Transcriptome analysis unveils regulatory landscape of *Spirochaeta africana*

Kotikoski Juho BSc , Trapp Vilma MSc, Belogurov Georgi A. PhD
Department of Life Technologies, University of Turku

MOLECULAR SYSTEMS BIOLOGY

UNIVERSITY OF TURKU

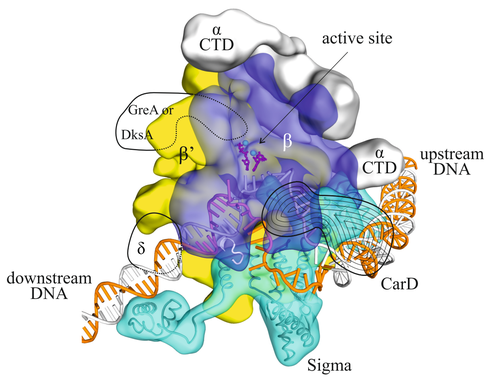
Introduction

Transcription is the first and most regulated step in gene expression. Transcription is catalyzed by RNA polymerase (RNAP), a complex multisubunit enzyme. RNAP can read regulatory signals encoded in the DNA and respond to changes in ribonucleoside triphosphate (NTP) concentration. Most regulatory inputs come to RNAP via accessory protein factors and regulatory RNAs. The bacterial transcription system consists of RNAP and the regulatory factors that affect it. The regulators can activate or repress transcription. The transcription systems of spirochetes, the bacteria that look and move like a cork-screw, are poorly understood. *Spirochaeta africana* is a uniquely good model for studying spirochaetal transcription because it is non-pathogenic, is easy to cultivate in the laboratory, and is related to several spirochetes that cause severe disease in humans.

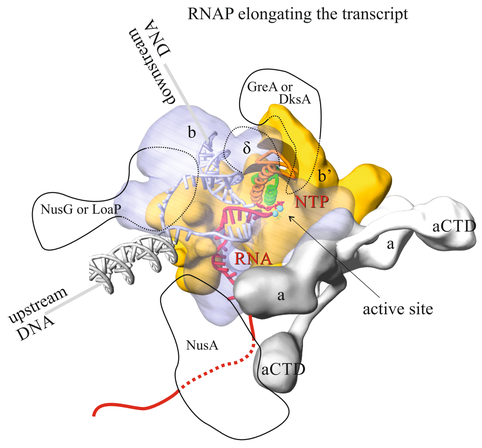
Transcription has three distinct phases: 1) initiation 2) elongation 3) termination. Transcription initiation begins when RNAP binds to the promoter region of the DNA. The promoter contains specific sequences recognized by RNAP, including the -10 and -35 regions. RNAP holoenzyme, composed of the core (α2ββ'ω) and sigma factor (σ), recognizes and binds to the promoter sequence. The sigma factor assists RNAP in locating and binding to the promoter region. Once bound, RNAP unwinds a short stretch of DNA (~14 base pairs) to form an open complex. In *S. africana*, CarD stimulates the initiation and DksA generally inhibits it.

During elongation, RNAP moves along the template DNA strand while synthesizing RNA from NTPs in the 5' to 3' direction. As RNAP advances, it unwinds the DNA ahead and rewinds it behind the transcription bubble.
Many transcription factors have an effect during the elongation phase. The Nus proteins (NusA, NusB, NusG, and LoaP) promote effective elongation and termination, while GreA and DksA can modulate pausing and increase the fidelity of transcription.

