

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

Ramaniuk O.^{1,3}, Černý, M.²., Krásný, L.¹, Vohradsky, J.^{2*}

¹ Laboratory of Microbial Genetics and Gene Expression

² Laboratory of Bioinformatics

Institute of Microbiology ASCR, v.v.i., Videnska 1083, 14220 Prague, Czech Republic

³ Department of Genetics and Microbiology, Faculty of Science, Charles University, Víničná 5, CZ-12843 Prague 2, Czech Republic

* corresponding author

e-mail:

ramaniuk@biomed.cas.cz

martin.cerny@biomed.cas.cz

krasny@biomed.cas.cz

vohr@biomed.cas.cz

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 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 are 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 \mathbf{y} represents the measured mRNA concentration time series proportional value and $\tilde{\mathbf{y}}$ represents the time series computed using the model equation 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{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.

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].

Table 1. List of strains and plasmids

Strain/Plasmid	Relevant characteristics	Source
<i>B. subtilis</i>		
BSB1	wt BaSysBio	[3]
MH5636	<i>rpoC</i> -His10	[22]
<i>E. coli</i>		
LK22	BL21 pCD2/Bsu_ sigA	[23]
LK1	pRLG770 with Pveg (-38/-1,+1G)	[24]

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.

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].

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_S1. Putative promoter sequences were PCR-amplified using wt *B. subtilis* gDNA as the template. The only exception was the control Pveg promoter, which was amplified from the LK1 plasmid containing a cloned Pveg fragment. The wt Pveg 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.

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- μ l 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 MgCl₂, 1 mM dithiothreitol (DTT), 0.1 mg/ml BSA (bovine serum albumin) and 60 mM NaCl. ATP, CTP, and GTP were at 200 μ M, and UTP was at 10 μ M plus 2 μ M radiolabelled [α -³²P]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 (Figure 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.

1. Results

1.1. 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 Methods.

The list of all SigA regulon genes analysed, together with the model parameters and analysis results, are shown in Supplemental_Table_S2. 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 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 Figure 1).

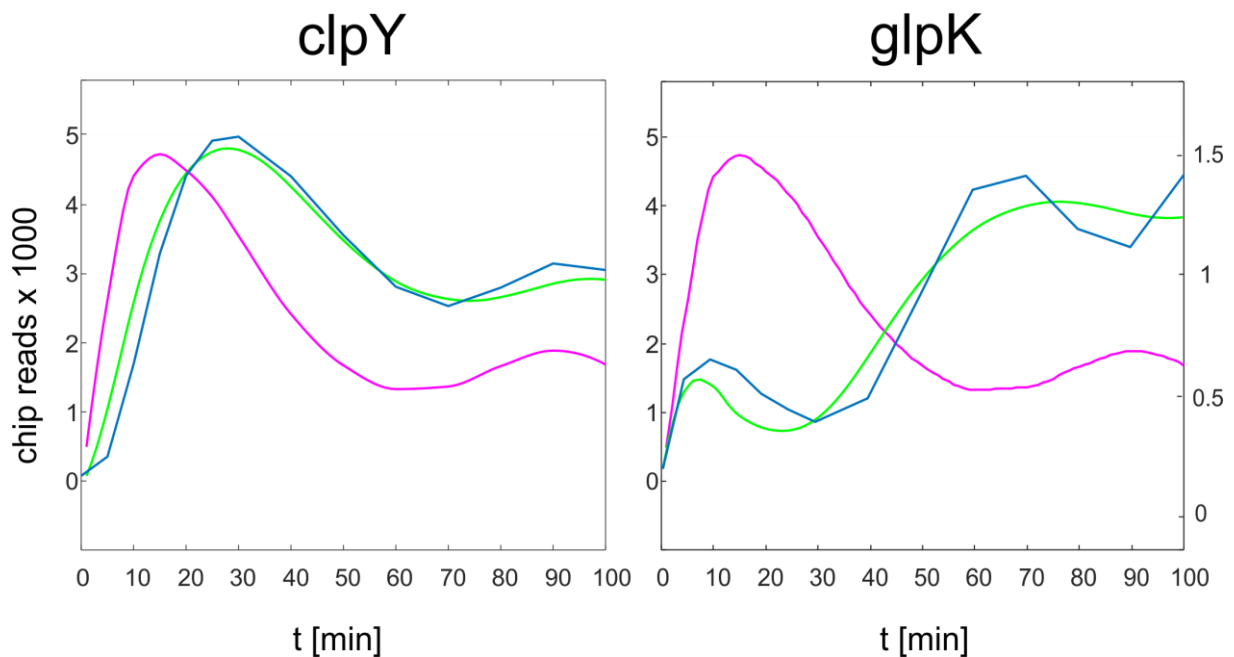


Figure 1. An example of positively (left, *clpY*) and negatively (right, *glpK*) controlled genes during germination and outgrowth. Red - *sigA* mRNA profile; blue - primary values (relative RNA levels) for target gene profiles; green - modelled profiles of the target gene.

