AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Bacteriology®

RESEARCH ARTICLE

σ^I from *Bacillus subtilis*: Impact on Gene Expression and Characterization of σ^I-Dependent Transcription That Requires New Types of Promoters with Extended −35 and −10 Elements

Olga Ramaniuk,[a,b] Martin Převorovský,[c] Jiří Pospíšil,[a] Dragana Vítovská,[a] Olga Kofroňová,[d] Oldřich Benada,[d] Marek Schwarz,[e] Hana Šanderová,[a] Jarmila Hnilicová,[a] Libor Krásný[a]

[a]Laboratory of Microbial Genetics and Gene Expression, Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic

[b]Department of Genetics and Microbiology, Faculty of Science, Charles University, Prague, Czech Republic

[c]Department of Cell Biology, Faculty of Science, Charles University, Prague, Czech Republic

[d]Laboratory of Molecular Structure Characterization, Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic

[e]Laboratory of Bioinformatics, Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic

ABSTRACT The σ^I sigma factor from *Bacillus subtilis* is a σ factor associated with RNA polymerase (RNAP) that was previously implicated in adaptation of the cell to elevated temperature. Here, we provide a comprehensive characterization of this transcriptional regulator. By transcriptome sequencing (RNA-seq) of wild-type (wt) and σ^I-null strains at 37°C and 52°C, we identified ~130 genes affected by the absence of σ^I. Further analysis revealed that the majority of these genes were affected indirectly by σ^I. The σ^I regulon, i.e., the genes directly regulated by σ^I, consists of 16 genes, of which eight (the *dhb* and *yku* operons) are involved in iron metabolism. The involvement of σ^I in iron metabolism was confirmed phenotypically. Next, we set up an *in vitro* transcription system and defined and experimentally validated the promoter sequence logo that, in addition to −35 and −10 regions, also contains extended −35 and −10 motifs. Thus, σ^I-dependent promoters are relatively information rich in comparison with most other promoters. In summary, this study supplies information about the least-explored σ factor from the industrially important model organism *B. subtilis*.

IMPORTANCE In bacteria, σ factors are essential for transcription initiation. Knowledge about their regulons (i.e., genes transcribed from promoters dependent on these σ factors) is the key for understanding how bacteria cope with the changing environment and could be instrumental for biotechnologically motivated rewiring of gene expression. Here, we characterize the σ^I regulon from the industrially important model Gram-positive bacterium *Bacillus subtilis*. We reveal that σ^I affects expression of ~130 genes, of which 16 are directly regulated by σ^I, including genes encoding proteins involved in iron homeostasis. Detailed analysis of promoter elements then identifies unique sequences important for σ^I-dependent transcription. This study thus provides a comprehensive view on this underexplored component of the *B. subtilis* transcription machinery.

KEYWORDS RNAP, RNA-seq, iron metabolism, promoter, sigma factor

Bacillus subtilis is a Gram-positive microorganism that belongs to the bacterial phylum *Firmicutes*. Its main habitat is soil, but it is also found in the gastrointestinal microbiomes of many organisms, including that of humans. In these habitats, *B. subtilis*

Received 26 April 2018 **Accepted** 9 June 2018

Accepted manuscript posted online 18 June 2018

Citation Ramaniuk O, Převorovský M, Pospíšil J, Vítovská D, Kofroňová O, Benada O, Schwarz M, Šanderová H, Hnilicová J, Krásný L. 2018. σ^I from *Bacillus subtilis*: impact on gene expression and characterization of σ^I-dependent transcription that requires new types of promoters with extended −35 and −10 elements. J Bacteriol 200:e00251-18. https://doi.org/10.1128/JB.00251-18.

Editor Tina M. Henkin, Ohio State University

Address correspondence to Libor Krásný, krasny@biomed.cas.cz.

encounters a wide range of conditions. Adaptation to these conditions depends on changes in gene expression. The first step in gene expression is transcription of DNA into RNA, which is mediated by RNA polymerase (RNAP) (1). The bacterial RNAP ($\alpha_2\beta\beta'\omega$) consists of several subunits. The α dimer holds together β and β', which form the catalytic center. The small ω subunit mainly binds to the C-terminal part of β' and helps with assembly and structural integrity of the enzyme (2, 3). Furthermore, in *Firmicutes* RNAP associates with additional small proteins, δ and ε, which are considered to be additional subunits, although the role of δ has recently been debated (4–6). The core enzyme ($\alpha_2\beta\beta'\omega\delta\varepsilon$) must associate with a σ factor to initiate transcription. σ factors endow RNAP with the ability to recognize promoters—specific sequences in DNA where transcription initiates. Depending on diverse environmental signals, bacterial gene expression is controlled by various σ factors that control different sets of genes (7, 8).

B. subtilis has one primary σ factor, σ^A, responsible for transcription of housekeeping genes. The σ^A factor consists of four domains, of which domain 1.1 autoinhibits this factor in free form, preventing its binding to promoter DNA (9, 10). Furthermore, *B. subtilis* has 17 alternative σ factors and one σ-like factor (11–14). While most *B. subtilis* σ factors have been systematically studied, σ^I remains poorly characterized, with a nearly unknown regulon.

The σ^I factor was discovered in 2001 by Zuber and colleagues (15). It is encoded by the nonessential gene *sigI* (*ykoZ*) that is cotranscribed with *rsgI*, which encodes the cognate anti-σ factor RsgI. Transcription of the *sigI-rsgI* operon is driven by a σ^I-dependent promoter (16) and also by a σ^A-dependent promoter (17). The predicted molecular weight of σ^I is 29.04 kDa, and it has a theoretical isoelectric point of 8.31. RsgI is a transmembrane protein that sequesters σ^I under favorable growth conditions and releases it when appropriate stimuli—for example, heat shock—appear (16). Degradation of RsgI by proteases ClpC and ClpP then provides a posttranslational layer of regulation of σ^I activity (18).

A σ^I deletion strain was reported to be unable to grow on agar plates at 54° to 55°C (15). This temperature-sensitive phenotype of strains lacking σ^I was also implied in other studies (16, 19–21). Moreover, it was shown that, along with heat shock response, σ^I was involved in cold shock response (20).

Currently, there are only seven genes, organized in five operons known to be σ^I dependent in *B. subtilis*: the *sigI-rsgI* operon itself (16), *bcrC*, the *mreBH-ykpC* operon (21), *lytE* (22), and *gsiB* (15). Expression of some of these genes was shown to be stimulated by WalR. WalR is the response regulator of the WalRK two-component system that controls cell wall metabolism (23).

BcrC helps protect the cell against bacitracin (24) and paraquat (25), and it is also needed for production of the carrier lipid for cell wall synthesis (26). MreBH participates in formation of straight rod-shaped cells, and its depletion or overexpression in *B. subtilis* leads to the appearance of malformed cells (27–30). Moreover, MreBH is required for the lytic activity of a cell wall hydrolase, LytE, that is important for cell elongation and separation (31–33). The function of the *ykpC* gene, which is located in the same operon as the *mreBH* gene, is still unknown. GsiB is a general stress protein that prevents inactivation of cellular enzymes upon freeze-thaw treatments and is involved in responses to nutrient deprivation (34, 35).

In summary, information concerning σ^I in *B. subtilis* is still fragmentary and incomplete. As *B. subtilis* is an extensively studied and industrially important organism, we decided to fill this gap in our knowledge. In this report, we use transcriptome sequencing (RNA-seq) to reveal the impact of *B. subtilis* σ^I on gene expression. Based on the transcriptomic data, we subsequently demonstrate that σ^I is involved in iron metabolism. Finally, using recombinant σ^I, we set up a σ^I-dependent *in vitro* transcription system to describe its promoter specificity, as well as to test the promoter regions of selected genes for their dependence on σ^I.

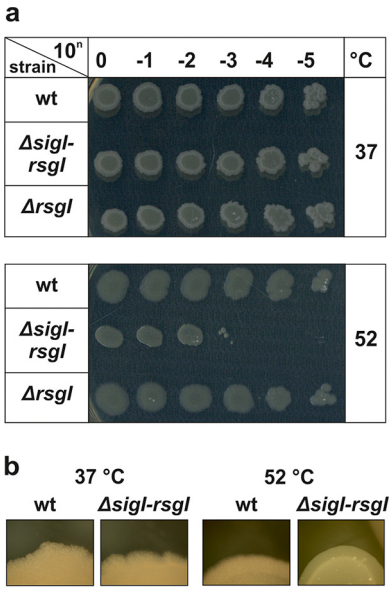


FIG 1 Spot assays of *B. subtilis* wt, Δ*sigI-rsgI*, and Δ*rsgI* strains on agar plates at 37°C and 52°C. (a) Serial dilutions of mid-logarithmic phase cultures (OD_{600}, ~0.45) of wt, Δ*sigI-rsgI*, and Δ*rsgI* strains were spotted on LB agar plates and incubated at 37°C and 52°C for 40 h. The experiment was repeated three times with the same result. (b) Colony morphology of wt and Δ*sigI-rsgI* strains.

RESULTS

The σ^I factor is important for growth at elevated temperature. We used *B. subtilis* 168 trp+ (BaSysBio) as the genetic background to create Δ*sigI-rsgI* and Δ*rsgI* strains (for details, see Materials and Methods). Under standard conditions (aerobic cultivation in LB broth at 37°C), Δ*sigI-rsgI* and Δ*rsgI* strain growth was indistinguishable from that of the wild-type (wt) strain (see Fig. S1 in the supplemental material).

Subsequently, we determined whether the Δ*sigI-rsgI* strain was sensitive to elevated temperature. We grew Δ*sigI-rsgI*, Δ*rsgI*, and wt strains in LB at 37°C, spotted serial dilutions of cell suspensions on agar plates, and let them incubate at 37°C and 52°C for 40 h (Fig. 1). All three strains displayed the same pattern of growth when cultivated at 37°C. In contrast, the Δ*sigI-rsgI* strain displayed impaired growth at 52°C, while wt and Δ*rsgI* strains cultivated at this temperature had the same pattern of growth as that at 37°C. Thus, σ^I was essential for efficient growth when the cells were chronically exposed to elevated temperature. This result was in agreement with previously published observations (15, 16, 34).

Deletion of σ^I changes cell morphology during heat stress. During the spot assays described above, we noticed that the Δ*sigI-rsgI* colonies grown at 52°C displayed different morphology in comparison to that of the wt and Δ*rsgI* strains (Fig. 1b). Hence, we looked at the cells in close detail by scanning electron microscopy (SEM) of mid-logarithmic phase Δ*sigI-rsgI*, Δ*rsgI*, and wt strains cultivated at 37°C and 52°C (Fig. 2). The results correlated with those of the spot assays; cells of all three strains grown at 37°C had the typical rod shape of *B. subtilis*. In contrast, the Δ*sigI-rsgI* cells grown at 52°C displayed a previously unreported phenotype; that is, cells were bent and their shape was irregular even compared to that of the same cells grown at 37°C, as well as to that of Δ*rsgI* and wt cells grown at 37°C and 52°C. Finally, we tested whether the loss of MreBH expression could be responsible for the growth defect at 52°C. We performed the same type of experiment as in Fig. 1 using a Δ*mreBH* strain. The absence of *mreBH* did not result in impaired growth at any temperature tested when the cells were cultivated on LB agar plates (see Fig. S2 in the supplemental material).

σ^I influences expression of 131 genes. To identify the genes regulated by σ^I, we used Δ*sigI-rsgI* and wt strains and RNA-seq as the first approach. We cultivated the strains at 37°C and 52°C to mid-logarithmic phase and performed Illumina-based

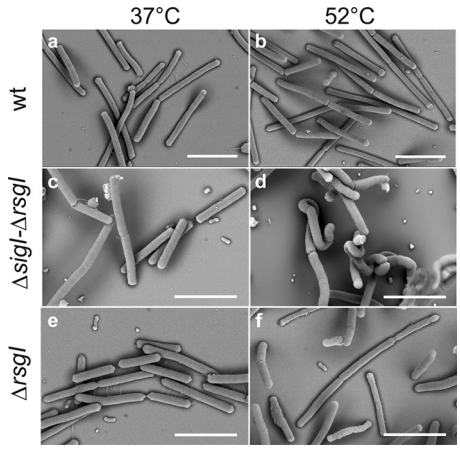


FIG 2 Cell shapes of the *B. subtilis* wt, Δ*sigI-rsgI*, and Δ*rsgI* cells imaged by scanning electron microscopy (SEM). (a) wt cells grown in LB at 37°C. (b) wt cells grown in LB at 52°C. (c) Δ*sigI-rsgI* cells grown in LB at 37°C. (d) Δ*sigI-rsgI* cells grown in LB at 52°C. (e) Δ*rsgI* cells grown in LB at 37°C. (f) Δ*regI* cells grown in LB at 52°C. The experiment was performed twice with identical results. Bars, 5 μm.

RNA-seq, with subsequent comparison of the Δ*sigI-rsgI* and wt transcriptomes to identify differentially expressed genes (DEGs).

First, we identified DEGs between Δ*sigI-rsgI* and wt cells grown at 37°C (>1.5-fold difference, <5% false discovery rate). Expression of 13 genes was decreased in the Δ*sigI-rsgI* strain relative to that of the wt (Fig. 3a, blue circle [i.e., genes positively regulated by σ^I]). Proteins encoded by these genes are mainly involved in transcription, translation, and coping with stress (see Fig. S3 and Table S2 in the supplemental material). Moreover, expression of 63 genes was increased (upregulated) in the Δ*sigI-rsgI* strain (Fig. 3b, blue circle [i.e., genes negatively regulated by σ^I]). They are mainly involved in the regulation of cell metabolism and in coping with stress (Fig. S3 and Table S2). Thus, at 37°C, σ^I positively affects 13 genes and negatively affects 63 genes.

Second, at 52°C we detected 55 downregulated and eight upregulated genes in the Δ*sigI-rsgI* strain compared to in the wt (Fig. 3, red circles). The majority of the affected genes are involved in the regulation of the cell metabolism (e.g., iron metabolism). In addition, upregulated genes included those involved in cell wall turnover (see Fig. S4 and Tables S4 and S5 in the supplemental material). Thus, at 52°C, σ^I positively affected 55 genes and negatively affected eight genes. A considerable portion of these genes encoded membrane-associated proteins (Fig. S3 and S4).

Six genes were downregulated at both temperatures, namely, *lytE*, *sigI*, *rsgI*, *mreBH*, *ykpC* (previously known to be σ^I dependent), and *spoVD*. Two genes—*ykrP*

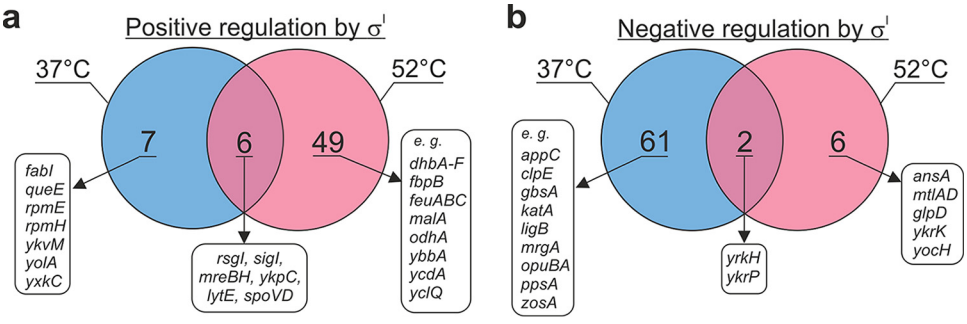


FIG 3 Genes in *B. subtilis* affected by σ^I. *B. subtilis* Δ*sigI-rsgI* and wt strains were grown at 37°C and 52°C in LB broth to an OD_{600} of ~0.45. RNA was extracted and libraries were prepared for transcriptome sequencing (RNA-seq). (a) Genes positively regulated by σ^I. These genes were downregulated in the Δ*sigI-rsgI* strain compared to in the wt strain. (b) Genes negatively regulated by σ^I. These genes were upregulated in Δ*sigI-rsgI* compared to wt.

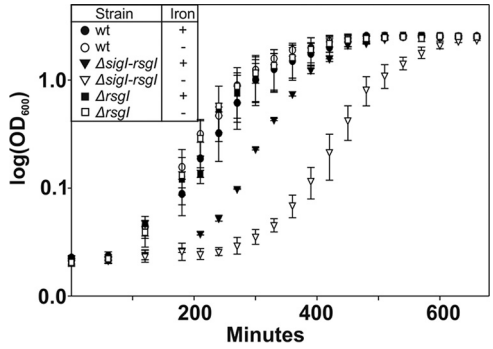


FIG 4 The σ^I factor is involved in iron metabolism in *B. subtilis*. Growth curves of the Δ*sigI-rsgI* (triangles), Δ*rsgI* (squares), and wt (circles) strains grown in defined MOPS medium at 37°C in the presence (filled shapes) or absence (empty shapes) of $FeCl_3$. The experiment was repeated three times. The error bars show ± standard deviation (SD).

and *yrkH*—were upregulated at both temperatures in the Δ*sigI-rsgI* strain compared to those in the wt.

Taking these results together, based on the RNA-seq data, we find that σ^I affected expression of 131 genes organized in 90 operons (Fig. 3; see also Tables S2 to S5 in the supplemental material). We note here that the identified DEGs could be regulated by σ^I either directly (the bona fide σ^I regulon) or indirectly.

Genes affected by σ^I in the context of the heat stimulon. To assess the importance of σ^I for the cell, we compared genes affected by σ^I with the genes stimulated by heat in wt (52°C versus 37°C) and investigated where among these genes the σ^I-affected genes ranked (see Fig. S5 and Tables S6 and S7 in the supplemental material). According to our RNA-seq data, the wt heat stimulon contained >370 genes (DEGs, >2-fold difference). Seven genes that were stimulated by σ^I (either directly or indirectly) belonged among the top 10% of the heat-stimulated genes (e.g., *malA*, carbon metabolism; *mreBH*, cell shape; and *dhb* genes, iron metabolism). We concluded that σ^I was involved in the regulation of expression of genes whose stimulation was prominent at elevated temperature.

σ^I is involved in iron metabolism. The RNA-seq data revealed that a significant number of genes influenced by σ^I (23 genes organized in 11 operons) were involved in iron metabolism (see Tables S3 and S4 in the supplemental material). Two of these genes were misregulated at 37°C and 21 genes were misregulated at 52°C.

The absence of σ^I increased expression of *zosA* (*pfeT*) [Fe(II) efflux pump] and *mrgA* (iron storage protein) at 37°C. We speculated that a combination of increased iron efflux and increased iron retention by MrgA in the Δ*sigI-rsgI* strain could affect iron homeostasis under limiting iron conditions. To test this hypothesis, we cultivated wt, Δ*sigI-rsgI*, and Δ*rsgI* strains in a defined 3-(*N*-morpholino)propanesulfonic acid (MOPS) medium containing or lacking iron. While the doubling time of all strains was comparable, the results repeatedly showed that at 37°C, the Δ*sigI-rsgI* strain had a markedly prolonged lag phase in the absence of iron (Fig. 4).

The presence of σ^I increased expression of 21 genes associated with iron metabolism at 52°C (Table S4), including genes involved in iron uptake (e.g., enzymes participating in siderophore [iron-chelating compound] synthesis and ABC transporters of siderophores). However, the Δ*sigI-rsgI* strain did not grow at 52°C in the defined MOPS medium, regardless of the presence or absence of iron (see Fig. S6 in the supplemental material).

The σ^I regulon is small. To start distinguishing between direct and indirect effects of σ^I, we cloned and purified recombinant σ^I by nickel affinity chromatography via the introduced C-terminal His tag (see Fig. S9 in the supplemental material). Next, we set up an *in vitro* transcription system to assess the ability of σ^I to directly regulate selected genes. We used 25 promoter regions identified in our RNA-seq analysis (including three

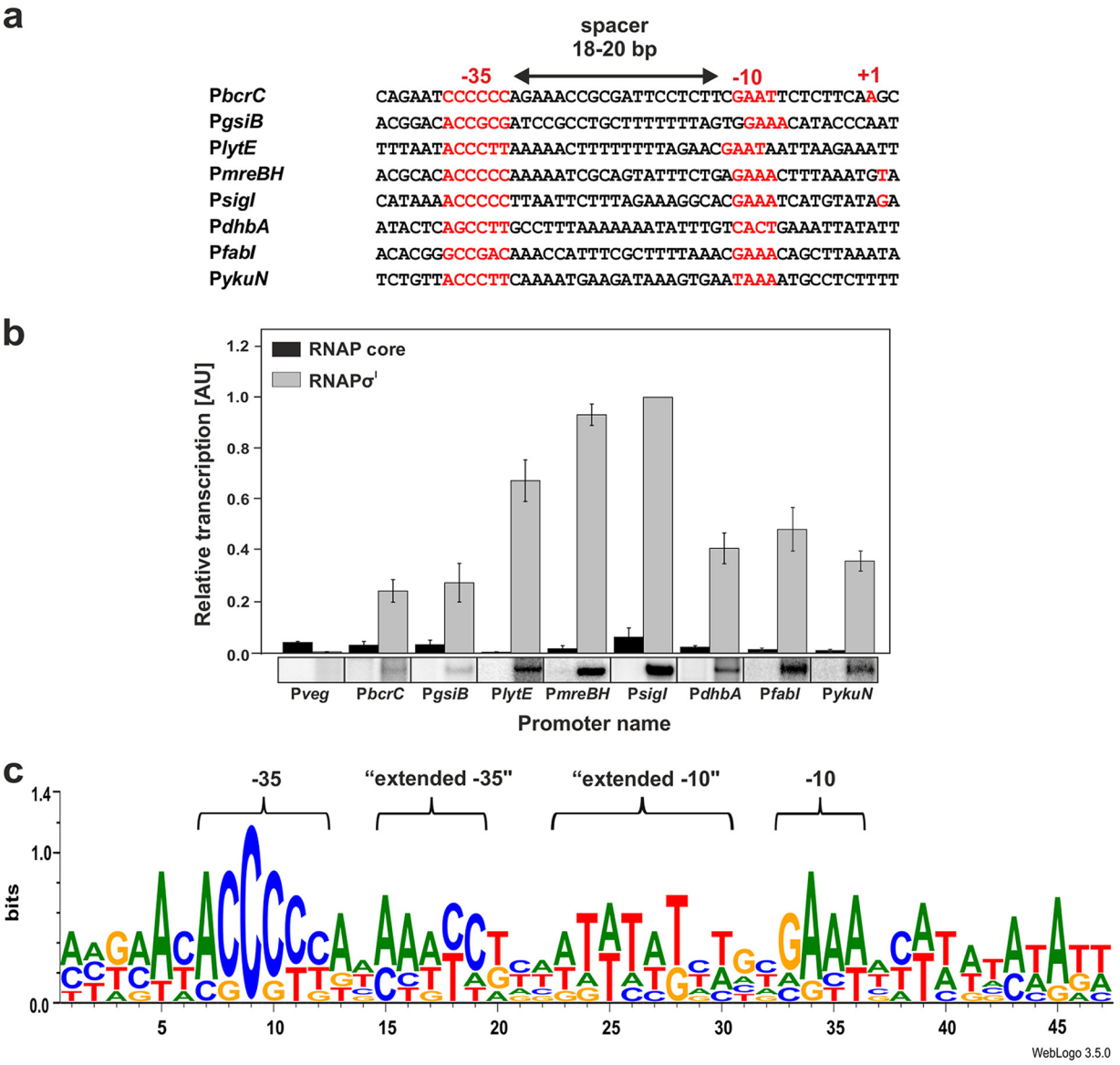


FIG 5 Multiple-round *in vitro* transcription assays with promoter regions of σ^I-regulated genes and RNAPσ^I. (a) Alignment of σ^I-dependent promoters. The −10 and −35 elements and +1 position for P*sigI* (16), P*mreBH*, and P*bcrC* (21) are in red. (b) Transcription was performed with the RNAPσ^I holoenzyme and the RNAP core. Transcription with the RNAP core was used to assess potential contamination of the RNAP core with σ factors. Promoter P*sigI* was used as a control and its transcription was set as 1. P*veg* (σ^A dependent) was used as a negative control for RNAPσ^I. Primary data (radioactively labeled transcripts resolved on polyacrylamide [PAA] gels) are shown below the graph. The error bars show averages from three independent experiments ± SD. (c) The σ^I consensus logo was created from the 8 promoter sequences shown in panel a. Conserved promoter elements are indicated above the logo.

promoter regions of genes previously demonstrated to be σ^I dependent) and the promoter regions of *bcrC* and *gsiB*, which were previously shown to be affected by σ^I but which did not appear in our transcriptomic screen (27 promoter regions in total). We performed transcription assays with these templates both with RNAPσ^A (see Fig. S7 in the supplemental material) and with RNAPσ^I (Fig. 5). RNAPσ^A was active on P*veg* (positive control) and also on promoter regions of the *lytE*, *sigI*, and *dhb* operon. These genes were previously known to contain σ^A-dependent promoters; our results confirmed these findings (Fig. S7). The experiments were done at 37°C. In addition, we also performed experiments with RNAPσ^I at 52°C, and the results were similar to those obtained at 37°C (data not shown).

RNAPσ^I was active on eight out of the 27 tested promoter regions, namely, *bcrC*, *gsiB*, *sigI-rsgI*, *lytE*, *mreBH-ykpC*, *dhb*, *yku*, and *fabI*. The first five promoter regions

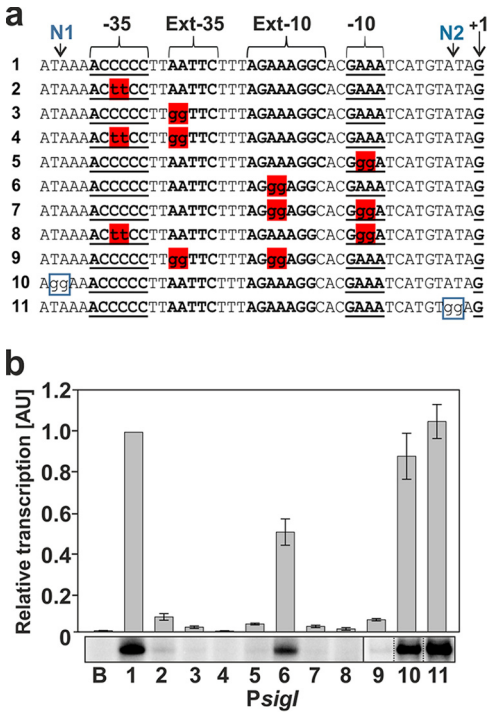


FIG 6 Multiple-round *in vitro* transcription assays with mutated P*sigI* promoter region. (a) Fragments of the P*sigI* promoter region used for evaluating the importance of the identified sequence elements. Fragment 1 contains the native P*sigI* promoter region. Fragments 2 to 9 contain mutations in −10, −35, extended −35 (Ext −35), extended −10 (Ext −10) elements, and their combinations. Mutations are highlighted in red. Fragments 10 and 11 contain control neutral (N) mutations highlighted in blue boxes. The −10 and −35 regions and +1 position are in bold and underlined. Extended elements are in bold. (b) *In vitro* transcription with specific double substitutions in P*sigI* −10 and −35 σ^I binding sites, extended −10, and extended −35 elements. Transcription was performed with the RNAP core (lane B, blank assay) and RNAPσ^I holoenzyme on PCR products as a template. The blank assay was used to demonstrate that the RNAP core was devoid of contaminating σ factors. Primary data (radioactively labeled transcripts resolved on PAA gel) are shown below the quantification graph; the vertical black line indicates the border between two PAA gels used for illustration. The dotted lines indicate lanes from the same gel electronically positioned next to each other. Error bars show averages from three independent experiments ± SD.

corresponded to genes/operons known to be regulated by σ^I. Thus, among the newly identified putative 22 promoter regions, RNAPσ^I was active on only three of them (promoter regions of *dhb* and *yku* operons and the *fabI* gene). Although we did not test the remaining 37 promoter regions that were positively affected by σ^I according to RNA-seq, this result strongly implied that the σ^I regulon is small.

Characterization of the σ^I promoter sequence. We aligned all of the known and newly identified σ^I promoter regions and created a σ^I promoter sequence logo (Fig. 5a and c). In comparison with the previously published σ^I promoter logo created from σ^I-dependent promoter regions from *Bacillales* (21, 36), in addition to the −10, −35 and "extended −35" elements, we also identified an "extended −10" element, which was dissimilar to the well-studied extended −10 element of σ^A-dependent promoters (37). A bioinformatic search for the σ^I consensus sequence among promoter regions of the σ^I-affected genes yielded no obvious additional hits.

As the σ^I logo was information rich (compared to that of, e.g., σ^A), we wished to determine the importance of the individual sequence elements for promoter activity. For this purpose, we used the P*sigI* promoter and systematically mutated these elements (Fig. 6). The results showed that double substitutions in −35 and −10 elements drastically reduced transcription (constructs 2 and 5, by ~90% and ~97%, respectively), and their combination (construct 8) virtually abolished transcription. Interestingly, the experiments revealed that the extended −35 element is highly

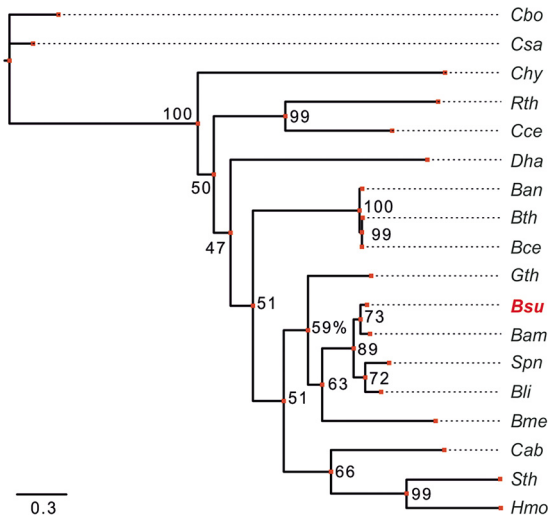


FIG 7 Phylogenetic tree of *B. subtilis* σ^I homologs. The list of the homologs is in Fig. S8 in the supplemental material. The phylogenetic tree was inferred with RAxML, and the best-scoring maximum likelihood tree is shown. Numbers denote bootstrap values (percent), as reported by RAxML. The scale bar represents expected number of substitutions per site. Cbo, *Clostridium botulinum*; Csa, *Clostridium saccharobutylicum*; Chy, *Carboxydothermus hydrogenoformans*; Rth, *Rumuniclostridium thermocellum*; Cce, *Clostridium cellulolyticum*; Dha, *Desulfitobacterium hafniense*; Ban, *Bacillus anthracis*; Bth, *Bacillus thuringiensis*; Bce, *Bacillus cereus*; Gth, *Geobacillus thermodenitrificans*; Bsu, *Bacillus subtilis*; Bam, *Bacillus amyloliquefaciens*; Spn, *Streptococcus pneumoniae*; Bli, *Bacillus licheniformis*; Bme, *Bacillus megaterium*; Cab, *Chlamydia abortus*; Sth, *Symbiobacterium thermophilum*; Hmo, *Heliobacterium modesticaldum*.

important, as its mutation almost completely canceled transcription (construct 3). The extended −10 element then led to an approximately 40% decrease in transcription activity (construct 6), suggesting that it is still required for optimal performance of the promoter (Fig. 6). Control "neutral" mutations without expected large effects on transcription confirmed that the "extended" elements are bona fide promoter regulatory sequences that are involved in σ^I-dependent transcription (Fig. 6b, constructs 10 and 11).

σ^I in different bacterial species. Finally, to identify σ^I homologs in other bacterial species, we performed BLAST (38) searches and created a phylogenetic tree (Fig. 7) from the most related amino acid sequences (see Fig. S8 in the supplemental material). The closest relatives of *B. subtilis* σ^I were proteins from *Bacillus amyloliquefaciens*, a putative σ^I from *Streptococcus pneumoniae*, and σ^I from *Bacillus licheniformis*.

DISCUSSION

In this study, we performed a comprehensive characterization of σ^I from *B. subtilis*. We determined the effect of its presence/absence on gene expression during exponential growth in a rich medium at 37°C and 52°C, identified its involvement in iron metabolism, and defined its transcriptional properties.

Effect of σ^I on gene expression. The σ^I factor affected expression of 131 genes. Nevertheless, the majority of these genes appear to be indirectly affected by σ^I. Of the seven genes that had been known or proposed to be regulated by σ^I, we identified five by RNA-seq, which gave us confidence in the credibility of the transcriptomic data. We also demonstrated the direct ability of RNAPσ^I to initiate transcription from upstream regions of these genes. In addition, we detected σ^I-dependent promoter activity for the *bcrC* and *gsiB* genes *in vitro*. However, we did not identify these genes in our RNA-seq analysis. This could be due to the very weak transcriptional activity of the σ^I-dependent promoters of these genes (Fig. 5b) and, perhaps, to the specific experimental conditions used; i.e., chronic heat stress (this study) versus temperature shift (15, 21).

Out of the 22 tested putative new promoter regions, *in vitro* transcription analysis revealed three σ^I-dependent promoters driving transcription of nine genes (*dhbA-dhbC-dhbE-dhbB-dhbF*, *ykuN-ykuO-ykuP*, and *fabI*). The first two operons are involved in

iron metabolism. The *dhb* operon is involved in bacillibactin siderophore synthesis (39). The *yku* operon encodes flavodoxins that replace ferredoxin under conditions of iron limitation and catalyze the O_2-dependent desaturation of the acyl chain of membrane phospholipids (40). In our transcriptomic data (in SigI-null relative to wt), we observed decreased expression of *ykuO* and *ykuP* genes at 52°C, but their σ^I-dependent promoter was identified in *in vitro* transcription within a DNA fragment preceding the first gene of the operon, *ykuN*. Careful inspection of the RNA-seq data revealed that also the *ykuN* gene was affected ($\sim 2\times \downarrow$) by the absence of σ^I in the same way as *ykuO* and *ykuP*, but this change in expression was not deemed significant by the DESeq2 algorithm, due to the low level of *ykuN* expression. The product of the *fabI* gene is enoyl-acyl carrier protein reductase, which participates in fatty acid biosynthesis (41). Expression of the *fabI* gene was affected at 37°C, suggesting that this effect is heat shock independent.

In summary, the σ^I regulon, as of now, is one of the smallest known *B. subtilis* regulons, currently containing 16 genes. A search for σ^I promoter sequences within the upstream regions of genes that did not function in *in vitro* transcriptions revealed, in all cases but one (*feuA*), the −10 GAAA motif and the absence of −35 regions. Hence, it is possible that some of these genes are still regulated by σ^I, and the presence of an unknown regulator is required. Moreover, as there are still 37 promoter regions with putative activation by σ^I untested, a few more genes may be added to the list in the future.

σ^I-dependent promoters. We used all eight known promoter regions to create the σ^I promoter sequence logo (Fig. 5). Our logo revealed that σ^I-dependent promoters are information rich in the spacer region; besides −35 and −10 elements, extended −35 and −10 elements were identified. Importantly, functional analysis of extended −35 and −10 elements demonstrated their significance for the promoter activity (Fig. 6). The importance of the spacer region sequence for efficient promoter utilization was demonstrated in a number of previous studies. In Gram-negative *Escherichia coli*, it was shown that an AT-rich spacer could both stimulate and inhibit transcription initiation, depending on the promoter (42, 76). In Gram-positive *B. subtilis* it was demonstrated for extracytoplasmic function (ECF) σ factor-dependent promoters that the homopolymeric T-tract motif, proximal to the −35 element, functioned in combination with the core promoter sequences to determine selectivity of ECF σ factors (43). This homopolymeric T-tract is reminiscent of, and might be analogous to, the σ^I extended −35 element described here. The sequence and position of the σ^I extended −10 element then differs from the TRTGn motif that was described for $\sigma^{A/70}$-dependent promoters (44), where it precedes the −10 element by 2 to 5 bp and interacts with domain 3 of $\sigma^{A/70}$ (45, 46). The extended −10 element of σ^I, however, is positioned 7 to 8 bp upstream of the −10 element, and may thus not interact with domain 3. It is possible that the A·T pairs within the "extended" promoter elements increase conformational flexibility (47, 48) of the relatively long spacer (18 to 20 bp; compared to ~17-bp σ^A promoters [49]) and thus contribute to promoter accommodation onto RNAP and to proper interactions of −35 and −10 elements with σ^I. Future studies will be required to address the detailed roles of the extended elements in recognition of these promoters by RNAP.

Cellular role of σ^I. Despite the relatively small number of genes directly regulated and indirectly affected by σ^I, this σ factor is important for proper functioning of the cell. No single gene directly regulated by σ^I is likely to be solely responsible for the growth defect at elevated temperature. Nevertheless, some of the indirectly affected genes could be responsible for this phenotype. The growth defect might be, at least in part, due to downregulation of genes involved in carbon uptake and central metabolism, such as *malA*, which encodes NAD(H)-dependent phospho-α-1,4-glucosidase (50) (the 3rd most upregulated gene in wt 37°C→52°C), and *odhA*, which encodes 2-oxoglutarate dehydrogenase, an essential enzyme of the tricarboxylic acid (TCA) cycle (51).

TABLE 1 List of strains and plasmids

Strain or plasmid	Relevant characteristics[a]	Source or reference
B. subtilis		
LK#1432 (BSB1)	BaSysBio wt 168 trp+	13
LK#1550	Δ*sigI-rsgI*::*spc*, BaSysBio	This work
MGNA-A781 (LK#1435)	Δ*rsgI*::MLS	72
LK#1456	Δ*rsgI*::MLS, BaSysBio	This work
MH5636 (LK#1275)	Bsu RNAP *rpoC*-10×His	66
DK5247 (LK#2181)	Δ*mreBH*::*kan*	73; gift from D. Kearns
LK#2191	Δ*mreBH*::*kan*, BaSysBio	This work
Escherichia coli		
LK#278	pET-22b, DH5α	This work
LK#475	pGEX-5X-3	This work
LK#1242	BL21/pSigI-6×His, BL21(DE3)	This work
LK#805 (LK22)	pCD2/Bsu_sigA, BL21(DE3)	67
LK#180 (pRLG770)	pRLG770, DH5α	74
LK#1177 (LK7558)	pRLG770 with P*veg* (−38/+1, +1G), DH5α	75
LK#1236	pRLG770 with P*bcrC*, DH5α	This work
LK#1230	pRLG770 with P*gsiB*, DH5α	This work
LK#1238	pRLG770 with P*lytE*, DH5α	This work
LK#1453	pRLG770 with P*mreBH*, DH5α	This work
LK#1452	pRLG770 with P*sigI*, DH5α	This work
LK#1766	pRLG770 with P*dhbA*, DH5α	This work

[a]MLS, macrolide-lincosamide-streptogramin B resistance; *kan*, kanamycin resistance gene; *spc*, spectinomycin resistance gene.

Furthermore, we observed a previously unreported malformed cell shape of Δ*sigI-rsgI* at 52°C. This malformation might be attributed to pronounced downregulation of *mreBH* and *lytE*, as it was previously demonstrated that both MreBH and LytE were required for proper cell shape (27, 52). Moreover, these two proteins were shown to interact in the cell (19, 27). The downregulation of both *mreBH* and *lytE* in the Δ*sigI-rsgI* strain at 52°C was quite pronounced (~16× ↓ for both genes), whereas at 37°C it was rather moderate (~3× ↓ for both genes), suggesting why the change in cell morphology was observed only at the higher temperature. The downregulation of *mreBH* alone, then, was not responsible for the impaired growth, as a Δ*mreBH* strain did not display this growth phenotype in spot assays (Fig. S2). In addition, *yfiY* and *yjeA*, which are involved in cell wall metabolism, were also downregulated, and this might have contributed to the distorted cell shape.

A significant fraction of the affected genes in the Δ*sigI* mutant were genes involved in iron metabolism. Accordingly, a growth defect, prolonged lag phase, was observed in the absence of iron at 37°C (Fig. 4). This could be due to altered iron homeostasis, as iron uptake during the lag phase is important for preparing the cell for subsequent exponential growth (53).

Finally, homologs of σ^I from *B. subtilis* exist in other *Bacillus* and *Clostridium* spp., as well as in some other species, such as *Heliobacterium* spp. and *Geobacillus* spp. (54). A phylogenetic tree of σ^I factor homologs is shown in Fig. 7. Interestingly, in different organisms, these factors regulate divergent sets of genes (thermotolerance [15, 22], virulence [55], and polysaccharide sensing [36]), although the promoter consensus sequences, at least between *Bacillus* and *Clostridium* species, are similar (36), illustrating, analogously to, e.g., eukaryotic Tcf7l1 (56), how different species tailored the use of a transcription factor to their specific needs.

MATERIALS AND METHODS

Bacterial strains and DNA manipulations. The strains and plasmids used in this study are listed in Table 1. For the study, we used *B. subtilis* 168 trp+ (BaSysBio) as the model organism (13). Competent *E. coli* cells were prepared as described previously (57); competent *B. subtilis* cells were prepared as described previously (58). PCR analysis was performed using the Expand High Fidelity PCR system (Roche). The list of the primers is shown in Table S1 in the supplemental material.

The Δ*sigI-rsgI*::spc knockout strain was prepared via double crossover. The upstream (primers 1478/sigI_rsgI_LA_F and 1479/sigI_rsgI_LA_R) and downstream (primers 1480/sigI_rsgI_RA_F and 1481/

sigI_rsgI_RA_R) regions of the *sigI-rsgI* operon were cloned into the pGEX-5X-3 plasmid vector bearing an inserted spectinomycin cassette (LK#475). The resulting plasmid (LK#1549) was transformed into BaSysBio competent cells. The final Δ*sigI-rsgI*::spc strain (LK#1550) was selected on spectinomycin plates. In the Δ*rsgI*::MLS strain (MGNA-A781; source, National BioResource Project [NBRP] *B. subtilis*, Japan), the *rsgI* gene is disrupted and not expressed. The Δ*rsgI*::MLS BaSysBio strain (LK#1456) was obtained by transformation of genomic DNA (gDNA) from the purchased MGNA-A781 strain into BaSysBio competent cells.

