trancriptional acivity between these batches of RNAP complicated the assesment of the significance of the observed effect. That said, it is certain that the effect is small if at all present.
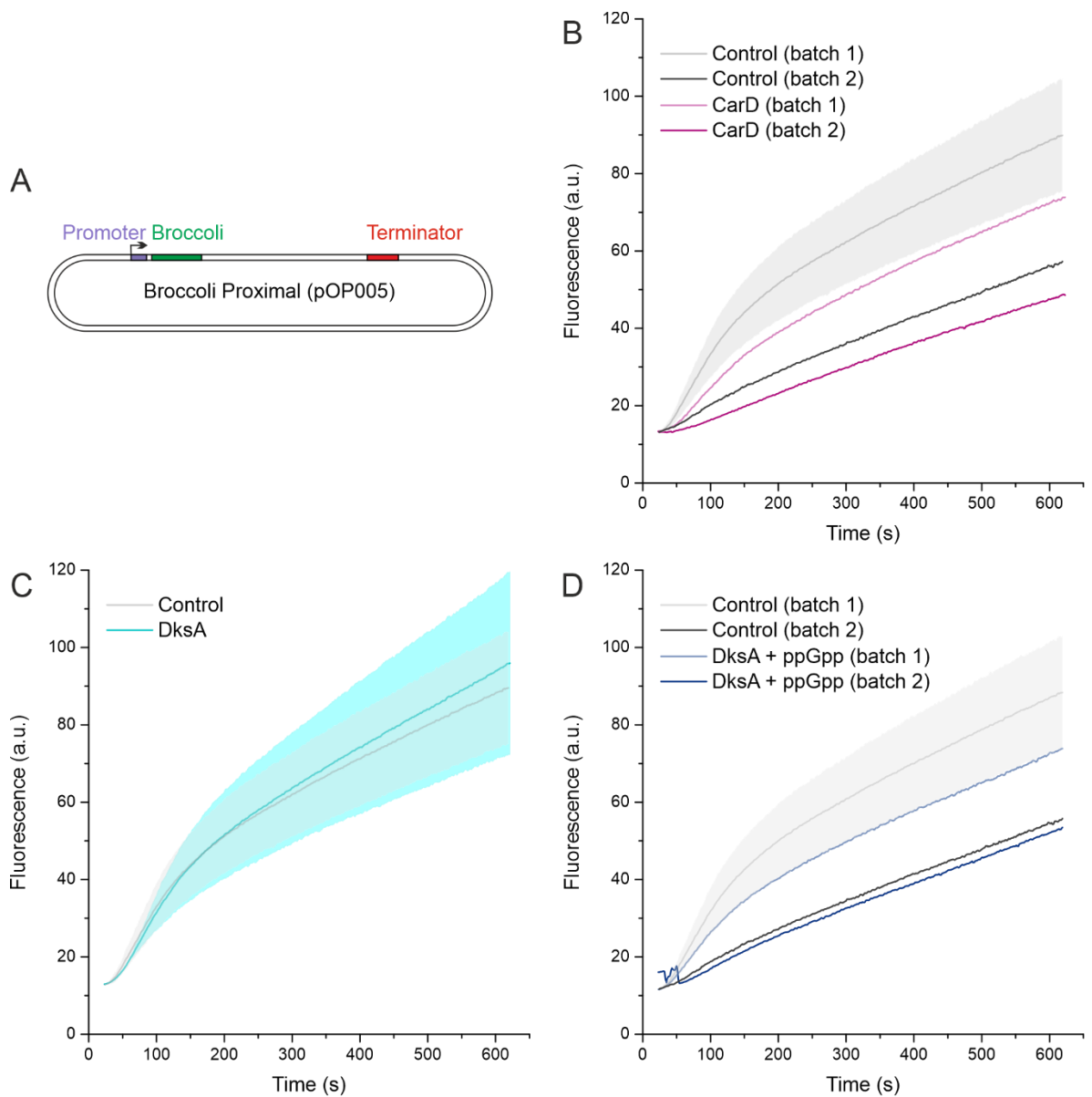
Figure 10 – Fluorescent data of various initiation factors. A) Broccoli Proximal template plasmid B) Effect of CarD. C) Effect of DksA. D) Effect of DksA in the presence of ppGpp. Batch numbers reference to purification batches of *S. africana* RNAP. Standard deviation was calculated for each duplicate and is indicated as the lighter colored area in the corresponding curve. In the case of CarD and DksA with ppGpp, experiments could not be done as duplicates with the same RNAP batch. Cuvette (3x3 mm, QS High Precision Cell, Article number: 105-251-15-40 (Hellma Analytics), [CarD] = 10 µM, [DksA] = 10 µM, [ppGpp] = 10 µM.

9.2 Elongation and termination factors

FLAP assays with elongation factors were done using Broccoli Distant plasmid since it has the Broccoli-FLAP positioned just prior to the terminator and it was constructed to reflect elongation. Following transcription factors were investigated: NusA, NusG, GreA, TPR, GreA and TPR, LoaP, EF-Tu, and Tex.

NusG and its paralog LoaP exhibited distinctively different effects on the kinetics of the increase in fluorescent signal. While both factors increased the duration of the lag phase in the beginning of the assays, LoaP had gentler, more robust and linear slope after the lag phase (Figure 11C) while NusG had delayed, more similar signal output relative to the control after the lag phase (Figure 11B).
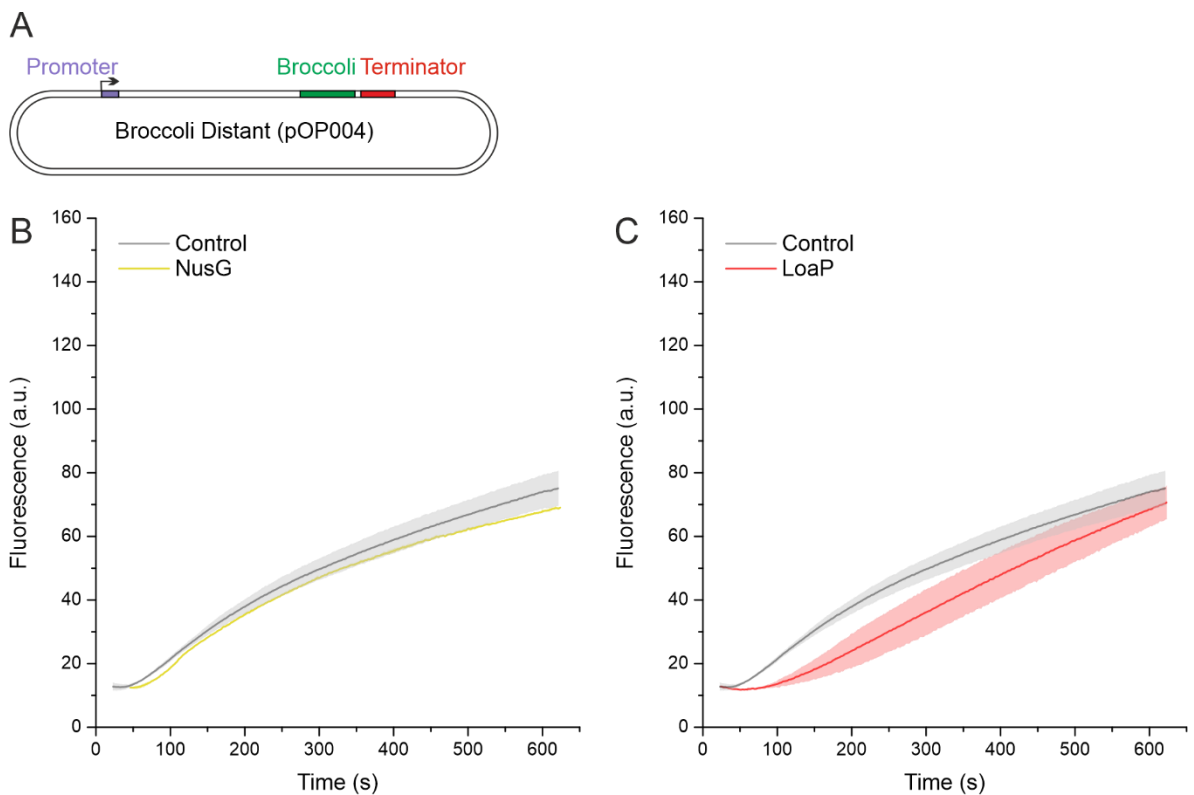
Figure 11 – Fluorescence data from FLAP assays with NusG and LoaP. A) Broccoli Distant template plasmid B) Effect of NusG. C) Effect of LoaP. Standard deviation was calculated for each duplicate and is indicated as the lighter colored area in the corresponding curve. Cuvette (3x3 mm, QS High Precision Cell, Article number: 105-251-15-40 (Hellma Analytics), [NusG] = 10 µM, [LoaP] = 10 µM.