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).

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 $w > 0$

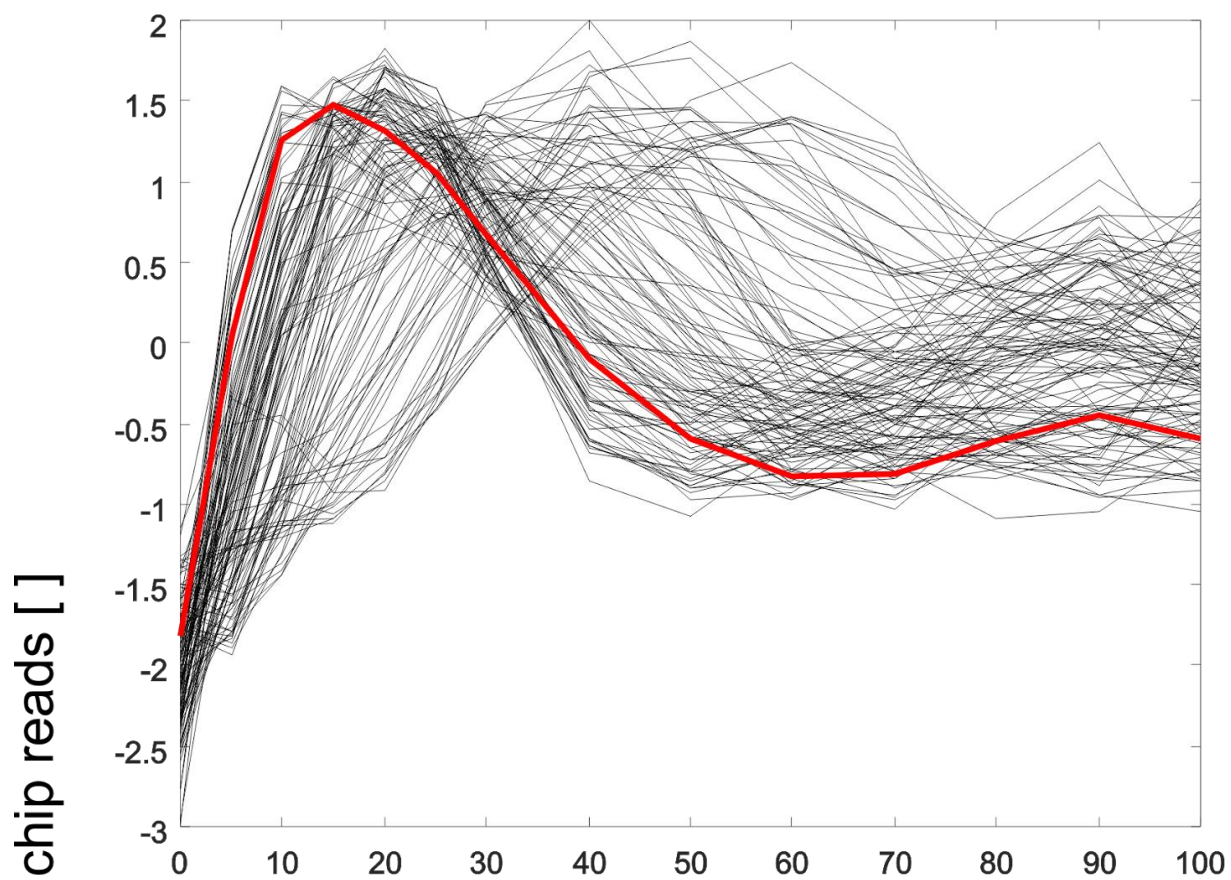
*adhA-yraA, aroF-aroB-aroH-trpE-trpD-trpC-trpF-trpB-trpA-hisC-tyrA-aroE, cggR-gapA-
 pgk-tpi-pgm-eno, citG, citZ-icd-mdh, ctsR-mcsA-mcsB-clpC-radA-yacK, codV-clpQ-
 clpY-codY, clpX, comA-yuxO, comEA-comEB-comEB, comQ-comX, cymR-yrvO-
 trmU, cspB, ctaA, drm-punA, epsA-epsB-epsC-yveN-epsE-epsF-epsG-epsH-epsI-
 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-rpml-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-foIB-foIK-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-foIC, ydjM, yfhQ-fabL-sspE, yfkJ-yfkl-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*