RNAP initiating transcription at the promoter



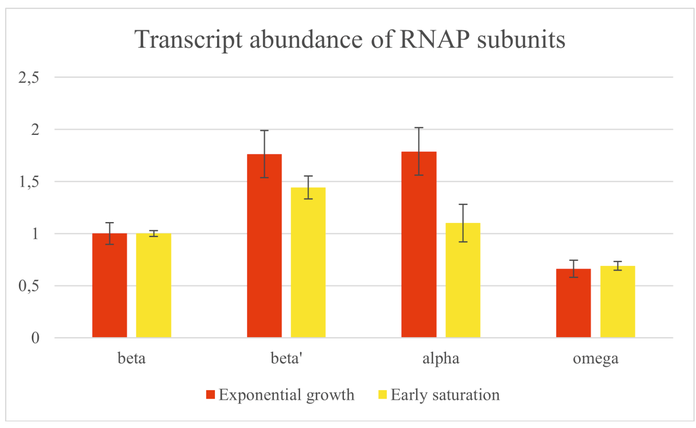
RNAP elongating the transcript



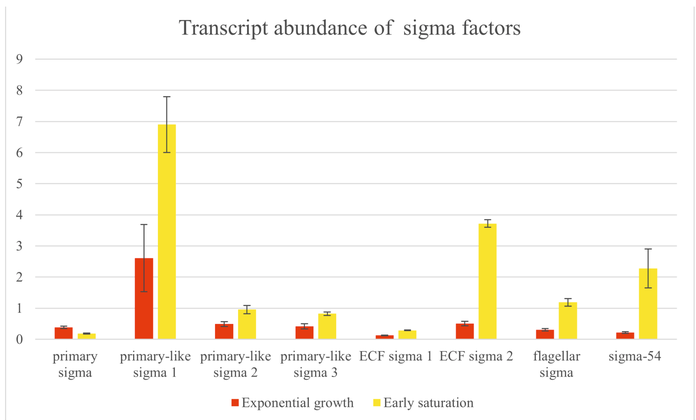
Termination is the final stage of transcription. There are two main types of termination in bacteria: Rho-dependent and Rho-independent. Rho-dependent termination involves the Rho factor, which recognizes rut sites in the RNA transcript. Rho binds to the transcript and releases the RNA and the RNA polymerase from the DNA template.

In Rho-independent termination, a sequence in the RNA transcript forms a stable hairpin, destabilizing the elongation complex, which leads to the dissociation of RNAP and the release of the transcript.

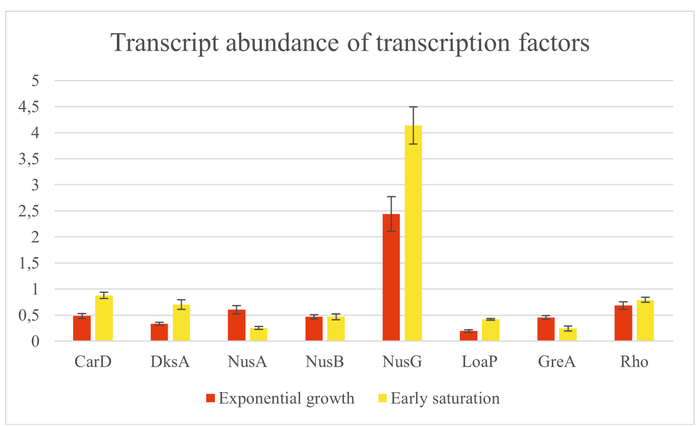
Results: *S. africana* transcriptome



The RNAP subunits were altogether well represented in the transcriptome. The omega subunit, however, was slightly underrepresented in our dataset compared to the stoichiometric ratio in the enzyme. Therefore, the beta subunits were used for the normalization of the entire dataset as the most consistent RNAP subunit. Because the total transcript abundance varied significantly between the logarithmic growth stage and the early saturation stage, respective beta subunits were used to normalize their transcriptomes to get relative transcription levels.