For overproduction of σ^I, the *sigI* gene was amplified with the primers 1164/sigI_For and 1166/sigI_Rev_His and cloned using NdeI and XhoI restriction enzymes into expression vector pET-22b (comprising the 6×His tag). The resulting plasmid, named pSigI-6×His (LK#1242), was transformed into BL21(DE3) competent cells. All promoter regions of σ^A- and σ^I-dependent genes were amplified from genomic DNA of *B. subtilis* BaSysBio with the primers 1067/PgsiB_For to 1897/PxlyA_R (Table S1) and cloned into the pRLG770 plasmid using EcoRI and HindIII restriction enzymes. Purified supercoiled plasmids for multiple-round *in vitro* transcription assays were obtained using the Wizard Plus Midipreps DNA purification system (Promega) with subsequent phenol-chloroform extraction and ethanol precipitation. The plasmids were dissolved in water. All constructs were verified by sequencing.

Media and growth conditions. For all experiments (with the exception of those testing iron requirement), precultures were inoculated from single colonies from LB agar plates and grown overnight in LB at 37°C. Overnight precultures were inoculated to fresh room temperature (RT) media to an optical density at 600 nm (OD_{600}) of 0.03 and grown at 37°C and 52°C, respectively. Precultures of Δ*sigI-rsgI*, Δ*rsgI*, and Δ*mreBH* strains were supplemented with spectinomycin (100 μg/ml), lincomycin (12.5 μg/ml) plus erythromycin (0.5 μg/ml), or kanamycin (10 μg/ml), respectively.

To monitor iron requirements, Δ*sigI-rsgI*, Δ*rsgI*, and wt strains were inoculated from a single colony on LB agar plates to MOPS medium containing or lacking iron and grown overnight at 37°C. Overnight precultures were inoculated to a fresh MOPS medium to an OD_{600} of 0.03; cultures were grown at 37°C for 24 h. Afterwards, cultures were inoculated into fresh MOPS medium at room temperature (RT) at an OD_{600} of 0.03 and grown at 37°C and 52°C. Growth was monitored at OD_{600}. MOPS medium containing iron consisted of 50 mM MOPS [3-(*N*-morpholino)propanesulfonic acid], 10 mM $(NH_4)_2SO_4$, 1 mM KH_2PO_4, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 5 μM $MnCl_2$, 0.5 μM $FeCl_3$, 50 μg/ml of each amino acid, and 0.4% glucose. MOPS medium lacking iron consisted of 50 mM MOPS, 10 mM $(NH_4)_2SO_4$, 1 mM KH_2PO_4, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 5 μM $MnCl_2$, 25 μM 2,2′-bipyridine, 50 μg/ml of each amino acid, and 0.4% glucose. To chelate iron, 2,2′-Bipyridine was added.

Spot assays. Δ*sigI-rsgI*, Δ*rsgI*, and wt strains were grown in LB broth at 37°C to mid-logarithmic phase (OD_{600} = 0.5). Serially diluted aliquots (1 μl; 10× dilutions between spots) were spotted on 1.5% LB agar plates lacking antibiotics and allowed to dry. Plates were incubated at 37°C and 52°C for 40 h. Bacterial colonies were visualized using an SZX10 stereomicroscope (Olympus) and photographed using an Olympus E-600 digital camera.

Electron microscopy. Cells were grown to exponential phase (OD_{600} = 0.5) at 37°C and 52°C and fixed in buffered 3% glutaraldehyde at 4°C. The extensively washed cells were then sedimented onto poly-L-lysine-coated glass coverslips at 4°C overnight. The coverslips were dehydrated through an alcohol series followed by absolute acetone and critical point dried in a K850 critical point dryer (Quorum Technologies Ltd., Ringmer, UK). The dried samples were sputter coated with 3 nm of platinum in a Q150T turbo-pumped sputter coater (Quorum Technologies Ltd., Ringmer, UK). The final samples were examined in an FEI Nova NanoSEM scanning electron microscope (FEI, Brno, Czech Republic) at 5 kV, using concentric backscatter (CBS) and through-the-lens (TLD) detectors. The cells were imaged according to a protocol described previously (59).

Bacterial growth for RNA sequencing. *Bacillus subtilis* Δ*sigI-rsgI* and wt strains (LK#1550 and LK#1432, respectively) were inoculated from a single colony to 10 ml of LB medium. Cultures were grown overnight at 37°C. The next day, the cultures were inoculated into 20 ml of LB medium (RT) to an OD_{600} of ~0.01 and grown at 37°C and 52°C, respectively. Cells were harvested in exponential phase (OD_{600}, ~0.5). Three ml of the culture were immediately treated with a double volume of RNAprotect bacteria reagent (Qiagen) for 5 min at RT to prevent degradation of RNA. Cells were pelleted and frozen immediately. In all steps of cultivation (with the exception of the last step of cultivation for total RNA isolation), the medium for the Δ*sigI-rsgI* strain was supplemented with spectinomycin (100 μg/ml). The experiment was repeated three times.

RNA extraction protocol. RNA was extracted using the RNeasy minikit 50 (Qiagen) according to the manufacturer's protocol. Finally, RNA was treated twice with DNase (Turbo DNA-free kit, Ambion).

RNA library construction. Two micrograms of total RNA were rRNA-depleted with RiboMinus transcriptome isolation kit, bacteria (Invitrogen). A strand-specific library was prepared for each sample with an Illumina-compatible NEXTflex rapid directional RNA-seq kit (Bioo Scientific) used according to the manufacturer's protocol.

Library sequencing. Transcriptome profiling with RNA-seq was performed at the Molecular Biology Laboratory (EMBL) Genomics Core Facility (Heidelberg, Germany). The pooled barcoded library (four samples in biological triplicates) was sequenced in a single lane using an Illumina HiSeq 2000 instrument (50 bp single end; ~8 to 16 million reads per sample regime). The quality of sequencing reads was checked with FastQC 0.11.2 (www.bioinformatics.babraham.ac.uk/projects/fastqc).

Alignment. Reads were aligned to the *Bacillus subtilis* subsp. *subtilis* strain 168 genome (NCBI nucleotide accession no. NC_000964) using Burrows-Wheeler Aligner (BWA) 0.7.9a-r786 (60) and samtools 0.1.19 (61). Alignment quality was checked using QualiMap 2.1.3 (62) and the Integrative

Genomics Viewer (IGV) 2.3 (63). *B. subtilis* genome annotation was obtained from the NCBI Assembly database (accession no. GCF_000009045.1). Analysis of differential gene expression was performed using R (www.r-project.org) and the Bioconductor package DESeq2 (64, 65) at a 5% false discovery rate (FDR). Data from RNA-seq can be found at https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6314.

Protein purification. *B. subtilis* RNAP with a 10×His-tagged β′ subunit was purified from the MH5636 strain, as described previously (66). σ^I protein purification was performed from the BL21/pSigI-6×His, BL21(DE3) strain (LK#1242) as described previously (66), with some modifications. Briefly, induction of σ^I was carried out at an OD_{600} of 0.6 with 0.05 mM IPTG (isopropyl-β-D-thiogalactopyranoside) for 3 h at RT. Cells were harvested by centrifugation, and the pellet was resuspended in P buffer (30 mM NaCl, 50 mM Na_2HPO_4, 3 mM 2-mercaptoethanol, and 1 mM PMSF [pH 7.3]) to avoid subsequent precipitation of the protein. The cells were then disrupted by sonication for 20 × 10 s on ice. Figure S9a in the supplemental material shows cell lysates before/after induction and soluble/insoluble fraction. The protein was purified from the soluble fraction by affinity chromatography using Ni-nitrilotriacetic acid (NTA)–agarose (Qiagen). After elution with 400 mM imidazole in P buffer, fractions containing σ^I were pooled and dialyzed against storage buffer (Fig. S9b in the supplemental material). The σ^A subunit of RNAP was overproduced from the pCD2 plasmid (67) (LK22) and purified as described previously (68). Proteins were dialyzed against storage buffer containing 50 mM Tris-HCl (pH 7.3 and 8.0 for σ^I and σ^A, respectively), 100 mM NaCl, 50% glycerol, and 3 mM 2-mercaptoethanol, and stored at −20°C. Proteins were visualized in NuPage 4 to 12% Bis-Tris gels (Invitrogen) with Novex Sharp prestained protein standard as a marker (Invitrogen).

***In vitro* transcription assays.** The *B. subtilis* RNA polymerase core was reconstituted with a saturating concentration of σ^A or σ^I (ratio 1:5) in storage buffer (50 mM Tris-HCl [pH 8.0], 0.1 M NaCl, and 50% glycerol) for 10 min at 37°C. Multiple-round transcription reactions were carried out in 10-μl reaction volumes with 30 nM RNAP holoenzyme and 100 ng of supercoiled plasmid DNA templates containing specific promoters (Fig. 5) or 100 ng of PCR-product templates containing promoter regions (Fig. 6). The transcription buffer contained 40 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 0.1 mg/ml bovine serum albumin (BSA), and 60 mM NaCl. ATP, CTP, and GTP were at 200 μM, and UTP was at 10 μM plus 2 μM radiolabeled [α-^{32}P]UTP. All transcription experiments were done at 37°C. Transcription reaction was started with RNAP and allowed to proceed for 15 min. Transcription was stopped with equal volumes (10 μl) of formamide stop solution (95% formamide and 20 mM EDTA [pH 8.0]). Samples were loaded onto 7 M urea-7% polyacrylamide gels and electrophoresed. Dried gels were scanned with Molecular Imager FX (Bio-Rad) and visualized and analyzed using Quantity One software (Bio-Rad). The strong constitutive σ^A-dependent P*veg* promoter cloned into pRLG770 was used as a control.

Sequence logo creation. The promoter sequence logo (Fig. 5d) was created from eight promoter regions (Fig. 5a) using the WebLogo 3 tool available online (http://weblogo.threeplusone.com/create.cgi).

Phylogenetic tree creation. Selection of σ^I homologs was based on the BLAST search (https://blast.ncbi.nlm.nih.gov/). The selected sequences were aligned using MUSCLE 3.8.31 (69) with default parameters (70). The phylogenetic tree was inferred using the RAxML BlackBox Web server (71); settings for protein sequences and maximum likelihood search were chosen, and other settings were left as default (Jones-Taylor-Thornton [JTT] substitution matrix).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/JB.00251-18.

SUPPLEMENTAL FILE 1, PDF file, 1.8 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

This work was funded by grants P305/12/G034 from the Czech Science Foundation (to L.K.), 17-29680A from the Czech Health Research Council (to L.K., bioinformatics), and 322815 from the Grant Agency of the Charles University (to O.R.). The work was also supported by the Czech research infrastructure for systems biology C4SYS (project LM2015055) and by the Charles University grants PRIMUS/MED/26 and UNCE 204013 (to M.P.). We also gratefully acknowledge the access to the electron microscopy facility supported by projects LO1509 of the Ministry of Education, Youth and Sports of the Czech Republic, and CZ.2.16/3.1.00/24023 of the Operational Program Prague-Competitiveness project, supported by the European Union.

We also thank Ivan Barvík for comments on the manuscript.

REFERENCES

1. Barvík I, Rejman D, Panova N, Šanderová H, Krásný L. 2017. Non-canonical transcription initiation: the expanding universe of transcription initiating substrates. FEMS Microbiol Rev 41:131–138. https://doi.org/10.1093/femsre/fuw041.

2. Minakhin L, Bhagat S, Brunning A, Campbell EA, Darst SA, Ebright RH, 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 U S A 98:892–897. https://doi.org/10.1073/pnas.98.3.892.
3. Weiss A, Moore BD, Tremblay MHJ, Chaput D, Kremer A, Shaw LN. 2017. The ω subunit governs RNA polymerase stability and transcriptional specificity in *Staphylococcus aureus*. J Bacteriol 199:e00459-16. https://doi.org/10.1128/JB.00459-16.
4. Lopez de Saro FJ, Woody AYM, Helmann JD. 1995. Structural analysis of the *Bacillus subtilis* δ factor: a protein polyanion which displaces RNA from RNA polymerase. J Mol Biol 252:189–202. https://doi.org/10.1006/jmbi.1995.0487.
5. Rabatinová A, Šanderová H, Matějčková JJ, Korelusová J, Sojka L, Barvík I, Veronika PapouŠková Sklenár V, Žídek L, Krásný L. 2013. The δ subunit of RNA polymerase is required for rapid changes in gene expression and competitive fitness of the cell. J Bacteriol 195:2603–2611. https://doi.org/10.1128/JB.00188-13.
6. Prajapati RK, Sengupta S, Rudra P, Mukhopadhyay J. 2016. *Bacillus subtilis* δ functions as a transcriptional regulator by facilitating the open complex formation. J Biol Chem 291:1064–1075. https://doi.org/10.1074/jbc.M115.686170.
7. Burgess RR. 1969. Separation and characterization of the subunits of ribonucleic acid polymerase. J Biol Chem 244:6168–6176.
8. Paget MS. 2015. Bacterial σ factors and anti-σ factors: structure, function and distribution. Biomolecules 5:1245–1265. https://doi.org/10.3390/biom5031245.
9. Dombroski AJ, Walter WA, Record MT, Slegele DA, Gross CA. 1992. Polypeptides containing highly conserved regions of transcription initiation factor σ^{70} exhibit specificity of binding to promoter DNA. Cell 70:501–512. https://doi.org/10.1016/0092-8674(92)90174-B.
10. Dombroski AJ, Walter WA, Gross CA. 1993. Amino-terminal amino acids modulate σ factor DNA-binding activity. Genes Dev 7:2446–2455. https://doi.org/10.1101/gad.7.12a.2446.
11. Gruber TM, Gross CA. 2003. Multiple σ subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol 57:441–466. https://doi.org/10.1146/annurev.micro.57.030502.090913.
12. Helmann JD. 2016. *Bacillus subtilis* extracytoplasmic function (ECF) σ factors and defense of the cell envelope. Curr Opin Microbiol 30:122–132. https://doi.org/10.1016/j.mib.2016.02.002.
13. Nicolas P, Mäder U, Dervyn E, Rochat T, Leduc A, Pigeonneau N, Bidnenko E, Marchadier E, Hoebeke M, Aymerich S, Becher D, Bisicchia P, Botella E, Delumeau O, Doherty G, Denham EL, Fogg MJ, Fromion V, Goelzer A, Hansen A, Härtig E, Harwood CR, Homuth G, Jarmer H. 2012. Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. Science 335:1103–1106. https://doi.org/10.1126/science.1206848.
14. McDonnell GE, Wood H, Devine KM, Mcconnell DJ. 1994. Genetic control of bacterial suicide: regulation of the induction of PBSX in *Bacillus subtilis*. J Bacteriol 176:5820–5830. https://doi.org/10.1128/jb.176.18.5820-5830.1994.
15. Zuber U, Drzewiecki K, Hecker M. 2001. Putative sigma factor SigI (YkoZ) of *Bacillus subtilis* is induced by heat shock. J Bacteriol 183:1472–1475. https://doi.org/10.1128/JB.183.4.1472-1475.2001.
16. Asai K, Ootsuji T, Obata K, Matsumoto T, Fujita Y, Sadaie Y. 2007. Regulatory role of RsgI in σ^I expression in *Bacillus subtilis*. Microbiology 153:92–101. https://doi.org/10.1099/mic.0.29239-0.
17. Salzberg LI, Powell L, Hokamp K, Botella E, Noone D, Devine KM. 2013. The WalRK (YycFG) and σ^I RsgI regulators cooperate to control CwlO and LytE expression in exponentially growing and stressed *Bacillus subtilis* cells. Mol Microbiol 87:180–195. https://doi.org/10.1111/mmi.12092.
18. Liu TY, Chu SH, Hu YN, Wang JJ, Shaw GC. 2017. Genetic evidence that multiple proteases are involved in modulation of heat-induced activation of the sigma factor SigI in *Bacillus subtilis*. FEMS Microbiol Lett 364:fnx054. https://doi.org/10.1093/femsle/fnx054.
19. Huang WZ, Wang JJ, Chen HJ, Chen JT, Shaw GC. 2013. The heat-inducible essential response regulator WalR positively regulates transcription of *sigI*, mreBH *and lytE* in *Bacillus subtilis* under heat stress. Res Microbiol 164:998–1008. https://doi.org/10.1016/j.resmic.2013.10.003.
20. Schirner K, Errington J. 2009. The cell wall regulator σ^I specifically suppresses the lethal phenotype of *mbl* mutants in *Bacillus subtilis*. J Bacteriol 191:1404–1413. https://doi.org/10.1128/JB.01497-08.
21. Tseng C-L, Shaw G-C. 2008. Genetic evidence for the actin homolog gene *mreBH* and the bacitracin resistance gene *bcrC* as targets of the alternative sigma factor SigI of *Bacillus subtilis*. J Bacteriol 190:1561–1567. https://doi.org/10.1128/JB.01497-07.
22. Tseng C-L, Chen J-T, Lin J-H, Huang W-Z, Shaw G-C. 2011. Genetic evidence for involvement of the alternative sigma factor SigI in controlling expression of the cell wall hydrolase gene *lytE* and contribution of LytE to heat survival of *Bacillus subtilis*. Arch Microbiol 193:677–685. https://doi.org/10.1007/s00203-011-0710-0.
23. Dubrac S, Bisicchia P, Devine KM, Msadek T. 2008. A matter of life and death: Cell wall homeostasis and the WalKR (YycGF) essential signal transduction pathway. Mol Microbiol 70:1307–1322. https://doi.org/10.1111/j.1365-2958.2008.06483.x.
24. Podlesek Z, Comino A, Herzog-Velikonja B, Zgur-Bertok D, Komel RGM. 1995. *Bacillus licheniformis* bacitracin-resistance ABC transporter: relationship to mammalian multidrug resistance. Mol Microbiol 16:969–976. https://doi.org/10.1111/j.1365-2958.1995.tb02322.x.
25. Cao M, Moore CM, Helmann JD. 2005. *Bacillus subtilis* paraquat resistance is directed by σ^M, an extracytoplasmic function σ factor, and is conferred by YqjL and BcrC. J Bacteriol 187:2948–2956. https://doi.org/10.1128/JB.187.9.2948-2956.2005.
26. Cao M, Helmann JD. 2002. Regulation of the *Bacillus subtilis bcrC* bacitracin resistance gene by two extracytoplasmic function σ factors. J Bacteriol 184:6123–6129. https://doi.org/10.1128/JB.184.22.6123-6129.2002.
27. Carballido-López R, Formstone A, Li Y, Ehrlich SD, Noirot P, Errington J. 2006. Actin homolog MreBH governs cell morphogenesis by localization of the cell wall hydrolase LytE. Dev Cell 11:399–409. https://doi.org/10.1016/j.devcel.2006.07.017.
28. Soufo HJD, Graumann PL. 2003. Actin-like proteins MreB and Mbl from *Bacillus subtilis* are required for bipolar positioning of replication origins. Curr Biol 13:1916–1920. https://doi.org/10.1016/j.cub.2003.10.024.
29. Kawai Y, Asai K, Errington J. 2009. Partial functional redundancy of MreB isoforms, MreB, Mbl and MreBH, in cell morphogenesis of *Bacillus subtilis*. Mol Microbiol 73:719–731. https://doi.org/10.1111/j.1365-2958.2009.06805.x.
30. Jones LJF, Carballido-López R, Errington J. 2001. Control of cell shape in bacteria: helical, actin-like filaments in *Bacillus subtilis*. Cell 104:913–922. https://doi.org/10.1016/S0092-8674(01)00287-2.
31. Margot P, Wahlen M, Gholamhuseinian A, Piggot P, Karamata D. 1998. The *lytE* gene of *Bacillus subtilis* 168 encodes a cell wall hydrolase. J Bacteriol 180:749–752.
32. Bisicchia P, Noone D, Lioliou E, Howell A, Quigley S, Jensen T, Jarmer H, Devine KM. 2007. The essential YycFG two-component system controls cell wall metabolism in *Bacillus subtilis*. Mol Microbiol 65:180–200. https://doi.org/10.1111/j.1365-2958.2007.05782.x.
33. Hashimoto M, Ooiwa S, Sekiguchi J. 2012. Synthetic lethality of the *lytE cwlO* genotype in *Bacillus subtilis* is caused by lack of D, L-endopeptidase activity at the lateral cell wall. J Bacteriol 194:796–803. https://doi.org/10.1128/JB.05569-11.
34. Völker U, Engelmann S, Maul B, Riethdorf S, Völker A, Schmid R, Mach H, Hecker M. 1994. Analysis of the induction of general stress proteins of *Bacillus subtilis*. Microbiology 140:741–752. https://doi.org/10.1099/00221287-140-4-741.
35. Mueller JP, Bukusoglu G, Sonenshein AL. 1992. Transcriptional regulation of *Bacillus subtilis* glucose starvation-inducible genes: control of *gsiA* by the ComP-ComA signal transduction system. J Bacteriol 174:4361–4373. https://doi.org/10.1128/jb.174.13.4361-4373.1992.
36. Muñoz-Gutiérrez I, Ortiz de Ora L, Rozman Grinberg I, Garty Y, Bayer EA, Shoham Y, Lamed R, Borovok I. 2016. Decoding biomass-sensing regulons of *Clostridium thermocellum* alternative sigma-I factors in a heterologous *Bacillus subtilis* host system. PLoS One 11:e0146316. https://doi.org/10.1371/journal.pone.0146316.
37. Jarmer H, Larsen TS, Krogh A, Saxild HH, Brunak S, Knudsen S. 2001. σ^A recognition sites in the *Bacillus subtilis* genome. Microbiology 147:2417–2424. https://doi.org/10.1099/00221287-147-9-2417.
38. Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E, Bourexis D, Brister JR, Bryant SH, Canese K, Charowhas C, Clark K, Dicuccio M, Dondoshansky I, Federhen S, Feolo M, Funk K, Geer LY, Gorelenkov V, Hoeppner M, Holmes B, Johnson M, Khotomlianski V, Kimchi A, Kimelman M, Kitts P, Klimke W, Krasnov S, Kuznetsov A, Landrum MJ, Landsman D, Lee JM, Lipman DJ, Lu Z, Madden TL, Madej T, Marchler-Bauer A, Karsch-Mizrachi I, Murphy T, Orris R, Ostell J, O'sullivan C, Panchenko A, Phan L, Preuss D, Pruitt KD, Rodarmer K, Rubinstein W, Sayers E, Schneider V, Schuler GD, Sherry ST, Sirotkin K, Siyan K, Slotta D, Soboleva A, Soussov V, Starchenko G, Tatusova TA, Todorov K, Trawick BW, Vakatov D, Wang Y, Ward M, Wilbur WJ, Yaschenko E, Zbicz K. 2016. Database

resources of the national center for biotechnology information. Nucleic Acids Res 44:D7–D19. https://doi.org/10.1093/nar/gkv1290.
39. May JJ, Wendrich TM, Marahiel MA. 2001. The *dhb* operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin. J Biol Chem 276:7209–7217. https://doi.org/10.1074/jbc.M009140200.
40. Baichoo N, Wang T, Ye R, Helmann JD. 2002. Global analysis of the *Bacillus subtilis* Fur regulon and the iron starvation stimulon. Mol Microbiol 45:1613–1629. https://doi.org/10.1046/j.1365-2958.2002.03113.x.
41. Heath RJ, Su N, Murphy CK, Rock CO. 2000. The enoyl-[acyl-carrier-protein] reductases FabI and FabL from *Bacillus subtilis*. J Biol Chem 275:40128–40133. https://doi.org/10.1074/jbc.M005611200.
42. Hook-Barnard IG, Hinton DM. 2009. The promoter spacer influences transcription initiation via σ^{70} region 1.1 of *Escherichia coli* RNA polymerase. Proc Natl Acad Sci U S A 106:737–742. https://doi.org/10.1073/pnas.0808133106.
43. Gaballa A, Guariglia-Oropeza V, Dürr F, Butcher BG, Chen AY, Chandrangsu P, Helmann JD. 2018. Modulation of extracytoplasmic function (ECF) sigma factor promoter selectivity by spacer region sequence. Nucleic Acids Res 46:134–145. https://doi.org/10.1093/nar/gkx953.
44. Voskuil MI, Chambliss GH. 1998. The −16 region of *Bacillus subtilis* and other Gram-positive bacterial promoters. Nucleic Acids Res 26:3584–3590. https://doi.org/10.1093/nar/26.15.3584.
45. Ruff E, Record M, Artsimovitch I. 2015. Initial events in bacterial transcription initiation. Biomolecules 5:1035–1062. https://doi.org/10.3390/biom5021035.
46. Mitchell JE, Zheng D, Busby SJW, Minchin SD. 2003. Identification and analysis of "extended −10" promoters in *Escherichia coli*. Nucleic Acids Res 31:4689–4695. https://doi.org/10.1093/nar/gkg694.
47. Johnson S, Chen YJ, Phillips R. 2013. Poly(dA:dT)-rich DNAs are highly flexible in the context of DNA looping. PLoS One 8:e75799. https://doi.org/10.1371/journal.pone.0075799.
48. Okonogi TM, Alley SC, Reese AW, Hopkins PB, Robinson BH. 2002. Sequence-dependent dynamics of duplex DNA: the applicability of a dinucleotide model. Biophys J 83:3446–3459. https://doi.org/10.1016/S0006-3495(02)75344-7.
49. Helmann JD. 1995. Compilation and analysis of *Bacillus subtilis* σ^{A}-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. Nucleic Acids Res 23:2351–2360. https://doi.org/10.1093/nar/23.13.2351.
50. Thompson J, Pikis A, Ruvinov SB, Henrissat B, Yamamoto H, Sekiguchi J. 1998. The gene *glvA* of *Bacillus subtilis* 168 encodes a metal-requiring, NAD(H)-dependent 6-phospho-α-glucosidase. J Biol Chem 273:27347–27356. https://doi.org/10.1074/jbc.273.42.27347.
51. Carlsson P, Hederstedt L. 1989. Genetic characterization of *Bacillus subtilis odhA* and *odhB*, encoding 2-oxoglutarate dehydrogenase and dihydrolipoamide transsuccinylase, respectively. J Bacteriol 171:3667–3672. https://doi.org/10.1128/jb.171.7.3667-3672.1989.
52. Domínguez-Cuevas P, Porcelli I, Daniel RA, Errington J. 2013. Differentiated roles for MreB-actin isologues and autolytic enzymes in *Bacillus subtilis* morphogenesis. Mol Microbiol 89:1084–1098. https://doi.org/10.1111/mmi.12335.
53. Rolfe MD, Rice CJ, Lucchini S, Pin C, Thompson A, Cameron ADS, Alston M, Stringer MF, Betts RP, Baranyi J, Peck MW, Hinton JCD. 2012. Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. J Bacteriol 194:686–701. https://doi.org/10.1128/JB.06112-11.
54. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733–D745. https://doi.org/10.1093/nar/gkv1189.
55. Kim JGY, Wilson AC. 2016. Loss of σ^{I} affects heat-shock response and virulence gene expression in *Bacillus anthracis*. Microbiology 162:564–574. https://doi.org/10.1099/mic.0.000236.
56. Mašek J, Machoň O, Kořínek V, Taketo MM, Kozmik Z. 2016. Tcf7l1 protects the anterior neural fold from adopting the neural crest fate. Development 143:2206–2216. https://doi.org/10.1242/dev.132357.
57. Hanahan D. 1983. Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol 166:557–580. https://doi.org/10.1016/S0022-2836(83)80284-8.
58. Dubnau D, Davidoff-Abelson R. 1971. Fate of transforming DNA following uptake by competent *Bacillus subtilis*: I. Formation and properties of the donor-recipient complex. J Mol Biol 56:209–221.
59. Seydlová G, Pohl R, Zborníková E, Ehn M, Šimák O, Panova N, Kolář M, Bogdanová K, Večeřová R, Fišer R, Šanderová H, Vítovská D, Sudzinová P, Pospíšil J, Benada O, Křížek T, Sedlák D, Bartůněk P, Krásný L, Rejman D. 2017. Lipophosphonoxins II: design, synthesis, and properties of novel broad spectrum antibacterial agents. J Med Chem 60:6098–6118. https://doi.org/10.1021/acs.jmedchem.7b00355.
60. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
61. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
62. Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32:292–294. https://doi.org/10.1093/bioinformatics/btv566.
63. Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192. https://doi.org/10.1093/bib/bbs017.
64. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8.
65. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Corrada Bravo H, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Ole AK, Pages H, Reyes A, MM. 2015. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods 12:115–121. https://doi.org/10.1038/nmeth.3252.
66. Qi Y, Hulett FM. 1998. Pho∼P and RNA polymerase σ^{A} holoenzyme are sufficient for transcription of Pho regulon promoters in *Bacillus subtilis*: PhoP∼P activator sites within the coding region stimulate transcription in vitro. Mol Microbiol 28:1187–1197. https://doi.org/10.1046/j.1365-2958.1998.00882.x.
67. Chang B, Doi R. 1990. Overproduction, purification, and characterization of *Bacillus subtilis* RNA polymerase σ^{A} factor. J Bacteriol 172:3257-3263. https://doi.org/10.1128/jb.172.6.3257-3263.1990.
68. Juang Y-L, Helmann JD. 1994. A promoter melting region in the primary σ factor of *Bacillus subtilis*. J Mol Biol 235:1470–1488. https://doi.org/10.1006/jmbi.1994.1102.
69. Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
70. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191. https://doi.org/10.1093/bioinformatics/btp033.
71. Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst Biol 57:758–771. https://doi.org/10.1080/10635150802429642.
72. Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P, Boland F, Brignell SC, Bron S, Bunai K, Chapuis J, Christiansen LC, Danchin A, Debarbouille M, Dervyn E, Deuerling E, Devine K, Devine SK, Dreesen O, Errington J, Fillinger S, Foster SJ, Fujita Y, Galizzi A, Gardan R, Eschevins C, Fukushima T, Haga K, Harwood CR, Hecker M, Hosoya D, Hullo MF, Kakeshita H, Karamata D, Kasahara Y, Kawamura F, Koga K, Koski P, Kuwana R, Imamura D, Ishimaru M, Ishikawa S, Ishio I, Le Coq D, Masson A, Mauel C, Meima R, Mellado RP, Moir A, Moriya S, Nagakawa E, Nanamiya H, Nakai S, Nygaard P, Ogura M, Ohanan T, O'Reilly M, O'Rourke M, Pragai Z, Pooley HM, Rapoport G, Rawlins JP, Rivas LA, Rivolta C, Sadaie A, Sadaie Y, Sarvas M, Sato T, Saxild HH, Scanlan E, Schumann W, Seegers JFML, Sekiguchi J, Sekowska A, Seror SJ, Simon M, Stragier P, Studer R, Takamatsu H, Tanaka T, Takeuchi M, Thomaides HB, Vagner V, van Dijl JM, Watabe K, Wipat A, Yamamoto H, Yamamoto M, Yamamoto Y, Yamane K, Yata K, Yoshida K, Yoshikawa H, Zuber U, Ogasawara N. 2003. Essential *Bacillus subtilis* genes. Proc Natl Acad Sci U S A 100:4678–4683. https://doi.org/10.1073/pnas.0730515100.
73. Koo BM, Kritikos G, Farelli JD, Todor H, Tong K, Kimsey H, Wapinski I,

Galardini M, Cabal A, Peters JM, Hachmann AB, Rudner DZ, Allen KN, Typas A, Gross CA. 2017. Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis*. Cell Syst 4:291–305. https://doi.org/10.1016/j.cels.2016.12.013.

74. Ross W, Thompson JF, Newlands JT, Gourse RL. 1990. *E.coli* Fis protein activates ribosomal RNA transcription *in vitro* and *in vivo*. EMBO J 9:3733–3742.
75. Krásný L, Gourse RL. 2004. An alternative strategy for bacterial ribosome synthesis: *Bacillus subtilis* rRNA transcription regulation. EMBO J 23: 4473–4483. https://doi.org/10.1038/sj.emboj.7600423.
76. Kolasa IK, Łoziński T, Wierzchowski KL. 2003. Effects of distortions by A-tracts of promoter B-DNA spacer region on the kinetics of open complex formation by *Escherichia coli* RNA polymerase. Acta Biochim Pol 50:909–920.

BBA - Gene Regulatory Mechanisms 1860 (2017) 894–904

Contents lists available at ScienceDirect

BBA - Gene Regulatory Mechanisms

journal homepage: www.elsevier.com/locate/bbagrm

Kinetic modelling and meta-analysis of the *B. subtilis* SigA regulatory network during spore germination and outgrowth

O. Ramaniuk[a,c], M. Černý[b], L. Krásný[a], J. Vohradský[b,*]

[a] Laboratory of Microbial Genetics and Gene Expression, Institute of Microbiology CAS, v.v.i., Videnska 1083, 14220 Prague, Czech Republic
[b] Laboratory of Bioinformatics, Institute of Microbiology CAS, v.v.i., Videnska 1083, 14220 Prague, Czech Republic
[c] Department of Genetics and Microbiology, Faculty of Science, Charles University, Viničná 5, CZ-12843 Prague 2, Czech Republic

ARTICLE INFO

Keywords:
Sigma A
Kinetic modelling
Regulatory network
Gene expression
Bacillus subtilis

ABSTRACT

This study describes the meta-analysis and kinetic modelling of gene expression control by sigma factor SigA of *Bacillus subtilis* during germination and outgrowth based on microarray data from 14 time points. The analysis computationally models the direct interaction among SigA, SigA-controlled sigma factor genes (*sigM*, *sigH*, *sigD*, *sigX*), and their target genes. Of the > 800 known genes in the SigA regulon, as extracted from databases, 311 genes were analysed, and 190 were confirmed by the kinetic model as being controlled by SigA. For the remaining genes, alternative regulators satisfying kinetic constraints were suggested. The kinetic analysis suggested another 214 genes as potential SigA targets. The modelling was able to (i) create a particular SigA-controlled gene expression network that is active under the conditions for which the expression time series was obtained, and where SigA is the dominant regulator, (ii) suggest new potential SigA target genes, and (iii) find other possible regulators of a given gene or suggest a new mechanism of its control by identifying a matching profile of unknown regulator(s). Selected predicted regulatory interactions were experimentally tested, thus validating the model.

1. Introduction

Bacillus subtilis is a model soil-dwelling, spore-forming organism. Its gene expression is controlled by various sigma factors [1] that, depending on the conditions, associate with RNA polymerase (RNAP). Sigma factors allow RNAP to recognize specific promoter sequences to initiate transcription of target genes. *B. subtilis* contains 19 different sigma factors, including the main sigma factor, SigA [2].

The past several years have witnessed an explosion in the amount of available experimental evidence from transcriptomic studies [3]. Much of the information contained in these studies is still under-exploited, and our understanding of regulatory networks in *B. subtilis* is far from complete. Mining transcriptomic databases offers an opportunity to provide new insights into sigma factor-controlled networks.

A crucial task involved in inferring gene regulatory networks in bacteria is the identification of the target genes of sigma factors. There are two main methods to discover such target genes: chromatin immunoprecipitation (ChIP) experiments and gene expression analysis. ChIP methods (ChIP-chip and ChIP-seq) test for physical interactions between sigma factors and gene promoter sequences. However, it has been shown that this static binding information may also include silent binding events that do not directly enhance transcription [4,5]. To increase certainty, ChIP experiments are complemented with RNAseq experiments in strains with deletions in the sigma factors of interest. However, these deletions are not possible for essential sigma factors, for which a different approach must be employed.

One such approach is the kinetic modelling of gene expression, which is a highly useful tool for discovering regulatory networks. Various methods to infer gene regulatory networks from gene expression data have been suggested, based on ordinary and stochastic differential equations, neural networks, dynamic Bayesian networks, and information theoretic- or correlation-based methods, which have been reviewed by Bansal et al., Penfold and Wild [6,7] and Bar-Joseph [8]. Similar to ChIPseq, kinetic modelling alone is not sufficient to reliably determine regulation networks, and multiple sources of information have to be combined.

Currently, in addition to experimental papers, several databases of regulatory interactions based on literature mining have emerged. These databases are non-specific in the sense of particular developmental processes of bacteria. They collect information about regulatory interactions between the regulator and its targets, regardless of the conditions under which the particular experiment was made. A combination

* Corresponding author.
E-mail addresses: ramaniuk@biomed.cas.cz (O. Ramaniuk), martin.cerny@biomed.cas.cz (M. Černý), krasny@biomed.cas.cz (L. Krásný), vohr@biomed.cas.cz (J. Vohradský).

http://dx.doi.org/10.1016/j.bbagrm.2017.06.003
Received 21 March 2017; Received in revised form 9 June 2017; Accepted 20 June 2017
Available online 22 June 2017
1874-9399/

of kinetic expression modelling obtained for particular experimental conditions with static and databased data may provide new insight into the kinetics of the control of sigma factors and their target genes and consequently allow modelling of gene expression kinetics in the sigma factor-controlled network.

In this study, we focused on the regulatory network of the *B. subtilis* primary sigma factor SigA, which has been reported to control over 800 genes, including genes for 8 alternative sigma factors. SigA is the main, vegetative, and essential sigma factor that is active mainly in the exponential phase and is responsible for transcription of housekeeping genes [9]. Here, based on 14 data points from germination and subsequent outgrowth of *B. subtilis* cells [10] we created a kinetic model of the SigA-controlled regulatory network under these conditions. The goal was to provide insight into how many and which genes are directly regulated by this sigma factor without the need to invoke additional regulatory layers. We also predicted new potential targets of SigA and provided experimental verification for some of them. For genes that did not match the SigA kinetic profile (mRNA encoding SigA), we suggested alternative sigma factors that were capable of modelling their expression profiles.

2. Materials and methods

2.1. Data acquisition

2.1.1. Time series of gene expression

We downloaded the *B. subtilis*transcriptomic microarray data from 14 time points (0, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90 and 100 min) obtained during germination and outgrowth as previously reported [10] from GEO http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE6865. Briefly, the generation of *Bacillus subtilis* 168 spores was induced by the depletion of defined MOPS medium during 4 days of shaking at 37 °C. Subsequently, spores were activated in the germination medium by thermal treatment at 70 °C for 30 min. The release of dipicolinic acid in the medium during spore germination was monitored using the terbium fluorescence assay. During germination and outgrowth, samples for RNA isolation were drawn at regular intervals. RNA was isolated from spores and outgrowing spores and then reverse transcribed to cDNA. Cy-labelled cDNA was made by the direct incorporation of Cy-labelled dUTP. Samples were hybridized to microarray slides, and microarrays were scanned using an Agilent G2505 scanner. Data were averaged over repeated samples. For further processing, the original log2-based data were exponentiated.

2.1.2. Sigma factor regulons

The sigA regulon genes were downloaded from SubtiWiki (http://www.subtiwiki.uni-goettingen.de/) [2] and DBTBS (dbtbs.hgc.jp) [11]. The database contains a collection of experimentally validated gene regulatory relations of *B. subtilis* genes constructed by surveying literature references. Among the SigA target genes (850), eight other sigma factors were found (SigD, SigH, SigM, SigX, YlaC, SigE, SigF and SigG). SigE, SigF, SigG and YlaC were excluded from the analysis because their expression profiles were too low and could therefore be subject to high experimental variance, which could lead to misinterpretations of the modelling results. For the other alternative sigma factors, their regulons were downloaded from SubtiWiki — SigD (73 genes), SigH (48 genes), SigM (84 genes) and SigX (31 genes). Some genes were members of more than one regulon. Altogether, a list of 1087 genes was compiled. The dataset contains time series of 4008 genes.

2.2. Kinetic model of gene expression

A kinetic model of gene expression controlled by a sigma factor that was originally developed by Vohradsky [12] and further revised and extended [13–15] was used in this study. The model was derived from the assumption that the mRNA level of a gene controlled by a sigma factor is determined by the concentration of the sigma factor binding in complex with RNA polymerase to the promoter region. The probability of the sigma factor binding to the gene promoter is determined by the sigma factors' binding strength and the number of molecules around the promoter. Transcription is a discontinuous process that depends on the actual binding of the holoenzyme to the promoter. When the number of sigma factors molecules is low, the probability of triggering transcription of a given gene is also low. With increasing amounts of sigma factor molecules, the probability of a gene transcription event increases until the promoter is saturated and the expression rate becomes constant. The relation between the accumulation of transcribed mRNA and sigma factor concentration can thus be described mathematically by a sigmoid with parameters reflecting the strength of binding, reaction delay and mRNA degradation rate. The sigmoidal shape of the function was also confirmed by the results of stochastic simulations (e.g., [16–19]). The model used in this study has the following form:

$$\frac{dy_i}{dt} = \frac{k_{1i}}{1 + \exp[-w_i R(t + \Delta t)_j + b_i]} - k_{2i} y_i \quad (1)$$

where y_i represents the concentration of the genes mRNA and R_j is the concentration of the j-th sigma factor modulated by parameter w_i, which corresponds to binding strength to the promoter. The b_i and Δt parameters correspond to the reaction delay. The accumulation of the mRNA of the gene i is diminished by degradation described by the term $k_{2i}y_i$.

Since the expression data were noisy, we smoothed the data prior to computation with a piecewise cubic spline with 6 knots (the best number of knots was determined empirically) [20]. By smoothing, the results were more robust with respect to the low-frequency phenomena expected in gene expression data. A further advantage of smoothing is that it lets us subsample the fitted curve at arbitrary resolution. We subsampled the profiles at 1-minute time steps, which allowed us to integrate (Eq. (1)) accurately with a computationally cheap Euler method. The parameters of the model for individual sigma factor-transcribed gene combinations were optimized using a simulated annealing scheme by minimizing an objective function

$$E = \sqrt{\sum (\mathbf{y} - \tilde{\mathbf{y}})^2} \quad (2)$$

where y represents the measured mRNA concentration time series proportional value and $\tilde{y}$ represents the time series computed using the model Eq. (1). For each profile, optimization was repeated 100 times with random values as estimates of the initial parameters, and those parameters that gave the smallest E were selected from the 100 runs. The expression values of the sigma factor-transcribed genes from the 14 time points were provided. This data set was subsequently analysed. The goal was to identify parameters that would give the best fit of the model to the actual profile of a given regulated gene with the sigA (or other considered sigma factor) profile as the regulator. The regulatory interaction between a sigma factor and a gene was accepted; i.e., the control of the transcribed gene by the given sigma factor was considered possible, if the profile $\tilde{\boldsymbol{y}}$ computed with the best set of parameters was within the confidence interval of the measured profile (y) in at least 12 measured time points of the profile (this constraint was chosen to minimize the influence of the first and last time points that have the highest experimental and spline fitting errors). When the confidence interval could not be determined from the experimental data, a flat value of 20% of profiles maximum was used as the confidence interval.