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

A



B

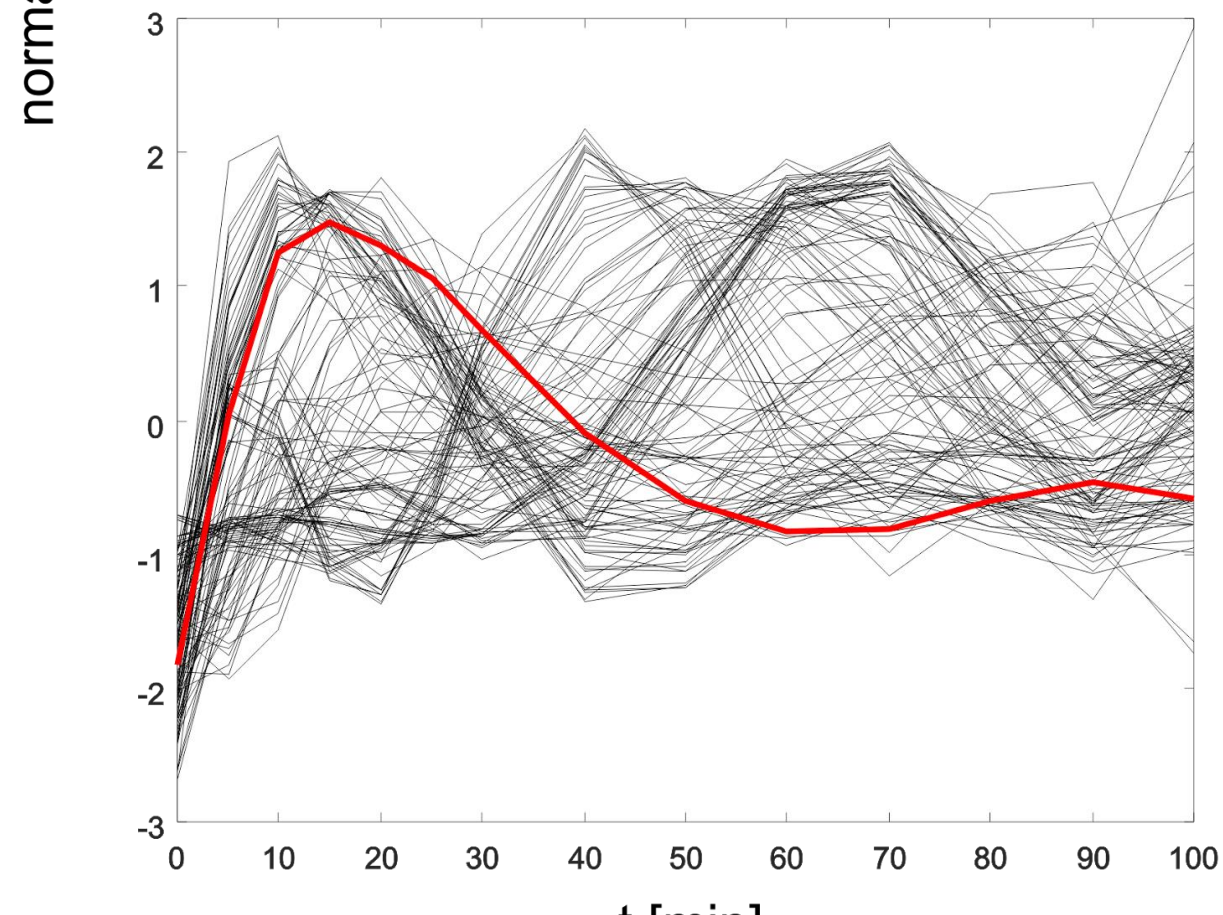


Figure 2. *Normalized expression profiles of genes A. found to be under control of SigA (red thick line), B. genes that might not be under control of SigA.*

Either the expression profiles of the positively controlled group (Figure 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, Figure 2B), the profiles in Figure 2A were consistent in shape. Interestingly, Figure 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.