NusA and Tex exhibited negative fluorescent signal output relative to the control while EF-Tu had higher signal output after the lag phase relative to the control. NusA, Tex and EF-Tu increased the duration of lag phase in the beginning of the assays. NusA decreased the overall signal output 2-fold relative to the control (Figure 12B). Tex delayed the signal output relative to the control due to a longer lag phase (Figure 12C). EF-Tu had steeper slope than the control after the lag phase, surpassing the signal output of the control 175 seconds after the initiation of the assay (Figure 12D).
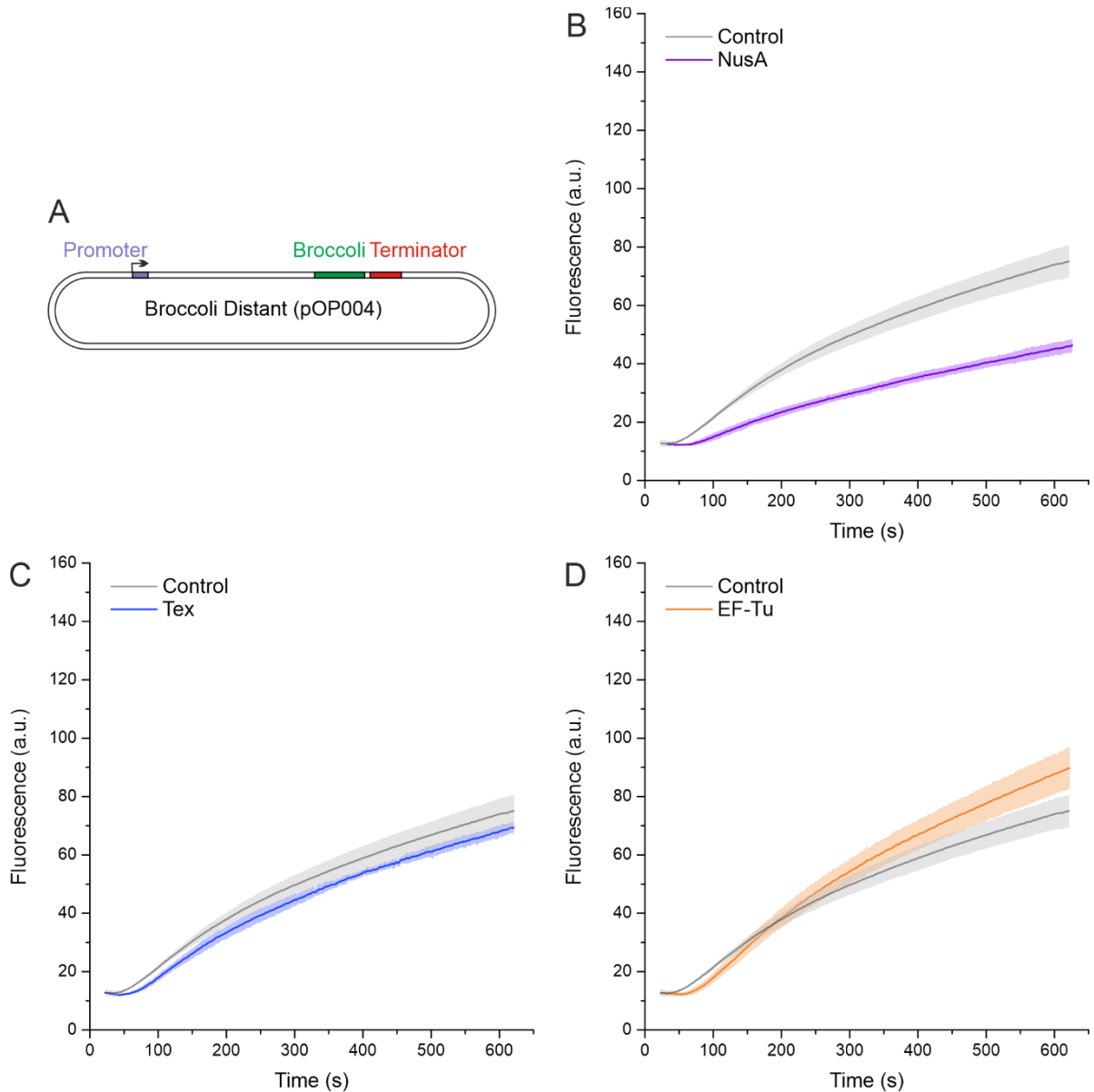
Figure 12 – Fluorescence data from FLAP assays with NusA and Tex. A) Broccoli Distant template plasmid B) Effect of NusA. C) Effect of Tex. D) Effect of EF-Tu. Standard deviation was calculated for each duplicate and is indicated as the lighter colored area in the

corresponding curve. Cuvette (3x3 mm, QS High Precision Cell, Article number: 105-251-15-40 (Hellma Analytics), [NusA] = 10 µM, [Tex] = 10 µM, [EF-Tu] = 6.8 µM.

TPR, GreA, and GreA in the presence of TPR had overall higher fluorescence signal output relative to the control. Duration of the lag phase with TPR (Figure 13B) and with GreA in the presence of TPR (Figure 13D) were relatively similar to the control. While GreA independently had slightly longer lag phase relative to the control, it increased the overall signal output ~2-fold relative to the control (Figure 13C). GreA had very steep, near linear slope after the lag phase. GreA in the presence of TPR had increased average signal output relative to the control, but measurements had quite substantial standard deviation.