2.3. Experimental validation

In the following two sub-sections, the strains, plasmids and experimental conditions used for experimental verification of computational predictions are listed.

Table 1
List of strains and plasmids.

Strain/plasmid	Relevant characteristics	Source
B. subtilis		
BSB1	wt BaSysBio	[3]
MH5636	*rpoC*-His10	[22]
E. coli		
LK22	BL21 pCD2/Bsu_*sigA*	[23]
LK1	pRLG770 with P*veg* (−38/−1, +1G)	[24]

2.3.1. Bacterial strains and plasmids

Strains and plasmids are listed in Table 1.

Competent *E. coli* cells, protein-expression strain BL21 (DE3), were prepared according to methods described by Hanahan [21].

2.3.2. Media and growth conditions

For protein purification, the strains were cultured in Luria-Bertani (LB) medium at 37 °C with continuous shaking.

2.3.2.1. Proteins purification. *Bacillus subtilis* RNAP with a His10-tagged β′ subunit was purified from the MH5636 strain. The purifications were performed as described previously [22].

SigA was overexpressed from plasmid pCD2 [23] and purified as described previously [25].

2.3.2.2. Transcription template preparation

Linear PCR products of putative promoter sequences were used as templates for *in vitro* transcription assays. The primers are listed in Table 1 and in Supplemental Table 1. Putative promoter sequences were PCR-amplified using wt *B. subtilis* gDNA as the template. The only exception was the control P*veg* promoter, which was amplified from the LK1 plasmid containing a cloned P*veg* fragment. The wt P*veg* promoter starts with an adenosine. Here, we used a variant that starts with guanosine. This alteration does not change the properties of the promoter [24]. All PCR reactions were performed using the Expand High Fidelity System (Roche). The purification of PCR constructs was performed using Agencourt AMPure XP beads (Beckman Coulter) according to the manufacturer's protocol. All constructs were verified by sequencing.

2.3.2.3. In vitro transcription assays. The RNAP holoenzyme was reconstituted with a saturating concentration of SigA. Reconstitutions were performed in a glycerol storage buffer (50 mM Tris–HCl, pH 8.0, 0.1 M NaCl, 50% glycerol) for 10 min at 37 °C.

Multiple round transcription reactions were carried out in 10-µl reaction volumes with 30 nM RNAP holoenzyme and 50 ng of linear DNA template. The transcription buffer contained 40 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 0.1 mg/ml BSA (bovine serum albumin) and 60 mM NaCl. ATP, CTP, and GTP were at 200 µM, and UTP was at 10 µM plus 2 µM radiolabelled [α-^{32}P]UTP.

All transcription experiments were done at 37 °C. Transcription was induced by adding reconstituted RNAP holoenzyme and allowed to proceed for 15 min. Transcription was stopped with equal volumes (10 µl) of formamide stop solution (95% formamide, 20 mM EDTA [pH 8.0]). Samples were loaded onto 7 M urea–7% polyacrylamide gels and electrophoresed. The dried gels were scanned with Molecular Imager FX (Bio-Rad) and were visualized and analysed using the Quantity One software (Bio-Rad).

2.4. Creation of promoter sequence logos

The promoter sequence logos (Fig. 7) were created using WebLogo 3 tool available online (http://weblogo.threeplusone.com/create.cgi). Sequences that served as templates for the logos are in Supplemental Tables 6 and 7.

3. Results

3.1. SigA regulon

To begin describing the SigA regulatory network in *B. subtilis*, we used transcriptomic microarray data from 14 time points (0, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90 and 100 min) obtained during germination and outgrowth as previously reported [10]. Subsequently, we extracted the known SigA regulon genes from SubtiWiki and DBTBS, and created and analysed their kinetic profiles based on the gene expression time series from the work of Keijser et al. [10] and the model of gene expression presented in Materials and methods.

The list of all SigA regulon genes analysed, together with the model parameters and analysis results, are shown in Supplemental Table 2. The analysis was performed on respective mRNAs, but for more intuitive understanding, we refer in the text to the regulators (sigma factors) as proteins. Prior to the modelling step, the expression profiles of the SigA regulon were checked for consistency, and genes with low values of maximal expression were excluded. This was necessary because profiles with low expression values had high experimental variance, and further analysis of these profiles could lead to misinterpretation. Additionally, 65 genes with a flat profile, i.e., those that were constitutively expressed, were excluded. Altogether, we excluded 539 genes. These genes may or may not be controlled by SigA, but we could not perform a reliable kinetic analysis on them.

The remaining 311 profiles were subjected to a kinetic analysis, as described in the Materials and methods. The parameters of the model allowed us to define the mode of action of SigA (and other studied sigma factors) according to the value and sign of the parameter w. The parameter w characterizes how much and in which direction (positive, negative) a given sigma factor affects the expression of a target gene. The sign of w determines how the controlled gene responds to the changing expression level of the sigma factor. If the sign is positive, the expression level of the controlled gene increases with increasing levels of the sigma factor; if the sign is negative, the expression level of the controlled gene decreases, and the expression profile is inversely correlated with the expression profile of the sigma factor (for examples, see Fig. 1).

Mechanistic explanations can be arbitrary, and include other regulatory effects that occur pre- or post-transcriptionally. Although such mechanisms can, in some cases, be included in the model if their nature is known, such events are generally not covered by the model and are only reflected in the sign of the parameter w. Possible models of negative control ($w < 0$) are discussed for specific cases below. The fit of the model to the measured expression of the controlled gene was assessed for goodness of fit, and those model parameters fitting the experimental profile of the regulated gene within the confidence interval were selected. The results of the kinetic analysis revealed two principal groups of genes: (i) those whose profiles could be modelled with SigA as the regulator (the model expression profile fitted the measured expression profile within confidence interval, Table 2) and (ii) those which could not (the model could not fit the measured expression profile, Table 3). A total of 190 genes associated with 113 operons could be modelled using SigA as the regulator, and 121 (from 56 operons) could not be modelled. Of the 190 genes that were found to be controlled by SigA, 156 genes (organized in 89 operons) were found to be positively controlled by SigA, and 35 genes (organized in 22 operons) were negatively controlled by SigA (see Table 2).

The expression profiles of those genes found to be positively controlled by SigA are shown in Fig. 2A, those that were found not to be controlled by SigA are shown in Fig. 2B. The profiles in Fig. 2 were, for display purposes, normalized to have zero mean and the same standard deviation.

Either the expression profiles of the positively controlled group

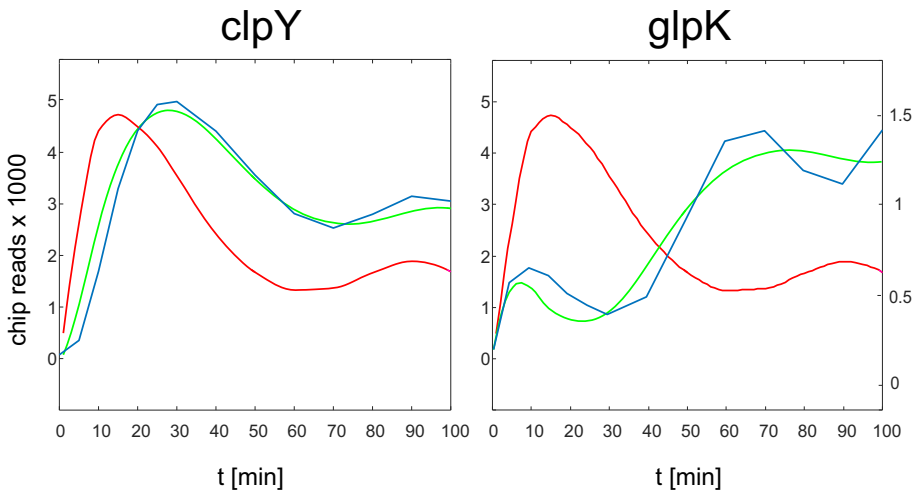


Fig. 1. An example of positively (left, clpY) and negatively (right, glpK) controlled genes during germination and outgrowth. Red — sigA mRNA profile; blue — primary values (relative RNA levels) for target gene profiles; green — modelled profiles of the target gene.

Table 2
List of two groups of genes found to be controlled by SigA consistent with the literature and the results of modelling. The two groups differ in the sign of the parameter w (positive or negative). Complete operons are listed; genes found by modelling are shown in bold. The genes that are not in bold did not satisfy the modelling constraints.

Positive control $w > 0$

*adhA-**yraA**, aroF-aroB-aroH-trpE-trpD-trpC-trpF-trpB-**trpA-hisC-tyrA**-aroE, **cggR**-gapA-pgk-tpi-pgm-eno, **citG**, citZ-**icd-mdh**, ctsR-mcsA-mcsB-**clpC**-radA-yacK, **codV-clpQ-clpY-codY**, **clpX**, **comA**-yuxO, comEA-**comEB**-comEB, **comQ-comX**, cymR-yrvO-**trmU**, **cspB**, **ctaA**, **drm-punA**, epsA-epsB-epsC-**yveN**-epsE-epsF- -epsG-epsH-epsI-epsJ-epsK-epsL-epsM-epsN-epsO, fadF-acdA-**rpoE**, **rpsD**, **ftsA-ftsZ**, **ftsH**, **fabHA-fabF**, **fbp**, **fur**, **galE**, **gcaD-prs**-ctc, **glnR**-glnA, **gltA-gltB**, **gudB**, **gyrA**, **gyr-recF-yaaB**-gyrB, hemA-hemX-**hemC-hemD-hemB-hemL**, **htpG**, ilvB-ilvH-**ilvC**-leuA-leuB-leuC-leuD, **infC-rpmI-rplT**-ysdA, **lepA-hemN**-hrcA-**grpE**- -dnaK-**dnaJ-yqeT-yqeU-yqeV**, lcfA-**ysiA-ysiB-etfB**-etfA, **lonA-ysxC**, **lytA**-lytB-**lytC**, **lytR**, med-**comZ**, **menF-menD**-ytxM-**menB**-menE-menC, **mtrA-mtrB**, **murE-mraY-murD-spoVE-murG-murB-divIB**-ylxW-**ylxX**-sbp, nifS-**yrxA**, **nfrA**-ywcH, opuAA-**opuAB-opuAC**, pabB-pabA-pabC-**sul-folB**-folK-yazB-yacF-**lysS**, **pheS**-pheT, **ppiB**, **pssA-ybfM-psd**, ptsG-**ptsH-ptsI**, **pucA**-pucB-pucC-pucD-pucE, **pyrG**, **recA**, **resA**-resB-resC-resD-resE, **rnc**-smc-**ftsY**, **rpmH**, **rpoB**, secA-**prfB**, **sigX-rsiX**, **sigM-yhdL**, sinI-**sinR**, **sipS**, **speE-speB**, gapB-**speD**, **spo0A**, **spo0B-obg**, **spo0E**, spoIIIJ-**jag**, spoVFA-spoVFB-asd-**dapG-dapA**, **thrS**, **thyB-dfrA-ypkP**, **veg**, valS-**folC**, **ydjM**, **yfhQ-fabL**-sspE, **yfkJ-yfkI-yfkH**, **ykuJ-ykuK-ykzF-ykuL**-ccpC, **ykuN-ykuO-ykuP**, **yocH**, ylpC-plsX-fabD-**fabG**-acpA, ylxM-**ffh**, ylxS-nusA-ylxR-ylxQ-**infB**-ylxP-**rbfA**, **ymaA-nrdE-nrdF-ymaB**, ypuE-ribD-ribE-ribA-ribH-**ribT**, **yusM**-yusL-yusK-yusJ, **yuxH**, **yqxD-dnaG-sigA**, **yvcE***

Negative control $w < 0$

*arsR-**yqcK**-arsB-**arsC**, **citB**, **citZ**-icd-mdh, ctsR-mcsA-**mcsB**-clpC-radA-yacK, **cysH-cysP-sat-cysC-ylnD-ylnE-ylnF**, glpF-**glpK**, iolR-**iolS**, **odhA-odhB**, **pbpD**-yuxK, **pckA**, pheS-**pheT**, pucA-**pucB**-pucC-pucD-pucE, **pucR**-pucJ-pucK-pucL-pucM, **purT**, **rsbR**-rsbS-rsbT-rsbU-rsbV-rsbW-sigB-rsbX, **sdhC-sdhA-sdhB**, **xylA**-xylB, ycdH-ycdI-**yceA**, **ykrT-ykrS**, ytrG-ytrA-ytrB-ytrC-ytrD-**ytrE-ytrF**, yusM-yusL-**yusK**-yusJ, **yvgR-yvgQ***

Table 3
List of genes of the SigA regulon identified in the literature that could not be modeled with SigA as regulator. Complete operons are listed, and the genes found by modelling are shown in bold. The genes that are not in bold did not satisfy the modelling constraints.

ackA**, **ahpC-ahpF**, **ald**, alsS-**alsD**, **ansA-ansB**, **clpE**, **clpP**, gcaD-prs-**ctc**, **ctsR-mcsA**-mcsB-clpC-radA-**yacK**, **cysK**, degS-**degU**, **dnaA-dnaN**, gltX-cysE-cysS-**yazC**-yacO-yacP, **glyA**, **groES-groEL**, **gtaB**, **guaA**, **guaD**, **hemA-hemX**-hemC-hemD-hemB-hemL, **hemZ**, lepA-hemN-**hrcA**-grpE-dnaK-dnaJ-yqeT-yqeU-yqeV, **lmrA-lmrB**, **mecA**, murE-mraY-murD-spoVE-murG-murB-divIB-**ylxW**-ylxX-**sbp**, **nadE**, **opuAA**-opuAB-opuAC, **pabB-pabA-pabC**-sul-folB-**folK-yazB**-yacF-lysS, **pta**, **ptsG**-ptsH-ptsI, pucR-pucJ-pucK-**pucL**-pucM, **purA**, **purE-purK-purB-purC-purS-purQ-purL-purF-purM-purN-purH-purD**, **pyrR-pyrP**- -**pyrB-pyrC-pyrAA-pyrAB-pyrK-pyrD-pyrF-pyrE**, rbsR-rbsK-rbsD-rbsA-rbsC-**rbsB**, rnc-**smc**-ftsY, rsbR-**rsbS-rsbT-rsbU-rsbV-rsbW-sigB-rsbX**, **secA**-prfB, **tagA-tagB**, **tagD-tagE-tagF**, **xpt-pbuX**, **ycdA**, **ycdH-ycdI**-yceA, **yceC-yceD**-yceE-yceF-yceG-yceH, **yhaG**, **yhcL**, **yjbC-yjbD**, **yjeA**, **ylxM**-ffh, **ylxS-nusA-ylxR-ylxQ**-infB-**ylxP**-rbfA, **yoeB**, **ylpC-plsX-fabD**-fabG-acpA, ypuE-**ribD-ribE-ribA-ribH**-ribT, **yrrT-mtnN-yrhA-yrhB-yrhC**, **yvbA-yvaZ**, **yxeK-yxeL-yxeM-yxeN-yxeO-yxeP**-yxeQ, wapA-wapI-wapA-**yxxG

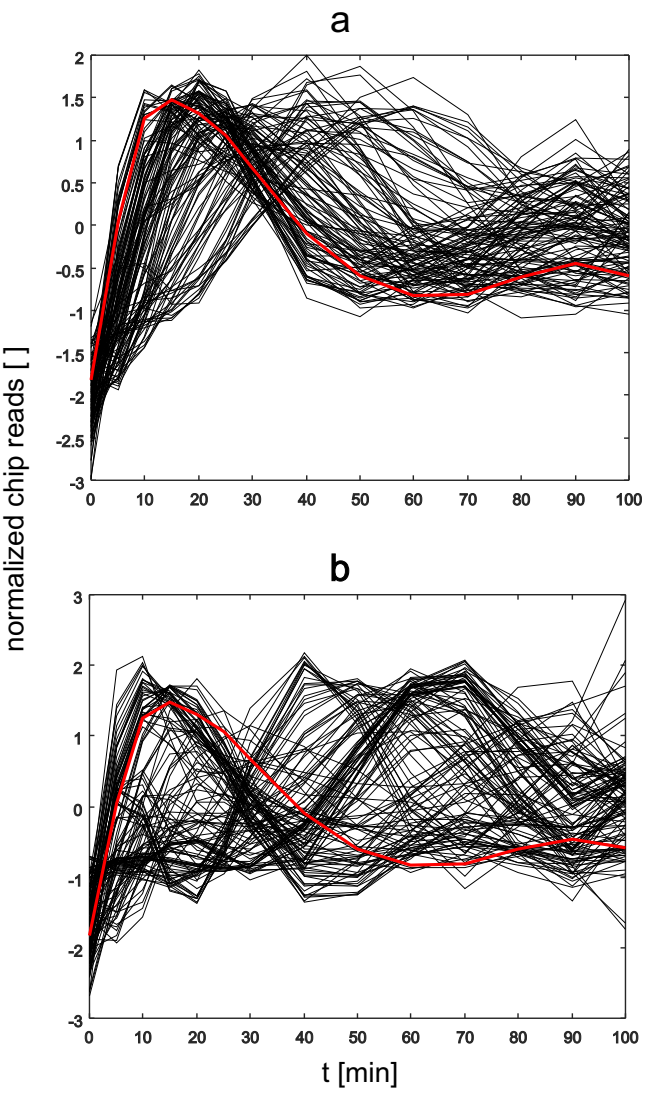


Fig. 2. Normalized expression profiles of (A) genes found to be under control of SigA (red thick line), and (B) genes that might not be under control of SigA.

(Fig. 2A) followed the profile of SigA mRNA, or their reaction to the SigA mRNA concentration was prolonged or delayed. Compared to genes of the set that might not be controlled by SigA (121 genes, Fig. 2B), the profiles in Fig. 2A were consistent in shape. Interestingly,

Fig. 2B also shows that some genes had profiles consistent with the SigA mRNA only during roughly the first half of growth and could thus be modelled and consequently may be controlled in this period by SigA. Although their control by SigA was potentially possible, such genes were excluded from the list of SigA-controlled genes because the control did not affect the whole period. The 121 genes (organized in 56 operons) that might not be controlled by SigA are listed in Table 3.

From the kinetic point of view, the *sigA* gene itself could be regulated by SigI. Whether this interaction is possible needs to be tested experimentally. A literature search did not confirm or refute this hypothesis.

3.2. New SigA targets

To identify novel putative SigA-controlled genes, we identified all genes that were not identified in databases as members of the SigA regulon and extracted their kinetic profiles. Altogether, we selected 3158 genes. Then, we modelled their expression profile with SigA (sigA mRNA) as the regulator. Of the 3158, 2582 gene expression profiles were excluded because their expression profiles were too low to make reliable conclusions. Of the remaining profiles, 575 satisfied general goodness of fit criteria (the model curve was within the confidence interval of the measured expression profile of the given gene). From this set, 214 gene expression profiles with a minimal sum of square differences between measured and modelled expression profile were selected as the most likely novel targets of SigA. The genes of this set, sorted according to increasing error of fit are shown in Table 4.

For the selected genes from Table 4, we analysed their ontologies and identified functional characteristics that were more abundant among these genes than in the whole set of genes. The results are summarized in Fig. 3 (lists of specific genes in individual categories are in Supplemental Table 3).

Fig. 3 shows that most of the newly predicted SigA-dependent genes are involved in essential functions of the cell, such as DNA repair, transcription, and translation. The enriched functional groups of predicted and confirmed groups of genes were mostly similar and differed mainly in the groups related to “nucleotide metabolism,” “cell envelope and cell division,” and “phosphoproteins”. For these functional groups, the number of genes in the predicted group was only about half of those in the known SigA regulon. “Homeostasis” genes were present only among the newly predicted SigA regulated genes. These predicted homeostasis genes are mainly involved in iron transport (e.g., [26]), and for one of them, *ytpQ*, a putative SigA-dependent promoter was previously proposed, consistent with our prediction [27]. Additionally, the genes of the “genetics” group were more abundant in the predicted set than in the confirmed SigA regulon. The important cellular functions of the newly predicted SigA-dependent genes from this group were consistent with their regulation by SigA (e.g., genes involved in DNA synthesis and manipulation: *polC* — DNA polymerase III (alpha subunit) [28], *ligA* — DNA ligase [29], or *topB* — DNA topoisomerase III [30]).

Table 4
Potential target genes of SigA. Genes whose expression profiles were best fitted by the model with SigA as the regulator. Whole operons are listed, and genes that were identified by modelling are shown in bold. The genes that are not in bold did not satisfy the modelling constraints.

yqzB-yqfL, ***fbaA-ywjH***, ***yyaF***-*rpsF-ssbA*-***rpsR***, ***acpA***, ***yprA***, ***yloC***, ***ahrC-recN***, ***efp***, ***ymfK***, ***smpB***, ***fmt***, ***prpC-prkC-yloQ***, ***proS***, ***hepS***, ***ruvA-ruvB***, ***ykpA***, ***rpmGA***, ***yrrC***, ***ypjQ***, ***yrhP***, ***ykaA-pit***, ***yloI***, ***dinG***, ***yloN***, ***mrpA***-*mrpB-mrpC*-***mrpD***-*mrpE-mrpF-mrpG*, ***yycC-yycB***, ***alsT***, ***yloH***, ***yugI***, ***yprB***, ***recQ***, ***yrrS***, ***ywiB***, ***yneF***, ***rnpA***, ***priA***, ***yncD***, ***mutS-mutL***, ***yloU-yloV***, *mobA*-***moeB-moeA-mobB***-*moaE*-***moaD***, ***ylzA***, ***def***, ***yaaD-yaaE***, ***rpsO***, ***recG***, ***ynbA***, ***aroA***, ***nudF***, ***ylmG-ylmH***, ***ylmB***, ***rpsT***, ***ylbB-ylbC***, ***ylbG***, ***ylbH***, ***ylbF***, ***ylbI***, ***yqhQ-yqhP***, ***yqhL***, ***ypbD***, ***rluB***, ***menA***, ***menH***, ***ywzC***, ***yfkB***, ***yclN-yclO-yclP***, ***yfmL***, ***ygaF***, ***pnpA***, ***yqfF***, ***accD-accA***, ***opuD***, ***ybbP***, ***gmk***, ***ytpQ***, ***secDF***, ***yvoF***, ***yocJ***, ***yvcI***-*yvcJ-yvcK-yvcL-crh*-***yvcN***, ***tgt***, ***yobF***, ***yrzD***, ***sdaAB-sdaAA***, ***ypdA***, ***ytzA***, ***ymcB***, ***ripX***, ***ask***, ***hepT***, ***ytxC***, ***yncF***, ***ywiE***, ***yrpC***, ***yrzL***, ***ypmA***, ***ybaL***, ***yabA-yabB-yazA-yabC***, ***mutSB***, ***pcrB***-*pcrA*-***ligA***-*yerH*, ***alaR-alaT***, *rpsB-tsf-pyrH*-***frr***, ***tmk***, ***metS***, ***ytcI***, ***senS***, ***dltA***-*dltB-dltC-dltD-dltE*, ***ytpA***, ***ytpS***, ***ytpT***, ***zwf***, ***hisJ***, ***yqfO***, ***yqxC***, ***divIC***, ***yhaJ***, ***yneS***, ***rocE-rocF***, ***panD***, ***yccF***, ***yoaA***, ***dgk***, ***ytxB***, ***polC***, ***ytoP***, ***yrzC***, ***mbl***, ***yybN***, ***ylaI***, ***ysmB***, ***ytlQ***, ***ytqI***, ***atpI***, ***yjzD***, ***kamA***, ***ykvY***, ***ykkD***, ***yheA***, ***ypmP***, ***proA***, ***yojG***, ***ykuQ***, ***yqzC***, ***gid***, ***yrrB***, ***yckJ***, ***yloS***, ***ymfL***, ***ykbA***, ***yhaK***, ***yfkA***, ***yusV***, ***ykvL***, ***mlpA***, ***yhfQ***, ***ytkP***, ***queA***, ***mutM***, ***yrvC***, ***serA***, ***yshE***, ***ytiA***, ***ytbJ***, ***ywdF***, ***ykvK***, ***yerB***, ***topB***, ***uppS***, ***ygxA***, ***parE***, ***rpmB***, ***dxs***-*yqxC-ahrC-recN*, ***penP***, ***spo0J***, ***yqfN***, ***ykvM***, ***yfhC***, ***asnB***, ***rnmV-ksgA***, ***ymzA***, ***birA***, ***ymfC***, ***ymfG***, ***ypmT***, ***leuS***

3.3. Alternative regulators of the SigA regulon

For the 121 genes that were found not to be controlled by SigA by kinetic modelling, we tried to find possible alternative sigma factors that could control their expression. These alternative sigma factors were SigB, SigD, SigH, SigI, SigL, SigM, SigW, SigX, and SigY. Other sigma factors were not considered because either their profiles were missing in the time series dataset or their overall expression was too low to be reliably used in the model. In some cases, equivalent fits for the expression profiles of particular target genes were achieved with more than one sigma factor. In these cases, we listed all sigma factors for given target genes that could fit the expression profile of the gene. Such regulations (sigma factor-target gene interactions) should be considered as equivalent. A summary of the results, including model parameters, is given in Supplemental Table 4. The results showed that SigB was the most frequently predicted alternative regulator, regulating 81 genes, 16 of which were negatively regulated by SigB (we found experimental evidence for control of the *comQ-comX* operon by SigB; data not shown). The second most frequent alternative regulator was SigX (77 genes). The other regulators were SigL (73 genes), SigM (46 genes), SigW (43 genes), SigH (39 genes), SigD (37 genes), SigY (36 genes), and SigI (33 genes).

Literature searches for the alternative interactions suggested in Supplemental Table 4 provided evidence for the following sigma factor-regulated gene pairs:

SigB (18 genes): *gtaB* [31], *nadE* [32], *yceC* [33], *ctc* [34], *yjbD* [35], *guaD* [36], *mcsA* [37], *rsbV* [33], *rsbW* [33], *rsbX* [33], *yceD* [36], *yjbC* [38], clpC [37], *clpP* [39], *mcsB* [37], *yfkH*, *yfkI* and *yfkJ* [40]. SigM (2 genes) — *ylxW* and *ylxX* [41]. SigW (2 genes) — *yceD* [42], *yjbD* [38].

3.4. Sigma factors from the SigA regulon

The SigA regulon also contains several sigma factors, SigE, SigD, SigH, SigM, SigX, SigF (SigH regulon), YlaC, and SigG (SigF regulon). SigE, SigG, SigF and YlaC were excluded from the analysis because their expression profiles had very low overall values and a high variance that might have caused false results during modelling. The expression profiles (respective mRNAs) of the *sigA*, *sigD*, *sigH*, *sigM*, and *sigX* genes are shown in Fig. 4, illustrating the prominent role of SigA. We analysed the regulons of these sigma factors and modelled the kinetic interactions between individual sigma factors and their regulons. The expression profiles of each gene and regulators suggested by the model, together with the parameters of the model and regulators found by the model, are summarized in Supplemental Table 5 for all genes mentioned below. Fig. 5 shows examples of genes/operons confirmed to be controlled by a given sigma factor, together with profiles where the control was not confirmed. In the next paragraph, we summarize the modelling results for the above sigma factors. Detailed descriptions of individual genes of the respective regulons together with literature references supporting our findings can be found in Supplemental file 4.

3.4.1. SigM regulon

SigM is an ECF-type sigma factor and is required for adaptation to inhibitors of peptidoglycan synthesis and survival at high salt concentration [43,44]. SigM has been reported to be controlled by SigA and by itself in a feedback loop [44]. Kinetic modelling confirmed control of SigM by SigA. Control of SigM by itself is, from the point of view of the model, trivial. We analysed the kinetics of 72 of the 84 SigM-regulated

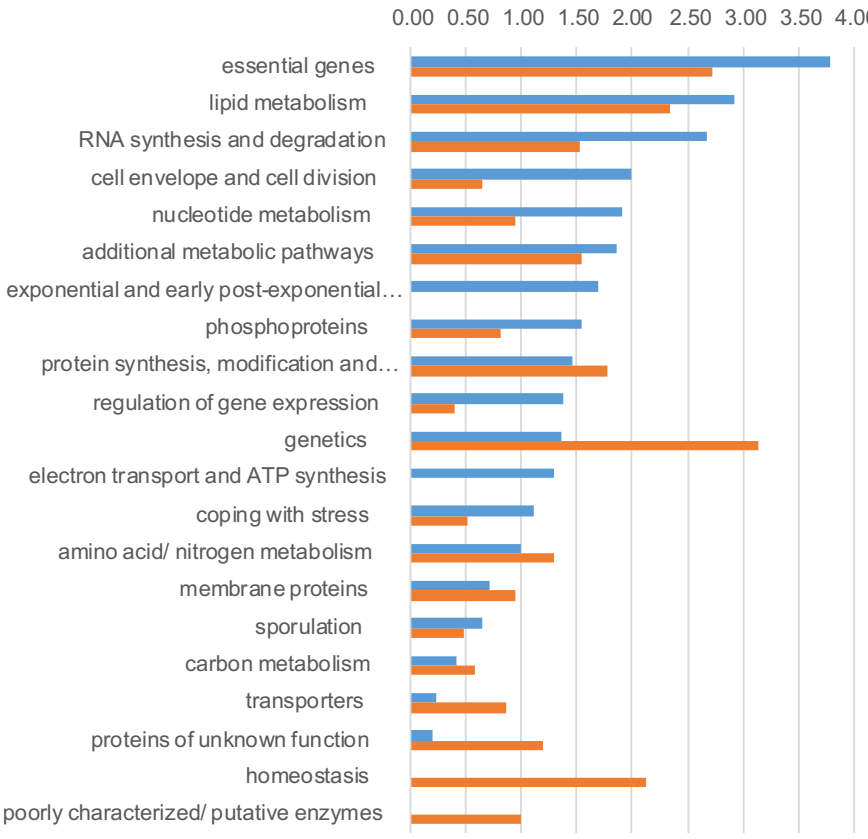


Fig. 3. Gene ontologies for potential SigA targets (red) compared with the group of genes known to be positively regulated with SigA (blue, Table 2). The bars show relative enrichment of the respective functional group within the subset compared to the number of genes in this functional group among all genes.

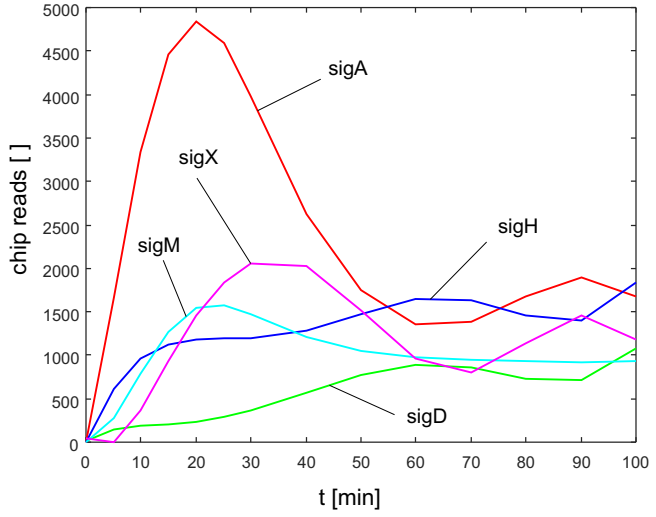


Fig. 4. Expression profiles (mRNAs) of sigma factors of the SigA regulon during germination and outgrowth. Red — SigA, green — SigD, blue — SigH, cyan — SigM, magenta — SigX. Horizontal axis — time in minutes, vertical axis — microarray units.

genes. The expression profiles of the sigM regulon exhibited quite a large variety of profiles, and not all of them could be modelled with the sigM mRNA profile as the regulator because the peaks of expression of SigM and the incident genes differed too much to be explained by posttranscriptional events (e.g., for the *yceC* operon, the difference was ca. 40 min); in some cases, the peak of a regulated gene preceded the peak of the regulator (e.g., secDF or yebC, see Supplemental Table 5). We found that of the 72 analysed kinetic profiles, we found parameters for 24 that allowed us to model the profiles with SigM mRNA. The other profiles could be modelled with profiles of sigma factors other than sigM (see Supplemental Table 5, sheet sigM). For most of the large operons, we observed a phenomenon mentioned also for other regulons; i.e., the overall expression (parameter k1) decreased with the distance from the promoter. This phenomenon was particularly apparent for the *dltA* operon.

3.4.2. SigH regulon

SigH in *Bacillus subtilis* directs the transcription of genes in early stationary phase and is essential for sporulation [45,46]. SigH itself has been reported to be controlled by SigA [47], and we found that the fit of its profile could be achieved only if SigD is taken as a second regulator controlling the later phase of growth. According to SubtiWiki, the SigH regulon is formed by 42 genes in 12 operons. Of the 42 genes suggested, only 8 were confirmed by kinetic modelling. These genes were *hbs*, *spoIIAB*, *spo0M*, *lytE*, *phrC*, *phrI*, *rapK*, and *spoVG*. Possible alternative regulators for the remaining genes are listed in Supplemental Table 5, sheet sigH.

3.4.3. SigD regulon

SigD is involved in regulation of flagella synthesis, motility, and chemotaxis [48,49]. *sigD* has been reported to be controlled by itself [50] and SigA [51]. However, for the regulation of *sigD* by SigA, we could not find parameters that would make a satisfactory fit for the data set used; the shapes of SigA and SigD mRNA profiles are almost opposite (see Fig. 4). However, *sigD* expression is also regulated by CodY [52]. CodY is a transcription factor that responds to intracellular GTP concentration. CodY is a repressor, and at higher [GTP], it dissociates from DNA, thus relieving the repression [53]. Cells going into stationary phase and sporulation have markedly decreased GTP levels, which increase during outgrowth [54]. The CodY-dependent regulation of SigD may, in part, explain the absence of its immediate response to SigA.

According to SubtiWiki, the SigD regulon is formed by 70 genes in 12 operons, exhibiting two principal profiles: the first similar to the profile of the SigD, and the second exhibits a peak at different time

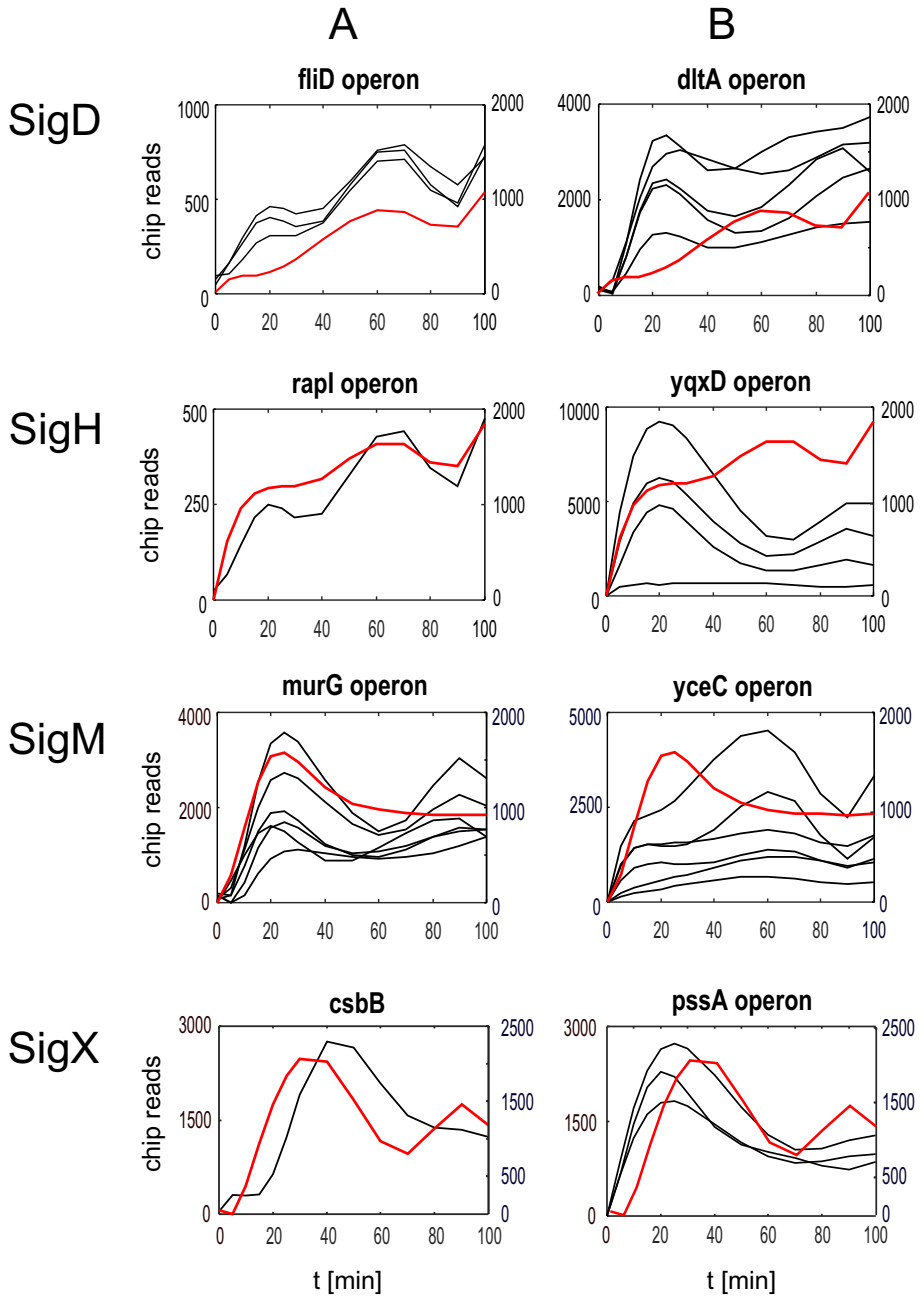


Fig. 5. Examples of genes and operons that were reliably modelled using respective sigma factor mRNA expression profile (column A) and those that could not be modelled (column B).

points (for an example see Fig. 5). The profiles of these genes were successfully modelled with SigD for 42 of the 70 genes in its regulon. For the remaining genes, alternative regulators were suggested (see Supplemental Table 5, sheet sigD). For most of the operons, the parameter k1 decreased with the distance from promoter.

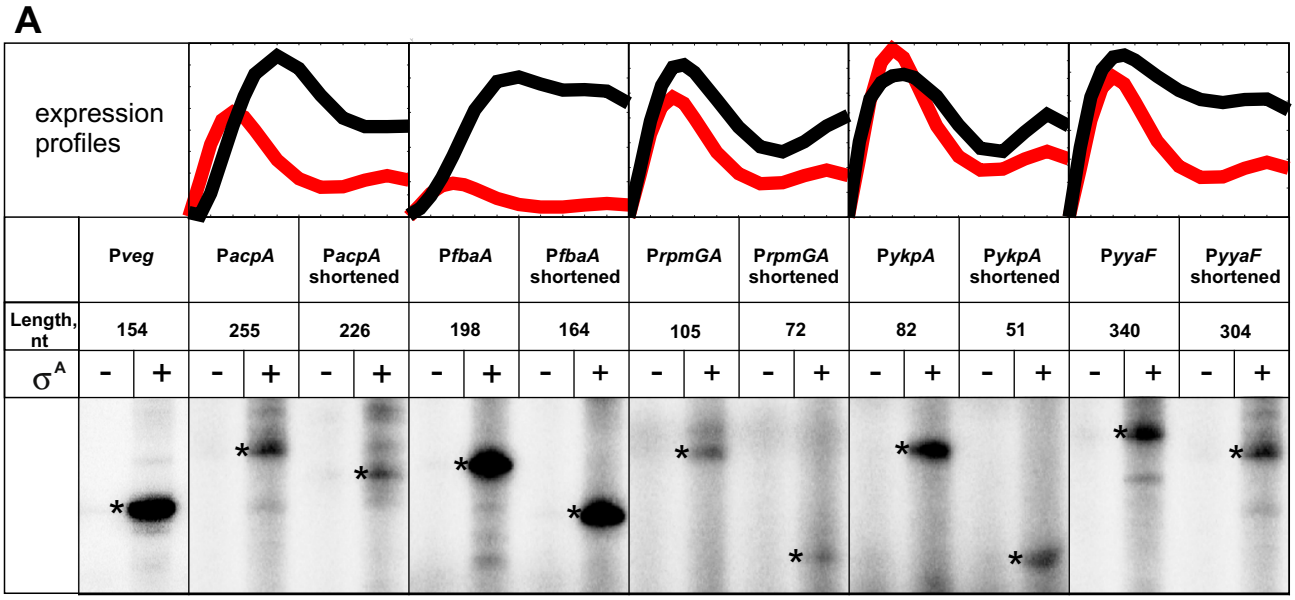
3.4.4. SigX regulon

SigX is an extracytoplasmic sigma factor (ECF) that helps the bacterial cell survive at high temperature, participates in the regulation of peptidoglycan synthesis and turnover, and coordinates antibiotic stress response [55–57]. *sigX* has been reported to be controlled by SigA [55]. Modelling confirmed this observation, with a relatively low decay rate constant in the model.

According to SubtiWiki, the SigX regulon is formed by 29 genes, including 6 operons. The kinetics of the SigX regulon genes are quite diverse, and a brief comparison with the SigX profile shows that many of the genes cannot be controlled by SigX; their peak of expression precedes that of SigX (see Fig. 5). This simple observation was also confirmed by modelling, which showed that of the 29 analysed expression profiles of the databased SigX regulon, 25 could be modelled better with a profile other than that of SigX (see Supplemental Table 5, sheet sigX).