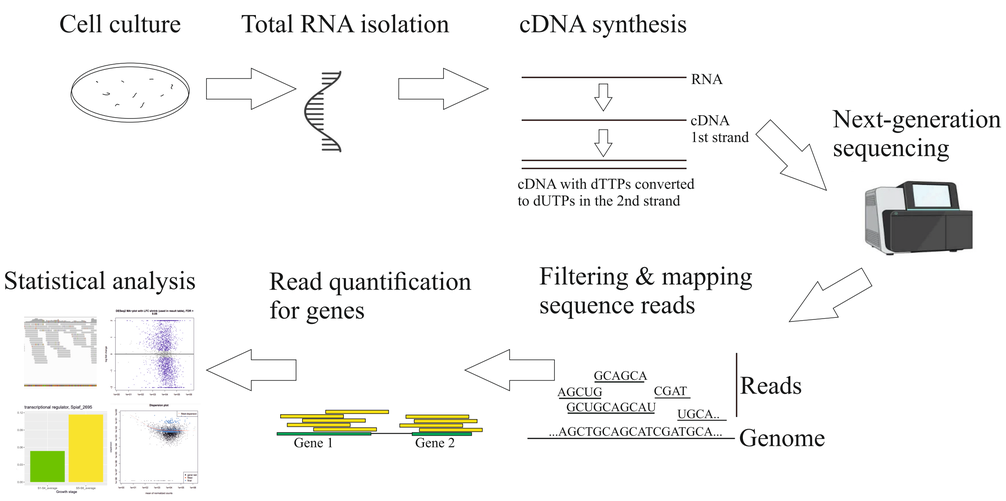


Sigma factors play a key role in the RNAP holoenzyme. They help the RNAP core in targeting and binding to the DNA promoter. We studied all the sigma factors in *S. africana*. Surprisingly, the primary sigma factor contributed under 10 % of the sigma factor transcripts in the exponential growth phase and 1 % in early saturation. Alternative sigma factors were all expressed at a higher level in early saturation, while the primary sigma was more abundant in the exponential growth phase. This indicates that alternative sigma factors may have a larger role in spirochaetal transcription regulation than previously thought.



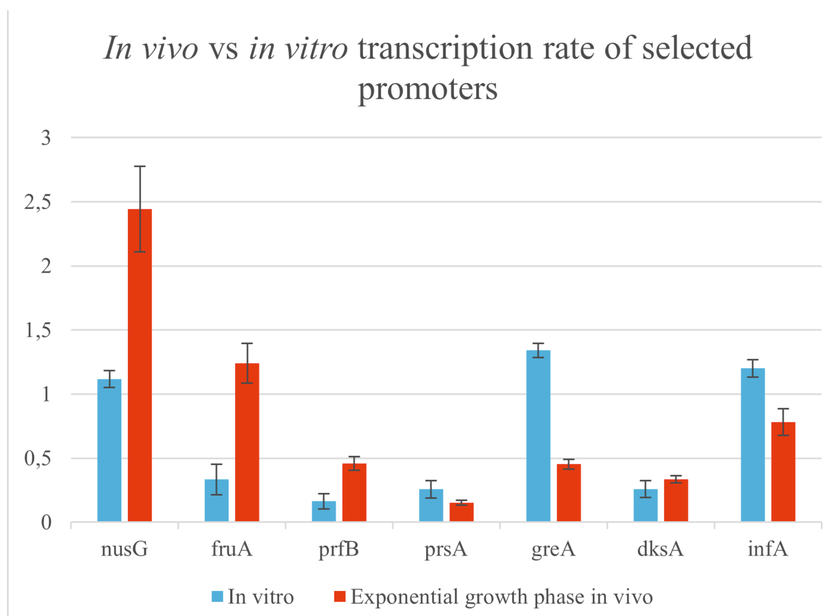
All well-characterized transcription factors (TFs) were present in the transcriptome at measurable levels. CarD and DksA were expressed in similar ratios between the exponential growth phase and early saturation; notably, they were both expressed at higher levels at the later growth stage. The increased production rate of CarD in the later growth stage may relate to prioritized transcription initiation. In contrast, NusA, an indicator of transcription elongation, was expressed at a higher rate in the logarithmic phase.

Workflow



Total RNA was extracted from six cultures of *S. africana* at two separate growth stages using a total RNA isolation protocol. Sequencing was conducted with Illumina NovaSeq 6000 SP v1.5. Paired-end FastQ file reads were trimmed with Trimmomatic 0.39 and aligned to the reference genome of *Spirochaeta africana* DSM8902 with Bowtie 2.5.3. Reads were counted for gene features of the reference genome using HTSeq 2.0.3. Read counts were normalized relative to the respective RNAP beta subunit amount for differential expression comparison.