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-foiB-foiK-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

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.

1.2. 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.

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, ylol, dinG, yloN, mrpA-mrpB-mrpC-mrpD-mrpE-mrpF-mrpG, yycC-yycB, alsT, yloH, yugl, 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, ylbl, 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, yvcl-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, ytcl, 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, ylal, ysmB, ytlQ, ytql, atpl, 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

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 Figure 3 (lists of specific genes in individual categories are in Supplemental_Table_S3).

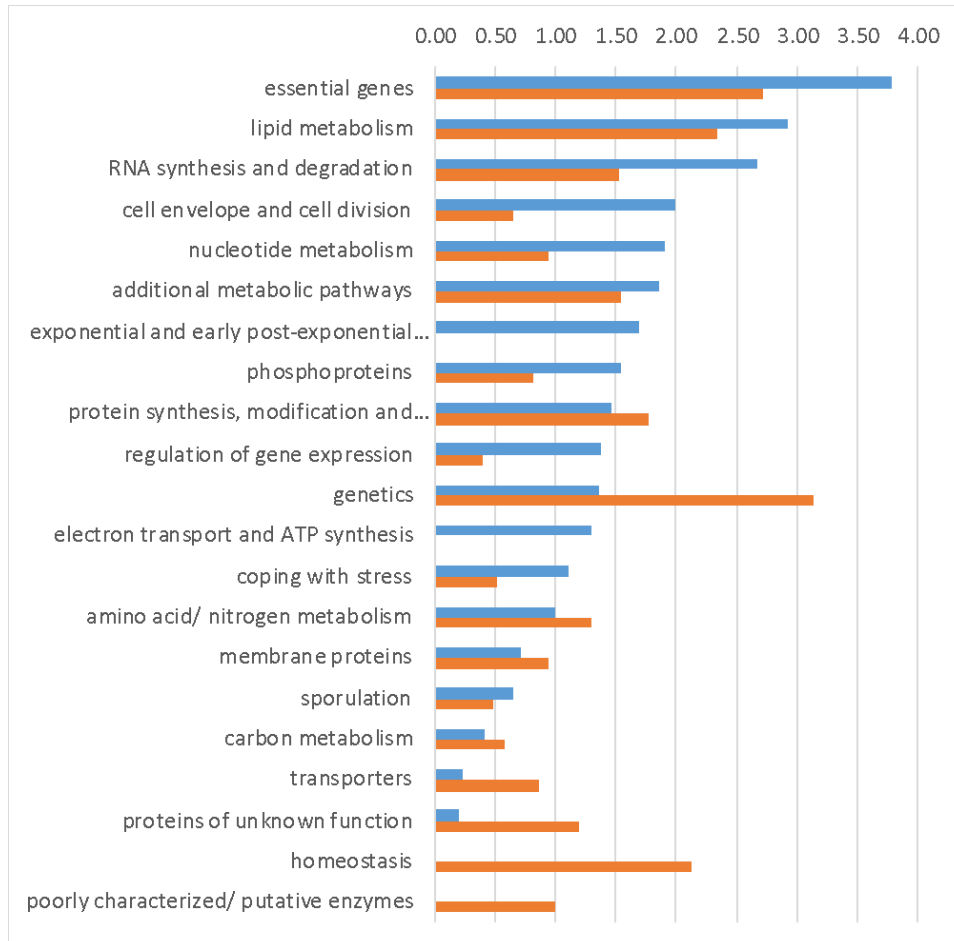


Figure 3. Gene ontologies for potential SigA targets (red) compared with the group of genes known to be positively regulated with SigA (blue, Table 2). The bars show relative enrichment of the respective functional group within the subset compared to the number of genes in this functional group among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

Figure 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]).

1.3. 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_S4. 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_S4 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].

1.4. 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 Figure 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_S5 for all genes mentioned below. Figure 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_text_S6.