3.5. Validation of selected predictions

We assessed the robustness of the model by experimentally testing selected interactions. To gauge the validity of these predictions, we selected the upstream regions of 10 genes that were predicted to be SigA-dependent and whose kinetic profiles correlated perfectly with the SigA profile (see Supplemental Table 1 for the full list). The upstream regions belonged to genes that were either monocistronic or positioned as the first genes of respective operons. Subsequently, for each upstream region, we prepared two PCR fragments. The shorter fragment lacked approximately 30 bp from the 3′ end in the direction of expected



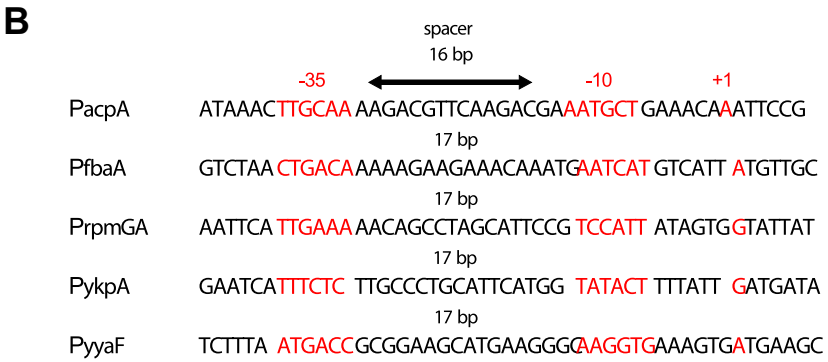


Fig. 6. *In vitro* multiple round transcriptions of DNA fragments derived from upstream regions of genes predicted to be SigA-dependent. (A) Representative primary data. Each reaction was performed with the RNAP core and SigA-containing holoenzyme to demonstrate that the core had not been contaminated with sigma factors prior to its reconstitution with SigA. Radioactively labelled samples were loaded onto polyacrylamide gels. P*veg* was used as a positive control. Transcript length was calculated with an RNA ladder (data not shown). The differences in length between the long and shortened fragment variants were in the 29–36 bp range. The distance between the long and shortened transcript variants in the gel differed for different promoters because each transcript had a unique length. Asterisks indicate the specific transcripts. The upper part of the figure shows respective kinetic modelling results. Red, SigA mRNA; black, specific gene. The scaling between boxes varies to accommodate the graphs as the levels of specific transcripts differed over a wide range. (B) Alignment of putative promoter sequences identified by *in vitro* transcription assays. The − 35 and − 10 hexamers and the transcription start sites (+ 1) are indicated in red. Spacer regions between − 35 and − 10 hexamers are indicated.

transcription. This was to distinguish the direction of transcription from the DNA fragments. Subsequently, we performed *in vitro* transcriptions in a defined system with *B. subtilis* RNAP-SigA holoenzyme with these fragments as templates.

As a positive control, we used the strong constitutive SigA-dependent P*veg* promoter [58]. As a negative control, we used the RNAP core without sigma factors. Transcription signals were then obtained with RNAP complexed with SigA for 5 of the 10 tested gene upstream regions, thus suggesting the presence of SigA-dependent promoters for *acpA*, *fbaA*, *rpmGA*, *ykpA*, and *yyaF*, in the expected orientation in each case (Fig. 6). The first two transcribed genes are involved in fatty acid biosynthesis [59] and carbon core metabolism [60]; the next encodes ribosomal protein L33a [61]; and the last two are a putative ABC transporter [62] and a putative GTP-binding protein [63], respectively.

The upstream region of the *acpA* gene was one of the 5 that were experimentally verified. According to DBTBS, the *acpA* gene was not present in the list of known SigA-dependent genes. However, we subsequently found that the *acpP* gene is synonymous with the *acpA*, gene for which a SigA-dependent promoter had been previously reported [64]. Indeed, we detected this promoter in our *in vitro* transcription experiments (data not shown). Nevertheless, we found an additional, so far unknown SigA-dependent promoter upstream of the already known promoter of the *acpA* gene (Fig. 6).

The lack of transcription from the remaining 5 upstream regions (data not shown) might be a consequence of a requirement for additional transcription factor(s), different reaction conditions or the absence of SigA-dependent promoters in these DNA fragments.

3.6. Analysis of promoter sequences

Finally, we performed sequence analysis of SigA-dependent promoters of genes that had been, or had not been verified by our kinetic analysis to be regulated by SigA during germination and outgrowth (Tables 2 and 3, Supplemental Table 6 and 7). A sequence logo, based on 70 known promoters for genes from Table 2 (those consistent with our kinetic analysis), is shown in Fig. 7B. A sequence log, based on 43 known promoters for genes from Table 3 (those inconsistent with our kinetic analysis) is shown in Fig. 7C. Both these logos resemble the M14 consensus sequence of SigA-dependent promoters as reported by Nicholas and colleagues [3] (Fig. 7A). A sequence logo created from the

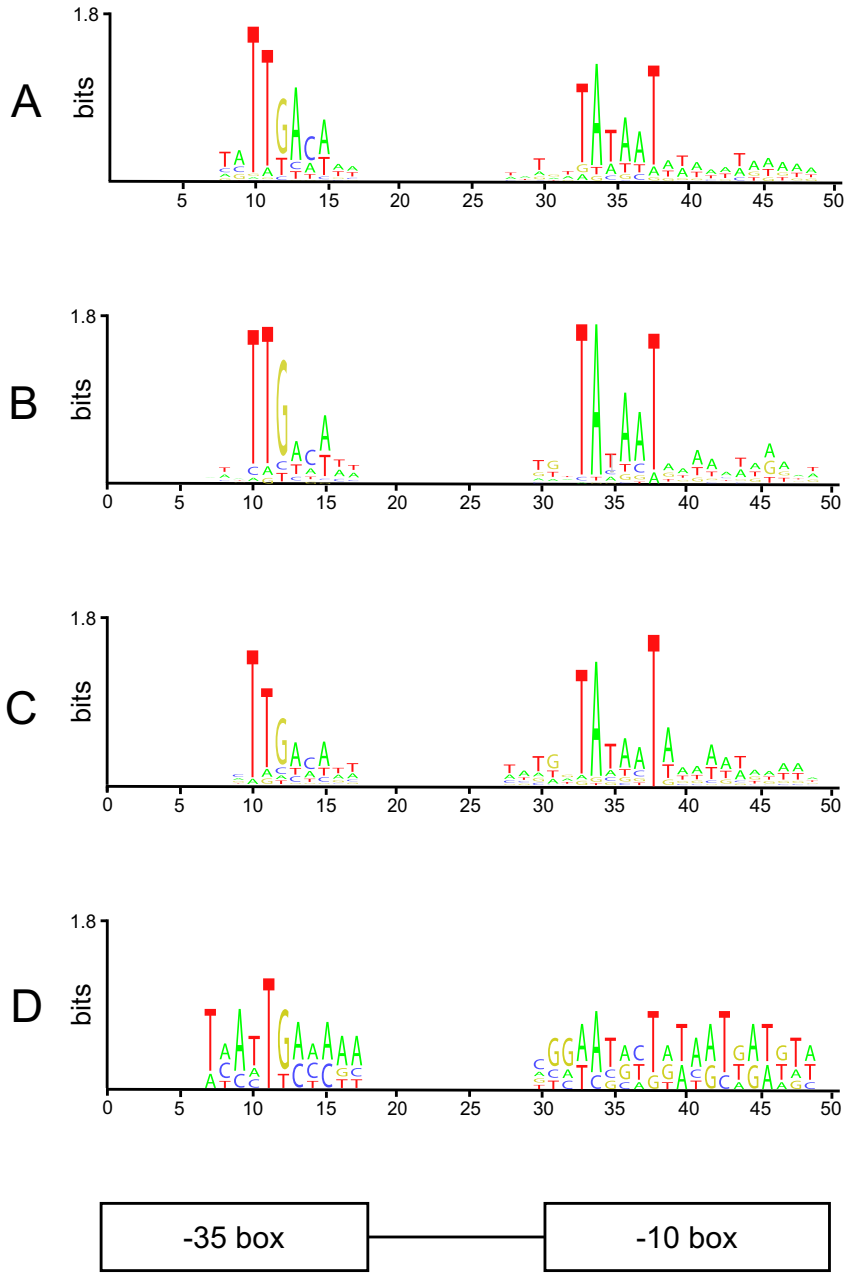


Fig. 7. Comparisons of sequence logos. (A) The M14 motif of SigA-dependent promoters according to Nicolas and colleagues [3]. (B) The sequence motif of known SigA-dependent promoters verified by the kinetic model. The logo was created from sequences listed in Table 2 (Supplemental Table 6). The numbering of the horizontal axis was done according to Nicolas and colleagues. (C) The sequence motif of known SigA-dependent promoters that were not verified by the kinetic model. The logo was created from sequences listed in Table 3 (Supplemental Table 6). (D) The sequence logo for the experimentally verified SigA dependent promoters from Fig. 6.

five newly identified promoters (Fig. 7C) is consistent with the M14 logo, but we note that the size of the sample was too small for reliable analysis.

4. Discussion

Here, we analysed the expression of *B. subtilis* genes during spore germination and outgrowth, as measured by microarrays in a unique time course experiment consisting of 14 time points spaced at 5- to 10-min intervals (see Materials and methods; [10]). This experiment allowed us to create a time series of gene expression for SigA, other sigma factors from the SigA regulon, and their target genes. The time series were then subjected to kinetic analysis based on a computational model of gene expression (see Materials and methods). The modelling was able to (i) discriminate for possible regulatory interactions among the sigma factors and their putative target genes, (ii) optimize parameters of the model that could be used to computationally simulate the accumulation of mRNA of a gene under the control of a specific sigma factor, (iii) find other possible regulators of specific genes or suggest a new mechanism of control of a gene by computing the profile of an unknown regulator that could explain the observed expression profile, (iv) suggest new SigA-dependent genes (selected predictions were subsequently validated experimentally), (v) create a specific SigA-controlled gene expression network that is active under the conditions measured by the microarray time series, and (vi) identify promoter sequence logos associated with the SigA-dependent promoters that are, or are not, dominantly regulated by SigA during germination and outgrowth.

Of the 850 genes in the SigA regulon suggested by SubtiWiki, we kinetically analysed 311 expression profiles, 190 of which were confirmed as possible target genes of SigA that also satisfied the kinetics-based criteria for the conditions covered by the microarray time series. Consequent analysis of the remaining time series data suggested another 214 genes as putative targets of SigA. Using *in vitro* transcription experiments we demonstrated that five of the ten newly predicted SigA-dependent genes are indeed transcribed by RNAP with SigA. The dependence of these genes on the housekeeping sigma factor was

consistent with their cellular roles (fatty acid biosynthesis, carbon core metabolism, and ribosomal protein) and core promoter sequence (Figs. 6B and 7D). The data complemented the information of the SigA regulation network in *B. subtilis* during germination and outgrowth and extended the list of known SigA-dependent genes.

Further, a promoter sequence logo was determined for SigA-dependent genes that were found to be controlled by SigA during germination and outgrowth (Fig. 7B), closely matching the M14 logo reported by Nicholas and colleagues [3]. However, the promoter logo of SigA dependent genes that were found not under control of SigA was almost identical (Fig. 7C), indicating additional level of control for the genes whose profiles could not be modeled with SigA. Taken together, the results show that for the SigA regulon, ~60% of the gene expression profiles analysed were consistent with the kinetic analysis and directly regulated by SigA without the requirement for additional regulators or modulators. The remaining 40% displayed more complex kinetics. This kinetics was not dependent on the core promoter sequence but rather could be explained by additional layers of regulation, such as alternative sigma or/and transcription factors (e.g. gene/regulator gtaB/SigB [31] *cysK*/CymR + Spx [65,66]; *ackA*/CcpA + CodY [67,68]). Thus, the kinetic analysis is clearly able to distinguish between genes where the dominant regulator is the sigma factor and where other regulators must be involved.

Of the alternative sigma factor regulons, the SigD regulon displayed the best fit between modelling and experimental evidence (42 of the 73 genes could be modelled satisfactorily). For the remaining sigma factors (SigM, H, X) the fit was less successful, pointing perhaps to other modes of regulation of these genes during germination and outgrowth. In the cases where a suggested sigma factor did not model the profiles of the genes of its regulon, alternative sigma factors satisfying the kinetics-based criteria were suggested. Nevertheless, the genes from the alternative sigma factors' regulons that were confirmed by our analysis suggest that the respective sigma factors play a role also during germination and outgrowth. Finally, for most of the large operons, we observed that the overall expression decreased with the distance from the promoter. This phenomenon was also observed for *Streptomyces coelicolor* and is discussed in the paper of Laing et al. [69].

A visualization of the regulatory network can be found online at https://cas-bioinf.github.io/bsubtilis-web/. The graph contains both the regulatory network derived from the literature and the network of regulations we determined as kinetically plausible.

Alternative modes of control such as anti- and anti-anti-sigma factors [70] or others (e.g., attenuation, additional transcription factors) were ignored in this study, although such modes of control could help explain some of the observed kinetics. However, for such complex analysis, additional information, which is not always available, would be required; this exceeds the scope of this study, which focused on the direct regulatory interactions between a regulator (SigA) and its target genes. Nevertheless, an alternative and thus far uncharacterized control is most likely involved in sigma factor-target gene interactions, where a negative value of the parameter w of the model was found. This implies a putative repressive role for sigma factors that will require further study to be understood in detail.

Finally, over the last several years, the amount of available data from massively parallel experiments has grown exponentially, and most data are evaluated only within the narrow focus of the paper for which they were prepared. The huge amount of information contained in data in public repositories remains unexplored, although studies are emerging that focus on their exploitation [71,72]. We believe that our analysis of the system described here will contribute to addressing this challenge.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbagrm.2017.06.003.

Formatting of funding sources

Funding: The work was supported by the Czech research infrastructure for systems biology C4SYS (project no LM2015055) (to JV), Science Foundation of Charles University (GA UK Grant No. 322815) (to OR), Czech Science Foundation Grant No. GA 13-16842S (to LK; *in vitro* experiments), and Czech Health Research Council Grant No. 17-29680A (to LK; bioinformatics).

Transparency document

The http://dx.doi.org/10.1016/j.bbagrm.2017.06.003 associated with this article can be found, in online version.

References

[1] M.S. Paget, Bacterial sigma factors and anti-sigma factors: structure, function and distribution, Biomol. Ther. 5 (2015) 1245–1265.
[2] R.H. Michna, F.M. Commichau, D. Todter, C.P. Zschiedrich, J. Stulke, SubtiWiki-a database for the model organism *Bacillus subtilis* that links pathway, interaction and expression information, Nucleic Acids Res. 42 (2014) D692–D698.
[3] P. Nicolas, U. Mäder, E. Dervyn, T. Rochat, A. Leduc, N. Pigeonneau, E. Bidnenko, E. Marchadier, M. Hoebeke, S. Aymerich, D. Becher, P. Bisicchia, E. Botella, O. Delumeau, G. Doherty, E.L. Denham, M.J. Fogg, V. Fromion, A. Goelzer, A. Hansen, E. Härtig, C.R. Harwood, G. Homuth, H. Jarmer, Condition-dependent transcriptome architecture in *Bacillus subtilis*, Science (2012) 1103–1106.
[4] K.L. MacQuarrie, A.P. Fong, R.H. Morse, S.J. Tapscott, Genome-wide transcription factor binding: beyond direct target regulation, Trends Genet. 27 (2011) 141–148.
[5] C.C. To, J. Vohradsky, Measurement variation determines the gene network topology reconstructed from experimental data: a case study of the yeast cyclin network, FASEB J. 24 (2010) 3468–3478.
[6] M. Bansal, G.D. Gatta, D. di Bernardo, Inference of gene regulatory networks and compound mode of action from time course gene expression profiles, Bioinformatics 22 (2006) 815–822.
[7] C.A. Penfold, D.L. Wild, How to infer gene networks from expression profiles, revisited, Interface Focus 1 (2011) 857–870.
[8] Z. Bar-Joseph, A. Gitter, I. Simon, Studying and modelling dynamic biological processes using time-series gene expression data, Nat. Rev. Genet. 13 (2012) 552–564.
[9] Y.D. Wen, C.T. Liao, K.M. Liou, W.H. Wang, W.C. Huang, B.Y. Chang, Structural and functional properties of a *Bacillus subtilis* temperature-sensitive sigma(A) factor, Proteins 40 (2000) 613–622.
[10] B.J. Keijser, A. Ter Beek, H. Rauwerda, F. Schuren, R. Montijn, H. van der Spek, S. Brul, Analysis of temporal gene expression during *Bacillus subtilis* spore germination and outgrowth, J. Bacteriol. 189 (2007) 3624–3634.
[11] N. Sierro, Y. Makita, M. de Hoon, K. Nakai, DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information, Nucleic Acids Res. 36 (2008) D93–D96.
[12] J. Vohradsky, Neural network model of gene expression, FASEB J. 15 (2001) 846–854.
[13] C.C. To, J. Vohradsky, Supervised inference of gene-regulatory networks, BMC Bioinf. 9 (2008) 2.
[14] T.T. Vu, J. Vohradsky, Nonlinear differential equation model for quantification of transcriptional regulation applied to microarray data of *Saccharomyces cerevisiae*, Nucleic Acids Res. 35 (2007) 279–287.
[15] T.T. Vu, J. Vohradsky, Inference of active transcriptional networks by integration of gene expression kinetics modeling and multisource data, Genomics 93 (2009) 426–433.
[16] A.S. Ribeiro, O.P. Smolander, T. Rajala, A. Hakkinen, O. Yli-Harja, Delayed stochastic model of transcription at the single nucleotide level, J. Comput. Biol. 16 (2009) 539–553.
[17] A. Ribeiro, R. Zhu, S.A. Kauffman, A general modeling strategy for gene regulatory networks with stochastic dynamics, J. Comput. Biol. 13 (2006) 1630–1639.
[18] M.R. Roussel, R. Zhu, Stochastic kinetics description of a simple transcription model, Bull. Math. Biol. 68 (2006) 1681–1713.
[19] J. Vohradsky, Stochastic simulation for the inference of transcriptional control network of yeast cyclins genes, Nucleic Acids Res. 40 (2012) 7096–7103.
[20] J. Lundgren, http://www.mathworks.com/matlabcentral/fileexchange/13812-splinefit, (2010).
[21] D. Hanahan, Studies on transformation of *Escherichia coli* with plasmids, J. Mol. Biol. 166 (1983) 557–580.
[22] Y. Qi, F.M. Hulett, PhoP-P and RNA polymerase sigmaA holoenzyme are sufficient for transcription of Pho regulon promoters in *Bacillus subtilis*: PhoP-P activator sites within the coding region stimulate transcription in vitro, Mol. Microbiol. 28 (1998) 1187–1197.
[23] B.Y. Chang, R.H. Doi, Overproduction, purification, and characterization of *Bacillus subtilis* RNA polymerase sigma A factor, J. Bacteriol. 172 (1990) 3257–3263.
[24] L. Sojka, T. Kouba, I. Barvik, H. Sanderova, Z. Maderova, J. Jonak, L. Krasny, Rapid changes in gene expression: DNA determinants of promoter regulation by the concentration of the transcription initiating NTP in *Bacillus subtilis*, Nucleic Acids Res. 39 (2011) 4598–4611.

[25] Y.L. Juang, J.D. Helmann, The delta subunit of *Bacillus subtilis* RNA polymerase. An allosteric effector of the initiation and core-recycling phases of transcription, J. Mol. Biol. 239 (1994) 1–14.
[26] A.M. Zawadzka, Y. Kim, N. Maltseva, R. Nichiporuk, Y. Fan, A. Joachimiak, K.N. Raymond, Characterization of a *Bacillus subtilis* transporter for petrobactin, an anthrax stealth siderophore, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 21854–21859.
[27] P. Zuber, S. Chauhan, P. Pilaka, M.M. Nakano, S. Gurumoorthy, A.A. Lin, S.M. Barendt, B.K. Chi, H. Antelmann, U. Mader, Phenotype enhancement screen of a regulatory spx mutant unveils a role for the ytpQ gene in the control of iron homeostasis, PLoS One 6 (2011) e25066.
[28] E. Dervyn, C. Suski, R. Daniel, C. Bruand, J. Chapuis, J. Errington, L. Janniere, S.D. Ehrlich, Two essential DNA polymerases at the bacterial replication fork, Science 294 (2001) 1716–1719.
[29] M.A. Petit, S.D. Ehrlich, The NAD-dependent ligase encoded by yerG is an essential gene of *Bacillus subtilis*, Nucleic Acids Res. 28 (2000) 4642–4648.
[30] H.B. Thomaides, E.J. Davison, L. Burston, H. Johnson, D.R. Brown, A.C. Hunt, J. Errington, L. Czaplewski, Essential bacterial functions encoded by gene pairs, J. Bacteriol. 189 (2007) 591–602.
[31] D. Varon, S.A. Boylan, K. Okamoto, C.W. Price, *Bacillus subtilis* gtaB encodes UDP-glucose pyrophosphorylase and is controlled by stationary-phase transcription factor σ(B), J. Bacteriol. 175 (1993) 3964–3971.
[32] H. Antelmann, R. Schmid, M. Hecker, The NAD synthetase NadE (OutB) of *Bacillus subtilis* is a σ(B)-dependent general stress protein, FEMS Microbiol. Lett. 153 (1997) 405–409.
[33] A. Petersohn, M. Brigulla, S. Haas, D. Jörg, U. Völker, M. Hecker, Global analysis of the general stress response of *Bacillus subtilis* global analysis of the general stress response of *Bacillus subtilis*, J. Bacteriol. 183 (2001) 5617–5631.
[34] J.F. Ollington, W.G. Haldenwang, T.V. Huynh, R. Losick, Developmentally regulated transcription in a cloned segment of the *Bacillus subtilis* chromosome, J. Bacteriol. 147 (1981) 432–442.
[35] A. Petersohn, J. Bernhardt, U. Gerth, D. Höper, T. Koburger, U. Völker, M. Hecker, R.G. Bernhardt, U.L.F. Gerth, Identification of ς B-dependent genes in *Bacillus subtilis* using a promoter consensus-directed search and oligonucleotide hybridization identification of B-dependent genes in *Bacillus subtilis* using a promoter consensus-directed search and oligonucleotide hybridization, J. Bacteriol. 181 (1999) 5718–5724.
[36] D. Höper, U. Völker, M. Hecker, Comprehensive characterization of the contribution of individual SigB-dependent general stress genes to stress resistance of *Bacillus subtilis*, 187 (2005) 2810–2826.
[37] E. Kruger, T. Msadek, M. Hecker, Alternate promoters direct stress-induced transcription of the *Bacillus subtilis* clpC operon, Mol. Microbiol. 20 (1996) 713–723.
[38] H. Antelmann, C. Scharf, M. Hecker, Phosphate starvation-inducible proteins of *Bacillus subtilis*: proteomics and transcriptional analysis, J. Bacteriol. 182 (2000) 4478–4490.
[39] U. Gerth, E. Krüger, I. Derré, T. Msadek, M. Hecker, Stress induction of the *Bacillus subtilis* clpP gene encoding a homologue of the proteolytic component of the Clp protease and the involvement of ClpP and ClpX in stress tolerance, Mol. Microbiol. 28 (1998) 787–802.
[40] C.W. Price, P. Fawcett, H. Cérémonie, N. Su, C.K. Murphy, P. Youngman, Genome-wide analysis of the general stress response in *Bacillus subtilis*, Mol. Microbiol. 41 (2001) 757–774.
[41] W. Eiamphungporn, J.D. Helmann, The *Bacillus subtilis* σM regulon and its contribution to cell envelope stress responses, Mol. Microbiol. 67 (2008) 830–848.
[42] M. Cao, P.A. Kobel, M.M. Morshedi, M.F.W. Wu, C. Paddon, J.D. Helmann, Defining the *Bacillus subtilis* sigma(W) regulon: a comparative analysis of promoter consensus search, run-off transcription/macroarray analysis (ROMA), and transcriptional profiling approaches, J. Mol. Biol. 316 (2002) 443–457.
[43] M. Cao, T. Wang, R. Ye, J.D. Helmann, Antibiotics that inhibit cell wall biosynthesis induce expression of the *Bacillus subtilis* sigma(W) and sigma(M) regulons, Mol. Microbiol. 45 (2002) 1267–1276.
[44] M.J. Horsburgh, A. Moir, Sigma M, an ECF RNA polymerase sigma factor of *Bacillus subtilis* 168, is essential for growth and survival in high concentrations of salt, Mol. Microbiol. 32 (1999) 41–50.
[45] H.L. Carter III, C.P. Moran Jr., New RNA polymerase sigma factor under spo0 control in *Bacillus subtilis*, Proc. Natl. Acad. Sci. U. S. A. 83 (1986) 9438–9442.
[46] H.L. Carter 3rd, L.F. Wang, R.H. Doi, C.P. Moran Jr., rpoD operon promoter used by sigma H-RNA polymerase in *Bacillus subtilis*, J. Bacteriol. 170 (1988) 1617–1621.
[47] M.A. Strauch, Delineation of AbrB-binding sites on the *Bacillus subtilis* spo0H, kinB, ftsAZ, and pbpE promoters and use of a derived homology to identify a previously unsuspected binding site in the bsuB1 methylase promote, J. Bacteriol. 177 (1995) 6999–7002.
[48] J.D. Helmann, L.M. Márquez, M.J. Chamberlin, Cloning, sequencing, and disruption of the *Bacillus subtilis* sigma 28 gene, J. Bacteriol. 170 (1988) 1568–1574.
[49] J.D. Helmann, F.R. Masiarz, M.J. Chamberlin, Isolation and characterization of the *Bacillus subtilis* sigma 28 factor, J. Bacteriol. 170 (1988) 1560–1567.
[50] M.Z. Gilman, J.S. Glenn, V.L. Singer, M.J. Chamberlin, Isolation of sigma-28-specific promoters from *Bacillus subtilis* DNA, Gene 32 (1984) 11–20.
[51] W. Estacio, S.S. Anna-Arriola, M. Adedipe, L.M. Marquez-Magana, Dual promoters are responsible for transcription initiation of the fla/che operon in *Bacillus subtilis*, J. Bacteriol. 180 (1998) 3548–3555.
[52] Q.O. Ababneh, J.K. Herman, CodY regulates SigD levels and activity by binding to three sites in the fla/che operon, J. Bacteriol. 197 (2015) 2999–3006.
[53] V.M. Levdikov, E.V. Blagova, V.L. Young, B.R. Belitsky, A. Lebedev, A.L. Sonenshein, A.J. Wilkinson, Structure of the branched chain amino acid and GTP sensing global regulator, CodY, from *Bacillus subtilis*, J. Biol. Chem. (2016).
[54] J.M. Lopez, A. Dromerick, E. Freese, Response of guanosine 5′-triphosphate concentration to nutritional changes and its significance for *Bacillus subtilis* sporulation, J. Bacteriol. 146 (1981) 605–613.
[55] X. Huang, et al., The *Bacillus subtilis* sigma-x protein is an extracytoplasmic function sigma factor contributing to survival at high temperature, J. Bacteriol. 179 (1997) 2915–2921.
[56] X. Huang, J.D. Helmann, Identification of target promoters for the *Bacillus subtilis* sigma X factor using a consensus-directed search, J. Mol. Biol. 279 (1998) 165–173.
[57] M. Cao, J.D. Helmann, Regulation of the *Bacillus subtilis* bcrC bacitracin resistance gene by two extracytoplasmic function sigma factors, J. Bacteriol. 184 (2002) 6123–6129.
[58] T. Fukushima, S. Ishikawa, H. Yamamoto, N. Ogasawara, J. Sekiguchi, Transcriptional, functional and cytochemical analyses of the veg gene in *Bacillus subtilis*, J. Biochem. 133 (2003) 475–483.
[59] H.R. Morbidoni, D. de Mendoza, J.E. Cronan Jr., *Bacillus subtilis* acyl carrier protein is encoded in a cluster of lipid biosynthesis genes, J. Bacteriol. 178 (1996) 4794–4800.
[60] H. Ludwig, G. Homuth, M. Schmalisch, F.M. Dyka, M. Hecker, J. Stulke, Transcription of glycolytic genes and operons in *Bacillus subtilis*: evidence for the presence of multiple levels of control of the gapA operon, Mol. Microbiol. 41 (2001) 409–422.
[61] G. Akanuma, H. Nanamiya, Y. Natori, K. Yano, S. Suzuki, S. Omata, M. Ishizuka, Y. Sekine, F. Kawamura, Inactivation of ribosomal protein genes in *Bacillus subtilis* reveals importance of each ribosomal protein for cell proliferation and cell differentiation, J. Bacteriol. 194 (2012) 6282–6291.
[62] Y. Quentin, G. Fichant, F. Denizot, Inventory, assembly and analysis of *Bacillus subtilis* ABC transport systems, J. Mol. Biol. 287 (1999) 467–484.
[63] N. Verstraeten, M. Fauvart, W. Versees, J. Michiels, The universally conserved prokaryotic GTPases, Microbiol. Mol. Biol. Rev. 75 (2011) 507–542 (second and third pages of table of contents).
[64] M.A. Martinez, D. de Mendoza, G.E. Schujman, Transcriptional and functional characterization of the gene encoding acyl carrier protein in *Bacillus subtilis*, Microbiology 156 (2010) 484–495.
[65] S. Even, P. Burguiere, S. Auger, O. Soutourina, A. Danchin, I. Martin-Verstraete, Global control of cysteine metabolism by CymR in *Bacillus subtilis*, J. Bacteriol. 188 (2006) 2184–2197.
[66] S. Nakano, M.M. Nakano, Y. Zhang, M. Leelakriangsak, P. Zuber, A regulatory protein that interferes with activator-stimulated transcription in bacteria, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 4233–4238.
[67] F.J. Grundy, D.A. Waters, S.H. Allen, T.M. Henkin, Regulation of the *Bacillus subtilis* acetate kinase gene by CcpA, J. Bacteriol. 175 (1993) 7348–7355.
[68] R.P. Shivers, S.S. Dineen, A.L. Sonenshein, Positive regulation of *Bacillus subtilis* ackA by CodY and CcpA: establishing a potential hierarchy in carbon flow, Mol. Microbiol. 62 (2006) 811–822.
[69] E. Laing, V. Mersinias, C.P. Smith, S.J. Hubbard, Analysis of gene expression in operons of *Streptomyces coelicolor*, Genome Biol. 7 (2006) R46.
[70] E.A. Campbell, L.F. Westblade, S.A. Darst, Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective, Curr. Opin. Microbiol. 11 (2008) 121–127.
[71] M.L. Arrieta-Ortiz, C. Hafemeister, A.R. Bate, T. Chu, A. Greenfield, B. Shuster, S.N. Barry, M. Gallitto, B. Liu, T. Kacmarczyk, F. Santoriello, J. Chen, C.D.A. Rodrigues, T. Sato, D.Z. Rudner, A. Driks, R. Bonneau, P. Eichenberger, An experimentally supported model of the *Bacillus subtilis* global transcriptional regulatory network, Mol. Syst. Biol. 11 (2015) 839.
[72] S.A. Leyn, M.D. Kazanov, N.V. Sernova, E.O. Ermakova, P.S. Novichkov, D.A. Rodionov, Genomic reconstruction of the transcriptional regulatory network in *Bacillus subtilis*, J. Bacteriol. 195 (2013) 2463–2473.

Published online 11 February 2014 *Nucleic Acids Research, 2014, Vol. 42, No. 8* ***5151–5163***
doi:10.1093/nar/gku113

Characterization of HelD, an interacting partner of RNA polymerase from *Bacillus subtilis*

Jana Wiedermannová[1,2], Petra Sudzinová[1], Tomáš Kovaľ[3], Alžbeta Rabatinová[1], Hana Šanderová[1], Olga Ramaniuk[1], Šimon Rittich[1], Jan Dohnálek[3,4], Zhihui Fu[5], Petr Halada[6], Peter Lewis[5] and Libor Krásný[1,*]

[1]Laboratory of Molecular Genetics of Bacteria, Institute of Microbiology, Academy of Sciences of the Czech Republic, Prague 14220, Czech Republic, [2]Department of Genetics and Microbiology, Faculty of Science, Charles University in Prague, Prague 12843, Czech Republic, [3]Department of Structure Analysis of Biomacromolecules, Institute of Macromolecular Chemistry, Academy of Sciences of the Czech Republic, Prague 16206, Czech Republic, [4]Laboratory of Structure and Function of Biomolecules, Institute of Biotechnology, Academy of Sciences of the Czech Republic, Prague 14220, Czech Republic, [5]School of Environmental and Life Sciences, University of Newcastle, Callaghan, NSW 2308, Australia and [6]Laboratory of Molecular Structure Characterization, Institute of Microbiology, Academy of Sciences of the Czech Republic, Prague 14220, Czech Republic

Received June 18, 2013; Revised January 14, 2014; Accepted January 16, 2014

ABSTRACT

Bacterial RNA polymerase (RNAP) is an essential multisubunit protein complex required for gene expression. Here, we characterize YvgS (HelD) from *Bacillus subtilis*, a novel binding partner of RNAP. We show that HelD interacts with RNAP-core between the secondary channel of RNAP and the alpha subunits. Importantly, we demonstrate that HelD stimulates transcription in an ATP-dependent manner by enhancing transcriptional cycling and elongation. We demonstrate that the stimulatory effect of HelD can be amplified by a small subunit of RNAP, delta. *In vivo*, HelD is not essential but it is required for timely adaptations of the cell to changing environment. In summary, this study establishes HelD as a valid component of the bacterial transcription machinery.

INTRODUCTION

RNA polymerase (RNAP) in bacteria is the key enzyme responsible for transcription of DNA into RNA. The bacterial RNAP core consists of $\alpha_2\beta\beta'\omega$ subunits and it is capable of transcription elongation but not initiation. Binding of an appropriate sigma factor to RNAP core enables the holoenzyme to recognize promoter DNA and initiate transcription (1). Unlike in gram-negative bacteria, RNAP from *Bacillus subtilis* and other gram-positive bacteria contains an additional subunit, δ. δ affects both transcription initiation and RNAP recycling, the latter depending on the ability of RNAP to be efficiently released from nucleic acids after transcription termination (2,3).

Regulation of RNAP is a complex process involving other factors besides the bona fide subunits. These factors (e.g. Gre, Nus and Rho factors) interact with RNAP and affect its function under various conditions and in various ways (4–6). Understanding the function of these factors and identification of new factors interacting with RNAP is imperative for understanding transcription and gene expression regulation.

Recently, HelD (YvgS) was identified as a binding partner of *B. subtilis* RNAP (7) and is the main copurifying band in preparations of RNAP from this organism. It is a putative helicase, and based on sequence homology, belongs to the superfamily I of DNA and RNA helicases, most closely related to HelIV helicases from gram-positive bacteria. The best characterized helicases, belonging to the same superfamily but only distantly related to HelD or HelIV, are UvrD and Rep helicases from *Escherichia coli*, or PcrA helicase from *Geobacillus stearothermophilus*. These helicases unwind DNA duplexes in an ATP-dependent manner, inchworming along the nucleic acid (8). HelD is strongly expressed during the exponential phase of growth with a further increase in expression in stationary phase (9). However, the cellular role(s) of HelD are poorly understood; it has been implicated in DNA repair and homologous recombination (10) but it has neither been characterized biochemically nor has its role(s) in transcription been investigated.

*To whom correspondence should be addressed. Tel: +420 241 063 208; Fax: +420 241 722 257; Email: krasny@biomed.cas.cz

In this study, we set out to characterize the function of HelD in transcription. We confirmed that HelD interacts with RNAP, and we identified the form of RNAP with which it interacts, and the region of RNAP to which it binds. Importantly, we found a functional link between HelD and δ and showed that these two proteins act synergistically to stimulate transcription.

MATERIALS AND METHODS

Bacterial strains and plasmids

Strains and plasmids are listed in Table 1. Competent *E. coli* cells [DH5α used for cloning or BL21 (DE3) used for overproduction of proteins] were prepared according to Hanahan (11). Competent *B. subtilis* cells were prepared as described (12).

All polymerase chain reactions (PCRs) were performed using the Expand High Fidelity System (Roche). All constructs were verified by sequencing. Wild-type *helD* was amplified by PCR from the genomic DNA of *B. subtilis* MH5636 (forward primer: 5′-caccatgaatcagcaggataagg-3′, reverse primer: 5′-tcattcagcaatctgatataag-3′), and cloned into the expression vector pET151/D-TOPO (Invitrogen) allowing in-frame fusion of a His_6 tag at the N-terminus of HelD. The resulting plasmid was named pHelD-His6 (LK800, see Table 1).

Table 1. List of strains and plasmids

Strain/plasmid	Relevant characteristics[a]	Source
B. subtilis		
MH5636	*rpoC*-His10	(13)
MGNA-A456	*helD*::MLS	(14)
LK637	*rpoC-10His, rpoE::kan*	(15)
LK782	*rpoC-10His, helD::MLS*	This work
LK1032	*rpoC-10His, rpoE::kan, helD::MLS*	This work
LK1401	*amyE::helD*-His6, *spc*	This work
BSB1	wt BaSysBio	(9)
E. coli		
LK22	BL21 pCD2/Bsu_*sigA*	(16)
RLG770	pRLG770	(17)
LK1	pRLG770 with P*veg* (−38/+1,+1G)	(18)
LK1109	pRLG770 with P*helD*	This work
LK888	pRLG770 with P*glpD*	This work
RLG7023	BL21/pFL31/*Bsu_rpoE*	(3)
LK800	BL21/pHelD-His6	This work
LK1413	pSG1721-HelD-His6	This work
pET151	pET151/D-TOPO	Invitrogen
pNG213	pETMCSIII/*His6-rpoA*	(6)
pNG490	pETMCSIII/*His6-rpoB*$_{1-608}$	(6)
pNG479	pETMCSIII/*His6-rpoB*$_{400-760}$	(6)
pNG480	pETMCSIII/*His6-rpoB*$_{750-1040}$	(6)
pNG481	pETMCSIII/*His6-rpoB*$_{950-1193}$	(6)
pNG482	pETMCSIII/*His6-rpoC*$_{1-433}$	(6)
pNG483	pETMCSIII/*His6-rpoC*$_{253-610}$	(6)
pNG484	pETMCSIII/*His6-rpoC*$_{600-915}$	(6)
pNG485	pETMCSIII/*His6-rpoC*$_{800-1199}$	(6)
pNG492	pETMCSIII/*His6-yloH* (ω_2)	(6)
pNG579	pETMCSIII/*His6-ykzG* (ω_1)	(6)
pNG613	pETMCSIII/*His6-rpoB*$_{750-846,\ 875-1040}$	(6)

[a]MLS, macrolide-lincosamide-streptogramin B resistance; kan, kanamycin; spc, spectinomycin.

The *B. subtilis helD*-null knockout strain (LK782) was prepared by transformation of *B. subtilis* strain MH5636 containing a His_{10}-tagged β' subunit (13) with chromosomal DNA from MGNA-A456, kindly provided by the National BioResource Project (Japan). A double knockout strain LK1032 (for *helD* and *rpoE* encoding the δ subunit of RNAP) was obtained by transformation of strain LK637 (15) with MGNA-A456 chromosomal DNA.

Supercoiled plasmids and linear DNA for *in vitro* transcription assays were obtained using the Wizard Midiprep Purification System (Promega) and subsequently phenol-chloroform extracted, precipitated with ethanol, and dissolved in water. The plasmids used in *in vitro* transcriptions contained promoter fragments cloned into p770 (17). Transcription terminated at a Rho-independent terminator. Linear DNA templates were prepared by PCR from the plasmid containing P*veg* (LK1). All linear templates started at −118 relative to the transcription start site. The template containing the Rho-independent terminator (at +145) ended at +255. The template without the Rho-independent terminator ended at +111. The template with the short transcribed region (Figure 6D) ended at +20.

For *in vitro* transcription assays, pRLG770 with P*veg* (−38/−1,+1G) was used (18) unless stated otherwise.

Media and growth conditions

For plasmid and protein purifications, appropriate strains were grown in Luria-Bertani (LB) medium at 37°C. For *in vivo* experiments, the cells were grown in defined 3-(N-morpholino)propanesulfonate (MOPS) - buffered medium (18) supplemented with 0.4% glucose and all 20 amino acids at 25 μg/ml.

Purification of proteins

Bacillus subtilis RNAP with a His_{10}-tagged β' subunit or His_6-HelD was purified from the strain LK782 (strain without *helD*), LK1032 (strain without *helD* and *rpoE*) or LK1401 (strain with HelD-His6). The purifications were performed as described (13). Induction of HelD-His6 in strain LK1401 was carried out at $OD_{600} = 0.5$ with 0.08% xylose for 2 h.

Plasmid pHelD-His6 was transformed into *E. coli* BL21 (DE3) and the production of HelD-His_6 induced following the addition of 1 mM IPTG for 2 h at room temperature. Cells were harvested and protein was purified by affinity chromatography as described for RNAP.

The σ^A subunit of RNAP was overproduced from the pCD2 plasmid (16) and purified as described (2).

The δ protein was purified from RLG7023 as described (3). Proteins were dialyzed against storage buffer containing 50 mM Tris–Cl, pH 8.0, 100 mM NaCl, 50% glycerol, 3 mM 2-mercaptoethanol and stored at −20° C. Proteins were visualized on NuPAGE 4-12% Bis-Tris gels (Invitrogen) with Novex Sharp Pre-Stained Protein Standard as a marker.

Determination of experimental pI

The experimental pI of HelD was determined by isoelectric focusing (IEF) using precast 5% polyacrylamide

Vertical Novex® IEF Mini Gels pH 3–10 and XCell SureLock™ mini-cell electrophoresis system (Life Technologies Corp.). IEF was performed according to the manufacturer's instructions.

Western blotting

Proteins were analyzed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie blue staining (SimplyBlue, Invitrogen) and detected by Western blotting using mouse monoclonal antibodies to σ^{70} [2G10] or to the β subunit of RNAP [8RB13] (both from Santa Cruz) and secondary antibodies conjugated with horseradish peroxidase. Signal was created using SuperSignal West Femto Chemiluminiscent Substrate (Thermo Scientific) and exposing blots to photographic film.

Far western blotting

Purified proteins were biotinylated using the EZ-Link sulfo-NHS-biotinylation system (Thermo Scientific). Far western blots were performed as detailed by Yang *et al.* (6), except protein–protein interactions were detected using Horseradish peroxidase-conjugated streptavidin and the Opti4CN system (BioRad). Binding affinity to RNAP fragments was determined from digitized scans of blots using ImageJ (NIH) where maximal binding (100%) was set as the intensity of the signal of HelD bound to the $\beta'_{600–915}$ fragment. HelD binding sites on RNAP were mapped onto the *B. subtilis* RNAP homology model (19) and visualized using PyMol (Schrödinger).