Figure 13 – Fluorescence data from FLAP assays with TPR, GreA, and GreA with TPR.
A) Broccoli Distant plasmid template. B) Effect of TPR. C) Effect of GreA. D) Effect of Gre
in the presence of TPR. Standard deviation was calculated for each duplicate and is indicated
as the lighter colored area in the corresponding curve. Cuvette (3x3 mm, QS High Precision
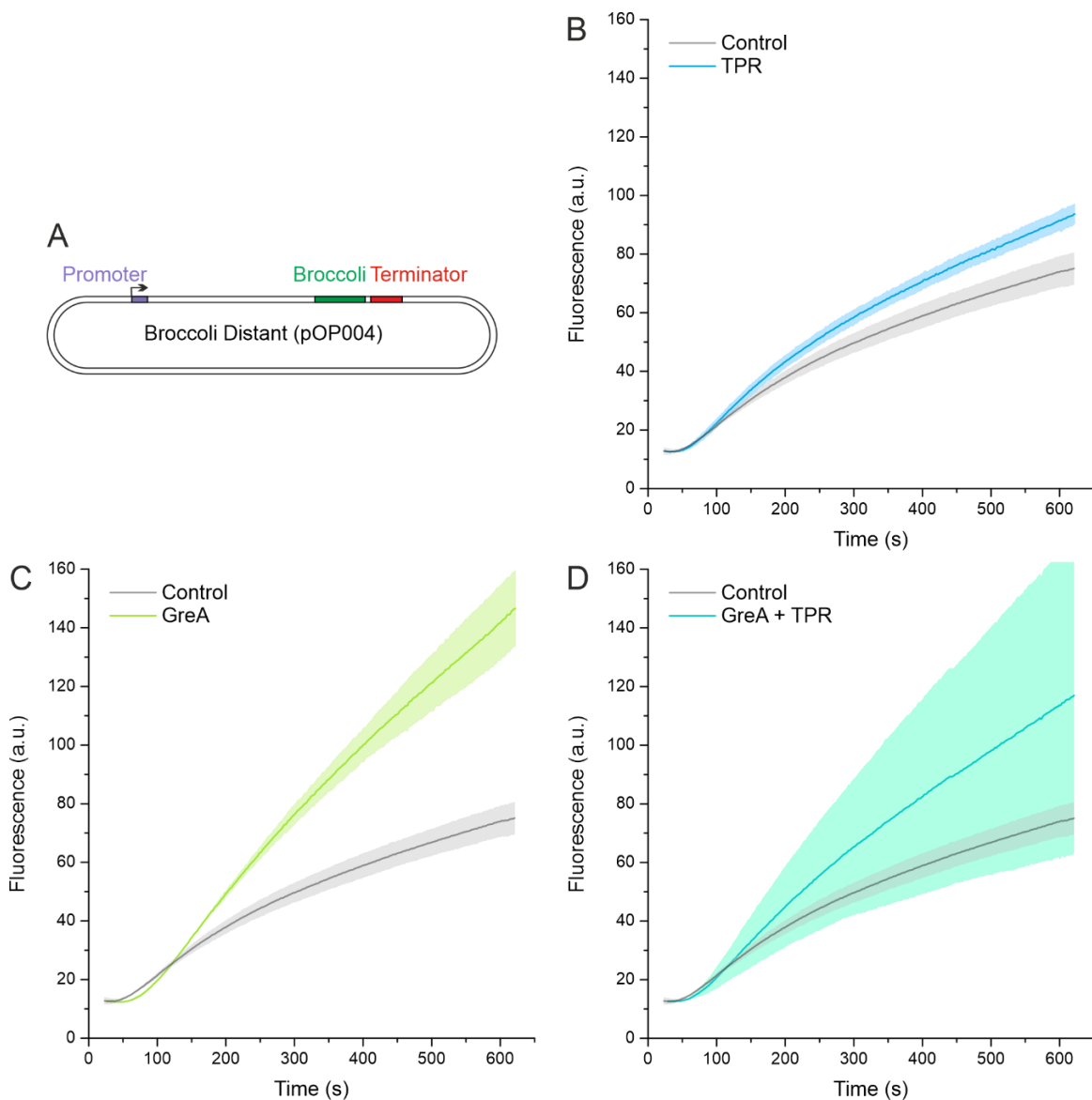Cell, Article number: 105-251-15-40 (Hellma Analytics), [TPR] = 10 μM, [GreA] = 6 μM,
[GreA] = 3.6 μM in the presence of TPR.

Termination factor, Rho, was studied with Broccoli Post termination plasmid which had the
Broccoli-FLAP positioned after the terminator (Figure 14B). Consequently, these assays had
very low baseline fluorescent signal compared to the assays with the other plasmids. The
addition of Rho significantly decreased the signal output compared to the control and

possibly increased the duration of the lag phase before the signal started increasing. These effects are consistent with Rho terminating transcription prematurely before the Broccoli aptamer is synthesized.
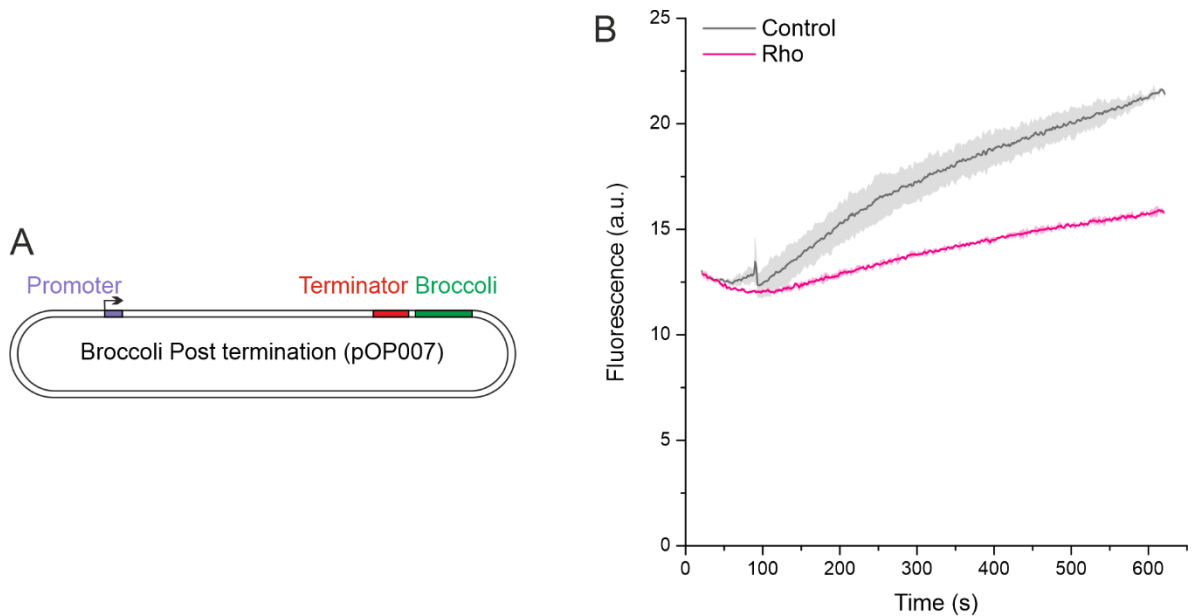
Figure 14 – Fluorescence data from FLAP assays with Rho. A) Broccoli Post termination plasmid template. B) Effect of Rho. Standard deviation was calculated for each duplicate and is indicated as the lighter colored area in the corresponding curve. Cuvette (3x3 mm, QS High Precision Cell, Article number: 105-251-15-40 (Hellma Analytics), [Rho] = 1.1 µM.

9.3 Investigating the effect of *S. africana* GreA

GreA of the spirochetes differs in its size from the Gre factors found in other bacteria. *S. africana* GreA has the typical Gre module that has the globular domain and coiled-coil domain that reaches into the active site of RNAP but contrary to the usual Gre proteins, spirochetal GreA has two additional α-helical domains (predicted to adopt tetratricopeptide fold by AlphaFold) connected via a CarD-like domain domain (Figure 15A). Experiments with *S. africana* GreA exhibited notable increase in the fluorescent signal compared to control or other transcription factors. To further delve into the effect of GreA, truncated *S. africana* GreA's were designed: one with the N-terminal α-helical domain deleted called

GreMidEnd (Figure 14B), and other one containing only GreA module called GreEnd (Figure 14C).
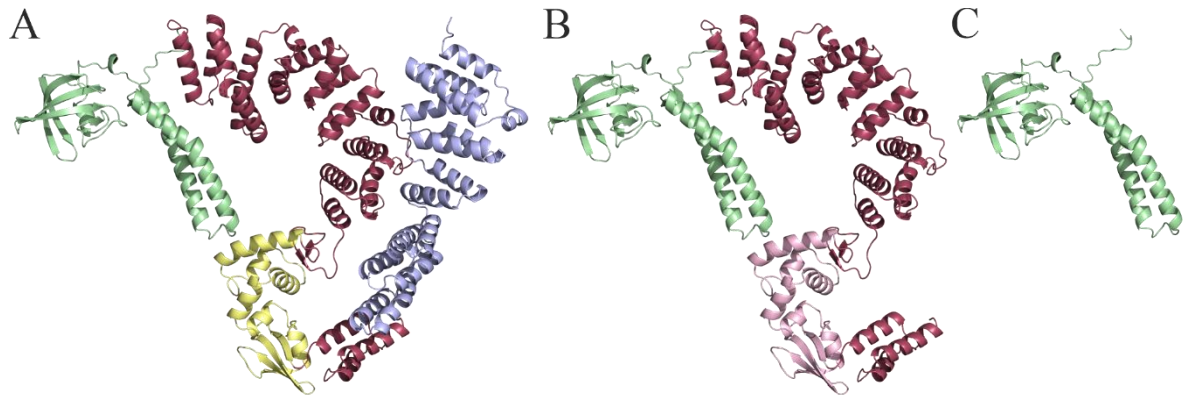
Figure 15 – Structure predictions of different sized *S. africana* GreAs (Jumper et al. 2021, Varadi et al. 2022). A) Full-sized GreA with the additional α-helical domains (light red and light blue). CarD-like domain illustrated in light yellow. B) Illustration of GreMidEnd. CarD-like domain illustrated in light pink. C) Illustration of GreEnd.