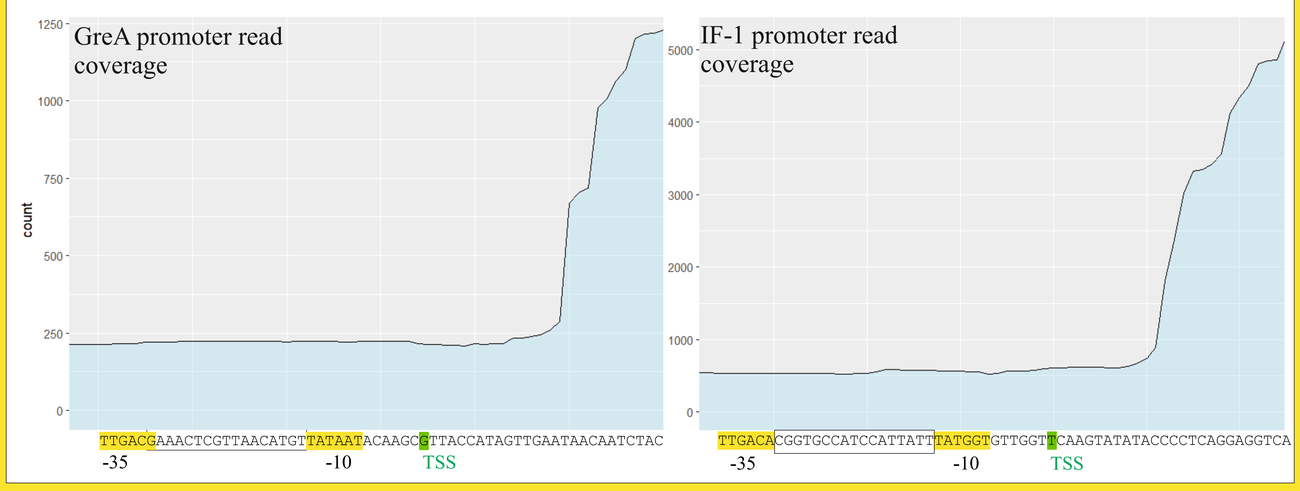
Results: *S. africana* promoters



To complement and verify the transcriptomic data, we performed *in vitro* expression assays with promoters from 7 genes in *S. africana*. RNAP, primary sigma factor, and CarD from *S. africana* were expressed in and purified from recombinant *Escherichia coli*. *In vitro* promoter activity was measured at pH 9 using a broccoli assay.

In comparison to *in vivo* expression, the promoters assayed *in vitro* had a narrower spread. The promoters in *greA* and *infA* genes had strong activity *in vitro*. Other medium to strong promoters, such as *fruA* and *prfB* had significantly lower activity than expected.

The transcriptomics dataset had a sufficiently deep coverage, which enabled us to also study the promoter structures of genes. Based on canonical promoter motifs, the transcriptome consistently showed an increase in read count 10-20 base pairs downstream of the predicted transcription start site. *In vitro* DNA extension assays were conducted to verify the start sites, which matched the predictions except for a single distinct promoter.



Conclusions

S. africana produces vastly more RNAs encoding alternative sigma factors than the primary sigma factor during exponential growth phase. Our future investigations will study the capacity of alternative sigma factors to produce transcripts *in vitro*.

All canonical transcription factors encoded by *S. africana* genome are measurably transcribed. RNAs encoding canonical transcription factors were less abundant than RNAP-encoding transcripts except for NusG that was transcribed at 2-4 times the level of RNAP.

Many promoters that produce abundant transcripts *in vivo* were also relatively strong *in vitro*. However, several promoters displayed perplexingly low activities *in vitro*. These promoters may utilize an alternative sigma factor or an unidentified activator.

While our transcriptome sequencing method was not designed to precisely determine the transcription start sites (TSS), the very deep coverage made it possible to estimate the locations of promoters with high accuracy. In all but one case the TSS mapped *in vitro* matched the TSS suggested by the transcriptome data.