Native PAGE assays

Five picomoles of RNAP and 25 pmol of HelD, T4 DNA ligase (TaKaRa) or MLV reverse transcriptase (Promega), respectively, were used. Proteins tested for mutual interactions were incubated for 15 min at 30°C in 10 µl in the storage buffer. After incubation samples were mixed with 3 µl of Native PAGE 4× Sample buffer (Invitrogen) and loaded onto the Native PAGE 4-16% Bis-Tris Gel (Invitrogen) and electrophoresed. The gels were subsequently Coomassie stained.

Enzymatic digestion for mass spectrometry

Coomassie blue-stained protein bands were excised from the gel, cut into small pieces and destained using 50 mM 4-ethylmorpholine acetate (pH 8.1) in 50% acetonitrile (MeCN). The proteins were further reduced with 30 mM tris(2-carboxyethyl)phosphine (TCEP) in 100 mM Tris–HCl (pH 8.0) at 65°C for 30 min and alkylated with 30 mM iodacetamide in 100 mM Tris–HCl (pH 8.0) for 60 min in the dark. The gel was washed with water, shrunk by dehydration in MeCN and re-swelled again in water. The supernatant was removed and the gel was partly dried in a SpeedVac concentrator. The gel pieces were then incubated overnight at 37°C in cleavage buffer containing 25 mM 4-ethylmorpholine acetate, 5% MeCN and trypsin (100 ng; Promega). The resulting peptides were extracted into 40% MeCN/0.3% trifluoroacetic acid (TFA). An aqueous 50% MeCN/0.1% TFA solution of α-cyano-4-hydroxycinnamic acid (5 mg/ml; Sigma) was used as a MALDI matrix. One microliter of the peptide mixture was deposited on the MALDI plate, allowed to air-dry at room temperature and overlaid with 0.4 µl of the matrix.

MALDI mass spectrometry and protein identification

Mass spectra were measured on an Ultraflex III MALDI-TOF instrument (Bruker Daltonics, Bremen, Germany) in the mass range of 700–4000 Da and calibrated internally using the monoisotopic $[M+H]^+$ ions of trypsin autoproteolytic fragments (842.5 and 2211.1 Da). The peak lists created using flexAnalysis 3.3 were searched using an in-house MASCOT search engine against the SwissProt 2013_09 database subset of *B. subtilis* proteins with the following search settings: peptide tolerance of 30 ppm, missed cleavage site value set to one, variable carbamidomethylation of cysteine, oxidation of methionine and protein N-term acetylation. Proteins with MOWSE scores over the threshold value of 56 calculated for the used settings were considered as being positively identified. If the score was lower or only slightly higher than the threshold value, the identity of protein candidate was confirmed by tandem mass spectrometry analysis.

ATPase activity

ATPase activity was measured by the hydrolysis of inorganic phosphate from [γ-^{32}P] ATP. The reaction mixture in 110 µl of 50 mM Tris–Cl, pH 8; 5 mM $MgCl_2$, 1 M KCl contained 550 pmol HelD or 550 pmol bovine serum albumin (BSA) or 275 pmol *Bacillus stearothermophilus* EF-Tu G-domain and 3450 pmol [γ-^{32}P] ATP (specific activity 1800 cpm/pmol). The reaction was performed at 30°C and followed kinetically for 90 min at 30°C. Aliquots of 20 µl were withdrawn at appropriate time intervals (0, 30, 60 and 90 min) and liberated Pi determined by the charcoal method (20). Five microliter of the product was spotted on Whatman 3MM filter paper, dried and scanned using a Molecular Imager_FX (Bio-Rad). The amounts were quantified with QuantityOne software (Bio-Rad). Blank samples were run simultaneously to determine background values of ATP hydrolysis. The experiments were repeated three times. The amounts of hydrolyzed ATP were calculated and plotted as the function of time.

In vitro transcription assays

Initiation competent enzyme was reconstituted using RNAP isolated from LK782 or LK1032 with a saturating concentration of σ^A in storage buffer (50 mM Tris–HCl, pH 8.0, 0.1 M NaCl, 50% glycerol) for 15 min at 30°C.

Multiple round transcription assays were carried out essentially as described by (13,18) unless stated otherwise. Briefly, reactions were carried out in 10 µl: 30 nM holoenzyme ($RNAP\sigma^A$), 2.5 nM supercoiled or 50 nM linear DNA template unless stated otherwise, transcription buffer (40 mM Tris–HCl, pH 8.0, 10 mM $MgCl_2$, 0.1 mg/ml BSA and 1 mM dithiothreitol (DTT)), 150 mM KCl

and NTPs (ATP, CTP and GTP were 200 μM; UTP was 10 μM plus 2 μM of radiolabeled [α-^{32}P]-UTP).

RNAP was reconstituted with HelD at a 1:4 ratio unless stated otherwise. The RNAP:δ ratio used in experiments where δ was added was 1:4 (saturating concentration). Reconstitution experiments were carried out in storage buffer for 15 min at 30°C. When denatured HelD was used, native HelD protein was denatured at 90°C for 5 min.

All transcription reactions were allowed to proceed for 15 min at 30°C (unless stated otherwise) and stopped with equal volumes of formamide stop solution (95% formamide, 20 mM EDTA, pH 8.0). Transcription assays on 20 nt linear templates were allowed to proceed for 30 min.

In single round transcription assays, and in the multiple round assays (Figure 6), RNAP was preincubated with plasmid DNA and HelD and δ were subsequently added and incubated for 10 min at 30°C. Reactions were carried out in transcription buffer supplemented with 150 mM KCl and started by the addition of NTPs (concentration of NTPs was the same as in multiple round *in vitro* transcription assays) together with 600 nM double-stranded DNA (dsDNA) competitor. The competitor dsDNA with a full-consensus promoter sequence (21) was used to sequester free RNAP and allow only one round of *in vitro* transcription (22). Stock dsDNA competitor was prepared by annealing equimolar amounts of complementary primers (LK 923: 5′-ccggaattcaaatatttgttgttaactcttgacaaaagtgttaaattgtgctatactgtattggttctcaagcttccg-3′ and LK 924: 5′-cggaagcttgagaaccaatacagtatagcacaatttaacacttttgtcaagagttaacaacaaatatttgaattccgg-3′) in 10 mM Tris–HCl, pH 8.0, 1 mM EDTA, 50 mM NaCl, which were denatured at 95°C for 5 min and then cooled down to 30°C (1°C per min). In negative controls, whole reaction mix with plasmid DNA and competitor dsDNA was started with RNAP to ensure that all RNAPs were sequestered by the dsDNA competitor.

In vitro transcription restart assays (Figure 5E and F) were carried out in two steps. The first 15-min step was basically the same as described above for multiple round reactions; KCl was used at 100 mM concentration. At the beginning of step II, the main compounds (water, NTPs, template DNA, δ, HelD, both δ and HelD, RNAP) were added to respective reactions in the same amounts as they were at the beginning of step I. Reactions were then allowed to proceed for another 15 min and were stopped with 10 μl of stop solution. In experiments with two templates (Figure 6A), 10 nM final concentrations of the plasmids were used.

Samples were loaded onto 7 M Urea 7% polyacrylamide gels and electrophoresed. The dried gels were scanned with a Molecular Imager_FX (BioRad). The amount of transcript (originating from the cloned promoters) was quantified with QuantityOne software (BioRad). All calculations and data fitting were carried out using SigmaPlot (Jandel Scientific).

In vivo experiments: outgrowth from the lag phase

Wild type *B. subtilis*, δ and HelD knockout, and the double knockout strains (MH5636, LK637, LK782 and LK1032, respectively) were cultivated in LB medium for 24 h at 37°C under continuous shaking to ensure entry into stationary phase and then diluted into fresh LB medium to $OD_{600} = 0.03$. OD_{600} was measured during outgrowth to compare the duration of the lag phase of the respective strains.

Protein sequence and domain analysis

Protein sequence searches were performed with the BLAST protein–protein service (23,24). Sequence alignments were carried out with Clustalx, using the Gonnet 250 weight matrix (25).

RESULTS

HelD copurifies with RNAP

In our preparations of *B. subtilis* RNAP by affinity chromatography, we regularly observed a major contaminating band of ~90 kDa (Figure 1A). This band did not appear in control experiments where the lysate made from *B. subtilis* cells containing no His-tagged protein was incubated with the affinity matrix, indicating that the protein does not have an intrinsic ability to bind the matrix (Figure 2A). This protein could not be removed by gel filtration and its level in the preparation decreased only after ion exchange chromatography [(26) and data

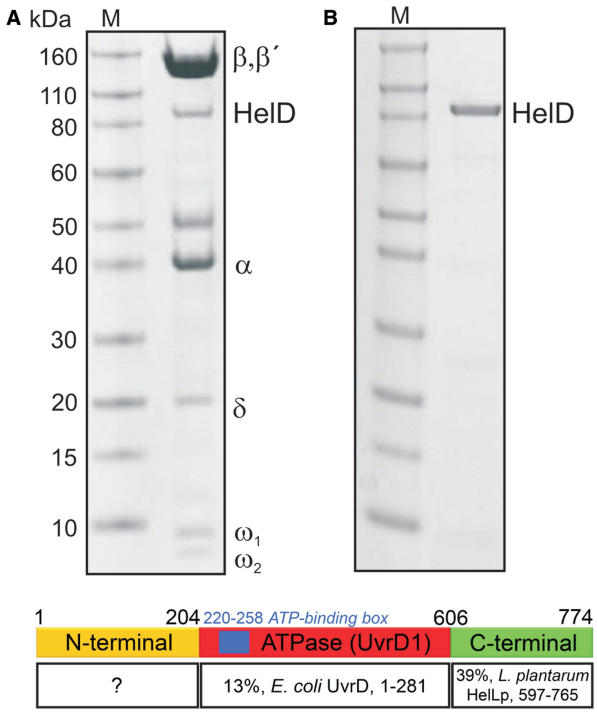


Figure 1. HelD copurifies with RNAP. (**A**) RNAP purified by Ni-affinity chromatography via the 10× N-terminal histidines on the β' subunit. Subunits of RNAP and the band corresponding to HelD are indicated. (**B**) HelD purified by Ni-affinity chromatography via the 6× N-terminal histidines. The apparent molecular weights are shown by Novex Sharp Pre-Stained Protein Standard, M. Panel C, Putative domain structure of *B. subtilis* HelD. Amino acid identity with respective protein fragments is indicated. The percent values show the level of sequence identity to the marked protein segment.

not shown]. This band was identified by MALDI mass spectrometry as the HelD protein, which was recently reported as a binding partner of RNAP (7).

The HelD protein is encoded by the *yvgS* (*helD*) gene and consists of 774 amino acids (aa). Based on its aa sequence, HelD is a putative UvrD-like helicase. Helicase from *Lactobacillus plantarum* HelLp, with a sequence identity of 39%, is the closest related protein for which any 3D structure (C-terminal domain, PDB ID 3DMN, unpublished) is available. The closest structurally related helicase of this class, UvrD from *E. coli*, for which the complete 3D structure has been determined (27), shows only 12% sequence identity with HelD. Based on domain arrangement of HelLp and UvrD and sequence comparisons of these two proteins with HelD, it is postulated that HelD comprises three domains (Figure 1C). The first, N-terminal domain (residues 1–204) may be involved in DNA binding. The second domain (205–606) contains the ATP-binding box and is related to the ATPase domain of UvrD helicase (residues 1–281 in UvrD, domain marked as UvrD1). The C-terminal domain (607–774) is related to the domain of UvrD helicase involved in DNA unwinding and shares significant similarity with the C-terminal domain of HelLp. However, in the case of UvrD this domain is formed by two distant parts of the protein chain, whereas the HelLp C-domain is formed by a single section of the protein chain. In summary, HelD is a unique protein only part of which resembles proteins characterized to date.

For further studies, we prepared recombinant HelD. The recombinant protein was purified by nickel affinity chromatography via the introduced N-terminal 6xHis tag (Figure 1B) and the experimentally determined pI of the protein under native conditions was ~7.1 (theoretical pI = 6.1). HelD was able to bind DNA (Supplementary Figure S1), and as it was predicted to be a helicase, we attempted to detect this activity in strand displacement assays. Using an array of DNA templates [5′ or 3′ overhangs, forked DNA or blunt-ended DNA; (28)] we detected no strand-displacement activity for this protein (data not shown).

Interaction of HelD with RNAP

To confirm that HelD binds to RNAP and to establish to which form (core or holoenzyme), we performed *in vivo* and *in vitro* experiments. *In vivo*, we overproduced C-terminally His-tagged HelD in *B. subtilis* cells (strain LK1401) and subsequently purified it via the His-tag. In parallel, we purified RNAP that was His-tagged at β' (MH5636) and we also performed a control purification from a strain without any His-tagged protein (BSB1). Figure 2A shows that RNAP core subunits copurified with the His-tagged HelD, and that RNAP is the main interacting partner of this protein. Western blot analysis (Figure 2B) showed that HelD interacts predominantly with the core form of RNAP as virtually no σ^A was detected.

To verify that purified HelD binds to RNAP *in vitro* we performed gel-shift experiments under native conditions in nondenaturing PAGE gels. Figure 2C lane 1 shows that HelD forms oligomers in solution as also seen by size exclusion chromatography (Supplementary Figure S2). The addition of HelD to RNAP resulted in reduced migration of RNAP, suggesting binding of HelD to

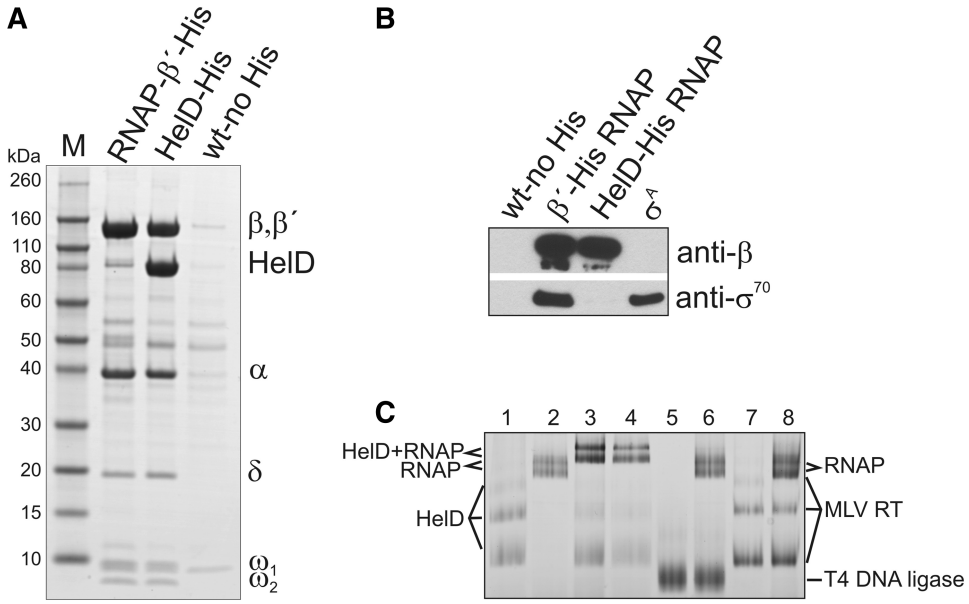


Figure 2. HelD interacts with RNAP core. (**A**) Coomassie stained gel of proteins purified by nickel affinity chromoatography. Lane 1—proteins purified from strain MH5636 containing His-tagged β' subunit of RNAP; Lane 2—proteins purified from strain LK1401 containing His-tagged HelD; Lane 3—proteins purified from strain (BSB1) without any His-tagged protein. (**B**) Western blot of the gel from (A). The antibodies used are indicated. In the last lane purified σ^A was used as a marker. (**C**) Band shift assay on native PAGE gel. RNAP was incubated with potential interacting partners: HelD, T4 DNA ligase (68 kDa), MLV reverse transcriptase (71 kDa). Samples were separated on 4–16% gradient native PAGE Bis–Tris gels. Lane 1—HelD, lane 2—RNAP, lane 3—HelD with RNAP, lane 4—DNaseI-treated HelD + DNaseI-treated RNAP, lane 5—T4 DNA ligase, lane 6—RNAP with T4 DNA ligase, lane 7—MLV reverse transcriptase, 6—RNAP with MLV reverse transcriptase. In all experiments, RNAP was prepared from the Δ*helD* strain (LK782).

RNAP (Figure 2C, lane 3). To confirm results from the gel-shift experiment, we excised the two bands predicted to be RNAP from lane 2 and lane 3 and determined the protein identity by mass spectrometry. The analysis confirmed the presence of the β, β' and α subunits as well as HelD in the bands in lane 3. It is not clear why RNAP and RNAP-HelD complexes ran as two bands on the gel (Figure 2C, lanes 2 and 3), but they appeared identical by mass spectrometry analysis, and we speculate that they may represent two conformational states of RNAP, both capable of HelD binding. We also verified by mass spectrometry that the bands in lane 1 consist of HelD (probably in different oligomeric states). As an additional control to exclude the possibility that HelD binds to RNAP via traces of DNA remaining in complex with the proteins, DNaseI was added to the RNAP and HelD preparations before the experiment and we obtained the same result (Figure 2C, lane 4). As specificity controls, we used T4 ligase and murine reverse transcriptase and neither of these proteins bound to RNAP (Figure 2C, lanes 5–8).

We conclude that HelD interacts with RNAP core *in vivo* and that purified HelD interacts directly with RNAP *in vitro*.

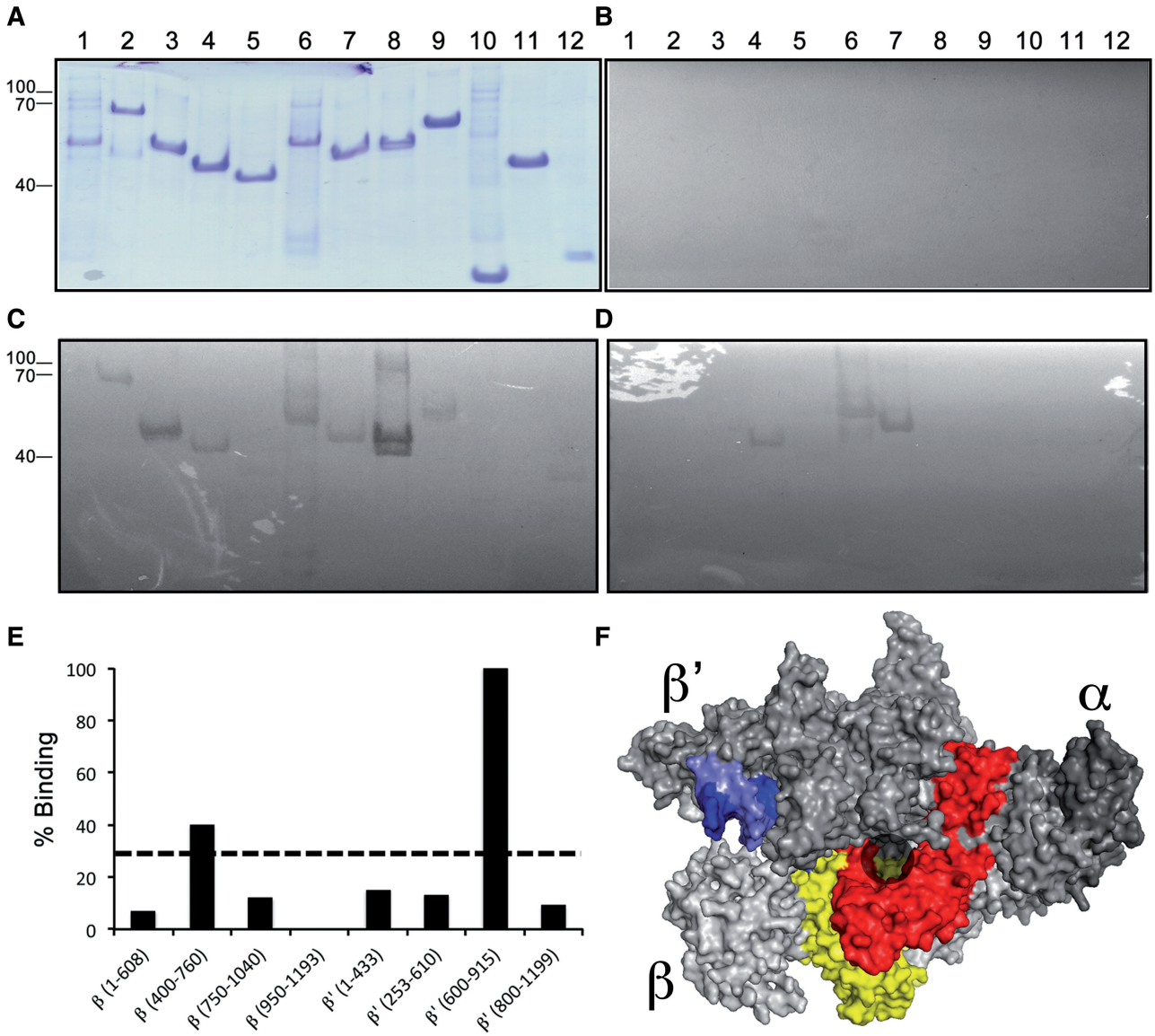


Figure 3. HelD binds on the downstream side of RNAP. (**A**) Coomassie blue stain of RNAP subunits and fragments showing approximate equal loading in all lanes. (**B**) Negative control blot containing all components except biotinylated ligand. (**C**) Blot performed using biotinylated HelD. (**D**) Control blot using biotinylated σ^A. Molecular weights (kDa) are shown down the left hand side. Lane 1—α; lane 2—$\beta_{1–606}$; lane 3—$\beta_{400–760}$; lane 4—$\beta_{750–1040}$; lane 5—$\beta_{950–1193}$; lane 6—$\beta'_{1–433}$; lane 7—$\beta'_{253–610}$; lane 8—$\beta'_{600–915}$; lane 9—$\beta'_{800–1199}$; lane 10—ω_2; lane 11—$\beta_{750–1040}$ Δflap; lane 12—ω_1. (**E**) Quantification of binding signal of HelD to β and β' fragments normalized to the strongest binding fragment, $\beta'_{600–915}$ (set at 100%). Dotted line represents 30%. (**F**) Surface rendered model of *B. subtilis* RNAP with the strongest HelD binding fragments labeled in yellow ($\beta_{400–760}$) and red ($\beta'_{600–915}$). α Subunits are shown in dark gray, β subunit in light gray and β' subunit in medium gray. DNA is shown in blue; template strand in light blue and nontemplate strand in dark blue. The approximate location of the secondary channel is indicated by the transparent gray circle.

Localization of HelD on RNAP

Next, we addressed the topology of the binding of HelD to RNAP. We used far western blot analysis using fragments of subunits of *B. subtilis* RNAP probed with biotinylated HelD (Figure 3). Controls were performed with no protein (Figure 3B) and with biotinylated σ^A that bound to the N-terminus of the β' subunit as well as the fragment of the β subunit containing the β-flap tip (Figure 3D). These data are consistent with previous far western blot results using anti-σ^A antibodies (6) and structural data (29). When using HelD as a probe, signal was observed for most of the fragments of the β and β' subunits (Figure 3C), but as shown in Figure 3E and F, it bound most strongly to two fragments that form the rim of the secondary channel by which it is believed that NTPs enter the active site (gray circle, Figure 3F). Mapping of the fragments to which HelD bound weakly enabled us to identify portions that all triangulated to the surface of RNAP close to the secondary channel (not shown) consistent with HelD forming extensive contacts with RNAP on the downstream side in close proximity to DNA (light and dark blue, Figure 3F). Assuming HelD binds on the downstream side of RNAP it is unlikely there is significant steric hindrance preventing σ from being simultaneously bound.

HelD affects transcription *in vitro*

Because HelD directly interacts with RNAP, we wished to determine whether it was important for the DNA-dependent synthesis of RNA. Therefore, we performed a panel of *in vitro* transcription experiments with RNAP purified from a HelD knockout strain, and studied the effect of addition of HelD. As a model promoter, we used P*veg*, a strong constitutive promoter that is well characterized (18). From this promoter RNAP transcribes a 145 nt transcript and transcription is terminated at a Rho-independent terminator (17,30). In *in vitro* multiple round transcription assays HelD displayed a stimulatory effect on transcription (Figure 4A, black columns; Supplementary Figure S3). However, a large stoichiometric excess of HelD over RNAP abolished transcription, possibly by nonspecific binding of HelD to DNA, resulting in the formation of transcription roadblocks (31). Negative control experiments using heat-denatured HelD failed to produce any stimulatory effect on transcription activity (Figure 4A; gray columns). We also tested the effect of HelD with two other templates containing two other promoters, P*helD* and P*glpD* (32), and we obtained similar results to those obtained with P*veg* (Supplementary Figure S3AB). Moreover the effect of HelD on transcription was salt-concentration dependent, as increasing the amount of salt in the reaction rendered the effect more pronounced (Supplementary Figure S4).

We conclude that HelD in low stoichiometric excess over RNAP stimulates transcription.

The effect of HelD is ATP-dependent

As HelD contains an ATP binding motif, we investigated whether the effect of HelD on transcription depends on ATP. First, we tested whether HelD possesses an ATPase activity. While none of the two control proteins, the GTPase domain of *B. stearothermophilus* elongation factor Tu or BSA, was able to hydrolyze ATP, HelD displayed significant ATPase activity (Figure 4B and C). This effect was independent of the presence of RNAP (data not shown). Next, we performed transcription assays with increasing concentrations of ATP either in the presence or absence of HelD. As a control, we performed the same type of experiment with increasing concentrations of CTP. When ATP or CTP concentrations were increased (from 1 to 200 μM), the other NTP concentrations were kept constant (100 μM). Figure 4D and E shows that by increasing the concentration of ATP from 1 to 20 μM (where the effect plateaued) the level of transcription increased ~6-fold in the presence of HelD. When HelD was not present, this stimulation was absent. Furthermore, increasing the CTP concentration had no significant effect on transcription either in the presence or absence of HelD. The overall higher level of transcription in the presence of HelD when CTP concentration was being varied was due to the 100 μM ATP that was present in the reaction.

We conclude that HelD is an ATPase and the stimulatory effect of HelD on transcription is ATP-dependent.

HelD and *δ* have a synergistic stimulatory effect on transcription cycling

The next question was which part of the process of RNA synthesis is affected by HelD. Even though Figure 2B shows that HelD binds to core and not holoenzyme, we wanted to test whether it has any effects on steps important for transcription initiation—RNAP association with σ^A and promoter binding. Unsurprisingly, the presence or absence of HelD had no effect either on the affinity of σ^A for RNAP or the affinity of RNAP for promoter DNA (Supplementary Figure S5A and B).

However, another putative helicase, the RapA protein of *E. coli*, has been shown to associate with RNAP and stimulate transcription by promoting cycling of RNAP by enhancing the release of RNAP from nucleic acids after termination and also by stimulating the dissociation of RNA–DNA hybrids (30). HelD and RapA are not sequence homologs; RapA belongs to the Swi2 family of helicases, but we speculated that these two proteins may be functional analogs. As the δ subunit of RNAP was previously also shown to enhance cycling of RNAP, we included δ as a positive control. We used RNAP from a double knockout strain (δ-null HelD-null, LK1032) in *in vitro* transcription assays from a P*veg*-based supercoiled template. First, we tested the effects of HelD and δ separately and observed moderate stimulatory effects for each (2-fold increase) (Figure 5A and B; multiple rounds). Surprisingly, when we combined HelD and δ together, we detected a strong increase in RNA synthesis. The ~10-fold stimulation by HelD and δ together was more than the sum of the stimulatory effects of HelD and δ alone, suggesting a synergistic effect of these proteins.

Next, we decided to directly test whether the synergistic effect of HelD and δ is due to a more efficient cycling of transcription. If this was true, then limiting transcription

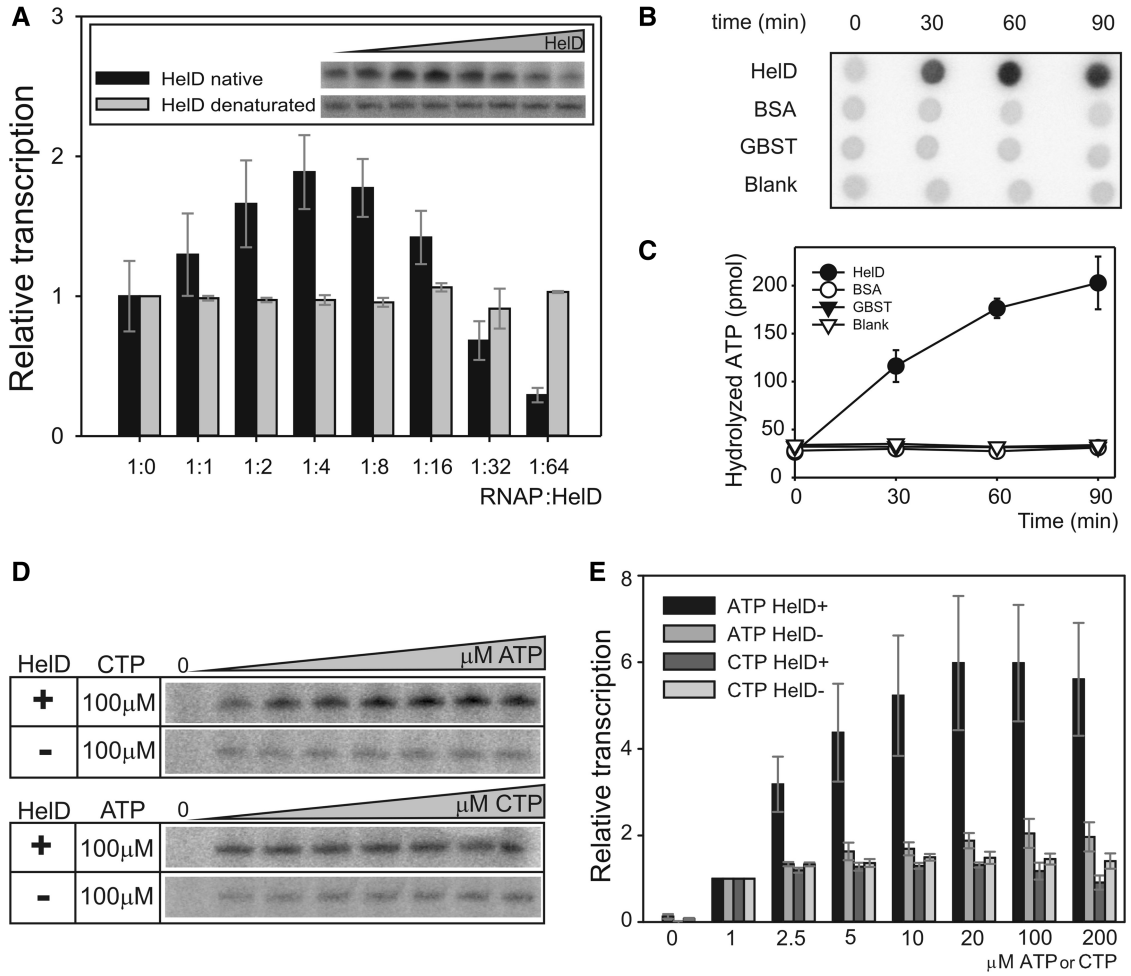


Figure 4. Effect of HelD on *in vitro* transcription. (**A**) RNAP (from HelD knockout strain LK782) was reconstituted with saturating concentrations of σ^A. Holoenzyme was incubated with increasing amounts (molar ratio from 1:0 to 1:64) of HelD (black bars) or heat denatured HelD (gray bars) and used to initiate transcription. Primary data are shown in the inserted box and show a representative experiment. The data were normalized to the 1:0 ratio set as 1. In this and the following experiments, the graphs represent data from three independent experiments ±SD. (**B**) HelD hydrolyzes ATP. Primary data show spots with hydrolyzed γ-^{32}P (for details see 'Materials and Methods' section). (**C**) Graph showing the quantification, with the symbol key presented in the graph. (**D**) The effect of HelD is ATP-dependent- representative primary data. RNAP from HelD knockout strain (LK 782) was tested in *in vitro* multiple round transcription assays in the presence/absence of HelD and increasing amounts of ATP (from 1 to 200 μM total ATP). As a control, CTP was varied. The concentration of CTP was 100 μM when [ATP] was varied and *vice versa*. (**E**) The effect of HelD is ATP-dependent quantification of the data normalized to transcription at 1 μM ATP or CTP. The color coding of the bars is defined in the graph.

to a single round should abolish the effect. As shown in Figure 5A and B, the pronounced stimulatory effect of these factors was cancelled, consistent with the stimulatory effect of HelD and δ being due to cycling of transcription.

To examine the cycling in more detail, we performed multiple round transcription experiments where we followed transcription in the presence/absence of δ, HelD or δ and HelD as a function of time. In the absence of HelD or δ, transcription stopped at about the 5-min time point. The addition of either protein modestly prolonged this time. The addition of both proteins then markedly increased the time with transcription still continuing at the final 25-min point (Figure 5C and D).

This experiment raised two principal questions: (i) Why does the cycling of transcription stop after 5 min in the absence of δ and HelD? and (ii) Why is the cycling more efficient with δ and HelD? We address these questions in the following two sections.

Transcription cycling stops because the template is not functional

There are three possibilities to explain why transcription stops after 5 min: first, depletion of NTPs prevents RNA synthesis; second, RNAP is in complex with RNA and/or DNA and inactive; third, the template DNA is not functional either because it is blocked by RNAP, or because it forms interactions with RNA [such as R-loops (33)].

To distinguish between these possibilities we allowed the *in vitro* transcription assays to proceed for 15 min without HelD and δ and then divided the mixture into several aliquots. Next, equal amounts of the main components (NTPs, template DNA, RNAP) were added as at the

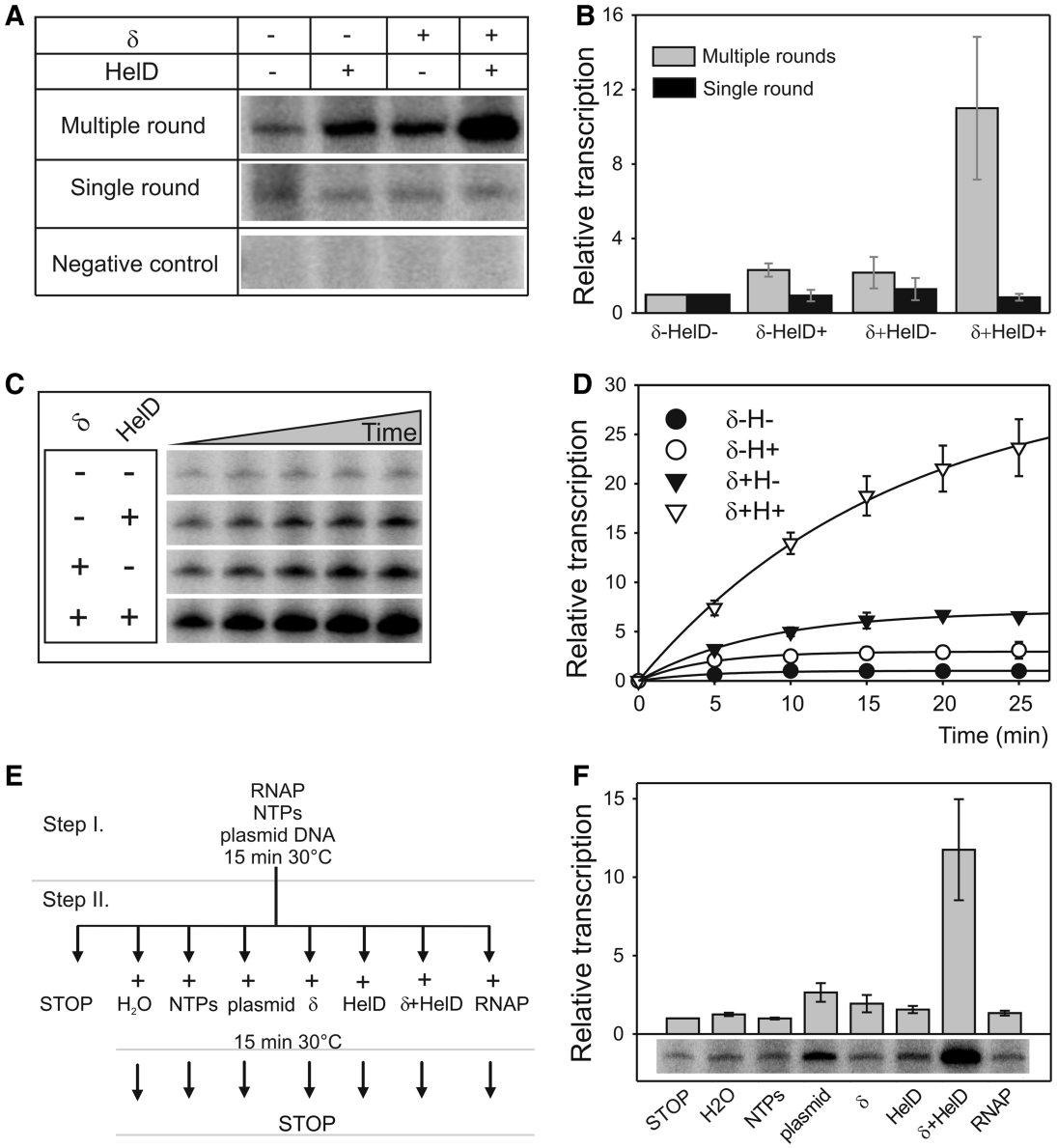


Figure 5. HelD and δ stimulate cycling of transcription. (**A**) Multiple round transcription assays were conducted in the absence of a competitor (dsDNA), whereas single round assays were conducted in its presence. In single round assays, the competitor was added to preform open complexes together with NTPs that were used to initiate transcription. The negative control shows the efficiency of the dsDNA competitor at preventing transcription of the 145-nt product when added to assay mixtures before formation of the open complex. (**B**) Quantification of data from multiple round (gray bars) and single round (black bars) transcription assays. Data from three independent experiments were quantified; the graph shows averages ±SD. (**C**) Multiple round *in vitro* transcription is limited in time—representative data. RNAP was preincubated with/without δ and/or HelD. Reactions were started at time zero by NTPs. Aliquots were withdrawn at 5, 10, 15, 20 and 25 min, respectively, and stopped by formamide stop solution. (**D**) Multiple round *in vitro* transcription is limited in time—quantification. Data were quantified from two independent experiments normalized to the value of transcription without δ and HelD after 15 min (set as 1). The bars indicate the range. (**E**) HelD and δ restart halted *in vitro* transcription—a schematic drawing of the experiment. In step I, all the indicated components of the transcription reaction were combined together and transcription reactions were allowed to proceed for 15 min. At this time point, the mixture was distributed to eight tubes containing the indicated components, and the reactions (Step II) were allowed to proceed for further 15 min. (**F**) HelD and δ restart halted *in vitro* transcription—a representative result and quantitation of the data. The quantified data were normalized to lane 1, set as 1. The graph represents data from three independent experiments ±SD.

beginning of the reaction, or HelD, δ or HelD and δ were added, and the reaction allowed to proceed for another 15 min. From the results (Figure 5E and F) we can exclude the first hypothesis: the addition of NTPs did not restart transcription (Figure 5F, lane 3). We can also exclude inactivation of all RNAP present in the experiment (second hypothesis) as addition of RNAP did not restart transcription (Figure 5F, lane 8). This is in agreement with the fact that there was 12-fold molar excess of RNAP over the template DNA already at the beginning of the reaction. Finally, the addition of plasmid DNA resulted in ~2-fold increase in transcription, suggesting it is the lack of an available template DNA that prevents further transcription (third hypothesis).

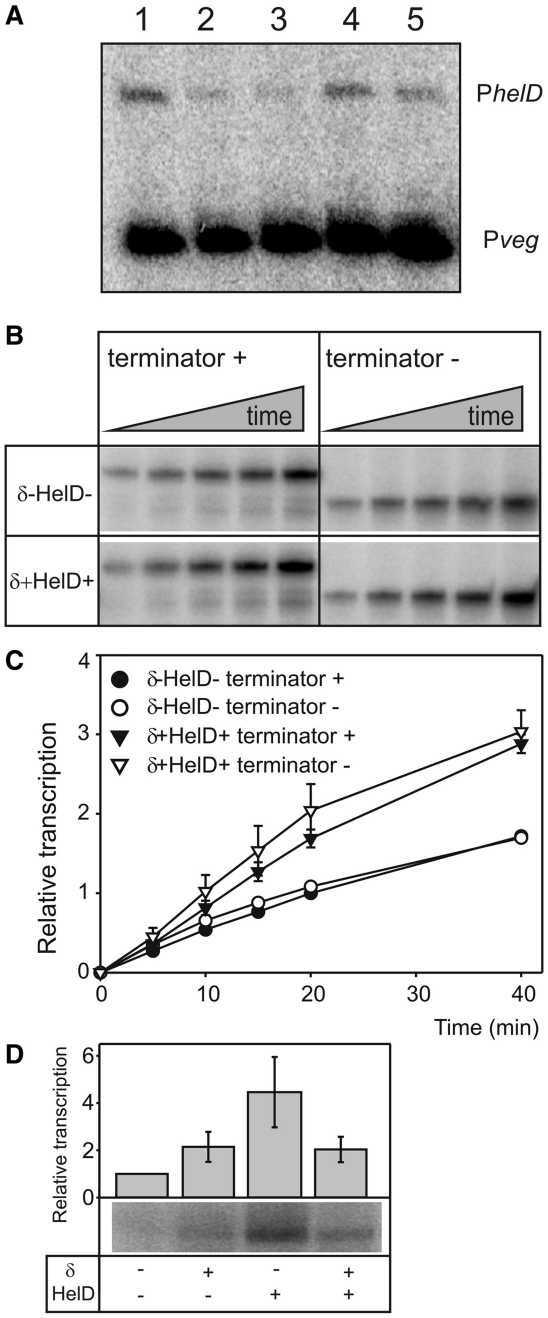


Figure 6. HelD liberates RNAP and affects elongation. (**A**) HelD liberates sequestered RNAPs—primary data. Transcription assays were performed with two supercoiled templates, either P*veg* (145-nt transcript) or P*helD* (245 nt). P*veg* was in all reactions from the beginning. P*helD* was added at the beginning (lane 1) or after 15 min (lanes 2–5) and reactions were allowed to proceed for another 15 min. RNAP contained no HelD or δ. These proteins were added after 15 min (lane 2—no protein; lane 3—δ; lane 4—HelD; lane 5—HelD+δ). (**B**) HelD and δ increase transcriptional rate on linear templates regardless of the presence/absence of a Rho-independent terminator—representative primary data. Two templates both containing the P*veg* promoter were used (see 'Materials and Methods' section for details). One template contained a Rho-independent terminator, while the other lacked this terminator. With both templates, the experiment was conducted ± HelD+δ. (**C**) Quantification of the data (plus two more repeats) from the experiment shown in (**B**). Panel (**D**), HelD stimulates transcription of 20 nt RNA. Multiple round *in vitro* transcription reactions on a linear template with a 20 bp template were performed in the presence/absence of HelD and/or δ. The radioactively labeled transcripts were resolved on 20% polyacrylamide gels. The graph shows a representative result and quantification of the data. The quantified data were normalized to lane 1, set as 1. The graph represents data from three independent experiments ± SD.