GreMidEnd and GreEnd were purified and their effect on transcription was investigated with a FLAP assay using Broccoli Distant construct as a template. The addition of GreEnd to the transcription reaction resulted in fluorescence curves that superimposed well with the control curves obtained in the absence of elongation factors (Figure 16B). GreMidEnd eluted as two well separated peaks in most chromatographies so both peaks were purified and tested individually. Both preparations of GreMidEnds' slightly increased the fluorescence output

(Figure 16C). However, the effects of GreMidEnd preparations were by far smaller than that of the full-sized GreA.

A

Promoter　　　　　　　Broccoli Terminator

Broccoli Distant (pOP004)
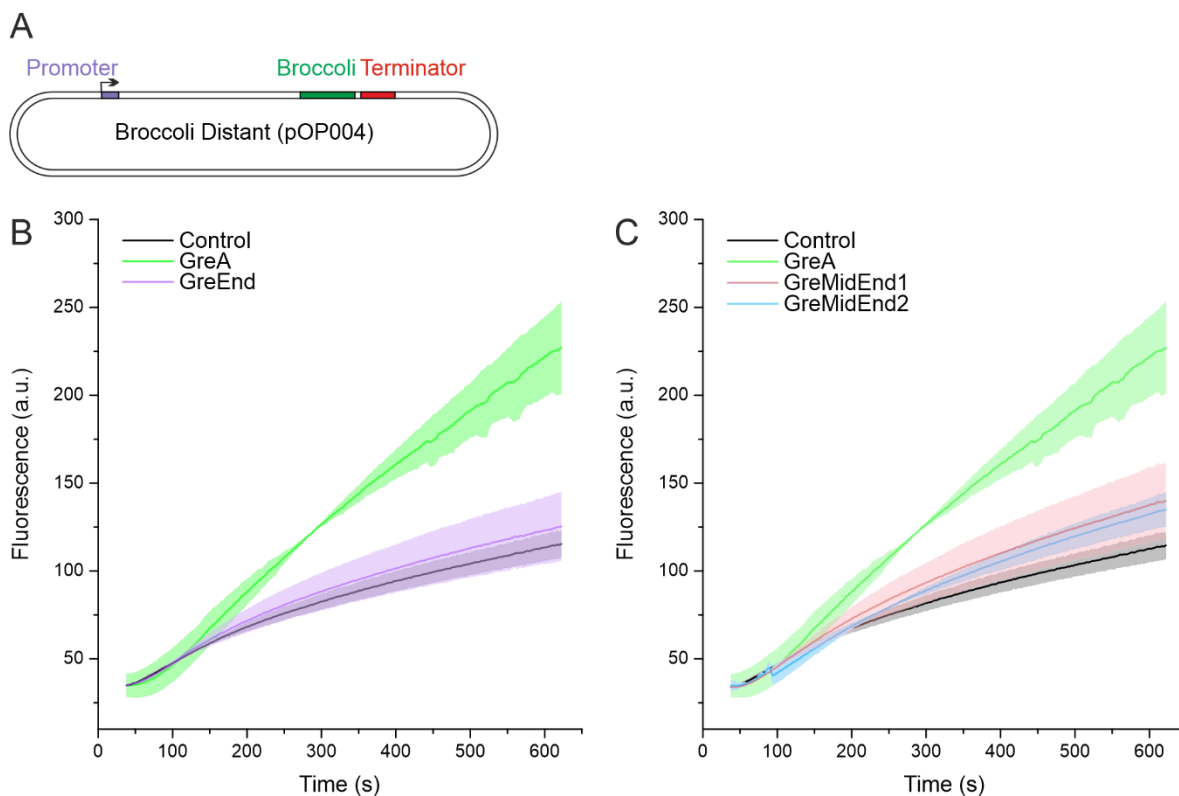
Figure 16 – Effect of the truncated Gre factors. A) Effect of GreEnd compared to the full-sized GreA. B) Effect of GreMidEnd's compared to the full-sized GreA. Standard deviation was calculated for each duplicate and is indicated as the lighter colored area in the corresponding curve. Cuvette (3x3 mm, QS High Precision Cell, Article number: 105-251-15-40 (Hellma Analytics), [GreA] = 6 µM, [GreMidEnd1] = 10 µM, [GreMidEnd2] = 10 µM, [GreEnd] = 10 µM.

The increased fluorescence output in the presence of full-length GreA may result from the increase in transcription output or from the effect of GreA on the folding or fluorescence of the Broccoli aptamer. To directly access the effect of GreA on the transcription output we performed reactions as in the fluorescence assay, stopped the reactions after 10 min with a formamide containing buffer and analyzed the RNA products on the agarose gel containing EtBr. (Figure 17). Two bands occurred in both reactions, one closer to the well and the other one that migrated longer. The band closer to the well was presumably the DNA plasmid and the other one the RNA transcript. Reaction with GreA appeared to have considerably more RNA transcript than the control which suggests that GreA increases the transcription output

45

rather than affect the transcript fluorescence. It also appeared as if there was less DNA template in the reaction with GreA, but when comparing the signal from the wells, it appeared that some of the DNA template in the GreA reaction stuck in the well. Overall, there was a lot of background signal on the whole gel.
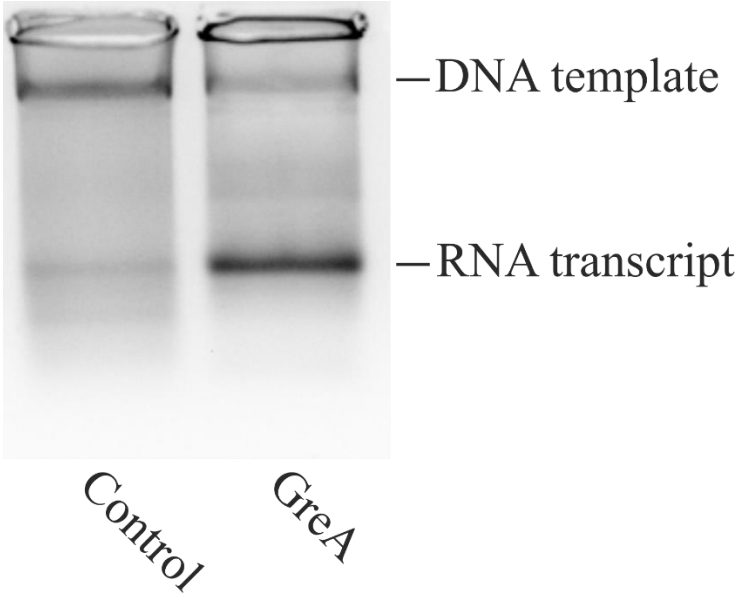
Figure 17 – Agarose gel electrophoresis of transcription products in the absence and the presence of GreA. Gel contained EtBr for band detection.

Discussion

10 Understanding the FLAP assays

10.1 Different plasmid constructs exhibit signature fluorescence curves without additional transcription factors with *E. coli* RNAP.

In the beginning of the FLAP assay, the holoenzyme starts transcribing plasmid DNA and at some point RNAP encounters Broccoli-FLAP encoding sequence and transcribes it. Transcribed Broccoli-FLAP adopts complex 3D fold and can then bind a fluorophore, which when bound begins to emit fluorescence. This cycle then repeats, and the increase of fluorescence can be monitored. Based on the mechanism of the FLAP assays, we can interpret the curves produced by the three plasmid constructs, which had distinctively different fluorescent curves in the assays (Figure 6) with *E. coli* RNAP. Broccoli Proximal had immediate increase of fluorescence after the beginning of the measurement while Broccoli Distant and Post termination had longer delay before the beginning of the fluorescence. The three plasmid constructs are differentiated by the varying position of the sequence encoding Broccoli-FLAP. Consequently, the upwards slope most notably present in the fluorescent curve of Broccoli Proximal reflects the initiation rate of the transcription and the lag phase in the beginning of the assays reflects the folding rate of the Broccoli-FLAP and the elongation rate of transcription could be measured comparing Broccoli Proximal to other plasmids where the sequence encoding Broccoli-FLAP is many hundreds or couple of thousands of base pairs distance from promoter. Broccoli Post termination had significantly lower fluorescence output from the three plasmids, which is caused by the positioning of the sequence coding Broccoli-FLAP after the terminator.