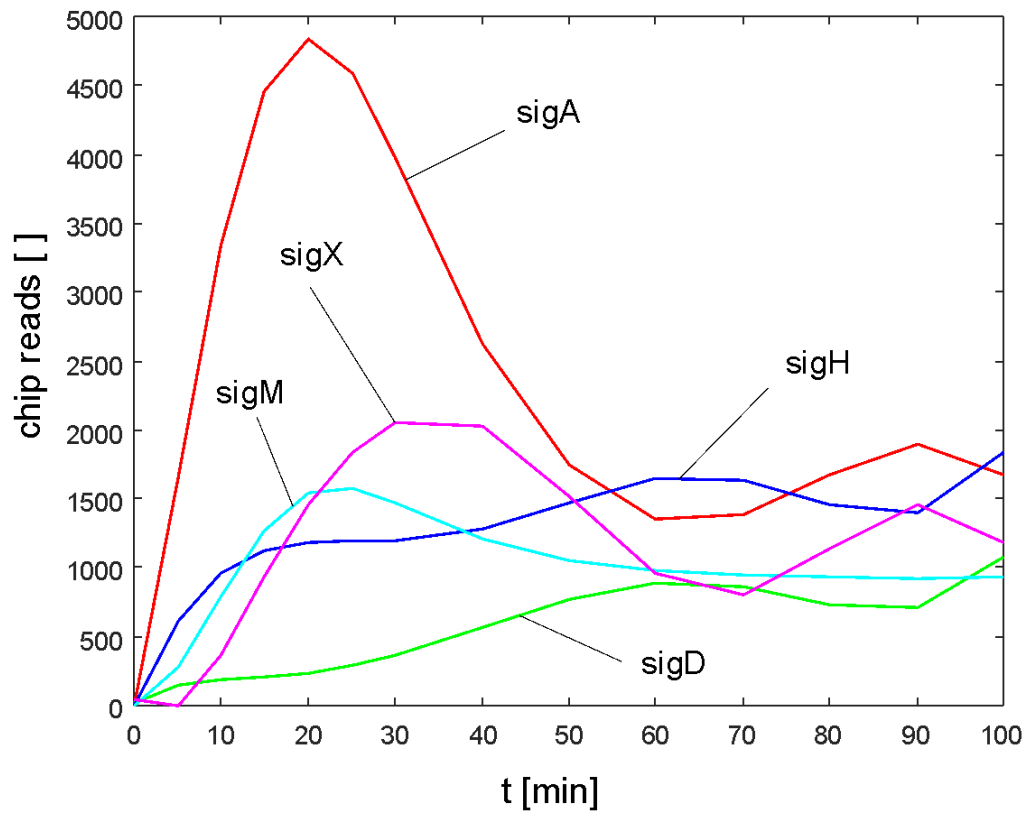


Figure 4. Expression profiles (mRNAs) of sigma factors of the SigA regulon during germination and outgrowth. Red - SigA, green - SigD, blue - SigH, cyan - SigM, magenta - SigX. Horizontal axis - time in minutes, vertical axis - microarray units.