The addition of HelD or δ alone had again a moderate stimulatory effect on transcription. However, a pronounced stimulatory effect was observed on the addition of HelD and δ together (Figure 5F, lane 7). This demonstrated that the lack of transcription-competent template was not caused by, for example, degradation or irreversible alteration of DNA.

HelD and *δ* release RNAP from DNA and affect elongation

The reason for the transcriptional inactivity of the plasmid in the previous experiment could be due to the presence of stalled (nonproductive) RNAP complexes either on the P*veg* template or elsewhere on the plasmid. Another possibility is an alteration of the plasmid by transcription (e.g. a change of supercoiling), which would suggest an effect of HelD on elongation. Finally, the plasmid could be in complex with RNA (e.g. R-loops) and thus not functional in transcription. Previous experiments have already ruled out such possibilities as affecting promoter affinity (Supplementary Figure S5B) or affinity of σ^A for RNAP (Supplementary Figure S5A).

To test whether the presence of δ and HelD releases nonproductive RNAP from DNA we performed *in vitro* transcription assays with two templates producing RNAs of different length (from P*veg* and P*helD* promoters). A subsaturating concentration of RNAP (with respect to the plasmid template) was used. When both templates were added at time zero, both were transcribed by RNAP in the absence of δ and HelD (Figure 6A, lane 1). When transcription started only from P*veg*, and P*helD* was added after 15 min, free RNAP appeared to be unavailable, as only a small amount of transcript from P*helD* was generated (Figure 6A, lane 2), suggesting that RNAP may be sequestered on the P*veg* template DNA. This effect was alleviated when HelD or HelD with δ were added together with the second template (Figure 6A, lanes 4 and 5). The presence of δ and HelD enabled RNAP to be reused for transcription from both templates. We note that δ alone did not display a stimulatory effect with the P*helD* template (Figure 6A, lane 3). This could be due to an effect of δ on transcription initiation at P*helD*, as δ is known to decrease transcription from promoters that form relatively unstable open complexes (important intermediates during transcription initiation) (15).

Next, we tested whether HelD and δ were able to release RNAP from posttermination complexes as suggested for RapA (30,34). We performed *in vitro* transcription reactions with linear templates containing or lacking Rho-independent terminators. Figure 6B and C shows that the presence or absence of the terminator had no effect on the stimulatory effect of HelD and δ. Interestingly,

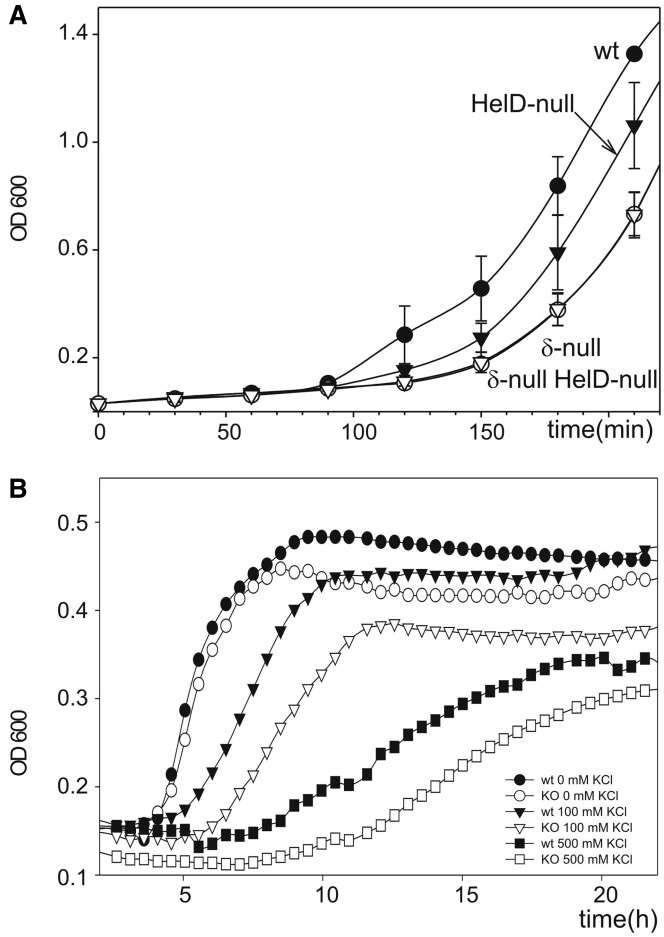


Figure 7. HelD and δ affect the growth phenotype. **(A)** wt, wild type *B. subtilis* strain (MH5636): closed circles; HelD-null (LK782): closed triangles; δ-null (LK637): open circles; δ-null HelD-null (LK1032): open triangles. The strains were grown for 24 h at 37°C. After 24 h, the cells were diluted into fresh LB medium at $OD_{600} = 0.03$ and the growth curves after the dilution are shown. (**B**) Growth in defined MOPS medium supplemented with all 20 amino acids in the presence 0, 0.5 or 1.0 M KCl as stated in the figure legend. Cells were cultivated in 1 ml of volumes at 37°C in 24-well EMSA plates in a TECAN Infinite® 200 Pro and OD 600 was measured every 10 min.

transcription did not stop on the linear templates after 5 min in the absence of these two proteins but continued throughout the duration of the experiment. However, the overall level of transcription was less in the absence of HelD and δ. We tested the effect of HelD and δ on releasing RNAP from the linear template and detected only a modest (20%) stimulation of this process (Supplementary Figure S6A and B). Hence, the main stimulatory effect of HelD on linear templates is most consistent with an effect on elongation.

To test whether the elongation effect could be due to the inability of RNAP to read through complexes of DNA with RNA, as also suggested for RapA (34), we performed experiments with a short linear DNA template where the transcribed region was only 20 nt long. This short length of transcribed template does not permit significant interactions of the nascent RNA with DNA (35). As shown in Figure 6D, HelD and δ were still capable of stimulating transcription. Interestingly, the effect of δ with HelD was decreased relative to the effect of HelD alone. A possible explanation could be a parallel negative effect of δ on, for example, the affinity of RNAP for the promoter located on this short linear DNA fragment.

We conclude that the effect of HelD in conjunction with δ on transcriptional cycling on supercoiled templates appears to be caused, at least in part, by liberating nonproductive RNAPs from complexes with DNA. A parallel stimulatory effect of HelD on the rate of elongation that is more apparent with linear templates seems to also play a role.

HelD and *δ* display similar phenotypes *in vivo*

To assess the importance of HelD for the cell, we performed phenotypic experiments, using a HelD-null strain. In comparison with wild type, no difference in growth was observed in rich (LB) or nutritionally defined media (MOPS). Similarly, no difference was observed in response to various stresses, such as heat and ethanol (data not shown). However, the mutant strain displayed a prolonged lag phase for stationary cells diluted into fresh medium. This phenotype was reminiscent of the phenotype displayed by the δ-null strain (10,15). Therefore, we directly compared wild type, δ-null, HelD-null and the double knockout strain, HelD-null and δ-null (MH5636, LK637, LK782 and LK1032, respectively). The cells were grown for 24 h and then diluted into fresh LB medium. The outgrowth from lag phase of the δ-null strain was delayed compared with wild-type cells for ~30 min, as reported previously (Figure 7A). Growth of the HelD-null strain was delayed for ~15 min and the double knockout strain displayed an identical delay to the δ-null strain. The same phenotype was observed for wild-type versus HelD-null strain in defined rich medium (MOPS supplemented with all amino acids) and in poor M9 medium (Supplementary Figure S7A). Further, we also observed a minor effect on growth in the presence of increased amounts of salt (Figure 7B and Supplementary Figure S7B). Albeit relatively small, these phenotypes were highly reproducible. We concluded that the absence of HelD decreased the ability of the cell to rapidly adapt to nutritional changes in the environment.

DISCUSSION

In this study we confirmed HelD as a new interacting partner of RNAP from *B. subtilis*, we identified that it binds to the core form of the enzyme, localized its approximate binding region on RNAP and provided insights into the role of HelD in transcription. We have shown that HelD hydrolyzes ATP independently of RNAP and stimulates transcription in an ATP-dependent manner. This differs from RapA from *E. coli* where the ATPase activity of RapA without RNAP was minimal (36).On supercoiled templates, this effect can be enhanced by an additional subunit called δ and the two proteins promote more efficient cycling of RNAP. Furthermore, HelD appears to increase the transcription rate by stimulating

elongation. Finally, we demonstrated that the absence of HelD prolongs the lag phase of growth in a similar manner as the absence of δ does. This phenotype negatively affects the ability of the cell to rapidly react to nutritional changes in the environment.

Far western blotting experiments indicate that HelD binds to RNAP core on the downstream side, in close proximity to the secondary channel (Figure 3F). Interestingly, RapA, a putative helicase that stimulates transcription in an ATP-dependent manner in *E. coli* was found to cross-link to β' and α subunits on *E. coli* RNAP (37). These results are at low resolution but they are consistent with the possibility that HelD and RapA may bind to the same region of their respective RNAPs. Future studies will have to address the exact mode of binding of these two proteins.

RapA is a putative helicase, and, similarly to HelD, no helicase activity was detected for this protein (38). Further, both HelD from *B. subtilis* and RapA of *Pseudomonas putida* (39) bind to core RNAP although RapA of *E. coli* was shown to bind to the holoenzyme containing the main sigma factor (37,39). HelD and RapA, however, belong to different protein families with no significant amino acid sequence homology. Closer sequence analysis suggests that even if both proteins use ATP-binding domains, the structures of these domains and also of the whole proteins are likely entirely unrelated. The two proteins further differ in their effects during UV-induced DNA damage when in *B. subtilis* the lack of HelD was reported to have a negative effect (10), whereas the lack of RapA had no effect in *E. coli* (37). Nevertheless, both proteins stimulate transcription in an ATP-dependent manner, and HelD can be further assisted in this function by a small subunit of RNAP, δ, which is absent from gram-negative bacteria. Interestingly, both δ and HelD are conserved in the industrially and medically important Firmicutes. The stimulatory effect of RapA was shown to be caused by enhancing of the transcriptional cycling by releasing RNAP from the DNA of posttermination complexes (30,34), and possibly also by assisting with reading through RNA–DNA hybrids (34). HelD and δ enhance the release of stalled RNAP from DNA and this, at least partially, contributes to the more efficient cycling of transcription. The character of the RNAP–DNA complexes is not apparent but it seems that it is not a posttermination complex, as the presence or absence of the terminator had no effect (Figure 6B and C). Finally, it appears that HelD may also contribute to more efficient elongation. Interestingly, the level of expression of HelD inversely correlates with the level of expression of topoisomerases TopA and TopB (9) that are known to affect transcription elongation (9,40). The elongation effect is unlikely to be due to a more efficient reading through of RNA–DNA hybrids, such as R-loops, as experiments with short transcribed regions seem to argue against this possibility. Further studies will be required to address the mechanistic details of the interplay between HelD, δ, RNAP and nucleic acids.

The phenotype of the HelD-null strain is not dramatic but it is likely that the prolonged lag phase would adversely affect the competitive fitness of the cell in nature. Further, it is possible that some other proteins, such as the PcrA helicase (7,41), may have overlapping functions with HelD and future studies will be required to decipher the role of these RNAP-associated helicases or helicase-like proteins in transcription. In conclusion, HelD is a novel type of RNAP-associated protein with an important role in transcription.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

The authors thank the National BioResource Project (NIG, Japan) for the *B. subtilis* HelD knockout strain MGNA-A456.

FUNDING

Czech Science Foundation [P302/11/0855, P305/12/G034, 13-16842S]; Science Foundation of Charles University [GAUK 425311]; Ministry of Education, Youth and Sports of the Czech Republic [EE2.3.30.0029]; European Regional Development Fund [BIOCEV CZ.1.05/1.1.00/02.0109]; National Health and Medical Research Council, Australia [APP1021479 to P.L.]; Operational Program Prague—Competitiveness [Project No.: CZ.2.16/3.1.00/24023 toward 'Prague Infrastructure for Structure Biology and Metabolomics']. Funding for open access charges: Czech Science Foundation [P302/11/0855, P305/12/G034, 13-16842S].

Conflict of interest statement. None declared.

REFERENCES

1. Osterberg,S., del Peso-Santos,T. and Shingler,V. (2011) Regulation of alternative sigma factor use. *Annu. Rev. Microbiol.*, **65**, 37–55.
2. Juang,Y.L. and Helmann,J.D. (1994) The delta subunit of *Bacillus subtilis* RNA polymerase. An allosteric effector of the initiation and core-recycling phases of transcription. *J. Mol. Biol.*, **239**, 1–14.
3. Lopez de Saro,F.J., Woody,A.Y. and Helmann,J.D. (1995) Structural analysis of the *Bacillus subtilis* delta factor: a protein polyanion which displaces RNA from RNA polymerase. *J. Mol. Biol.*, **252**, 189–202.
4. Epshtein,V., Dutta,D., Wade,J. and Nudler,E. (2010) An allosteric mechanism of Rho-dependent transcription termination. *Nature*, **463**, 245–249.
5. Toulme,F., Mosrin-Huaman,C., Sparkowski,J., Das,A., Leng,M. and Rahmouni,A.R. (2000) GreA and GreB proteins revive backtracked RNA polymerase *in vivo* by promoting transcript trimming. *EMBO J.*, **19**, 6853–6859.
6. Yang,X., Molimau,S., Doherty,G.P., Johnston,E.B., Marles-Wright,J., Rothnagel,R., Hankamer,B., Lewis,R.J. and Lewis,P.J. (2009) The structure of bacterial RNA polymerase in complex with the essential transcription elongation factor NusA. *EMBO Rep.*, **10**, 997–1002.
7. Delumeau,O., Lecointe,F., Muntel,J., Guillot,A., Guedon,E., Monnet,V., Hecker,M., Becher,D., Polard,P. and Noirot,P. (2011) The dynamic protein partnership of RNA polymerase in *Bacillus subtilis*. *Proteomics*, **11**, 2992–3001.

8. Yang,W. (2010) Lessons learned from UvrD helicase: mechanism for directional movement. *Annu. Rev. Biophys.*, **39**, 367–385.
9. Nicolas,P., Mader,U., Dervyn,E., Rochat,T., Leduc,A., Pigeonneau,N., Bidnenko,E., Marchadier,E., Hoebeke,M., Aymerich,S. *et al.* (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science*, **335**, 1103–1106.
10. Carrasco,B., Fernandez,S., Petit,M.A. and Alonso,J.C. (2001) Genetic recombination in *Bacillus subtilis* 168: effect of DeltahelD on DNA repair and homologous recombination. *J. Bacteriol.*, **183**, 5772–5777.
11. Hanahan,D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.*, **166**, 557–580.
12. Dubnau,D. and Davidoff-Abelson,R. (1971) Fate of transforming DNA following uptake by competent Bacillus. subtilis. I. Formation and properties of the donor-recipient complex. *J. Mol. Biol.*, **56**, 209–221.
13. Qi,Y. and Hulett,F.M. (1998) PhoP-P and RNA polymerase sigmaA holoenzyme are sufficient for transcription of Pho regulon promoters in *Bacillus subtilis*: PhoP-P activator sites within the coding region stimulate transcription *in vitro*. *Mol. Microbiol.*, **28**, 1187–1197.
14. Kobayashi,K., Ehrlich,S.D., Albertini,A., Amati,G., Andersen,K.K., Arnaud,M., Asai,K., Ashikaga,S., Aymerich,S., Bessieres,P. *et al.* (2003) Essential *Bacillus subtilis* genes. *Proc. Natl Acad. Sci. USA*, **100**, 4678–4683.
15. Rabatinova,A., Sanderova,H., Matejckova,J.J., Korelusova,J., Sojka,L., Barvik,I., Papouskova,V., Sklenar,V., Zidek,L. and Krasny,L. (2013) The delta subunit of RNA polymerase is required for rapid changes in gene expression and competitive fitness of the cell. *J. Bacteriol.*, **195**, 2603–2611.
16. Chang,B.Y. and Doi,R.H. (1990) Overproduction, purification, and characterization of *Bacillus subtilis* RNA polymerase sigma A factor. *J. Bacteriol.*, **172**, 3257–3263.
17. Ross,W., Thompson,J.F., Newlands,J.T. and Gourse,R.L. (1990) *E. coli* Fis protein activates ribosomal RNA transcription *in vitro* and *in vivo*. *EMBO J.*, **9**, 3733–3742.
18. Krasny,L. and Gourse,R.L. (2004) An alternative strategy for bacterial ribosome synthesis: *Bacillus subtilis* rRNA transcription regulation. *EMBO J.*, **23**, 4473–4483.
19. Johnston,E.B., Lewis,P.J. and Griffith,R. (2009) The interaction of *Bacillus subtilis* sigmaA with RNA polymerase. *Protein Sci.*, **18**, 2287–2297.
20. Anborgh,P.H., Parmeggiani,A. and Jonak,J. (1992) Site-directed mutagenesis of elongation factor Tu. The functional and structural role of residue Cys81. *Eur. J. Biochem.*, **208**, 251–257.
21. Gaal,T., Ross,W., Estrem,S.T., Nguyen,L.H., Burgess,R.R. and Gourse,R.L. (2001) Promoter recognition and discrimination by EsigmaS RNA polymerase. *Mol. Microbiol.*, **42**, 939–954.
22. Barker,M.M., Gaal,T. and Gourse,R.L. (2001) Mechanism of regulation of transcription initiation by ppGpp. II. Models for positive control based on properties of RNAP mutants and competition for RNAP. *J. Mol. Biol.*, **305**, 689–702.
23. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2008) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
24. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
25. Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
26. Anthony,L.C., Artsimovitch,I., Svetlov,V., Landick,R. and Burgess,R.R. (2000) Rapid purification of His(6)-tagged *Bacillus subtilis* core RNA polymerase. *Protein Expr. Purif.*, **19**, 350–354.
27. Lee,J.Y. and Yang,W. (2006) UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. *Cell*, **127**, 1349–1360.
28. Cao,Z. and Julin,D.A. (2009) Characterization *in vitro* and *in vivo* of the DNA helicase encoded by *Deinococcus radiodurans* locus DR1572. *DNA Repair (Amst)*, **8**, 612–619.
29. Vassylyev,D.G., Sekine,S., Laptenko,O., Lee,J., Vassylyeva,M.N., Borukhov,S. and Yokoyama,S. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature*, **417**, 712–719.
30. Sukhodolets,M.V., Cabrera,J.E., Zhi,H. and Jin,D.J. (2001) RapA, a bacterial homolog of SWI2/SNF2, stimulates RNA polymerase recycling in transcription. *Genes Dev.*, **15**, 3330–3341.
31. Epshtein,V., Toulme,F., Rahmouni,A.R., Borukhov,S. and Nudler,E. (2003) Transcription through the roadblocks: the role of RNA polymerase cooperation. *EMBO J.*, **22**, 4719–4727.
32. Holmberg,C. and Rutberg,B. (1991) Expression of the gene encoding glycerol-3-phosphate dehydrogenase (glpD) in *Bacillus subtilis* is controlled by antitermination. *Mol. Microbiol.*, **5**, 2891–2900.
33. Gowrishankar,J., Leela,J.K. and Anupama,K. (2013) R-loops in bacterial transcription: their causes and consequences. *Transcription*, **4**, 153–157.
34. McKinley,B.A. and Sukhodolets,M.V. (2007) Escherichia coli RNA polymerase-associated SWI/SNF protein RapA: evidence for RNA-directed binding and remodeling activity. *Nucleic Acids Res.*, **35**, 7044–7060.
35. Vassylyev,D.G., Vassylyeva,M.N., Zhang,J., Palangat,M., Artsimovitch,I. and Landick,R. (2007) Structural basis for substrate loading in bacterial RNA polymerase. *Nature*, **448**, 163–168.
36. Yawn,B., Zhang,L., Mura,C. and Sukhodolets,M.V. (2009) RapA, the SWI/SNF subunit of *Escherichia coli* RNA polymerase, promotes the release of nascent RNA from transcription complexes. *Biochemistry*, **48**, 7794–7806.
37. Sukhodolets,M.V. and Jin,D.J. (2000) Interaction between RNA polymerase and RapA, a bacterial homolog of the SWI/SNF protein family. *J. Biol. Chem.*, **275**, 22090–22097.
38. Sukhodolets,M.V. and Jin,D.J. (1998) RapA, a novel RNA polymerase-associated protein, is a bacterial homolog of SWI2/SNF2. *J. Biol. Chem.*, **273**, 7018–7023.
39. Stec-Dziedzic,E., Lyzen,R., Skarfstad,E., Shingler,V. and Szalewska-Palasz,A. (2013) Characterization of the transcriptional stimulatory properties of the Pseudomonas putida RapA protein. *Biochim. Biophys. Acta*, **1829**, 219–230.
40. Baranello,L., Kouzine,F. and Levens,D. (2013) DNA Topoisomerases: beyond the standard role. *Transcription*, **4** doi: 10.4161/trns.26598.
41. Gwynn,E.J., Smith,A.J., Guy,C.P., Savery,N.J., McGlynn,P. and Dillingham,M.S. (2013) The conserved C-terminus of the PcrA/UvrD helicase interacts directly with RNA polymerase. *PLoS One*, **8**, e78141.

DR. LIBOR KRÁSNÝ (Orcid ID : 0000-0002-4120-4347)

Article type : Research Article

Title: Ms1 RNA increases the amount of RNA polymerase in *Mycobacterium smegmatis*

Running Title: Function of Ms1 RNA in mycobacteria

Michaela Šiková[1], Martina Janoušková[1,2], Olga Ramaniuk[1], Petra Páleníková[1], Jiří Pospíšil[1], Pavel Bartl[3], Agnieszka Suder[1], Petr Pajer[4], Pavla Kubičková[4], Ota Pavliš[4], Miluše Hradilová[5], Dragana Vítovská[1], Hana Šanderová[1], Martin Převorovský[6], Jarmila Hnilicová[1] and Libor Krásný[1]

[1] Department of Microbial Genetics and Gene Expression, Institute of Microbiology, Czech Academy of Sciences, Prague, Czech Republic

[2] Department of Genetics and Microbiology, Faculty of Science, Charles University, Prague, Czech Republic

[3] Department of Nuclear Chemistry, Faculty of Nuclear Science and Physical Engineering, Czech Technical University in Prague, Prague, Czech Republic

[4] Military Health Institute, Military Medical Agency, Prague, Czech Republic

[5] Department of Genomics and Bioinformatics, Institute of Molecular Genetics, Czech Academy of Sciences, Prague, Czech Republic

[6] Department of Cell Biology, Faculty of Science, Charles University, Prague, Czech Republic

Michaela Šiková and Martina Janoušková should be considered joint first author

Jarmila Hnilicová and Libor Krásný should be considered joint senior author and corresponding author

Contact Information: E-mail: krasny@biomed.cz (LK) and hnilicova@biomed.cas.cz (JH), Tel. +420 241 063 208

Keywords: Ms1, transcriptome, mycobacteria, sRNA, RNA polymerase, RNA-seq, PNPase

Acknowledgements: We thank Dr. Isabella Moll (Max F. Perutz Laboratories, Austria) for provision of antibody. We thank Brian Grogan for English language editing. This work used instruments provided by C4Sys infrastructure. This work was supported by grants from the Czech Science Foundation [13-27150P to J.H., P305/12/G034 to L.K.], Charles University [PRIMUS/MED/26 and UNCE 204013 to M.P.], the Grant Agency of Charles University [946216 to M.J.], Ministry of the Interior of the Czech Republic [VI20152020044], and Ministry of Education, Youth and Sports of the Czech Republic within project the Center for advanced applied science, project number CZ.02.1.01/0.0/0.0/16_019/0000778. The authors have no conflict of interest to declare.

Ms1 RNA increases the amount of RNA polymerase in *Mycobacterium smegmatis*

Summary

Ms1 is a sRNA recently found in mycobacteria and several other actinobacterial species. Ms1 interacts with the RNA polymerase (RNAP) core devoid of sigma factors, which differs from 6S RNA that binds to RNAP holoenzymes containing the primary sigma factor. Here we show that Ms1 is the most abundant non-rRNA transcript in stationary phase in *Mycobacterium smegmatis*. The accumulation of Ms1 stems from its high-level synthesis combined with decreased degradation. We identify the Ms1 promoter, P_{Ms1}, and cis-acting elements important for its activity. Furthermore, we demonstrate that PNPase (an RNase) contributes to the differential

accumulation of Ms1 during growth. Then, by comparing the transcriptomes of *wt* and Δ*Ms1* strains from stationary phase, we reveal that Ms1 affects the intracellular level of RNAP. The absence of Ms1 results in decreased levels of the mRNAs encoding β and β´ subunits of RNAP, which is also reflected at the protein level. Thus, the Δ*Ms1* strain has a smaller pool of RNAPs available when the transcriptional demand increases. This contributes to the inability of the Δ*Ms1* strain to rapidly react to environmental changes during outgrowth from stationary phase.

Keywords

Ms1, transcriptome, mycobacteria, sRNA, RNA polymerase, RNA-seq, PNPase

Introduction

Bacterial growth, adaptation, and survival depend on accurate responses to environmental changes. These responses result from changes in gene expression. A key enzyme responsible for this process is RNA polymerase (RNAP), which transcribes the genetic information from DNA into RNA.

The bacterial RNAP core (E) consists of two α subunits, two large catalytic subunits, β and β', and the small ω subunit. The RNAP core is capable of transcription elongation but cannot initiate transcription. The RNAP core binds a σ factor (this complex is called RNAP holoenzyme, Eσ). The σ factor, then, helps RNAP recognize the specific promoter sequence, open the transcription bubble, and synthesize the first few nucleotides of the transcript (Browning & Busby, 2016). The promoter specificity of RNAP thus depends on the associated σ factor. Bacteria contain one primary σ factor (such as σ^{70} in *Escherichia coli* or σ^{A} in *Mycobacterium smegmatis*), which is responsible for the majority of transcription during vegetative growth, and various numbers of alternative σ factors that recognize specific promoters necessary to cope with stress (Feklístov *et al.*, 2014). In mycobacteria, two essential regulators of transcription initiation, RbpA and CarD, interact with Eσ and stabilize its association with DNA (Rammohan *et al.*, 2016, Hu *et al.*, 2014,

Hubin *et al.*, 2015, Verma & Chatterji, 2014, Rammohan *et al.*, 2015, Srivastava *et al.*, 2013, Davis *et al.*, 2015).

Previously we identified a ~300 nt long small RNA (sRNA) in *M. smegmatis* termed Ms1 (Pánek *et al.*, 2011) that interacts with the RNAP core devoid of σ factors (Hnilicová *et al.*, 2014). An Ms1 homolog was also found in *M. tuberculosis* (a sRNA named MTS2823) (Arnvig *et al.*, 2011) and predicted in other mycobacterial and actinobacterial species such as *Nocardia* or *Rhodococcus* (Hnilicová *et al.*, 2014). Both Ms1 and MTS2823 accumulate during stationary phase of growth (Arnvig *et al.*, 2011, Hnilicová *et al.*, 2014).

Ms1 shares a similar predicted secondary structure with 6S RNA – the double stranded hairpin interrupted with the central bubble. In addition, Ms1 has two short hairpins at the 5´and 3´ends (Hnilicova *et al.*, 2014). The main difference between Ms1 and 6S RNA is that 6S RNA associates with RNAP holoenzyme - the complex of RNAP with the primary σ factor (Eσ) (Cavanagh & Wassarman, 2014, Trotochaud & Wassarman, 2005, Wassarman & Storz, 2000), whereas Ms1 binds the RNAP core devoid of σ factors (Hnilicová *et al.*, 2014). 6S RNA regulates expression of several hundred genes in *E. coli* (Cavanagh *et al.*, 2008, Neusser *et al.*, 2010) by binding to $E\sigma^{70}$. Previously, artificial MTS2823 overexpression in exponential phase in *M. tuberculosis* downregulated expression of 301 genes and upregulated 2 genes (Arnvig *et al.*, 2011). However, it is unclear how expression of Ms1 is regulated from its natural locus, and importantly, what the role(s) of Ms1 are in stationary phase when it is most abundant and physiologically relevant.

Here we identify the Ms1 promoter in *M. smegmatis* and *M. tuberculosis* and characterize Ms1 expression and degradation in different growth phases. Furthermore, with a *ΔMs1* strain we demonstrate the effect of Ms1 on bacterial survival and growth. Then, we compare the transcriptomes of the *ΔMs1* strain versus wild type (*wt*) both in exponential and stationary phase and we identify genes regulated in Ms1-dependent manner. We show that Ms1 affects gene expression of RNAP subunits β and β´. Based on Ms1‘s effect on mycobacterial transcription we discuss hypotheses of how Ms1 functions mechanistically.

Results

Ms1 is the most expressed non-rRNA gene in stationary phase

To start characterizing Ms1, we performed RNA-seq experiments. We compared transcriptomes of *wt M. smegmatis* in exponential and stationary phases to quantify Ms1 expression during mycobacterial growth.

We isolated RNA from mid-log exponential phase (OD_{600} 0.5) and from stationary phase after 24 hours of cultivation (OD_{600} 2.5, approximately 12 hours after the entry into stationary phase). In *wt M. smegmatis*, the most upregulated genes after the entry into the stationary phase were Ms1 and genes from the *MSMEG_1971-MSMEG_1979* cluster. Importantly, over 30 % of non-rRNA transcripts in stationary phase belonged to Ms1. Thus, Ms1 is the most abundant transcript (disregarding the ribosomal rRNAs) in stationary phase (Supporting Table 1). In exponential phase, the Ms1 amount decreased to ~0.1% of non-rRNA transcripts. Nevertheless, Ms1 was still in the Top-20 most expressed exponential phase genes (Supporting Table 1).

Based on RNA-seq experiment (where we used the same amount of RNA per sample), Ms1 expression increased >350-fold in stationary phase RNA sample compared to exponential phase RNA. However, to conclude that Ms1 expression increased >350-fold per cell, we have to assume that the overall yields of the RNA sample is identical per cell under different experimental conditions (Chen *et al.*, 2015). Therefore, we verified the Ms1 increase in stationary phase by quantitative PCR (qPCR) (Fig 1A). Immediately at the beginning of RNA isolation from a known number of cells, we added *in vitro* transcribed *B. subtilis* 6S RNA as the RNA recovery marker as control for the efficiency of RNA isolation or potential RNA degradation. Then, the amount of Ms1 was normalized to the cell number and RNA recovery marker. This approach yielded a ~115-fold increase in the Ms1 level in stationary phase. This suggests that to obtain the same amount of RNA from stationary phase and exponential phase, it is necessary to use 3-times as many stationary phase cells. In other words, approximately 3-fold less RNA is isolated from the same number of stationary phase *M. smegmatis* cells compared to exponential phase. Contrary to Ms1, the amount of the primary σ factor (σ^A) mRNA per cell

increased in exponential phase ~4.5-fold compared to stationary phase (Fig 1B). Also, the expression of the three main RNases, PNPase (*MSMEG_2656*), RNase J (*MSMEG_4626*) and RNase E (*MSMEG_2685*), per cell was higher in exponential phase compared to stationary phase (Fig 1B).

Ms1 is transcribed from a single promoter

Next, we wished to determine the mechanisms of the high intracellular accumulation of Ms1. As the accumulation is a combined effect of synthesis and degradation, we investigated these two processes, focusing first on synthesis. Previously we identified the first nucleotide of Ms1 and the putative Ms1 promoter (P_{Ms1}) in *M. smegmatis* (Hnilicová *et al.*, 2014). In *M. tuberculosis*, the 5´end of MTS2823 (Ms1 homolog) was mapped 15 nucleotides upstream compared to *M. smegmatis* (Arnvig *et al.*, 2011). The region upstream of the reported MTS2823 transcription start site contains no obvious match for a promoter consensus sequence; correspondingly, no promoter sequence for the Ms1 homolog in *M. tuberculosis* was reported (Arnvig *et al.*, 2011). Nevertheless, a sequence identical with the *M. smegmatis* putative Ms1 promoter is present in the same location also in *M. tuberculosis*.

Therefore we first tested in *M. smegmatis* (i) whether the predicted Ms1 core promoter (Hnilicová *et al.*, 2014) is functional; (ii) whether Ms1 has one or several promoters; (iii) whether any repressor/activator binds to the sequence upstream of the Ms1 promoter(s), and (iv) whether *M. smegmatis* and *M. tuberculosis* Ms1 transcripts have the same 5' ends, and therefore might be similarly regulated by their upstream sequences.

We cloned several *M. smegmatis* promoter region fragments (-38/+9 - the putative core promoter; -131/+9; -231/+9; -331/+9; and -491/+9, see Fig 2A) into the integrative pSM128 vector (Dussurget *et al.*, 1999) encoding the *lacZ* reporter gene and measured β-galactosidase activity in *M. smegmatis* in different growth phases (Fig 2B). The -38/+9 region supported significantly above-background β-galactosidase activities (Fig 2B), confirming that this region contained a core promoter sequence. The relative activities of other promoter fragments were then normalized to the activity of the -38/+9 core promoter in exponential phase, which

was set as 100 (Fig. 2B). By the β-galactosidase measurements, the activity of the P_{Ms1} core promoter was highest in late stationary phase (after 48 hours of growth). When we added 100 bp (of the native sequence) upstream to the core promoter (region -131/+9), promoter activity increased both in exponential and stationary phases ~10-fold. A further 100 bp extension of the promoter fragment (-231/+9) led to an additional ~8-fold increase promoter activity in both growth phases, indicating the presence of either general transcriptional activator(s) or promoter(s) similarly active in exponential and stationary phases. Yet another two sequential extensions by 100 bp (-331/+9 and -491/+9) had only a minor effect (max. ~1.5-fold transcription increase) that was not growth phase specific.

To distinguish whether the upstream region of the core promoter contained any other promoter(s) or whether it contained binding site(s) for transcription factors, we deleted a part of the core promoter sequence (fragment -491/-22) or mutated the -10 region of the core promoter (-491/+9mut) (Figs 2A and 2C). Both fragments (-491/-22, -491/+9mut) were unable to initiate transcription of *lacZ* (Fig 2B). Thus, no other promoter is located within the -491/-22 region. From this result, we can conclude that Ms1 is transcribed from the single core promoter and this promoter is regulated by additional transcription factors. In addition, the Ms1 promoter (P_{Ms1}) is very strong and its activity in stationary phase is comparable to the ribosomal *rrnB* promoter (-72/+20 fragment) (China *et al.*, 2010, Tare *et al.*, 2012) when it is most active in exponential phase (Fig 2B). We note that the background activity of the promoter-less pSM128 vector was relatively high (but significantly lower than the activity of the core promoter). In the Ms1 core promoter mutants, the cloning of DNA fragments completely devoid of promoter sequences likely disrupted the background-generating sequence around the cloning site.

Nevertheless, the β-galactosidase activity of the control *rrnB* promoter (-72/+20 fragment) remained high both in early and late stationary phases (Fig. 2B) which contradicts the well established knowledge from other bacterial species that expression of ribosomal RNAs decreases upon starvation and on entry into stationary phase (Krásný & Gourse, 2004, Paul *et al.*, 2004, Aviv *et al.*, 1996, Murray *et al.*, 2003). The high activity of the *rrnB* promoter in stationary phase measured by the β-galactosidase assay could have been caused by the increased stability of the

β-galactosidase (LacZ) enzyme during mycobacterial growth. Therefore, we measured by qPCR directly the amount of the relatively unstable *lacZ* mRNA that better reflects promoter activity changes (Krásný *et al.*, 2008). To do so, we isolated RNA from the known number of cells and we added *in vitro* transcribed *B. subtilis* 6S RNA as RNA recovery marker as control for the efficiency of RNA isolation or potential RNA degradation. Then, the amount of the *lacZ* mRNA was normalized to the cell number and RNA recovery marker. The activity of the *rrnB* promoter decreased ~60-fold in early stationary and ~300-fold in late stationary phase compared to exponential phase (Fig. 2D), consistent with the situation in other species mentioned above. This approach revealed that the activity of the P_{Ms1} largest promoter fragment (-491/+9) did not vary much throughout the growth cycle: it increased ~2-fold in early stationary phase and in late stationary phase returned to the exponential phase level (Fig. 2D), indicating almost constitutive expression. A comparison of P_{Ms1} with the *rrnB* promoter then revealed that P_{Ms1} is a strong promoter albeit somewhat less active than P_{rrnB} in exponential phase.

Further comparisons of the lacZ mRNA-based and β-galactosidase data revealed that the addition of upstream promoter elements (-491/+9) increased the activity of the Ms1 -38/+9 core promoter about 70-fold based on the *lacZ* mRNA level compared to the ~77-fold increase based on the LacZ protein level (Fig. 2B and 2D, exponential phase), indicating that both measurements were comparable, except for the LacZ protein accumulation during the mycobacterial growth.

Finally, to resolve the discrepancy between the transcription start sites (TSS) of Ms1 and its *M. tuberculosis* homolog MTS2823, we mapped the *M. tuberculosis* TSS. We treated total RNA with Terminator 5'-Phosphate-Dependent Exonuclease (TEX) to degrade monoposphorylated 5´ends and followed this by 5´RACE. We obtained one dominant MTS2823-specific PCR product (Fig 2F) that corresponded to the same transcription start site as Ms1 in *M. smegmatis* (Fig 2C). Thus, both Ms1 and MTS2823 have the same transcription start site and the promoter sequences we found in *M. smegmatis* are conserved also in *M. tuberculosis* (Fig 2E).

Ms1 transcription increases during ethanol stress and starvation

To determine what types of stress induce transcription of Ms1 in addition to the harsh nutritional conditions in stationary phase, we quantified the activity of the -491/+9-*lacZ* promoter fragment under various stress conditions: heat-stress (50°C), cold-stress (RT - room temperature), ethanol (5% v/v), oxidative stress (7.5 mM H_2O_2), hyperosmotic stress (0.5 M NaCl), hypoosmotic stress (H_2O), acidic stress (pH 4), alkaline stress (pH 9), and starvation (PBS). The cells were cultivated to exponential phase (OD_{600} 0.5, ~6 hrs of growth), stressed for 2 and 4 hours, and then β-galactosidase activity was measured (Fig 2G). The β-galactosidase activities of the -491/+9-*lacZ* construct under stress conditions were normalized to the control sample at time 0 (before stresses, Fig 2G) which was set as 1.

Ethanol stress induced *lacZ* transcription almost ~7-fold. The Ms1 promoter was also activated during growth in PBS (~1.6-fold), which simulates starvation. Thus, Ms1 transcription is enhanced especially under specific stress conditions (ethanol, starvation) and not generally by stress. Except for ethanol, no other stress induced Ms1 in a similar manner as the entry into stationary phase of growth (Figs 2B and 2G) and under no stress condition the P_{Ms1} activity was as high as in stationary phase (Fig 2G).

Ms1 is more stable in stationary phase

The Ms1 level increases ~115-fold upon entry into stationary phase (Fig 1A). However, the longest Ms1 promoter fragment (-491/+9) increased its activity only ~2-fold under the same conditions when we measured the mRNA level of *lacZ* gene (Fig 2D). Thus, the number of Ms1 molecules must be regulated also by another mechanism besides transcription. Therefore, we compared the stability of Ms1 in exponential and stationary phases. To efficiently detect Ms1 in exponential phase, we used a strain bearing a replicative plasmid, pJAM2-*rrnB-Ms1*, where Ms1 was highly expressed from the strong *rrnB* promoter in exponential phase (Hnilicová *et al.*, 2014). We added rifampicin to stop *de novo* transcription and isolated total RNA from several time points after the addition of rifampicin. RNA was then isolated from each time point and RNA recovery marker was added (in this case we used

Ms1notail – a shortened version of Ms1, see Experimental Procedures section) and the amount of Ms1 was determined by northern blotting (Figs 3A and 3B) and normalized to the RNA recovery marker (Fig 3C). The Ms1 half-life was 8 minutes in exponential phase and ~8 hours in stationary phase (Fig 3C), revealing that the increased stability in stationary phase contributes to Ms1 accumulation.

To test whether rifampicin affects the cell numbers and/or viability which could have potentially influenced RNA stability, we calculated CFU in rifampicin-treated cells and in control cells that were cultivated without the addition of rifampicin. In exponential phase, rifampicin treatment for 30 min had no effect on CFU: cells treated with rifampicin had CFU [1.7±1.4]*10^7, control cells [3.3±2.5]*10^7. This confirms that rapid degradation of Ms1 in exponential phase is not a secondary effect caused by bacterial death after rifampicin treatment. The 6 hours-long treatment with rifampicin in stationary phase decreased CFU ~3-fold: control cells without rifampicin had [1.7±1]*10^8 and rifampicin-treated cells had [6.3±3.8]*10^7 CFU. Thus, the observed Ms1 decay in stationary phase was probably partly affected by the decreased viability of cells during rifampicin treatment and the Ms1 half-life in stationary phase might be even somewhat longer than 8 hours (Fig 3C).