10.2 Fluorescence signal corresponds to the accumulation of RNAs containing Broccoli-FLAP in the assays

Gel analysis of the FLAP assays resulted (Figure 7) in two bands in the reactions where rNTPs were present and one band in the control reaction where rNTPs were not added to the reaction. The upper band represents the plasmid since no transcription occurs in the reaction where rNTPs were not added. Consequently, the lower band represents the transcribed RNA with Broccoli-FLAP since it occurred in the gel, stained with DFHBI-1T as well. By staining the other gel with the DFHBI-1T, we can confirm our hypothesis that the fluorescence signal increase in the assays is caused by the transcription of RNAs containing transcribed Broccoli-

FLAP. However, it appears that plasmid DNA can also bind the fluorophore in some amounts, but this should only result in higher background signal in the FLAP assays.

10.3 Effects of various reactant concentration changes to the FLAP assay

The main goal of the optimizations was to reduce the reactant usage and find optimal conditions for the eventual FLAP assays with *S. africana* RNAP and multiple RNAP-associated transcription factors. Optimization assays with *E. coli* RNAP revealed that reduced plasmid concentration had the highest negative effect to transcription output overall (Figure 8, red curve). Even though multiple RNAP molecules can transcribe the same plasmid, the strong negative effect can be explained with much higher concentration of RNAP holoenzyme in the reaction relative to the plasmid concentration. We also tested reducing the holoenzyme concentration in the assay with *E. coli* RNAP (Figure 8, black curve). Decreased holoenzyme concentration appeared to have near nonexistent effect on the transcription, indicating that decreasing holoenzyme concentration to 20-fold from 40-fold of the plasmid concentration is still sufficient for maximum transcription output with *E. coli* RNAP. Reduced rNTP concentrations with *E. coli* RNAP led to one-third decrease of the fluorescence signal while compared to the control (Figure 8, blue curve). Reducing rNTP concentrations to half from the original appeared to have notable negative effect to the transcription and causing less linear curve presumably due to decreasing rNTP pool while the assay proceeded. In the optimization assays with *S. africana* RNAP, rNTP concentrations were also experimented with and decreasing rNTP concentrations affected negatively to the transcription output similar to the assays with *E. coli* RNAP.

We concluded that NTPs should be kept high at ~1 mM to ensure good fluorescence signal output. The template concentration is the main limiting factor for the signal output in the examined concentration range (24 nM – 48 nM). The signal output scales proportionally to the template concentration and can be adjusted to the desired range by varying the template concentration. At the same time, holoenzyme concentration of 1 μM was saturating in the case of *E. coli* holoenzyme.

With *S. africana* RNAP, first thing was to investigate the effect of pH and [Mg^{2+}] for the transcription. Transcription buffer was modified to have higher pH (9.0) since *S. africana* is from basic Lake Magadi in Africa. From varying pH conditions (8.0-9.0), pH of 9.0 had the highest fluorescent output with 1 mM [Mg^{2+}]. Most of the reaction conditions were kept

similar to what was used with *E. coli* RNAP. Because *S. africana* RNAP had lower signal output than its *E. coli* counterpart, it was concluded that 1 mM holoenzyme concentration was the lowest that would be used for the assays with transcription factors. Increasing the amount of *S. africana* RNAP to counter the already lower signal output was not desired due to difficulties to express and purify large amounts of *S. africana* RNAP.

10.4 Pre-assembled holoenzymes-promoter complex might create burst of fluorescent signal in the beginning of the FLAP assays

An interesting effect appeared with *S. africana* RNAP optimizations with Broccoli Proximal compared to similar curve using the same plasmid with *E. coli* RNAP. While Broccoli Proximal with *E. coli* RNAP produced almost linear steep slope for the whole observation time, *S. africana* RNAP had very steep slope for the start which then declined to gentler slope after a while from the start of the measurements. The explanation why this effect could be only observed with *S. africana* RNAP might be that the efficiency of *S. africana* RNAP is worse than that of *E. coli* RNAP. Also, the measurement time with *E. coli* RNAP might have been too short with too high reactant concentrations that we could not yet observe the effect. A more intriguing question is why it occurs.

One possible explanation could be that the initial burst of transcription initiation observed as a steep slope in the fluorescence curve is a consequence of the pre-assembled holoenzyme-promoter complexes that are formed when holoenzyme is mixed with plasmid template DNA ahead of the reaction initiation by addition of the rNTPs. Due to pre-assembled holoenzyme-promoter complex, holoenzyme is ready to initiate transcription when rNTPs are added and this could explain the initial burst of fluorescence signal output monitored in the assays. As all the pre-assembled holoenzyme-promoter complexes have been exhausted in the initial burst, the assembly of new holoenzymes for the next cycles of transcription would explain the decreased initiation rate. However, this theory relies on the fact that the reaction simply runs out of readily assembled holoenzyme-promoter complexes.

Whether or not the same effect is observed with Broccoli Distant is unclear as the overall signal output is steadier throughout the assays, but it appears that there is a steeper slope at the beginning of the assays with Broccoli Distant, but it is much gentler than with Broccoli Proximal. Also, the later linear phase is less linear, which might be result of decreased NTP pool in the reaction or more likely related to the position of the sequence encoding Broccoli-

FLAP. Since the sequence encoding Broccoli-FLAP in the Broccoli Distant plasmid is prior to the terminator, there is a possibility that more transcribing RNAPs would encounter transcriptional pauses or termination before reaching the sequence encoding Broccoli-FLAP and consequently not transcribing Broccoli-FLAPs, leading into declining fluorescent curve.

11 RNAP-associated transcription factors of *S. africana*

11.1 Transcription initiation factors

Initiation factor FLAP assays were done using plasmid that had Gre promoter which drives expression of the TPR-GreA operon in *S. africana* genome. Using Broccoli Proximal plasmid that had Gre promoter, we found that *S. africana* CarD, DksA, and ppGpp with DksA behaved as expected from studies with these transcription factors from other bacteria. In our assay, CarD had a longer lag phase than the control, but transcription initiation rate occurred to be similar with the control. CarD might prolong the FLAP folding as the lag phase is the primary cause of FLAP folding with Broccoli Proximal (Huang et al. 2022). DksA individually occurred to increase the initiation rate of transcription compared to control with gre-promoter. DksA has been found to promote transcription from metabolic promoters which would explain the increased initiation rate. DksA with ppGpp on the other hand decreased the initiation rate when compared to DksA individually. However, all aforementioned initiation factors have been reported to have varying activity depending on the promoter sequence. In the future, transcription initiation FLAP assays using also a plasmid with a rRNA promoter would provide more complete picture of how these initiation factors work since they have been reported to behave differently based on the promoter type.

11.2 Transcription elongation factors

11.2.1 LoaP slows down transcription elongation but promotes transcription in long-term

LoaP is a NusG paralog that has been reported to facilitate anti-termination with intrinsic termination RNA hairpins and to promote transcription of long templates like antibiotic synthesis gene clusters (Goodson et al. 2017, Elghondakly et al. 2021). In our FLAP assays, LoaP (Figure 11C) had long lag phase in the beginning of the assay before the fluorescent signal started to rise robustly, which indicates that LoaP prevents the decline of transcription rate with time observed in the absence of LoaP or in the presence of its paralogue NusG. We

do not yet know for sure why transcription rate decreases with time. One possible explanation is that the transcription template is gradually blocked by stalled or paused RNAPs as multiround transcription proceeds. If so, LoaP can prevent RNAP pausing and stalling thereby ensuring the linear increase of the fluorescence signal over the entire measurement interval of 10 minutes. Such an explanation is consistent with the expected antipausing activity of LoaP, but the effect needs to be further investigated using templates of various lengths and containing regulatory pause sequences. If our explanation is correct, LoaP is expected to have a greater effect on longer templates and on the template containing regulatory pause sequences.