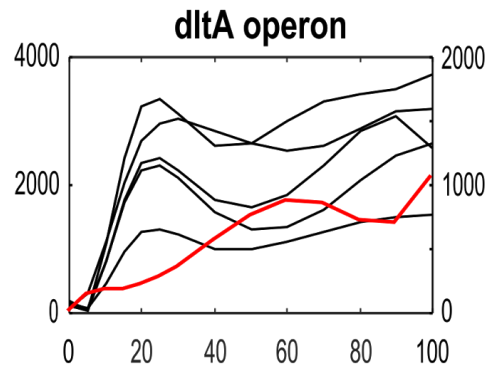
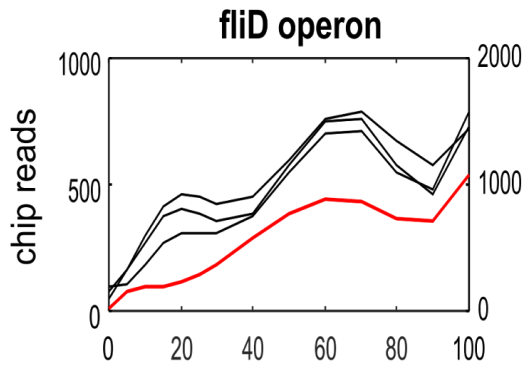
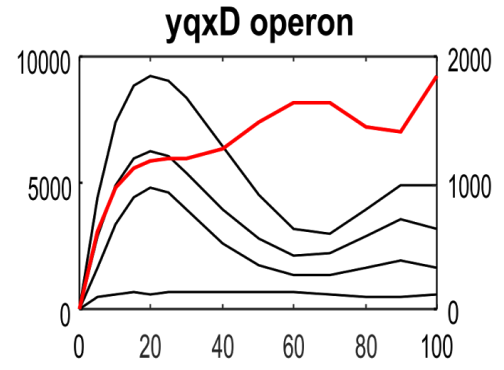
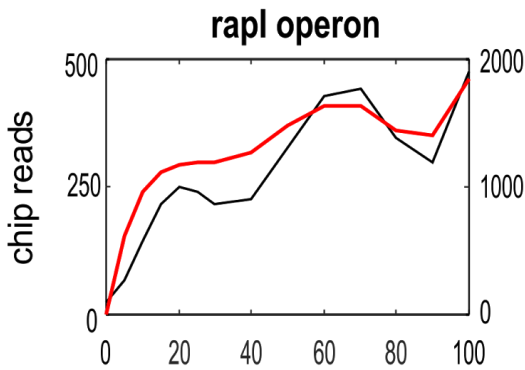
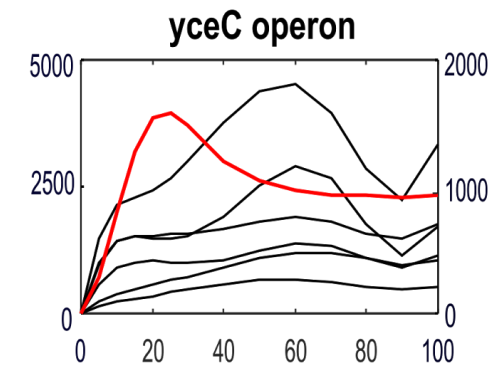
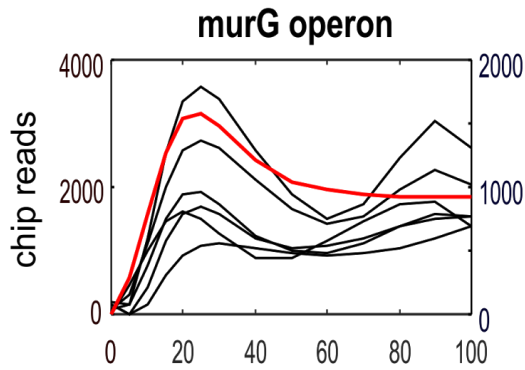
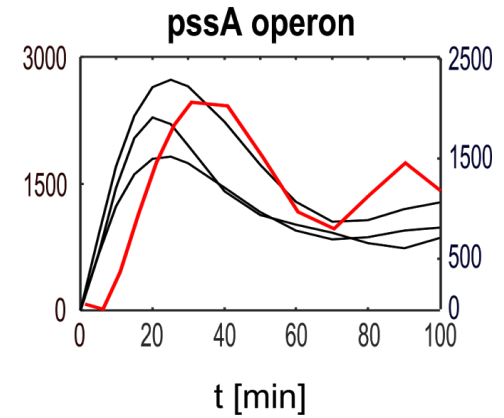
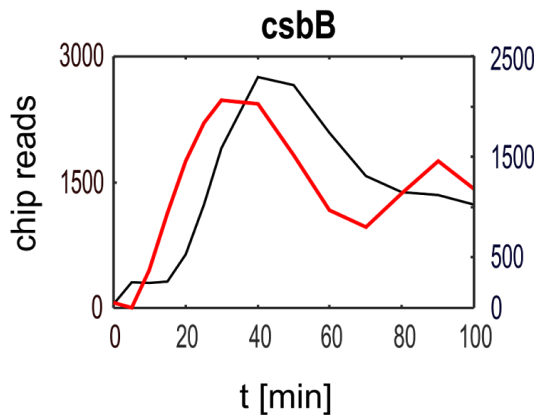
A**B****SigD****SigH****SigM****SigX**

Figure 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).

1.4.1. 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 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_S5). 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_S5, sheet sigM). For most of the large operons, we observed a phenomenon mentioned also for other regulons; i.e., the overall expression (parameter k_1) decreased with the distance from the promoter. This phenomenon was particularly apparent for the *dltA* operon.

1.4.2. 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*, *spollAB*, *spo0M*, *lytE*, *phrC*, *phrI*, *rapK*, and *spoVG*. Possible alternative regulators for the remaining genes are listed in Supplemental_Table_S5, sheet sigH.

1.4.3. 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 Figure 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 points (for an example see Figure 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_S5, sheet sigD). For most of the operons, the parameter k_1 decreased with the distance from promoter.

1.4.4. 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.