PNPase contributes to degradation of Ms1

Next, we investigated which RNase might play a role in Ms1 degradation. Previously, we identified PNPase (*MSMEG_2656*, 3´- 5´phosphorolytic activity) as a binding partner of Ms1 (Hnilicová *et al.*, 2014). PNPase expression increased ~20-fold in stationary phase compared to exponential phase (Fig 1B) indicating that this RNase might specifically degrade Ms1 in exponential phase. However, the expression of RNase E and RNase J is also elevated in exponential phase (~28-fold and ~8-fold, respectively, Fig 1B). To test the direct role of PNPase, RNase J and RNase E in Ms1 degradation *in vivo*, we prepared strains in which RNase expression can be reduced by CRISPR (Rock *et al.*, 2017) (strains *LK2197, LK2198, LK2219* and control strain *LK2261*, the control strain contains non-targeting sgRNA). The depletion of particular RNase was induced at the onset of exponential phase (OD_{600} ~0.2) and, after 3 hours of growth, cells were harvested at mid-log exponential phase (OD_{600} ~0.5). We isolated total RNA from each strain from the same amount of cells

and immediately at the beginning of isolation we added RNA recovery marker (6S RNA from *B .subtilis*) to control for the efficiency of RNA isolation. We quantified the efficiency of CRISPR knockdowns by qPCR, the mRNA levels of individual RNases were normalized to the RNA recovery marker. CRISPR efficiently depleted each RNase (the expression decreased >90 % in individual strains, Fig 3D). Then, we determined Ms1 amount by northern blotting (Figs 3E and 3F). RNase E and RNase J depletion did not elevate Ms1 level (Fig 3E). The depletion of PNPase in exponential phase increased the level of Ms1, confirming that PNPase contributes to the Ms1 degradation *in vivo*. However, the Ms1 increase was moderate, about 30 % (Fig 3F) – PNPase depletion did not completely stop Ms1 degradation (the Ms1 increase in stationary phase is ~115-fold, therefore we expected a higher increase in the Ms1 level in exponential phase after almost complete inhibition of Ms1 degradation). A possibility is that another exponential phase-dependent cofactor (e.g. RNA helicase) contributes to the degradation of the highly structured Ms1 RNA in exponential phase.

Therefore, we tested whether (i) PNPase alone is able to degrade Ms1 RNA *in vitro* and (ii) if Ms1 has any structural features that would block its degradation by PNPase. We overexpressed and purified His-tagged *M. smegmatis* PNPase (Fig 3G). We performed digestion assays *in vitro* with Ms1 and two Ms1 mutant variants: (i) "Ms1nb", lacking the central bubble (Hnilicová *et al.*, 2014), and (ii) "Ms1notail", lacking the tails at 5´and 3´ends (Fig 3H), and part of the RpoC mRNA as a control. PNPase significantly (but not completely) degraded both the unmodified Ms1 and Ms1nb (Fig 3H), indicating that this Ms1-interacting RNase might be responsible, at least in part, for Ms1 degradation and that the central bubble is not required for the degradation. To the contrary, the Ms1notail mutant was degraded only modestly by PNPase, indicating that the tail motif enhances Ms1 degradation. Control RNase A did not show any substrate specificity with respect to Ms1 variants.

Thus, PNPase degrades Ms1 both *in vivo* and *in vitro* and the degradation efficiency depends on the Ms1 sequence/secondary structure. However, another RNase or RNA helicase releasing Ms1 secondary structure is probably necessary for complete Ms1 degradation inside the cell.

ΔMs1 cells are viable

To determine whether Ms1 is essential for the cell, we deleted 255 nucleotides from the 304 nt-long *Ms1* gene (the +3 to +257 region was replaced with a hygromycin resistance cassette, see Fig 4A). Surprisingly, Δ*Ms1* cells reproducibly reached slightly higher OD_{600} at the end of exponential phase (Fig 4B). We did not detect any morphological changes in Δ*Ms1* cells by electron microscopy neither in exponential nor in stationary phase (data not shown).

Next, we examined the Δ*Ms1* growth in several stress conditions: ethanol stress (5% EtOH), high salt conditions (0.5M NaCl), oxidative stress (diamide) or acidic pH (pH 5). Mostly, no significant phenotypic effects (compared to *wt*) were detected (data not shown). However, cell survival after γ irradiation was significantly affected by the presence of Ms1. Δ*Ms1* stationary phase cells chronically irradiated with the dose of 150 Gy formed about 50% less colony forming units (CFU) than *wt*, while irradiation with 250 Gy was lethal for almost ~80% of Δ*Ms1* cells compared to *wt* (Fig 4C). Half of samples were chronically irradiated at room temperature (Fig 4C and 4E) and the second half was kept at room temperature without irradiation as a control of cell viability (Figs 4D and 4F, labelled as “mock controls”). No difference was observed between Δ*Ms1* and *wt* cells in exponential phase (Fig 4E) when Ms1 is present in low amounts or in non-irradiated controls (Figs 4D and 4F). Ms1 thus helped protect stationary phase cells from radioactive exposure.

Ms1 regulates the amount of RNA polymerase

Next, we compared transcriptomes of *wt* and Δ*Ms1* strains in exponential and stationary phases. Despite the fact that Ms1 is so abundant in *M. smegmatis*, RNA-seq revealed only minor changes in gene expression upon deletion of Ms1. In exponential phase of growth, we detected five statistically significant differentially expressed genes in Δ*Ms1* cells (FDR corrected P-value 0.05, see Supporting Table 2). In stationary phase of growth, fourteen genes altered their expression in the Δ*Ms1* strain (FDR corrected P-value 0.05, see Supporting Table 2) including *MSMEG_1367* and *MSMEG_1368* (*rpoB* and *rpoC*, respectively) genes that encode

two main RNAP subunits β and β´. In the RNA-seq data, the expression of β and β´ decreased by half in Δ*Ms1* compared to *wt*.

We decided to confirm the effect of Ms1 on the expression of genes identified by RNA-seq in stationary phase. We selected the following genes: *MSMEG_1205*, a cyclopropane-fatty-acyl-phospholipid synthase, playing a role in fatty acid metabolism (2x↑ in ΔMs1); *MSMEG_5540*, encoding an unknown protein (2x↓ in ΔMs1); *MSMEG_5556* encoding MFS permease (2x↓ in ΔMs1); and *rpoB* and *rpoC* genes encoding RNAP subunits β and β´ (both 2x↓ in ΔMs1).

For the validation experiments, we also created a complemented Δ*Ms1+Ms1* strain and compared it with *wt* and Δ*Ms1* strains. Δ*Ms1+Ms1* contained a plasmid (stably integrated into the genome) carrying Ms1 under its native promoter. The amount of Ms1 in stationary phase in the Δ*Ms1+Ms1* strain was comparable to the amount of Ms1 in *wt* (Fig 5A, the gel shows total RNA stained with GelRed). In Δ*Ms1*, no band corresponding to the size of Ms1 was visible and the Ms1 level decreased to ~ zero in RNA-seq data. Subsequently, we extracted RNA from *wt,* Δ*Ms1* and Δ*Ms1+Ms1* stationary phase cells and quantified the mRNA levels of selected genes by qPCR to validate the RNA-seq data. As a reference gene, we used the gene encoding σ^A that did not alter expression in stationary phase in the Δ*Ms1* transcriptome compared to *wt* according to the RNA-seq experiment. The qPCR results correlated with the RNA-seq data, and importantly, the complemented strain displayed a reversal of the levels of the tested genes to near *wt* values (Fig 5B). Therefore, the expression of these genes (including RNAP β and β´ subunits) was indeed dependent on the presence of Ms1. In other words, the changes in the expression were not caused by any kind of mutations in the gene promoters that could be randomly created during Δ*Ms1* strain preparation and selection.

To define if the lowered amount of RNAP is associated with a decreased protein level of the RNAP β subunit, we used quantitative western blotting (Fig 5C). The level of the RNAP β subunit in the Δ*Ms1* strain was at ~70% of its level in the *wt* strain. The addition of the Ms1 copy into the *ΔMs1* strain restored the level of the RNAP β subunit and we observed no difference between *wt* and Δ*Ms1+Ms1* strains (Fig 5C).

Ms1 accelerates outgrowth from stationary phase

As Ms1 increases the total amount of RNAP in stationary phase it suggests that Ms1 (and the higher amount of RNAP) could be beneficial during outgrowth. To test this hypothesis, we cultured Δ*Ms1* and *wt* strains to stationary phase (OD_{600} 2.5, approximately 12 hours after the end of exponential phase) and then diluted both strains into fresh medium and measured $OD_{600.}$ Both strains were diluted to the same OD_{600} (0.1) and their growth was monitored for 6 hours. Immediately before the dilution, we measured CFU of Δ*Ms1* and *wt.* There was no difference in viability and cell number of both strains: *wt* had [2.31±1]*10^8 and Δ*Ms1* [2.37±0.9]*10^8. The *wt* strain grew faster immediately after the dilution and the difference in OD_{600} between *wt* and Δ*Ms1* lasted for ~3 hours after the dilution (Fig 6A). The average doubling time in the first 2 hours after the dilution was 133±5 minutes for the *wt* strain and 146±3 min for the Δ*Ms1* and the difference was significant (P-value 0.006, T-test). There was no statistically significant difference in doubling times 2-6 hours after the dilution. No difference in growth was observed between *wt* and Δ*Ms1+Ms1* complemented strain (Fig 6B). In addition, we prepared the *LK2200* strain, where we could efficiently knockdown Ms1 by CRISPR (Fig 6E). The Ms1 knockdown was induced by anhydrotetrycycline (Rock *et al.*, 2017). We depleted Ms1 in stationary phase (Fig 6E) in *LK2200*, diluted the strain into a fresh medium and measured OD_{600} during the outgrowth (Fig 6C). *LK2200* with depleted Ms1 (with added anhydrotetracycline) grew more slowly than *LK2200* without anhydrotetracycline (no Ms1 knockdown was induced) and the difference in OD_{600} remained for ~4 hours after the dilution (Fig 6C). The average doubling time in the first two hours after the dilution was 162±8 min for the *LK2200* strain with anhydrotetracycline and 190±8 min for the *LK2200* strain without anhydrotetracycline; the difference was statistically significant (P-value 0.008, T-test). The doubling times did not significantly differ in the next four hours (hour 2-6 after the dilution). No difference during outgrowth was observed when we added anhydrotetracycline to *LK2261* control strain with non-targeting sgRNA oligo (Fig 6D). Thus, by an independent method (CRISPR induced Ms1 knockdown) we confirmed that Ms1 is important during the outgrowth.

Next, we monitored the correlation between the RNAP and Ms1 levels between Δ*Ms1* and *wt* strains during outgrowth. The Δ*Ms1* strain had a decreased level of the RNAP β protein subunit compared to *wt* at the time of dilution and the difference was still detectable 15 minutes after the dilution (Figs 6F and 6G). 30 minutes after the dilution however, the amount of RNAP in both strains was comparable. It is possible that at this time Ms1 was already degraded because Ms1 is highly unstable in exponential phase of growth (Fig 3A). Therefore, we quantified the degradation of endogenous Ms1 after dilution of *wt* cells from stationary phase into fresh medium. Ms1 was reduced to 50% within ~18 min after the dilution (Fig 6H), indicating that the degradation rate of endogenous Ms1 was comparable with the effect of Ms1 on the level of RNAP.

Previously, it was shown that the addition of rifampicin during the outgrowth stabilizes 6S RNA (Beckmann *et al.*, 2012, Wurm *et al.*, 2010). 6S RNA serves as a template for transcription of small 'product' RNAs (pRNAs), and active transcription is necessary for the release of 6S RNA from the primary RNAP holoenzyme and subsequent 6S RNA degradation (Wassarman & Saecker, 2006, Gildehaus *et al.*, 2007, Beckmann *et al.*, 2012, Beckmann *et al.*, 2011, Cavanagh *et al.*, 2012). Therefore we tested if inhibition of transcription by rifampicin could stabilize Ms1 during the outgrowth. We cultured *wt* cells to stationary phase (24 hours), diluted them to OD_{600} 0.1 into a fresh medium with or without rifampicin and measured the level of Ms1 by qPCR (Fig 6I). Contrary to 6S RNA, rifampicin had no effect on Ms1 degradation (Fig 6I). This further suggests that Ms1 and 6S RNA differ not only in binding RNAP core and holoenzyme, but both molecules have distinct mechanisms of operation in the bacterial cells.

Taken together, we show (i) that the number of RNAP molecules is positively correlated with the Ms1 presence both in stationary phase and shortly after the dilution into a fresh medium; and (ii) that the presence of Ms1 is advantageous for the mycobacterial cell during outgrowth.

Discussion

In this study, we characterize synthesis and degradation of Ms1 and the effects of Ms1 on gene expression and cell fitness. Ms1 is the most abundant non-rRNA transcript in stationary phase in *M. smegmatis* and its accumulation depends on its strong and almost constitutive promoter and differential RNA degradation. Our data further reveals that deletion of Ms1 affects expression of only several genes during steady state. However, Ms1 influences the RNAP level and likely sequesters RNAP cores in a cache of inactive enzymes that can be reactivated when needed. This appears to be crucial when the demand for RNAP increases during environmental changes such as outgrowth from stationary phase.

Regulation of Ms1 Accumulation: Promoter Activity and Ms1 Stability

Ms1 is transcribed from a single promoter in *M. smegmatis* (Fig 2B). The same promoter-like sequence is present also upstream of the Ms1 homolog MTS2823 in *M. tuberculosis* as we mapped the 5' end of MTS2823 to the same position of the 5' end of *M. smegmatis* Ms1 (Fig. 2F). The 3' end of Ms1 is then the result of intrinsic termination due to the presence of a stem-loop structure followed by a run of Us (Hnilicova *et al.*, 2014). This contrasts with 6S RNA in *E. coli* where the termination is Rho-dependent (Chae *et al.*, 2011).

The sequences of the -35 and -10 hexamers indicate that P_{Ms1} could be dependent on σ^A, similarly to 6S RNA in *E. coli* where there are two promoters, P1 and P2, recognized by the primary σ factor, σ^{70} (Kim & Lee, 2004). Nevertheless, transcription from the P2 promoter is both σ^{70}- and σ^S-dependent (Kim & Lee, 2004). The activity of P_{Ms1} is stimulated by upstream sequences that must contain binding sites for transcription factors, because no other promoter is located within the ~500 bp upstream of the Ms1 TSS (Fig 2B). *MSMEG_6173* (morphological differentiation associated protein, similar to HAD family hydrolase), a gene preceding Ms1 in the *M. smegmatis* genome, ends 114 bp upstream of the Ms1 transcription start site and it is in the same orientation as Ms1. Ms1 could be theoretically transcribed in one operon with *MSMEG_6173*; however, based on the RNA-seq data, *MSMEG_6173* is

much less expressed than Ms1 (normalized expression of *MSMEG_6173* is ~12 versus ~66,000 for Ms1 in stationary phase, Supporting Table 1).

The P_{Ms1} activity is roughly comparable to the activity of the *rrnB* promoter (-72/+20 promoter fragment) in exponential phase of growth. Similarly to P_{Ms1}, the P_{rrnB} core promoter is rather weak and is activated by the upstream elements more than 300-fold (Arnvig *et al.*, 2005). Thus, transcriptional activators are probably responsible for activation of both P_{rrnB} (~300-fold increase) and P_{Ms1} (~70-fold increase in exponential, ~360-fold increase in early stationary phase by addition of upstream elements to the core promoter, based on *lacZ* mRNA expression data, Fig 2D). This suggests that transciption factors strongly affect expression of the most abundant mycobacterial RNAs and are important gene expression activators in mycobacteria.

The activity of P_{Ms1} including upstream regulatory elements (-491/+9 region) increased ~2-fold in exponential versus stationary phase (Fig 2D). Even more notably, the Ms1 stability increased ~60-70-fold in stationary compared to exponential phase, which together with the elevated transcription initiation results in a predicted ~130-fold increase in the Ms1 level in stationary phase. This then corresponds to the actual ~115-fold increase in the Ms1 level as quantified by qPCR (Fig 1A). A similar mechanism - regulated sRNA degradation - could also influence the accumulation of other sRNAs such as 6S RNAs that has not been fully elucidated yet (Cavanagh & Wassarman, 2014).

Here, as illustrated by our promoter activity results obtained by two different approaches (β-galactosidase assays and direct measurements of the *lacZ* mRNA), we insert a cautionary note that β-galactosidase assay-derived promoter activities may be in some cases inacurate, especially when determining promoter activity decreases.

The promoter activity could not explain the Ms1 accumulation. The increased stability of Ms1 in stationary phase suggested that Ms1 accumulation might be regulated by an RNase. Therefore, we investigated the role of PNPase in Ms1 degradation because previously we had pulled down PNPase with Ms1 (Hnilicová *et al.*, 2014). First, we confirmed that PNPase degrades Ms1 in exponential phase *in vivo* (Fig 3E). PNPase is an exoribonuclease that catalyzes phosphorolytic

degradation of RNA from the 3´end and also the reverse reaction - polymerization of nucleoside diphosphates (Briani *et al.*, 2016). PNPase participates in maturation and quality control of stable RNA: tRNAs (Li *et al.*, 2002) and rRNAs (Cheng & Deutscher, 2003), sRNAs (Saramago *et al.*, 2014, Andrade *et al.*, 2012, De Lay & Gottesman, 2011, Andrade & Arraiano, 2008) and affects stability of numerous mRNAs (Mohanty & Kushner, 2003). In *E. coli*, PNPase is present either as homotrimer or in the complex with RNA helicase RhlB (Lin & Lin-Chao, 2005, Portier, 1975) or as a part of the degradosome, assembled from RNase E, PNPase, RhlB and enolase (Bandyra *et al.*, 2013). RhlB helps PNPase to overcome secondary structures in substrate RNA that would otherwise inhibit PNPase degradation activity.

In this study we isolated a recombinant PNPase without any accessory protein (Fig 3G) and showed that PNPase was able to degrade Ms1 *in vitro* (Fig 3H). Interestingly, the Ms1 “tails” (regions at 5´and 3´ends) were necessary for Ms1 degradation by PNPase. This indicated that the putative Ms1 terminator likely plays a role in modulation of Ms1 stability and degradation. Based on RNA-seq data, PNPase (*MSMEG_2656*) is the most expressed RNase in exponential phase in *M. smegmatis.* PNPase is even more abundant than RNase E (*MSMEG_4626*) or RNase J (*MSMEG_2685*) (Supporting Table 1) that are involved in maturation of ribosomal RNA in *M. smegmatis* (Taverniti *et al.*, 2011). Importantly, PNPase expression increases ~20-fold in exponential compared to stationary phase (Fig 1B), and this is consistent with the short biological half-life of Ms1 in this phase.

To conclude, Ms1 is transcribed from the strong P_{Ms1}promoter and its accumulation is achieved, at least in part, by differential degradation by PNPase.

Dissociation of Ms1 from RNAP

The mechanism of Ms1-RNAP complex disruption and its role in Ms1 protection is unknown. 6S RNA, which binds the RNAP-primary σ factor holoenzyme, serves as a template for transcription of pRNAs, which results in the dissociation of the 6S RNA:RNAP complex during outgrowth from stationary phase (Wassarman & Saecker, 2006, Gildehaus *et al.*, 2007, Beckmann *et al.*, 2012, Beckmann *et al.*,

2011, Cavanagh *et al.*, 2012). Inhibition of transcription by rifampicin stabilizes 6S RNA:RNAP complex and protects 6S RNA from degradation during the outgrowth (Beckmann *et al.*, 2012, Wurm *et al.*, 2010). However, rifampicin did not protect Ms1 from degradation during the outgrowth (Fig. 6I) indicating a different mechanism of Ms1 release from RNAP than active transcription. Previously, we showed that Ms1 could bind RNAP in exponential phase to a similar extent as in stationary phase (Hnilicova *et al.*, 2014). When we measured the stability of Ms1 in mid-exponential phase, we used rifampicin to block *de novo* transcription and Ms1 was again rapidly degraded (Fig. 3A). During the outgrowth and in exponential phase rifampicin did not affect Ms1 level. This suggests that either rifampicin does not stabilize Ms1-RNAP complex or the interaction with RNAP does not protect Ms1 from degradation.

Ms1 is highly unstable in exponential phase; therefore, the Ms1-RNAP interaction in exponential phase must be also unstable. Previously we showed that the overexpression of the primary σ factor, σ^A, in stationary phase decreased the amount of Ms1-RNAP (Hnilicova *et al.*, 2014). A hypothesis that needs to be examined in the future is whether the primary or alternative σ factors participate in dissociation of Ms1-RNAP complex during the outgrowth and in exponential phase of growth.

Ms1 affects the RNAP level

The low overall effect of Ms1 on transcription was surprising as we had expected that due to the sequestration of RNAP by Ms1, Δ*Ms1* would contain more free RNAPs that would widely influence gene expression. However, the total RNAP level decreases in Δ*Ms1* compared to *wt* (Fig 5C) and Δ*Ms1* thus lacks the extra reservoir of RNAP that is sequestered in complex with Ms1 in *wt.* It is then possible, that the relative amount of RNA polymerase available for transcription remains the same in Δ*Ms1* and *wt* (Fig 7). It is not clear, how the level of the mRNA encoding β and β' subunits is regulated.

What is the advantage to synthetize Ms1?

In exponential phase, Ms1 represents about 0.1 % of all non-rRNA. The expense of its synthesis and degradation in *wt* may account for the slightly improved growth of Δ*Ms1* at the end of exponential phase in the defined rich medium (Fig 4B). In stationary phase the Ms1 amount increases ~115-fold (Fig 1A) which constitutes more than 30 % of non-rRNA. What is the purpose of synthetizing such an enormous amount of this sRNA in the cell?

The presence of Ms1 increases the cell fitness in stationary phase where the Δ*Ms1* strain is less resistant to gamma irradiation. A contributing factor to this phenotype could be the lowered expression of *MSMEG_5556*, a major facilitator superfamily transporter (Li *et al.*, 2017). It may be involved in transport of detrimental metabolites out of the cell (Ramón-García *et al.*, 2009). Alternatively, higher amounts of RNAP in stationary phase *wt* cells might help *M. smegmatis* overcome consequences of radioactive exposure by an unknown mechanism. If γ irradiation destroyed a fraction of RNAP in the bacterial cell, a higher number of functional RNAP would remain in *wt* cells as these cells had a higher amount of RNAP before γ irradiation. Chronic γ irradiation produces complex oxidative stress for cells. In these conditions, cells have to respond to the resulting reactive oxygen species (ROS) that damage DNA and proteins. This can partially mimic growth in macrophages, which produce reactive oxygen and nitrogen intermediates (Burney *et al.*, 1999, Nathan & Shiloh, 2000). In *M. tuberculosis*, the Ms1 homolog MTS2823 is present at a high level during chronic stages of infection in lungs where it may contribute to the ability of the pathogen to survive in the hostile environment (Arnvig *et al.*, 2011, Arnvig & Young, 2012). The activity of Ms1 promoter also increased during ethanol stress (Fig 2G). In bacteria, ethanol damages cell wall and membrane integrity and causes increased level of reactive oxygen species (Cao *et al.*, 2017) that is similar to chronic γ irradiation stress. Ms1 could be therefore specifically connected with this type of stress.

Finally, it appears that Ms1 increases the pool of RNAP molecules that can become important during rapid changes in the environment such as outgrowth from stationary phase. This is consistent with the fact that the intracellular level of RNAP generally positively correlates with the growth rate (Klumpp & Hwa, 2008). Thus, it is likely advantageous for mycobacterial cells to have stored RNAP enzymes that can rapidly

start transcription of appropriate genes when surrounding conditions become favorable. This "jump start" is probably beneficial for the cell especially when it faces competition from other species or for pathogens that must deal with its host's immune system.

Experimental Procedures

Bacterial strains, growth conditions, plasmids

For detailed description of individual strains, see List of strains and plasmids in Supporting Data. *M. smegmatis* mc^2 155 (*wt*, *LK865*) were grown at 37°C in Middlebrook 7H9 medium with 0.2% glycerol and 0.05% Tween 80, and harvested in exponential (OD_{600} ~0.5; 6 hours of cultivation), early stationary (OD_{600} ~2.5–3, 24 hours of cultivation) or late stationary (OD_{600} ~2, 48 hours of cultivation) phase unless stated otherwise. When required, media were supplemented with hygromycin (50 µg/ml), kanamycin (20 µg/ml), spectinomycin (20 µg/ml for *M. smegmatis*, 100 µg/ml for *E. coli*), rifampicin (250 µg/ml) or anhydrotetracycline (100 µg/ml). Transformations of *M. smegmatis* mc^2 155 cells were performed by electroporation. To test the stress response, NaCl (final concentration 0.5 M), diamide (final concentration: 0-10 mM dilution series) and EtOH (5% final concentration) were pipetted directly into 7H9 medium; or 25 ml of culture was filtered using MF™ – Membrane Filters 0.45 µm HA (Millipore) and cells were resuspended either in 25 ml of preheated 7H9 medium with adjusted pH, PBS or distilled water. For γ irradiation, *wt* (*LK865*) and *ΔMs1 (LK1611)* strains were aliquoted, and transported at 4°C in sterilized cuvettes. Half of samples were chronically irradiated by gammacell 220 (Dose 40 Gy/hour) at room temperature (RT). The second half was kept at RT without irradiation as a control of cell viability. Every 2 hours, samples were collected (irradiated and non-irradiated), diluted and plated in triplicates. CFU/ml values were counted after 2 days of incubation at 37°C. The CFU/ml values for *wt* (*LK865*) and *ΔMs1 (LK1611)* in stationary phase and after rifampicin treatment and were calculated in duplicates and are shown as an average from three independent experiments ± SEM (standard error of the mean). The doubling times were calculated with ±SEM.

∆*Ms1 and* ∆*Ms1+Ms1 strains*

∆Ms1 *(LK1611)* strain was prepared by recombineering (van Kessel & Hatfull, 2007). Briefly, 100 ng of linear allelic exchange substrate (AES) was electroporated into *M. smegmatis* mc^2 pJV53 cells *(LK1321)*. To prepare AES, first the *hyg* resistance gene was amplified by PCR with Phusion High-Fidelity DNA Polymerase (NEB) from the pML1342 plasmid (Huff *et al.*, 2010) and cloned into pUC18 into PstI/XbaI (for primers see List of primers in Supporting Data). Then LA (left arm – upstream of the 5´end of Ms1) was PCR amplified from genomic DNA and cloned into HindIII/PstI; a short DNA linker was then cloned into XbaI to separate the close XbaI/BamHI restriction sites that were preventing the cloning. Subsequently, RA (right arm) was cloned into XbaI/BamHI. RA was commercially synthetized by GeneArt Gene Synthesis service (Invitrogen, for the sequence see Supporting Data). The construct was verified by sequencing. After transformation into *M. smegmatis* mc^2 pJV53 *kan* cells *(LK1321)*, colonies were selected at hygromycin/kanamycin plates; deletion of Ms1 was verified by PCR, and the absence of the Ms1 band in total RNA. Then, the ∆*Ms1* strain was cured from pJV53 by passaging without the kanamycin *(LK1611)*. For ∆*Ms1* strain complementation *(LK2223)*, *Ms1* gene plus 483 bp upstream and 50 bp downstream were cloned into NotI/PacI restriction sites of pTC-mcs vector (Klotzsche et al., 2009) and transformed into ∆*Ms1 (LK1611)* strain.

RNA isolation and RT-qPCR

25 ml cells from the indicated growth phases were quickly cooled on ice, pelleted and immediately frozen. The pellet was then suspended in 240 μl TE (pH 8.0) plus 60 μl lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM LiCl, 50 mM EDTA pH 8.0, 5% SDS) and 600 μl acidic phenol (pH~3):chloroform (1:1). Lysates were sonicated in a fume hood, centrifuged, the aqueous phase extracted three times with acidic phenol (pH~3):chloroform and precipitated with ethanol. RNA was dissolved in water, treated with DNase (TURBO DNA-free Kit, Ambion) and visualized on a 7 M urea 7% polyacrylamide gel by staining with GelRed (LabMark). RNA was reverse transcribed into cDNA (SuperScriptIII, Invitrogen) using random hexamers and amplified by quantitative reverse transcription PCR (RT-qPCR) in a LightCycler 480 System

(Roche Applied Science) in duplicate reactions containing LightCycler 480 SYBR Green I Master and 0.5 µM primers (each). Primers were designed with Primer3 software and their sequences are in the Supporting Data. Negative controls (no template reactions and reactions with RNA as a template to control for contamination with genomic DNA) were run in each experiment, the quality of the PCR products was determined by dissociation curve analysis and the efficiency of the primers determined by standard curves. The mRNA level was quantified on the basis of the threshold cycle (Ct) for each PCR product that was normalized to reference gene value according to the formula $2^{(Ct^{(ref)}-Ct^{(mRNA)})}$ and expression (E) normalized to *wt* strain ($E=E_{\Delta Ms1}/E_{wt}$).

β-*galactosidase assay and promoter constructs*

Annealed oligonucleotides (for the -38 +9 construct, -72/+20 *rrnB* promoter) or PCR amplified promoter fragments with Phusion High-Fidelity DNA Polymerase (NEB) (see List of primers in Supporting Data) were cloned via the ScaI site preceding the *lacZ* reporter gene in the pSM128 integrative vector (Dussurget *et al.*, 1999). Sequence verified constructs were transformed into *M. smegmatis* mc^2 155 (*wt*, *LK865*). 1 ml of bacterial culture (see growth conditions) was centrifuged (13200 x g, 10 min, 4°C) and pellets washed with 500 µl of Z buffer 2 (Z buffer with 2.7 µl of 2-mercaptoethanol/ml; Z buffer: 60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, pH 7.0), centrifuged again and resuspended in 500 µl of Z buffer 2. Cells were sonicated 3x 20 s (amplitude 50 %) on ice with 1 min pauses between sonications and centrifuged (17900 x g, 10 minutes, 4°C). 100 µl of sonicate and 900 µl of Z buffer were incubated for 5 min at 30 ºC and 200 µl of orthonitrophenyl-β-D-galactopyranoside (ONPG) solution was added (4 mg/ml ONPG in Z-buffer). After 20 min, the reaction was stopped with 500 µl of 1M Na_2CO_3, OD_{420} and OD_{550} were measured, and β-galactosidase activity (in arbitrary units) calculated: activity = $1000*(OD_{420}-1.75*OD_{550})/(v*t*c)$, where v is sample volume [ml], t time of reaction, c protein concentration [mg/ml] measured by Bradford protein assay. To determine *lacZ* mRNA amount, 20 ml of exponential phase and 10 ml of stationary phase cells were quickly cooled on ice, pelleted and immediately frozen. Paralelly, OD_{600} of cell cultures was measured. 40 ng of *in vitro* transcribed 6S RNA from *B. subtilis* was

added to each RNA isolation as an RNA recovery marker. RNA was isolated, treated with DNase, reverse transcribed into cDNA amplified by qPCR. The mRNA level was quantified on the basis of the threshold cycle (Ct) for each PCR product that was normalized to 6S RNA recovery marker value according to the formula 2^($Ct^{(6S\ RNA)}$-$Ct^{(mRNA)}$) and *lacZ* expression normalized to OD_{600} of individual cell cultures.

RNA-seq

Isolated RNA was dissolved in water and two times treated with DNase (TURBO DNA-free Kit, Ambion). Five μg of total RNA were rRNA depleted with a Ribo-Zero rRNA Removal Kit (Gram-Positive Bacteria) (Illumina), and strand specific libraries were prepared with Illumina compatible NEXTflex Rapid Directional RNA-Seq Kit (Bioo Scientific) according to manufacturer´s instructions. Pooled barcoded libraries (four samples in biological triplicates) were sequenced in a single lane at Illumina HiSeq 2000 in 50 bp single end regime at EMBL Genomics Core Facility, Heidelberg, Germany. Reads were aligned to *Mycobacterium smegmatis* str. mc^2 155 (NCBI Reference Sequence: NC_008596.1) using HISAT2 (Kim *et al.*, 2015) version 2.1.0 (with parameters "--rna-strandness R" and "--no-spliced-alignment") and samtools (Li *et al.*, 2009) version 1.3.1. and inspected using IGV browser 2.0.23 and the igvtools package (Thorvaldsdóttir *et al.*, 2013). Differentially expressed genes were detected by DESeq2 (Love *et al.*, 2014), Benjamini-Hochberg FDR corrected P-value 0.05 was used as a cutoff. RNA-seq data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-7004. Normalized gene expression was calculated as a number of mapped reads per gene divided by gene length and the size of non-rRNA library (sum of all mapped reads except for reads mapping to 3,819,833-3,825,494 and 5,024,129-5,031,105 that represents rRNA operons *MSMEG_3755-MSMEG_3757* and *MSMEG_4929-MSMEG_4931* plus surrounding intergenic regions). Normalized expression was averaged from 3 biological replicates and multiplied by 10^8.

Northern blotting, western blotting

RNAs were resolved on a 7% polyacrylamide gel and transferred onto an Amersham Hybond-N membrane according to the protocol described in (Pánek *et al.*, 2011).

Probes were 5′ ^{32}P-labeled oligonucleotides (anti-Ms1: 5′-TCGTGGCCGTCCGCTTTTCGAAACTACGC-3′; anti-5S: 5′-CTGGCAGGCTTAGCTTCCGGGTTCGGGATG-3′), and signals were captured in Fuji MS phosphor storage screens, scanned with a Molecular Imager_FX (Bio-Rad) and quantified with QuantityOne software (Bio-Rad). Alternatively, 5´end biotinylated oligonucleotides (with the same sequence) were hybridized to the membrane and detected with BrightStar BioDetect Kit (Ambion) according to manufacturer´s instructions. Northern blot signals were normalized either to 5S rRNA or to recovery marker RNA. Recovery markers RNA (Ms1notail; the recovery marker was a shorter version of the Ms1 allowing to distinguish these two according to molecular mass in gels or *B. subtilis* 6S RNA) were added at the beginning of RNA extraction as a control for RNA degradation or pipetting errors (Rabatinová *et al.*, 2013). Proteins were analyzed by SDS-PAGE and detected by western blotting using mouse monoclonal antibody against the β subunit of RNAP [clone name 8RB13] and secondary antibody conjugated with a fluorophore dye and quantified with an Odyssey reader (LI-COR Biosciences) or goat anti-S3 ribosomal protein conjugated with HRP (kindly provided by Dr. Isabella Moll, Max F. Perutz Laboratories Vienna, Austria). After the last wash in PBS/0.05% Tween 20, the membrane was washed three times in 100 mM Tris-HCl pH 9, 150 mM NaCl, 1 mM $MgCl_2$, then 66 µl NBT and 33 µl BCIP (both from Promega) were added into 10 ml of the same buffer and the colorimetric reaction was stopped by water. The signal was quantified using the QuantityOne (Bio-Rad) software.

CRISPR knockdown

sgRNA oligos (see Supporting data) were designed and cloned into PLJR962 according to the published protocol (Rock *et al.*, 2017). Sequence-verified constructs were electroporated into *wt* (*LK865*) *M. smegmatis* resulting in strains *LK2197* (*MSMEG_2685,* RNase E knockdown), *LK2198* (*MSMEG_4626,* RNase J knockdown), *LK2219* (*MSMEG_2656,* PNPase knockdown), *LK2200* (Ms1 knockdown) and *LK2261* - negative control, non-targeting control sgRNA (Rock *et al.*, 2017). The knockdowns of RNases were induced in exponential phase with anhydrotetracycline for 3 hrs, OD_{600} of cell cultures was measured and the same

amout of cells harvested. 200 ng of *in vitro* transcribed 6S RNA from *B.subtilis* was added to each RNA isolation as an RNA recovery marker. RNA was isolated and analysed by northern blotting or treated with DNase, reverse transcribed into cDNA amplified by qPCR. The RNase J, RNase E and PNPase mRNA level was quantified on the basis of the threshold cycle (Ct) for each PCR product that was normalized to 6S RNA recovery marker value according to the formula $2^{(Ct^{(6S\ RNA)}-Ct^{(mRNA)})}$.

5'RACE from M. tuberculosis H37Rv

M. tuberculosis H37Rv ATCC 27294 strain was cultivated for three weeks on Lowenstein-Jensen agar plates at 35.5°C in aerobic conditions. The cells were harvested and directly lysed in TRI Reagent (Sigma Aldrich) and RNA isolated according to manufacturer´s instructions. Five µg of total RNA was treated with 5U TEX (Terminator 5'-Phosphate-Dependent Exonuclease; Epicentre) for 1 h at 37°C. After the reaction, RNA was extracted with TRIzol and precipitated with ethanol. Purified RNA was treated with 1U of Tobacco Acid Pyrophosphatase (TAP; Epicentre) for 1 h at 37°C, extracted with TRIzol again, precipitated with ethanol, and a 5′-adaptor DNA/RNA oligonucleotide (5′-ATCGTaggcaccugaaa-3′, DNA in upper case letters) was ligated to the 5′ ends (1 h at 37°C). RNA was then extracted and reversely transcribed into cDNA (SuperScriptIII, Invitrogen) with an MTS2823 (Ms1 homolog) specific reverse primer (5′-CATCTGCTGTTCGCAATTAC-3′). The same reverse primer and the 5′-ATCGTAGGCACCTGAAA-3′ forward primer were used for polymerase chain reaction (PCR) with Taq DNA polymerase (Biotools). The PCR products were sequenced and mapped to the *M. tuberculosis* H37Rv (GenBank #AL123456.2).

Purification of PNPase-His6, in vitro digestion assay

Gene *MSMEG_2656* (*gpsI*) encoding PNPase was amplified by PCR with Phusion High-Fidelity DNA Polymerase (NEB) (for primers see Supporting Data) from *M. smegmatis* mc^2 155 genomic DNA, inserted into NdeI–NotI sites of pET-22b(+) and transformed into *E. coli* BL21 (DE3). Induction of PNPase-His6 was carried out at

OD_{600} = 0.5 with 1mM IPTG for 2 hours at room temperature. Rifampicin (f.c. 100 μg/ml) was added together with IPTG. Cells were harvested by centrifugation, washed, resuspended in buffer P (300 mM NaCl, 50 mM Na_2HPO_4, 5% glycerol, 3 mM 2-mercaptoethanol, 0.1 mM phenylmethylsulfonyl-fluoride) and disrupted by sonication. Cell debris was removed by centrifugation, the supernatant was mixed with 1 ml Ni-NTA Agarose (Qiagen) and incubated for 1 hour at 4°C with gentle shaking. Ni-NTA was washed with buffer P, then with buffer P supplemented with 30 mM (P+30) imidazole and PNPase eluted with buffer P+400 mM imidazole and dialysed against storage buffer (10 mM Tris-HCl, pH 8.0, 25 mM NH_4Cl, 100 μM EDTA, 10 mM $MgCl_2$, 30 μM DTT, 50% glycerol). $MgCl_2$ (f.c. 5 mM) was added and PNPase stored at −20° C. Proteins were separated on NuPAGE 4-12% Bis-Tris gels (Invitrogen) and visualized by Coomassie blue staining (SimplyBlue, Invitrogen). The identity of PNPase was verified by excising the corresponding band followed by mass spectrometry. Ms1, Ms1nb, Ms1notail DNA templates were designed based on secondary structure *in silico* modeling (Hnilicová *et al.*, 2014, Pánek *et al.*, 2011) and purchased from Invitrogen DNA (GeneArt Strings DNA Fragments) with T7 promoters at their 5′ ends; they were amplified by PCR with Phusion High-Fidelity DNA Polymerase (NEB) (Hnilicová *et al.*, 2014). The 371 nt-long *rpoC* template was amplified from genomic DNA with primers carrying a T7 promoter at the 5´end (see Supporting Data) and all RNAs were prepared with T7 RiboMAX Express Large Scale RNA Production System (Promega). RNA degradation assay (10 μl total volume) contained 400 ng RNA, 10 mM $MgCl_2$, 10 mM Na_2HPO_4, 20 mM Tris-HCl pH 7.6, 6.9 μM PNPase (or RNase A, Sigma Aldrich). The mixture was incubated for 30 min at 37°C (Nurmohamed *et al.*, 2011). Reaction was stopped with 10 μl of 200 mM EDTA; In control reactions, no enzyme was added and samples were incubated as described above. Samples were mixed 2:1 with sample buffer (95% formamide, 20 mM EDTA pH 8.0), heated for 1 min at 90°C and electrophoresed in 7 M urea 7% polyacrylamide gels. The gels were stained for 20 min with GelRed (LabMark) and visualized using a UV transilluminator and quantified with QuantityOne software (Bio-Rad).

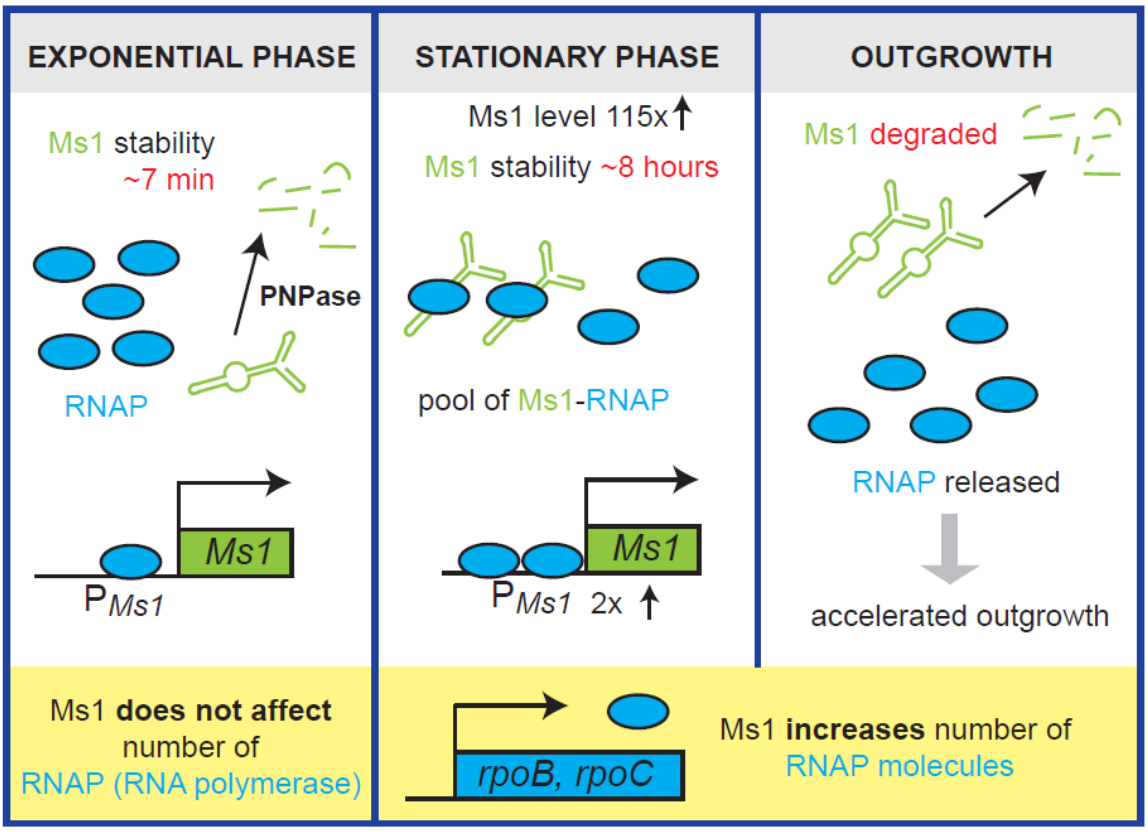
Acknowledgements

We thank Dr. Isabella Moll (Max F. Perutz Laboratories, Austria) for provision of antibody. We thank Brian Grogan for English language editing. This work used instruments provided by C4Sys infrastructure. This work was supported by grants from the Czech Science Foundation [13-27150P to J.H., P305/12/G034 to L.K.], Charles University [PRIMUS/MED/26 and UNCE 204013 to M.P.], the Grant Agency of Charles University [946216 to M.J.], Ministry of the Interior of the Czech Republic [VI20152020044], and Ministry of Education, Youth and Sports of the Czech Republic within project the Center for advanced applied science, project number CZ.02.1.01/0.0/0.0/16_019/0000778. The authors have no conflict of interest to declare.