The slope after the lag phase reflects into initiation rate, but due to LoaP reported to have anti-termination properties, one option is that the TECs, which have LoaP associated, transcribe template plasmids more than once creating robust increase in the monitored fluorescence signal. More reasonable explanation for the robust fluorescent signal is that LoaP, which has been reported to have similar association mechanism with RNAP than NusG, facilitates promoter escape similarly to NusG (Wang and Artsimovitch 2021) and consequently increases initiation rate that we can monitor with our assay. The same effect could not be observed with NusA. This would imply that LoaP is either more efficient facilitating transcription initiation than NusA and/or supports the possibility that LoaP promotes transcription over intrinsic RNA hairpin terminators due to its its anti-termination properties. Also, LoaP could be investigated with a plasmid construct with the sequence encoding FLAP positioned after the intrinsic terminator to evaluate if LoaP has antiterminator activity.

11.2.2 Full-size spirochetal GreA promotes transcription considerably

Spirochetes have unique Gre factors in that they are massive when compared to Gre factors from other bacteria. The usual bacterial Gre factors consist of two domains. These are a domain with two parallel α-helices that enters through RNAP's secondary channel to reach into the active site and another domain that consists of one β-barrel and α-helix is involved in promoting the association with RNAP. However, spirochetes have other domains linked to the N-terminus of the usual Gre factor core. Especially, *S. africana* GreA has two extra domains consisting mostly of α-helices separated by β-barrel, which adds together creating massive, 106 kDa GreA. When screening different factors with FLAP assay, GreA had

considerably higher transcription output when compared to other factors that were tested with the assay. Although GreA appeared to prolong the lag phase duration, it increased the overall transcription rate immensely. The increased transcription output in FLAP assays was confirmed with gel electrophoresis. Gel pictures confirmed that GreA increases the accumulation of RNAs considerably compared to the control reaction.

After the initial assays with GreA, we wanted to discover if truncated *S. africana* GreA's (Figure 16) would have the same effect as the full-sized Gre factor. FLAP assays with the truncated GreA factors produced only minor increases in fluorescent signal while compared to the control, signifying the importance of the additional domains.

S. africana GreA have not been the focus of intense research, consequently the assumptions of the function of GreA are reliant on the information gathered from other bacteria. An important function of GreA in other bacteria is the transcription anti-pause activity. GreA facilitates the cleavage of the nascent RNA in the active site of the RNAP in backtracked RNAPs (Nickels and Hochschild 2004). This allows transcription to continue by rescuing the backtracked RNAPs. However, as there has been no reports of *S. africana* GreA cleavage activity, we can only assume that it would possess this function and results from a cleavage assay would give more insight into such function with *S. africana* GreA. FLAP assays performed in this study demonstrate that only full-sized GreA increases transcription output considerably which signifies the importance of the additional domains that *S. africana* GreA has and suggests that the cleavage function alone is not sufficient for the observed effect. The full-sized *S. africana* GreA has CarD-like domain that is also present in the GreMidEnd. CarD has been generally considered to be an activator of transcription, but it can sometimes also act as an inhibitor. It can be speculated that it is this CarD-like domain that would activate the notable increase in the signal output with the full-sized GreA. However, as said GreMidEnd also has the CarD-like domain but does not activate the transcription. It occurs that combination of several domains is important to have the same effect than with full-sized GreA and the transcription activating nature cannot be attributed to any of the individual domains present in *S. africana* Gre factors. The function of the other additional domains is unknown. We speculate that the additional domains could associate with other parts of the RNAP and consequently facilitate the observed increase in transcription output. Future research with *S. africana* GreA and new Broccoli-FLAP templates are required for more complete understanding of the function of this unusual Gre factor.

Conclusions

Transcription is essential for all living organisms. It is the first step gene expression and consequently tightly regulated. Transcription regulation can be achieved with DNA binding transcription factors, activators and repressors, or in some cases with transcription factors that associate with RNAP and facilitate its function through specific interactions. RNAP-associated transcription factors can facilitate various stages of transcription. In this study, we developed a straightforward assay for studying transcription factors in context of the whole transcription cycle. Our assay is based on the novel discovery of FLAPs especially FLAP called Broccoli. Transcription of the Broccoli-FLAP allows monitoring of the transcription in real-time by measuring the fluorescence of Broccoli-FLAP -fluorophore complex.

In our study, sequence encoding Broccoli-FLAP was positioned in varied distances from strong bacterial promoter in a plasmid. Three different plasmids were constructed, and FLAP assays were optimized for *E. coli* and *S. africana* RNAPs. Initial assays were done with all three plasmid constructs and revealed the characteristics of each plasmid construct. Fluorescent curves of the FLAP assays can be interpreted to have following phases: lag phase in the beginning of the assay and the near linear phase after the lag phase. Lag phase can be used to measure the elongation rate of transcription and the linear phase appeared to reflect the initiation rate of transcription. Optimization assays provided the information that the [NTPs] should be in excess amounts at ~1 mM to ensure good signal output. The main limiting factor for the signal output was the DNA template concentration. Holoenzyme concentration was decided to be kept at 1 µM with *S. africana* holoenzyme due to decent signal output and difficulties of expressing and purifying *S. africana* RNAP. With *S. africana* RNAP, we also increased the transcription buffer pH to 9.0 due to physiological pH of *S. africana*.

RNAP-associated transcription factors of *S. africana* have not been studied before and functions of these factors in *S. africana* are unknown. In our study, these factors were investigated using optimized FLAP assay to understand how FLAP assays work and how they can be applied into studying transcription factors that facilitate transcription on various stages. Most of the studied transcription factors expressed similar functions that would be expected of them based on findings from other bacteria. However, LoaP and GreA had fascinating effects, most notably GreA, which was studied more extensively. LoaP had a

transcription reducing effect in short-term but increased transcription output robustly in long-term. GreA increased transcription output considerably and this kind of effect was not observed with any other transcription factors used in the study. Truncated Gre factors of *S. africana* were experimented with to explain the unique effect of full-sized GreA, but the truncated Gre factors did not express the same effect. Both LoaP and GreA, would require more research in the future with different plasmid templates to have a more complete picture of their function.

FLAP assays provide a novel, straightforward way to monitor transcription in real-time. Multi-round assay allows monitoring of the various stages of transcription, which is highly advantageous when studying RNAP-associated transcription factors transcription factors. In the future, more complete understanding of the RNAP-associated transcription factors of *S. africana* through FLAP assays could give insight into transcription in pathogenic spirochetes that cause diseases such as leptospirosis, Lyme disease, syphilis and intestinal spirochaetosis.

References

Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. & Block, S. M. (2005) Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**:460–465.

Akama, S., Yamamura, M. & Kigawa, T. (2012) A Multiphysics Model of In Vitro Transcription Coupling Enzymatic Reaction and Precipitation Formation. *Biophysical Journal* **102**:221–230.

Bar-Nahum, G., Epshtein, V., Ruckenstein, A. E., Rafikov, R., Mustaev, A. & Nudler, E. (2005) A Ratchet Mechanism of Transcription Elongation and Its Control. *Cell* **120**:183–193.

Basu, R. S., Warner, B. A., Molodtsov, V., Pupov, D., Esyunina, D., Fernández-Tornero, C., Kulbachinskiy, A. & Murakami, K. S. (2014) Structural Basis of Transcription Initiation by Bacterial RNA Polymerase Holoenzyme. *Journal of Biological Chemistry* **289**:24549–24559.

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

Boyaci, H., Chen, J., Jansen, R., Darst, S. A. & Campbell, E. A. (2019) Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding. *Nature* **565**:382–385.

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

Burgess, R. R., Travers, A. A., Dunn, J. J. & Bautz, E. K. F. (1969) Factor Stimulating Transcription by RNA Polymerase. *Nature* **221**:43–46.

Callaci, S., Heyduk, E. & Heyduk, T. (1998) Conformational Changes of Escherichia coli RNA Polymerase ς70 Factor Induced by Binding to the Core Enzyme. *Journal of Biological Chemistry* **273**:32995–33001.

Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L. & Darst, S. A. (2002) Structure of the Bacterial RNA Polymerase Promoter Specificity σ Subunit. *Molecular Cell* **9**:527–539.