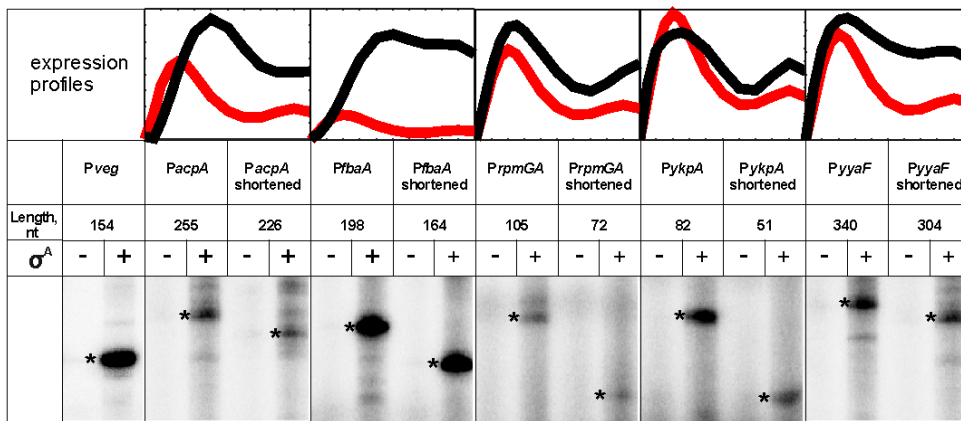
According to SubtiWiki, the SigX regulon is formed by 29 genes, including 6 operons. The kinetics of the SigX regulon genes are quite diverse, and a brief comparison with the SigX profile shows that many of the genes cannot be controlled by SigX; their peak of expression precedes that of SigX (see Figure 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_S5, sheet sigX).

1.5. 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_S1 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 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 (Figure 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.

A



B

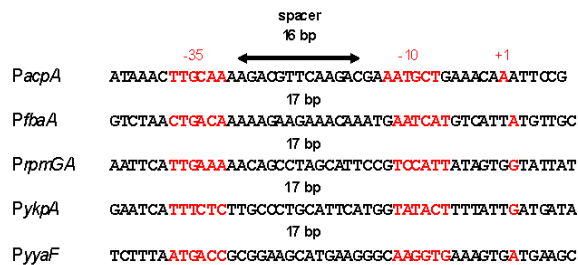


Figure 6. *In vitro* multiple round transcriptions of DNA fragments derived from upstream regions of genes predicted to be SigA-dependent. (A) Representative primary data. Each reaction was performed with the RNAP core and SigA-containing holoenzyme to demonstrate that the core had not been contaminated with sigma factors prior to its reconstitution with SigA. Radioactively labelled samples were loaded onto polyacrylamide gels. Pveg 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.

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 (Figure 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.

1.6. 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 (Table 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 Figure 7B. A sequence log, based on 43 known promoters for genes from Table 3 (those inconsistent with our kinetic analysis) is shown in Figure 7C. Both These logos resembles the M14 consensus sequence of SigA-dependent promoters as reported by Nicholas and colleagues [3] (Figure 7A). A sequence logo created from the five newly identified promoters (Figure 7C) is consistent with the M14 logo, but we note that the size of the sample was too small for reliable analysis.