Author contributions

JH, MS, LK designed the study, MS, MJ, JH, OR, PP, DV, JP, PB, AS, OP, PK, PP, HS and MH performed the experiments, JH, MP, MS, MJ and LK analyzed the data and JH, LK wrote the manuscript.

Graphical abstract



Abbreviated Summary

Ms1 RNA binds the RNA polymerase (RNAP) core and is the most abundant non-rRNA transcript in stationary phase in *Mycobacterium smegmatis*. This study identifies the mechanisms responsible for its high accumulation. Further, we show that Ms1 increases expression of two RNAP subunits, β and β'. The increased level of RNAP helps mycobacterial cells rapidly start transcription of appropriate genes when surrounding conditions become favorable.

References

Andrade, J.M., and Arraiano, C.M. (2008) PNPase is a key player in the regulation of small RNAs that control the expression of outer membrane proteins. *RNA* **14**: 543-551.

Andrade, J.M., Pobre, V., Matos, A.M., and Arraiano, C.M. (2012) The crucial role of PNPase in the degradation of small RNAs that are not associated with Hfq. *RNA* **18**: 844-855.

Arnvig, K., and Young, D. (2012) Non-coding RNA and its potential role in Mycobacterium tuberculosis pathogenesis. *RNA Biol* **9**: 427-436.

Arnvig, K.B., Comas, I., Thomson, N.R., Houghton, J., Boshoff, H.I., Croucher, N.J., Rose, G., Perkins, T.T., Parkhill, J., Dougan, G., and Young, D.B. (2011) Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis. *PLoS Pathog* **7**: e1002342.

Arnvig, K.B., Gopal, B., Papavinasasundaram, K.G., Cox, R.A., and Colston, M.J. (2005) The mechanism of upstream activation in the rrnB operon of Mycobacterium smegmatis is different from the Escherichia coli paradigm. *Microbiology* **151**: 467-473.

Aviv, M., Giladi, H., Oppenheim, A.B., and Glaser, G. (1996) Analysis of the shut-off of ribosomal RNA promoters in Escherichia coli upon entering the stationary phase of growth. *FEMS Microbiol Lett* **140**: 71-76.

Bandyra, K.J., Bouvier, M., Carpousis, A.J., and Luisi, B.F. (2013) The social fabric of the RNA degradosome. *Biochim Biophys Acta* **1829**: 514-522.

Beckmann, B.M., Burenina, O.Y., Hoch, P.G., Kubareva, E.A., Sharma, C.M., and Hartmann, R.K. (2011) In vivo and in vitro analysis of 6S RNA-templated short transcripts in Bacillus subtilis. *RNA Biol* **8**: 839-849.

Beckmann, B.M., Hoch, P.G., Marz, M., Willkomm, D.K., Salas, M., and Hartmann, R.K. (2012) A pRNA-induced structural rearrangement triggers 6S-1 RNA release from RNA polymerase in Bacillus subtilis. *EMBO J* **31**: 1727-1738.

Briani, F., Carzaniga, T., and Dehò, G. (2016) Regulation and functions of bacterial PNPase. *Wiley Interdiscip Rev RNA* **7**: 241-258.

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

Burney, S., Caulfield, J.L., Niles, J.C., Wishnok, J.S., and Tannenbaum, S.R. (1999) The chemistry of DNA damage from nitric oxide and peroxynitrite. *Mutat Res* **424**: 37-49.

Cao, H., Wei, D., Yang, Y., Shang, Y., Li, G., Zhou, Y., Ma, Q., and Xu, Y. (2017) Systems-level understanding of ethanol-induced stresses and adaptation in E. coli. *Sci Rep* **7**: 44150.

Cavanagh, A.T., Klocko, A.D., Liu, X., and Wassarman, K.M. (2008) Promoter specificity for 6S RNA regulation of transcription is determined by core promoter sequences and competition for region 4.2 of sigma70. *Mol Microbiol* **67**: 1242-1256.

Cavanagh, A.T., Sperger, J.M., and Wassarman, K.M. (2012) Regulation of 6S RNA by pRNA synthesis is required for efficient recovery from stationary phase in E. coli and B. subtilis. *Nucleic Acids Res* **40**: 2234-2246.

Cavanagh, A.T., and Wassarman, K.M. (2014) 6S RNA, A Global Regulator of Transcription in Escherichia coli, Bacillus subtilis, and Beyond. *Annu Rev Microbiol*.

Chae, H., Han, K., Kim, K.S., Park, H., Lee, J., and Lee, Y. (2011) Rho-dependent termination of ssrS (6S RNA) transcription in Escherichia coli: implication for 3' processing of 6S RNA and expression of downstream ygfA (putative 5-formyl-tetrahydrofolate cyclo-ligase). *J Biol Chem* **286**: 114-122.

Chen, K., Hu, Z., Xia, Z., Zhao, D., Li, W., and Tyler, J.K. (2015) The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. *Mol Cell Biol* **36**: 662-667.

Cheng, Z.F., and Deutscher, M.P. (2003) Quality control of ribosomal RNA mediated by polynucleotide phosphorylase and RNase R. *Proc Natl Acad Sci U S A* **100**: 6388-6393.

China, A., Tare, P., and Nagaraja, V. (2010) Comparison of promoter-specific events during transcription initiation in mycobacteria. *Microbiology* **156**: 1942-1952.

Davis, E., Chen, J., Leon, K., Darst, S.A., and Campbell, E.A. (2015) Mycobacterial RNA polymerase forms unstable open promoter complexes that are stabilized by CarD. *Nucleic Acids Res* **43**: 433-445.

De Lay, N., and Gottesman, S. (2011) Role of polynucleotide phosphorylase in sRNA function in Escherichia coli. *RNA* **17**: 1172-1189.

Dussurget, O., Timm, J., Gomez, M., Gold, B., Yu, S., Sabol, S.Z., Holmes, R.K., Jacobs, W.R., and Smith, I. (1999) Transcriptional control of the iron-responsive fxbA gene by the mycobacterial regulator IdeR. *J Bacteriol* **181**: 3402-3408.

Feklístov, A., Sharon, B.D., Darst, S.A., and Gross, C.A. (2014) Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol* **68**: 357-376.

Gildehaus, N., Neusser, T., Wurm, R., and Wagner, R. (2007) Studies on the function of the riboregulator 6S RNA from E. coli: RNA polymerase binding, inhibition of in vitro transcription and synthesis of RNA-directed de novo transcripts. *Nucleic Acids Res* **35**: 1885-1896.

Hnilicova, J., Matejckova, J.J., Sikova, M., Pospisil, J., Halada, P., Panek, J., and Krasny, L. (2014) Ms1, a novel sRNA interacting with the RNA polymerase core in mycobacteria. *Nucleic Acids Research* **42**: 11763-11776.

Hnilicová, J., Jirát Matějčková, J., Šiková, M., Pospíšil, J., Halada, P., Pánek, J., and Krásný, L. (2014) Ms1, a novel sRNA interacting with the RNA polymerase core in mycobacteria. *Nucleic Acids Res* **42**: 11763-11776.

Hu, Y., Morichaud, Z., Perumal, A.S., Roquet-Baneres, F., and Brodolin, K. (2014) Mycobacterium RbpA cooperates with the stress-response σB subunit of RNA polymerase in promoter DNA unwinding. *Nucleic Acids Res* **42**: 10399-10408.

Hubin, E.A., Tabib-Salazar, A., Humphrey, L.J., Flack, J.E., Olinares, P.D., Darst, S.A., Campbell, E.A., and Paget, M.S. (2015) Structural, functional, and genetic analyses of the actinobacterial transcription factor RbpA. *Proc Natl Acad Sci U S A* **112**: 7171-7176.

Huff, J., Czyz, A., Landick, R., and Niederweis, M. (2010) Taking phage integration to the next level as a genetic tool for mycobacteria. *Gene* **468**: 8-19.

Kim, D., Langmead, B., and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**: 357-360.

Kim, K.S., and Lee, Y. (2004) Regulation of 6S RNA biogenesis by switching utilization of both sigma factors and endoribonucleases. *Nucleic Acids Res* **32**: 6057-6068.

Klumpp, S., and Hwa, T. (2008) Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc Natl Acad Sci U S A* **105**: 20245-20250.

Krásný, L., and Gourse, R.L. (2004) An alternative strategy for bacterial ribosome synthesis: Bacillus subtilis rRNA transcription regulation. *EMBO J* **23**: 4473-4483.

Krásný, L., Tiserová, H., Jonák, J., Rejman, D., and Sanderová, H. (2008) The identity of the transcription +1 position is crucial for changes in gene expression in response to amino acid starvation in Bacillus subtilis. *Mol Microbiol* **69**: 42-54.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, G.P.D.P. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Li, P., Gu, Y., Li, J., Xie, L., Li, X., and Xie, J. (2017) Mycobacterium tuberculosis Major Facilitator Superfamily Transporters. *J Membr Biol* **250**: 573-585.

Li, Z., Reimers, S., Pandit, S., and Deutscher, M.P. (2002) RNA quality control: degradation of defective transfer RNA. *EMBO J* **21**: 1132-1138.

Lin, P.H., and Lin-Chao, S. (2005) RhlB helicase rather than enolase is the beta-subunit of the Escherichia coli polynucleotide phosphorylase (PNPase)-exoribonucleolytic complex. *Proc Natl Acad Sci U S A* **102**: 16590-16595.

Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.
Mohanty, B.K., and Kushner, S.R. (2003) Genomic analysis in Escherichia coli demonstrates differential roles for polynucleotide phosphorylase and RNase II in mRNA abundance and decay. *Mol Microbiol* **50**: 645-658.
Murray, H.D., Schneider, D.A., and Gourse, R.L. (2003) Control of rRNA expression by small molecules is dynamic and nonredundant. *Mol Cell* **12**: 125-134.
Nathan, C., and Shiloh, M.U. (2000) Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens. *Proc Natl Acad Sci U S A* **97**: 8841-8848.
Neusser, T., Polen, T., Geissen, R., and Wagner, R. (2010) Depletion of the non-coding regulatory 6S RNA in E. coli causes a surprising reduction in the expression of the translation machinery. *BMC Genomics* **11**: 165.
Nurmohamed, S., Vincent, H.A., Titman, C.M., Chandran, V., Pears, M.R., Du, D., Griffin, J.L., Callaghan, A.J., and Luisi, B.F. (2011) Polynucleotide phosphorylase activity may be modulated by metabolites in Escherichia coli. *J Biol Chem* **286**: 14315-14323.
Paul, B.J., Ross, W., Gaal, T., and Gourse, R.L. (2004) rRNA transcription in Escherichia coli. *Annu Rev Genet* **38**: 749-770.
Portier, C. (1975) Quaternary structure of polynucleotide phosphorylase from Escherichia coli: evidence of a complex between two types of polypeptide chains. *Eur J Biochem* **55**: 573-582.
Pánek, J., Krásny, L., Bobek, J., Jezková, E., Korelusová, J., and Vohradsky, J. (2011) The suboptimal structures find the optimal RNAs: homology search for bacterial non-coding RNAs using suboptimal RNA structures. *Nucleic Acids Res* **39**: 3418-3426.
Rabatinová, A., Šanderová, H., Jirát Matějčková, J., Korelusová, J., Sojka, L., Barvík, I., Papoušková, V., Sklenár, V., Žídek, L., and Krásný, L. (2013) The δ subunit of RNA polymerase is required for rapid changes in gene expression and competitive fitness of the cell. *J Bacteriol* **195**: 2603-2611.
Rammohan, J., Ruiz Manzano, A., Garner, A.L., Prusa, J., Stallings, C.L., and Galburt, E.A. (2016) Cooperative stabilization of Mycobacterium tuberculosis rrnAP3 promoter open complexes by RbpA and CarD. *Nucleic Acids Res* **44**: 7304-7313.
Rammohan, J., Ruiz Manzano, A., Garner, A.L., Stallings, C.L., and Galburt, E.A. (2015) CarD stabilizes mycobacterial open complexes via a two-tiered kinetic mechanism. *Nucleic Acids Res* **43**: 3272-3285.
Ramón-García, S., Martín, C., Thompson, C.J., and Aínsa, J.A. (2009) Role of the Mycobacterium tuberculosis P55 efflux pump in intrinsic drug resistance, oxidative stress responses, and growth. *Antimicrob Agents Chemother* **53**: 3675-3682.
Rock, J.M., Hopkins, F.F., Chavez, A., Diallo, M., Chase, M.R., Gerrick, E.R., Pritchard, J.R., Church, G.M., Rubin, E.J., Sassetti, C.M., Schnappinger, D., and Fortune, S.M. (2017) Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nat Microbiol* **2**: 16274.
Saramago, M., Bárria, C., Dos Santos, R.F., Silva, I.J., Pobre, V., Domingues, S., Andrade, J.M., Viegas, S.C., and Arraiano, C.M. (2014) The role of RNases in the regulation of small RNAs. *Curr Opin Microbiol* **18**: 105-115.
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., and Campbell, E.A. (2013) Structure and function of CarD, an essential mycobacterial transcription factor. *Proc Natl Acad Sci U S A* **110**: 12619-12624.
Tare, P., China, A., and Nagaraja, V. (2012) Distinct and contrasting transcription initiation patterns at Mycobacterium tuberculosis promoters. *PLoS One* **7**: e43900.

Taverniti, V., Forti, F., Ghisotti, D., and Putzer, H. (2011) Mycobacterium smegmatis RNase J is a 5'-3' exo-/endoribonuclease and both RNase J and RNase E are involved in ribosomal RNA maturation. *Mol Microbiol* **82**: 1260-1276.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178-192.

Trotochaud, A.E., and Wassarman, K.M. (2005) A highly conserved 6S RNA structure is required for regulation of transcription. *Nat Struct Mol Biol* **12**: 313-319.

van Kessel, J.C., and Hatfull, G.F. (2007) Recombineering in Mycobacterium tuberculosis. *Nat Methods* **4**: 147-152.

Verma, A.K., and Chatterji, D. (2014) Dual role of MsRbpA: transcription activation and rescue of transcription from the inhibitory effect of rifampicin. *Microbiology* **160**: 2018-2029.

Wassarman, K.M., and Saecker, R.M. (2006) Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science* **314**: 1601-1603.

Wassarman, K.M., and Storz, G. (2000) 6S RNA regulates E. coli RNA polymerase activity. *Cell* **101**: 613-623.

Wurm, R., Neusser, T., and Wagner, R. (2010) 6S RNA-dependent inhibition of RNA polymerase is released by RNA-dependent synthesis of small de novo products. *Biol Chem* **391**: 187-196.

Figure legends

Fig 1. Expression of Ms1 and RNases in exponential and stationary phase.

A. Expression of Ms1 measured by qPCR was normalized per cell. Ms1 level increases ~115-fold in stationary versus exponential phase. The average represents three independent biological experiments and error bars show ±SEM.

B. Expression of several RNases and the primary σ factor, σ^A, was measured by qPCR and normalized per cell. The average from three independent biological experiments is shown with ±SEM.

Fig 2. Identification of the Ms1 promoter.

A. Ms1 promoter region fragments differing in length of the upstream region were fused to *lacZ*. Shortened and mutated Ms1 promoter variants (-491/-22; -491/+9mut) were used to show that Ms1 was transcribed from a single promoter, P_{Ms1}.

B. The graph shows results of β-galactosidase assays for constructs shown in A. β-galactosidase activity was measured in exponential (exp, OD_{600} 0.5, ~6 hrs of growth), early stationary (OD_{600} 2-3, ~24 hrs of growth) and late stationary (OD_{600} 2,

~48 hrs of growth) phases. Data were normalized to the value for the core promoter (-38/+9 construct) in exponential phase which was set to 100. The averages from at least three independent biological experiments performed in duplicates are shown with ±SEM. The dotted horizontal line represents β-galactosidase activity measured in control cells with an empty pSM128 plasmid. *rrnB* - ribosomal promoter (-72/+20), a.u. - arbitrary units.

C. Sequences of the unmodified and mutated (mutations in -491/+9mut indicated with asterisks) Ms1 core promoter. -22 indicates the endpoint of the -491/-22 fragment.

D. The bar graph shows relative *lacZ* mRNA levels of β-galactosidase (LacZ) per cell for empty pSM128 plasmid, -38/+9 core promoter, -491/+9 largest Ms1 promoter fragment and ribosomal promoter (rrnB, -72/+20). Data were normalized to the value of the -38/+9 construct in exponential phase which was set to 100. The averages from three independent biological experiments performed in duplicates are shown with ±SEM.

E. Putative promoter elements of the Ms1 homolog in *M. tuberculosis* H37Rv are shown in boxes, the transcription start site (+1) identified by 5´RACE is indicated with the horizontal arrow.

F. The transcription start site of the Ms1 homolog (MTS2823) was identified by 5´RACE in *M. tuberculosis*. PCR products of 5´RACE were resolved on agarose gel and sequenced. The arrow indicates the specific band and its sequence from the 5´end.

G. The graph shows β-galactosidase activity of the -491/+9 Ms1 promoter fragment 2 and 4 hrs after the induction of various types of stress. β-galactosidase activity was normalized to the activity of a control sample at the time 0 (before stress). The graph shows at least three independent biological experiments performed in duplicates. Error bars represent standard deviation; a.u. - arbitrary units, statistically significant difference (P-value <0.05, T-test) is marked by asterisk.

Fig 3. Stability of Ms1 in exponential and stationary phase.

A. Northern blots of total RNA isolated from *pJAM2-rrnB-Ms1* (a strain overexpressing Ms1) in exponential phase probed with an oligonucleotide complementary to Ms1. Time shows minutes after rifampicin addition.

B. Northern blots of total RNA isolated from *pJAM2-rrnB-Ms1* (a strain overexpressing Ms1) in stationary phase probed with an oligonucleotide complementary to Ms1. Time shows hours after rifampicin addition.

C. The graph shows averaged data ±SD (standard deviation) for Ms1 stability from three biological replicates. The graph was used to calculate the half-lives of Ms1: 8 min in exponential and ~8 hrs in stationary phase.

D. The knockdown of RNase E (*LK2197* strain), RNase J (*LK2198*), PNPase (*LK2219*) was induced for 3 hours with anhydrotetracycline. A control strain (*LK2261*, that contains non-targeting control sgRNA) was used as a negative control and also treated for 3 hours with anhydrotetracycline. The cells were harvested at mid-exponential phase and total RNA isolated from the same amount of cells. *B. subtilis* 6S RNA was added directly at the beginning of RNA isolation as the RNA recovery marker that controls for efficient RNA isolation. The efficiency of RNase knockdown was determined by qPCR in individual strains. The amounts of RNase E mRNA in *LK2197*, RNase J mRNA in *LK2198*, PNPase mRNA in *LK2219* and all RNases in the control strain (*LK2261*) were normalized to the RNA recovery marker (*B. subtilis* 6S RNA). The graphs show relative decreases of mRNA levels in individual knockdown strains compared to the *LK2261* control strain. The averages represent two independent biological experiment, the error bars indicates ±SD.

E. RNA isolated from RNase E (*LK2197* strain), RNase J (*LK2198*), PNPase (*LK2219*) and control strain *LK2261* strain was resolved by northern blotting and probed with oligonucleotides complementary to Ms1 and *B. subtilis* 6S RNA. *B. subtilis* 6S RNA serves as a loading control. The experiment was repeated two times with the same result.

F. Quantification of Ms1 level from Fig 3E. Ms1 levels in RNase depleted strains (*LK2197*, *LK2198* and *LK2219*) were normalized to the Ms1 level in control (*LK2261*) strain. The bar graph shows averages from two independent experiments with ±SD.

H. *In vitro* digestion assays were performed with three variants of Ms1: „Ms1nb“ is a mutant variant lacking the central bubble; „Ms1notail“ is a mutant variant lacking the tails at 5´and 3´ends. Ms1, Ms1nb and Ms1notail predicted secondary structures are shown above the gel. A 371 nt fragment of the *rpoC* mRNA was used as a control. RNAs were incubated with PNPase, RNase A or no enzyme (negative control, „-“), resolved on a polyacrylamide gel and stained with GelRed. Full length RNAs are marked with asterisks.

Fig 4. Characterization of the Δ*Ms1* strain.

A. The Ms1 gene was replaced with a hygromycin resistance gene in *M. smegmatis wt* cells.

B. Δ*Ms1* and *wt* growth was monitored for 24 hours; statistically significant difference ($P<0.05$, T-test) in OD_{600} was observed after 12 hrs cultivation in 7H9 medium supplemented with glycerol. The data are averages from six independent experiments. The error bars show ±SEM.

C. Survival (colony forming units; CFU) of stationary phase chronically γ irradiated Δ*Ms1* and *wt* cells.

D. Survival (CFU) of stationary phase non-irradiated control Δ*Ms1* and *wt* cells.

E. Survival (CFU) of exponential phase chronically γ irradiated Δ*Ms1* and *wt* cells.

F. Survival (CFU) of exponential phase non-irradiated control Δ*Ms1* and *wt* cells. The mock controls represent non-irradiated samples that were otherwise treated as irradiated samples. The data points are averages from four independent experiments, error bars represent ±SD.

Fig 5. Ms1-regulated genes.

A. Total RNA from Δ*Ms1* and *wt* stationary phase cells was isolated, resolved on a polyacrylamide gel and stained with GelRed; in Δ*Ms1* strain, Ms1 is absent. In Δ*Ms1+Ms1* strain, *Ms1* under its native promoter is integrated into the genome and Ms1 expression is restored to the level of *wt*.

B. Expression of five selected genes was measured by qPCR in Δ*Ms1, wt* and Δ*Ms1+Ms1* strains. *MSMEG_1205, MSMEG_5540, MSMEG_5556* and *rpoB*, *rpoC* genes encoding β and β´ RNAP subunits altered their expression in Δ*Ms1* and these changes were reverted by the addition of the *Ms1* wt copy (Δ*Ms1+Ms1* strain). Expression levels were normalized in each sample to the reference gene encoding σ^A and then compared to *wt* (set as 1). The average value represents four biological replicates, error bars ±SEM. Statistically significant differences (P-value <0.05, T-test) are marked with asterisks.

C. The total protein level of the RNAP β subunit in stationary phase is significantly lower in Δ*Ms1* versus *wt* (P-value <0.05, T-test, marked by asterisk) and restored in the Δ*Ms1+Ms1* strain. Five μg of protein lysate was loaded and the amount of RNAP was quantified by western blotting with an anti-β subunit RNAP antibody. The S3 ribosomal protein represents loading control. Three independent experiments were performed; averages are shown, error bars represent ±SEM.

Fig 6. Outgrowth of Δ*Ms1* from stationary phase.

A. Δ*Ms1* and *wt* stationary phase cells (after 24 hrs cultivation) were diluted into a fresh medium to OD_{600} 0.1 and their growth was monitored for 6 hrs. During the first 3 hours, we detected a statistically significant delay in Δ*Ms1* growth (P-value <0.05, T-test, marked with asterisks). The data represent averages from seven independent experiments ±SEM.

B. Δ*Ms1+Ms1* and *wt* stationary phase cells (after 24 hrs cultivation) were diluted into a fresh medium to OD_{600} 0.1 and their growth was measured for 6 hrs. The data represent averages from three independent experiments ±SEM.

C. In the *LK2200* strain, Ms1 CRISPR knockdown can be induced by anhydrotetracycline. The *LK2200* strain was inoculated to OD_{600} 0.1; after 8 hrs of cultivation, the Ms1 knockdown was induced by anhydrotetracycline (ATC) or not (control) and the strain was cultivated for another 16 hrs to reach stationary phase. Then, cells were diluted into a fresh medium to OD_{600} 0.1 without anhydrotetracycline and their growth was measured for 6 hrs similarly as in experiment in Fig 6A. The cells, where Ms1 was depleted in stationary phase by anhydrotetracycline, displayed slower outgrowth than control cells without anhydrotetracycline. Three independent experiments were performed. Statistically significant delay in outgrowth between *LK2200* with anhydrotetracycline compared to *LK2200* without anhydrotetracycline (P-value <0.05, T-test) is marked with asterisks.

D. The *LK2261* strain (negative control that contains non-targeting control sgRNA) was cultivated as the *LK2200* strain in Fig 6C. Addition of anhydrotetracycline (ATC) had no effect on the outgrowth.

E. CRISPR knockdown in *LK2200* strain was verified by northern blotting. *LK2200* and *LK2261* control strains were inoculated to OD_{600} 0.1; after 8 hrs of cultivation anhydrotetracycline (ATC) was added and the strains were cultivated for another 16 hrs to reach stationary phase. Total RNA was isolated from stationary phase, resolved on polyacrylamide gels and analyzed by northern blotting with oligonucleotide probes complementary to Ms1 or 5S rRNA (loading control). In *LK2200*, Ms1 was depleted below the detection liming.

F. The RNAP level during outgrowth was determined by quantitative western blotting with an anti-β subunit RNAP antibody. The antibody against the S3 protein was used as a control. Time 0 was before dilution into fresh media. Five μg of protein lysate were loaded from indicated time points.

G. Quantification of RNAP amount from Fig 6F. The bar graph shows averages from four independent experiments with ±SEM; statistically significant differences (P-value <0.05, T-test) are marked with asterisks.

H. The level of endogenous Ms1 after dilution of stationary phase cells into a fresh medium was determined by northern blotting and normalized to 5S rRNA. The

experiment was performed in three biological replicates. The graph (averages ±SD) shows the decay of Ms1: a 50% reduction after ~18 min in exponential phase.

I. The level of endogenous Ms1 after dilution of stationary phase cells into a fresh medium was determined by qPCR and normalized to 16S rRNA in the presence or absence of rifampicin. The experiment was performed in two biological replicates, error bars represent ±SD.

Fig 7. RNAP level in *wt* and Δ*Ms1* strains

In exponential phase, Ms1 is unstable and rapidly degraded, PNPase contributed to the degradation. The level of RNAP is unchanged in Δ*Ms1* compared to *wt* strain.

In the stationary phase, Ms1 interacts with RNAP in the *wt* strain and a fraction of RNAP is sequestered by Ms1. In Δ*Ms1*, no Ms1 is present. When Ms1 is missing, a corresponding fraction of the cellular RNAP is absent. This suggests that the Ms1 absence is compensated for by a decreased amount of RNAP molecules and the pool of RNAPs dedicated to transcription remains similar both in *wt* and Δ*Ms1* strains.

During the outgrowth, the difference in RNAP level between Δ*Ms1* and *wt* remains similar for ~30 min and the growth Δ*Ms1* is slowed down for ~3 hours.

Figure 1

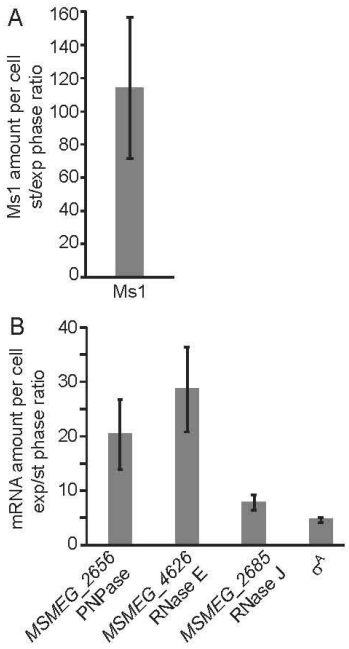


Figure 2

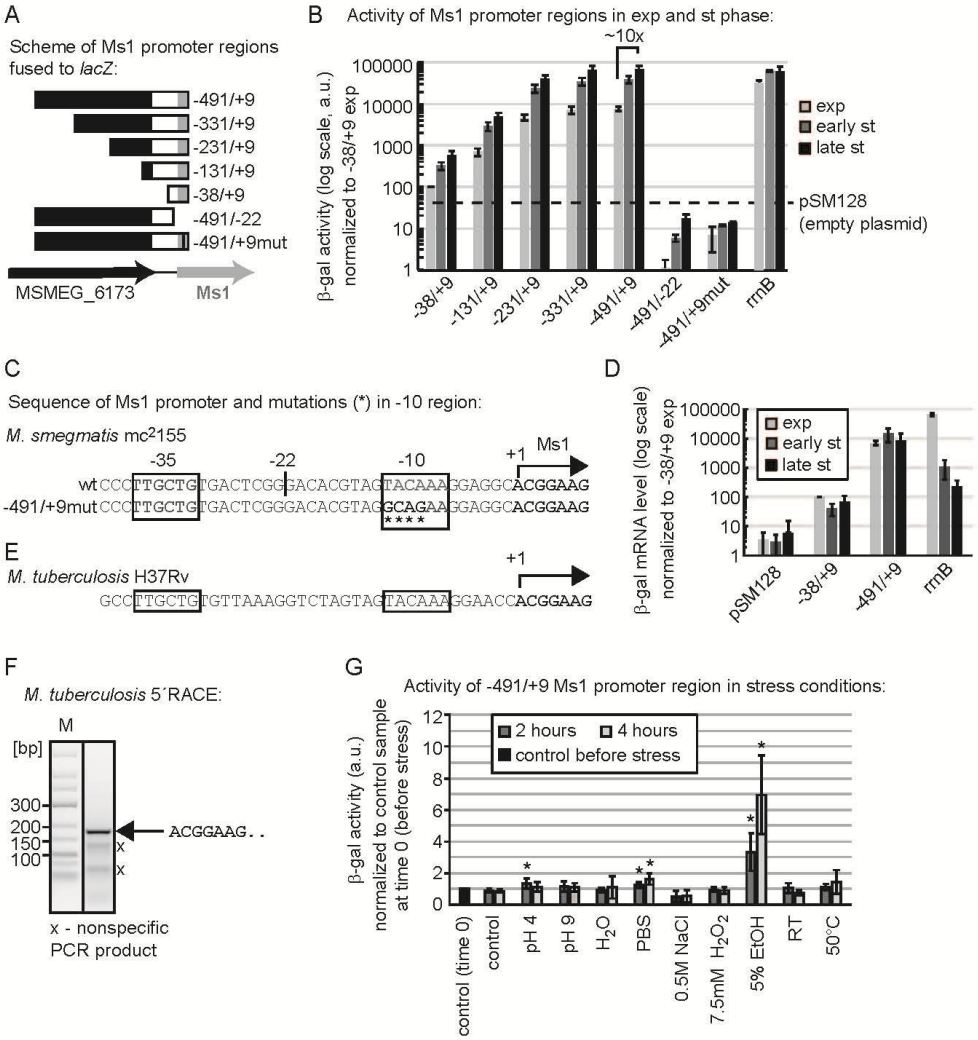


Figure 3

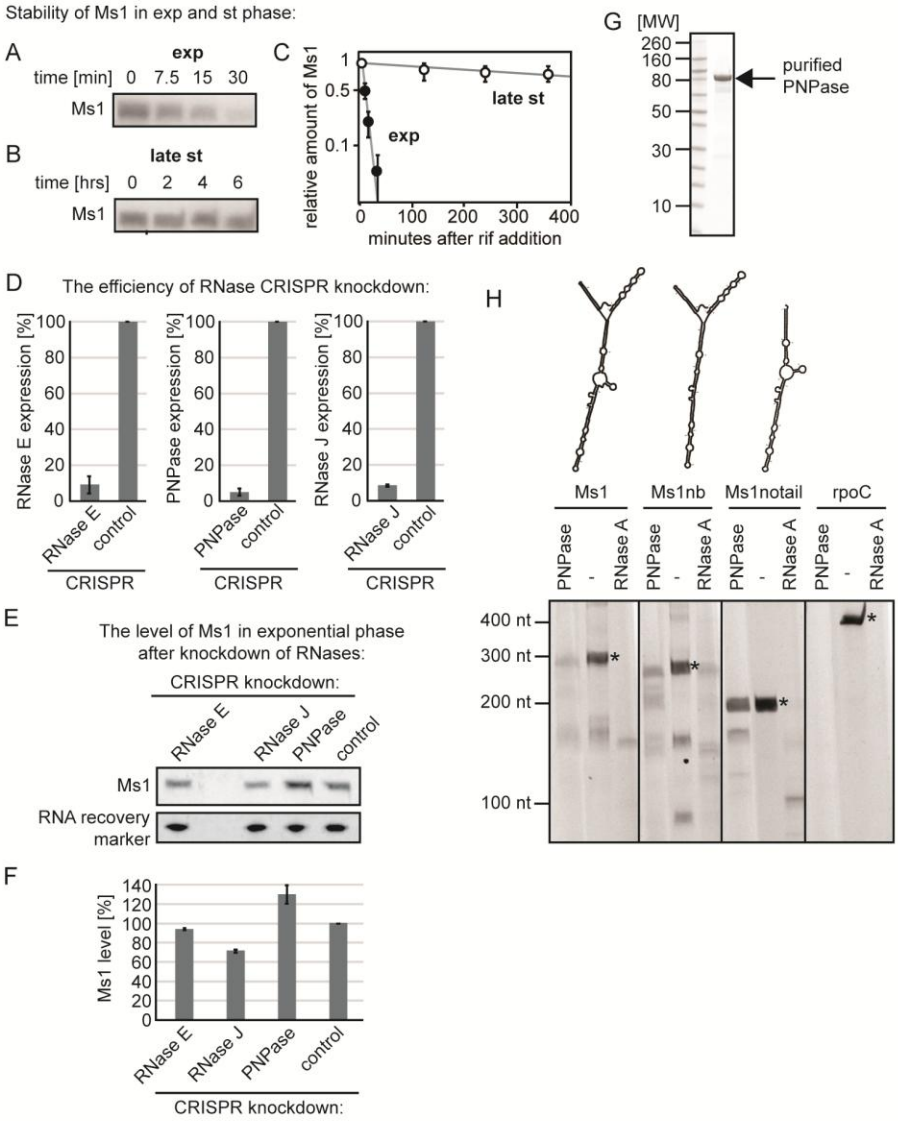


Figure 4

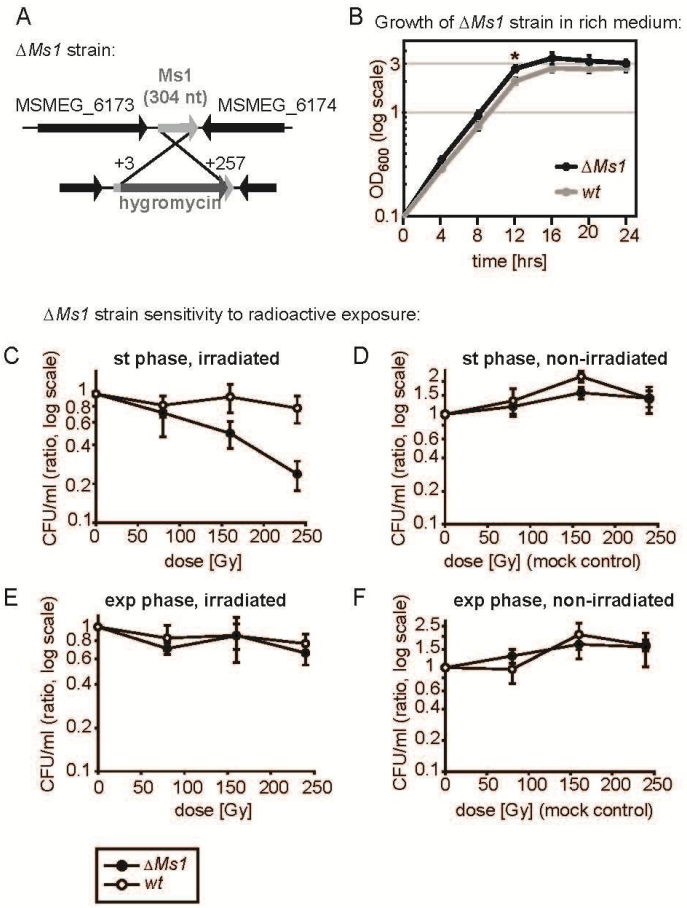


Figure 5

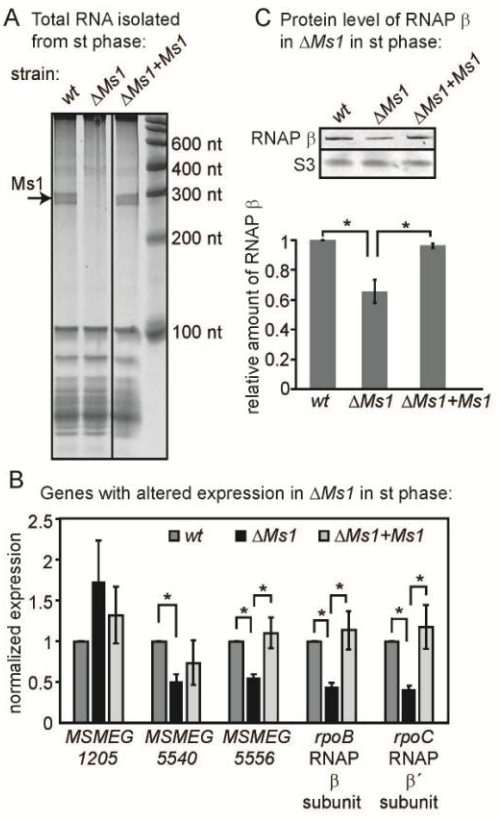


Figure 6

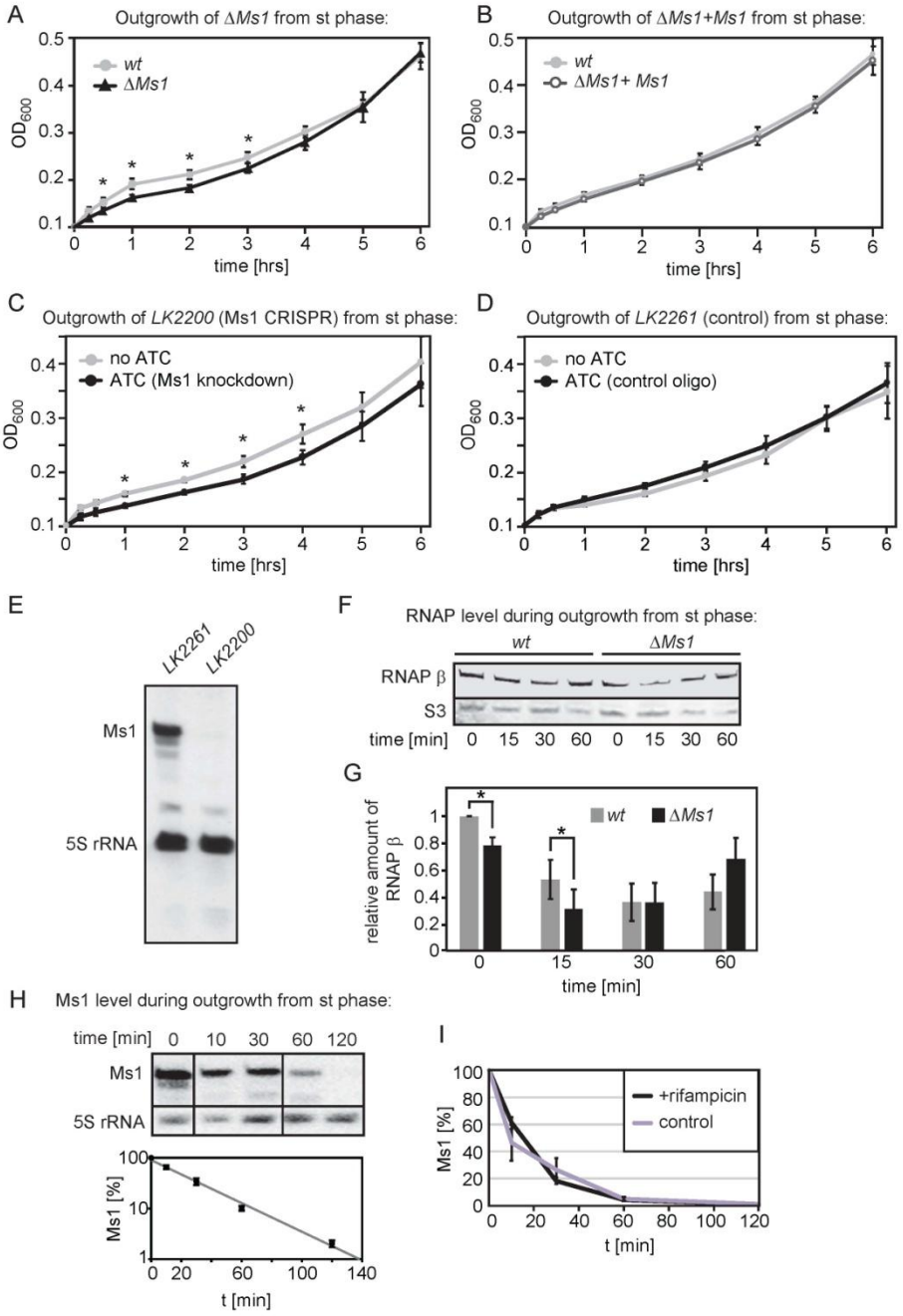


Figure 7