Chakraborty, A., Wang, D., Ebright, Y. W., Korlann, Y., Kortkhonjia, E., Kim, T., Chowdhury, S., Wigneshweraraj, S., Irschik, H., Jansen, R., Nixon, B. T., Knight, J., Weiss, S. & Ebright, R. H. (2012) Opening and Closing of the Bacterial RNA Polymerase Clamp. *Science* **337**:591–595.

Chambon, P. (1975) Eukaryotic Nuclear RNA Polymerases. *Annu Rev Biochem* **44**:613–638.

Chen, J., Chiu, C., Gopalkrishnan, S., Chen, A. Y., Olinares, P. D. B., Saecker, R. M., Winkelman, J. T., Maloney, M. F., Chait, B. T., Ross, W., Gourse, R. L., Campbell, E. A. & Darst, S. A. (2020) Stepwise Promoter Melting by Bacterial RNA Polymerase. *Molecular Cell* **78**:275-288.e6.

Chudakov, D. M., Matz, M. V., Lukyanov, S. & Lukyanov, K. A. (2010) Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues. *Physiological Reviews* **90**:1103–1163.

Cook, V. M. (2007) Strand Opening-deficient Escherichia coli RNA Polymerase Facilitates Investigation of Closed Complexes with Promoter DNA. *Journal of Biological Chemistry* **282**:21319–21326.

Cramer, P. (2002) Multisubunit RNA polymerases. *Current Opinion in Structural Biology* **12**:89–97.

Cramer, P., Bushnell, D. A. & Kornberg, R. D. (2001) Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. *Science* **292**:1863–1876.

Crick, F. (1970) Central Dogma of Molecular Biology. *Nature* **227**:561–562.

Daubendiek, S. L. & Kool, E. T. (1997) Generation of catalytic RNAs by rolling transcription of synthetic DNA nanocircles. *Nat Biotechnol* **15**:273–277.

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 USA* **104**:7833–7838.

Dombroski, A., Walter, W. & Gross, C. (1993) Amino-terminal amino acids modulate σ-factor DNA-binding activity. *Genes & Development* **7**:2446–2455.

Ebright, R. H. (2000) RNA Polymerase: Structural Similarities Between Bacterial RNA Polymerase and Eukaryotic RNA Polymerase II. *Journal of Molecular Biology* **304**:687–698.

Elghondakly, A., Wu, C. H., Klupt, S., Goodson, J. & Winkler, W. C. (2021) A NusG Specialized Paralog That Exhibits Specific, High-Affinity RNA-Binding Activity. *Journal of Molecular Biology* **433**:167100.

Epshtein, V., Cardinale, C. J., Ruckenstein, A. E., Borukhov, S. & Nudler, E. (2007) An Allosteric Path to Transcription Termination. *Molecular Cell* **28**:991–1001.

Epshtein, V., Dutta, D., Wade, J. & Nudler, E. (2010) An allosteric mechanism of Rho-dependent transcription termination. *Nature* **463**:245–249.

Feklistov, A. & Darst, S. A. (2011) Structural Basis for Promoter −10 Element Recognition by the Bacterial RNA Polymerase σ Subunit. *Cell* **147**:1257–1269.

Fernández-Coll, L., Potrykus, K., Cashel, M. & Balsalobre, C. (2020) Mutational analysis of Escherichia coli GreA protein reveals new functional activity independent of antipause and lethal when overexpressed. *Sci Rep* **10**:16074.

Filonov, G. S., Moon, J. D., Svensen, N. & Jaffrey, S. R. (2014) Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *Journal of the American Chemical Society* **136**:16299–16308.

Fuchs, T. M., Deppisch, H., Scarlato, V. & Gross, R. (1996) A new gene locus of Bordetella pertussis defines a novel family of prokaryotic transcriptional accessory proteins. *J Bacteriol* **178**:4445–4452.

Furth, J. J., Hurwitz, J. & Anders, M. (1962) The Role of Deoxyribonucleic Acid in Ribonucleic Acid Synthesis. *Journal of Biological Chemistry* **237**:2611–2619.

Garner, A. L., Rammohan, J., Huynh, J. P., Onder, L. M., Chen, J., Bae, B., Jensen, D., Weiss, L. A., Manzano, A. R., Darst, S. A., Campbell, E. A., Nickels, B. E., Galburt, E. A. & Stallings, C. L. (2017) Effects of increasing the affinity of CarD for RNA polymerase on Mycobacterium tuberculosis growth, rRNA transcription, and virulence. *Journal of Bacteriology* **199**.

Gentry, D., Xiao, H., Burgess, R. & Cashel, M. (1991) The omega subunit of Escherichia coli K-12 RNA polymerase is not required for stringent RNA control in vivo. *J Bacteriol* **173**:3901–3903.

Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. (2001) Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 Å Resolution. *Science* **292**:1876–1882.

Goodson, J. R., Klupt, S., Zhang, C., Straight, P. & Winkler, W. C. (2017) LoaP is a broadly conserved antiterminator protein that regulates antibiotic gene clusters in Bacillus amyloliquefaciens. *Nature Microbiology* **2**.

Gourse, R. L., Chen, A. Y., Gopalkrishnan, S., Sanchez-Vazquez, P., Myers, A. & Ross, W. (2018) Transcriptional Responses to ppGpp and DksA. *Annu Rev Microbiol* **72**:163–184.

Gralla, J. D., Carpousis, A. J. & Stefano, J. E. (1980) Productive and abortive initiation of transcription in vitro at the lac UV5 promoter. *Biochemistry* **19**:5864–5869.

Gusarov, I. & Nudler, E. (1999) The Mechanism of Intrinsic Transcription Termination. *Molecular Cell* **3**:495–504.

Hao, Z., Epshtein, V., Kim, K., Proshkin, A., Svetlov, V., Kamarthapu, V., Bharati, B., Mironov, A., Walz, T. & Nudler, E. (2021) Pre-termination Transcription Complex: Structure and Function. *Molecular Cell* **81**:281–292.

Harwood, C. S. & Canale-Parola, E. (1984) ECOLOGY OF SPIROCHETES. *Annu Rev Microbiol* **38**:161–192.

Haugen, S. P., Berkmen, M. B., Ross, W., Gaal, T., Ward, C. & Gourse, R. L. (2006) rRNA Promoter Regulation by Nonoptimal Binding of σ Region 1.2: An Additional Recognition Element for RNA Polymerase. *Cell* **125**:1069–1082.

Hawley, D. & McClure, W. (1983) Compilation and analysis of Escherichia coli promoter DNA sequences. *Nucleic Acids Research* **11**:2237–2255.

He, X., Thornton, J., Carmicle-Davis, S. & McDaniel, L. S. (2006) Tex, a putative transcriptional accessory factor, is involved in pathogen fitness in Streptococcus pneumoniae. *Microbial Pathogenesis* **41**:199–206.

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

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. *Journal of Biological Chemistry* **281**:12362–12369.

Huang, Y.-H., Trapp, V., Puro, O., Mäkinen, J., Metsä-Ketelä, M., Wahl, M. & Belogurov, G. (2022) Fluorogenic RNA aptamers to probe transcription initiation and co-transcriptional RNA folding by multi-subunit RNA polymerases. In *Methods in Enzymology*.

Hurwitz, J., Bresler, A. & Diringer, R. (1960) The enzymic incorporation of ribonucleotides into polyribonucleotides and the effect of DNA. *Biochemical and Biophysical Research Communications* **3**:15–19.

Igarashi, K. & Ishihama, A. (1991) Bipartite functional map of the E. coli RNA polymerase α subunit: Involvement of the C-terminal region in transcription activation by cAMP-CRP. *Cell* **65**:1015–1022.

Igarashi, K., Nobuyuki, F. & Ishihama, A. (1991) Identification of a subunit assembly domain in the alpha subunit of Escherichia coli RNA polymerase. *Journal of Molecular Biology* **218**:1–6.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., … Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583–589.

Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonjia, E., Weiss, S. & Ebright, R. H. (2006) Initial Transcription by RNA Polymerase Proceeds Through a DNA-Scrunching Mechanism. *Science* **314**:1144–1147.

Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A. & Darst, S. A. (2000) A Structural Model of Transcription Elongation. *Science* **289**:619–625.

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.

Landick, R. (2006) The regulatory roles and mechanism of transcriptional pausing. *Biochemical Society Transactions* **34**:1062–1066.

Liolos, K., Abt, B., Scheuner, C., Teshima, H., Held, B., Lapidus, A., Nolan, M., Lucas, S., Deshpande, S., Cheng, J.-F., Tapia, R., Goodwin, L. A., Pitluck, S., … Kyrpides, N. C. (2013) Complete genome sequence of the halophilic bacterium Spirochaeta africana type strain (Z-7692T) from the alkaline Lake Magadi in the East African Rift. *Stand Genomic Sci* **8**:165–176.

Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A. N., Niu, W., Ebright, Y. W., Levy, R. & Ebright, R. H. (2002) Structural Organization of

Bacterial RNA Polymerase Holoenzyme and the RNA Polymerase-Promoter Open Complex. *Cell* **108**:599–614.

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

Miropolskaya, N., Esyunina, D. & Kulbachinskiy, A. (2017) Conserved functions of the trigger loop and Gre factors in RNA cleavage by bacterial RNA polymerases. *Journal of Biological Chemistry* **292**:6744–6752.

Mishanina, T. V., Palo, M. Z., Nayak, D., Mooney, R. A. & Landick, R. (2017) Trigger loop of RNA polymerase is a positional, not acid–base, catalyst for both transcription and proofreading. *Proc Natl Acad Sci USA* **114**.

Murakami, K., Fujita, N. & Ishihama, A. (1996) Transcription factor recognition surface on the RNA polymerase alpha subunit is involved in contact with the DNA enhancer element. *The EMBO Journal* **15**:4358–4367.

Murakami, K., Kimura, M., Owens, J. T., Meares, C. F. & Ishihama, A. (1997a) The two α subunits of *Escherichia coli* RNA polymerase are asymmetrically arranged and contact different halves of the DNA upstream element. *Proc Natl Acad Sci USA* **94**:1709–1714.

Murakami, K., Owens, J. T., Belyaeva, T. A., Meares, C. F., Busby, S. J. W. & Ishihama, A. (1997b) Positioning of two alpha subunit carboxy-terminal domains of RNA polymerase at promoters by two transcription factors. *Proc Natl Acad Sci USA* **94**:11274–11278.

Nakamura, S. (2020) Spirochete Flagella and Motility. *Biomolecules* **10**:550.

Nickels, B. E. & Hochschild, A. (2004) Regulation of RNA Polymerase through the Secondary Channel. *Cell* **118**:281–284.

Nudler, E., Mustaev, A., Goldfarb, A. & Lukhtanov, E. (1997) The RNA–DNA Hybrid Maintains the Register of Transcription by Preventing Backtracking of RNA Polymerase. *Cell* **89**:33–41.

Paige, J. S., Wu, K. Y. & Jaffrey, S. R. (2011) RNA Mimics of Green Fluorescent Protein. *Science* **333**:642–646.

Perez-Rueda, E. (2000) The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic Acids Research* **28**:1838–1847.

Ray-Soni, A., Bellecourt, M. J. & Landick, R. (2016) Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu Rev Biochem* **85**:319–347.

Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. (2006) Abortive Initiation and Productive Initiation by RNA Polymerase Involve DNA Scrunching. *Science* **314**:1139–1143.

Richardson, L. V. & Richardson, John. P. (1996) Rho-dependent Termination of Transcription Is Governed Primarily by the Upstream Rho Utilization (rut) Sequences of a Terminator*. *The Journal of Biological Chemistry* **271**:21597–21603.

Roberts, J. W. (1969) Termination Factor for RNA Synthesis. *Nature* **224**:1168–1174.

Rosa, P. A., Tilly, K. & Stewart, P. E. (2005) The burgeoning molecular genetics of the Lyme disease spirochaete. *Nat Rev Microbiol* **3**:129–143.

Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K. & Gourse, R. L. (1993) A Third Recognition Element in Bacterial Promoters: DNA Binding by the α Subunit of RNA Polymerase. *Science* **262**:1407–1413.

Ross, W., Vrentas, C. E., Sanchez-Vazquez, P., Gaal, T. & Gourse, R. L. (2013) The Magic Spot: A ppGpp Binding Site on E. coli RNA Polymerase Responsible for Regulation of Transcription Initiation. *Molecular Cell* **50**:420–429.

Ruff, E. F., Drennan, A. C., Capp, M. W., Poulos, M. A., Artsimovitch, I. & Record, M. T. (2015) E. coli RNA Polymerase Determinants of Open Complex Lifetime and Structure. *Journal of Molecular Biology* **427**:2435–2450.

Said, N., Hilal, T., Sunday, N. D., Khatri, A., Bürger, J., Mielke, T., Belogurov, G. A., Loll, B., Sen, R., Artsimovitch, I. & Wahl, M. C. (2021) Steps toward translocation-independent RNA polymerase inactivation by terminator ATPase ρ. *Science* **371**:1673.

Simpson, R. (1979) The molecular topography of rna polymerase-promoter interaction. *Cell* **18**:277–285.

Skare, J. T. & Garcia, B. L. (2020) Complement Evasion by Lyme Disease Spirochetes. *Trends in Microbiology* **28**:889–899.

Srivastava, D. B., Leon, K., Osmundson, J., Garner, A. L., Weiss, L. A., Westblade, L. F., Glickman, M. S., Landick, R., Darst, S. A., Stallings, C. L. & Campbell, E. A. (2013) Structure and function of CarD, an essential mycobacterial transcription factor. *Proc Natl Acad Sci USA* **110**:12619–12624.

Steitz, T. A. (1998) A mechanism for all polymerases. *Nature* **391**:231–232.

Sutherland, C. & Murakami, K. S. (2018) An Introduction to the Structure and Function of the Catalytic Core Enzyme of Escherichia coli RNA Polymerase. *EcoSal Plus* **8**.

Sweetser, D., Nonet, M. & Young, R. A. (1987) Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc Natl Acad Sci USA* **84**:1192–1196.

Turtola, M. & Belogurov, G. A. (2016) NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. *ELife*.

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Žídek, A., Green, T., Tunyasuvunakool, K., … Velankar, S. (2022) AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* **50**:439–444.

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 Å resolution. *Nature* **417**:712–719.

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

Wang, B. & Artsimovitch, I. (2021) NusG, an Ancient Yet Rapidly Evolving Transcription Factor. *Front Microbiol* **11**:619618.

Watson, J. D. (Ed.) (2014) *Molecular biology of the gene*. (Seventh edition). Boston: Pearson.

Weijland, A., Harmark, K., Cool, R. H., Anborgh, P. H. & Parmeggiani, A. (1992) Elongation factor Tu: A molecular switch in protein biosynthesis. *Molecular Microbiology* **6**:683–688.

Weixlbaumer, A., Leon, K., Landick, R. & Darst, S. A. (2013) Structural Basis of Transcriptional Pausing in Bacteria. *Cell* **152**:431–441.

Wilson, K. & Von Hippel, P. (1995) Transcription termination at intrinsic terminators: The role of the RNA hairpin. *Proc Natl Acad Sci USA* **92**:8793–8797.

Winkelman, J. T., Winkelman, B. T., Boyce, J., Maloney, M. F., Chen, A. Y., Ross, W. & Gourse, R. L. (2015) Crosslink Mapping at Amino Acid-Base Resolution Reveals the Path of Scrunched DNA in Initial Transcribing Complexes. *Molecular Cell* **59**:768–780.

Zalenskaya, K., Lee, J., Gujuluva, C. N., Shin, Y. K., Slutsky, M. & Goldfarb, A. (1990) Recombinant RNA polymerase: Inducible overexpression, purification and assembly of Escherichia coli rpo gene products. *Gene* **89**:7–12.

Zaychikov, E., Denissova, L. & Heumann, H. (1995) Translocation of the Escherichia coli transcription complex observed in the registers 11 to 20: "jumping" of RNA polymerase and

asymmetric expansion and contraction of the "transcription bubble". *Proceedings of the National Academy of Sciences* **92**:1739–1743.

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.

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M. X., Arnold, E. & Ebright, R. H. (2012) Structural Basis of Transcription Initiation. *Science* **338**:1076–1080.

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. *International Journal of Systematic Bacteriology* **46**:305–312.