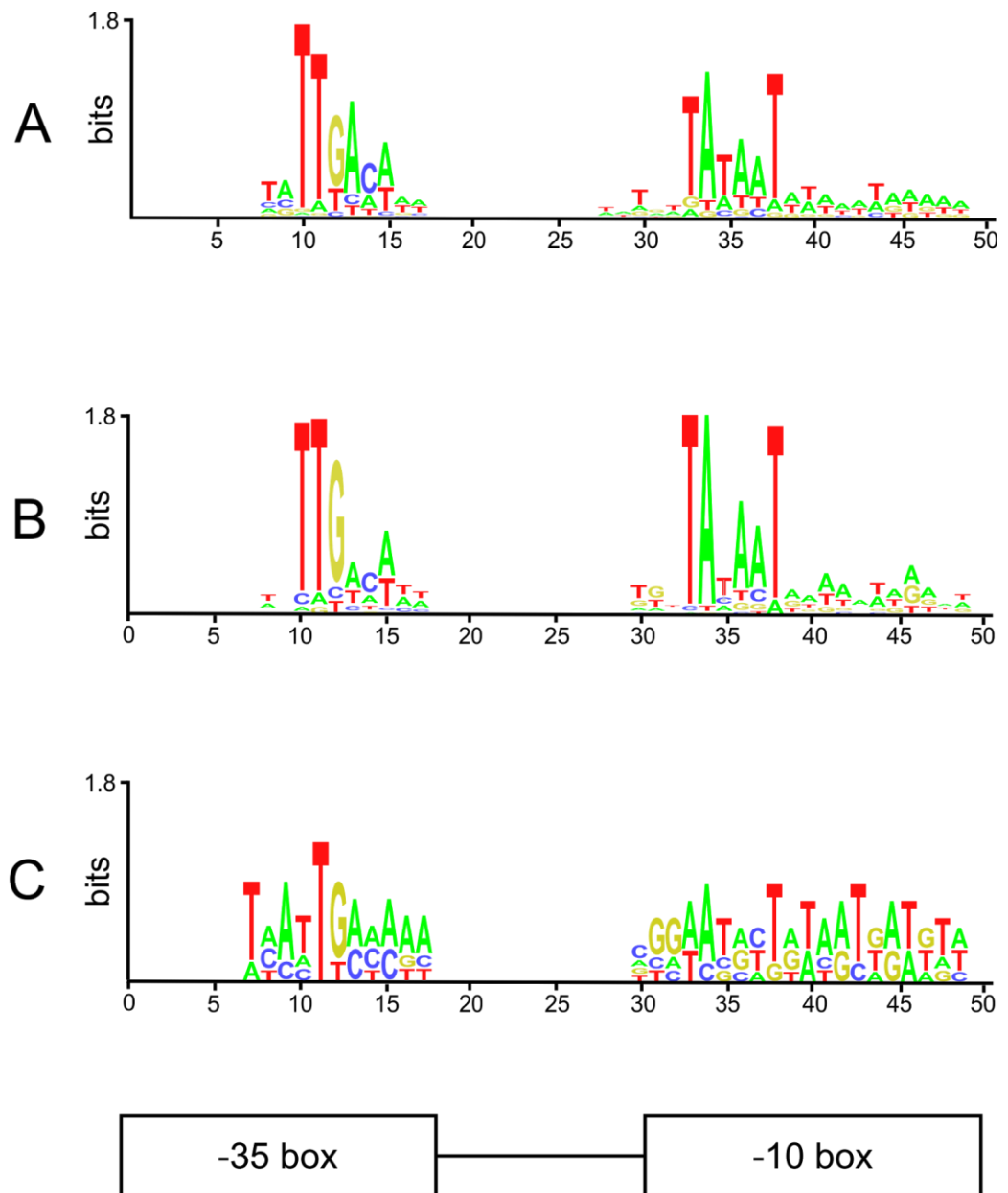


Figure 7. Comparisons of sequence logos. (A) The M14 motif of SigA-dependent promoters according to Nicolas and colleagues [3]. (B) The sequence motif of known SigA-dependent promoters verified by the kinetic model. The logo was created from sequences listed in Table

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 Figure 6.

2. 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 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 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 a 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 for 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 (Figures 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 (Figure 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/b subtilis-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.

3. 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 the 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).

References

- [1] M.S. Paget, Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution, *Biomolecules*, 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) D692-698.
- [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) D93-96.
- [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 Bioinformatics*, 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>, in, 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 *ypqQ* 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. Thomaidis, 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 , *Journal of Bacteriology*, 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 Microbiology Letters*, 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*, *Journal of bacteriology*, 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 σ^B -Dependent Genes in *Bacillus subtilis* Using a Promoter Consensus-Directed Search and Oligonucleo, *J .Bacteriol.*, 181 (1999) 5718-5724.
- [36] D. Höper, U. Völker, M. Hecker, Comprehensive Characterization of the Contribution of Individual σ^B -Dependent General Stress Genes to Stress Resistance of *Bacillus subtilis* Comprehensive Characterization of the Contribution of Individual σ^B -Dependent General Stress Genes to Stress R, 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, *Journal of Bacteriology*, 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, *Molecular Microbiology*, 28 (1998) 787-802.
- [40] C.W. Price, P. Fawcett, H. C  r  monie, N. Su, C.K. Murphy, P. Youngman, Genome-wide analysis of the general stress response in *Bacillus subtilis*., *Molecular microbiology*, 41 (2001) 757-774.
- [41] W. Eiamphungporn, J.D. Helmann, The *Bacillus subtilis* σ M regulon and its contribution to cell envelope stress responses, *Molecular Microbiology*, 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/microarray analysis (ROMA), and transcriptional profiling approaches., *Journal of molecular biology*, 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., *Molecular microbiology*, 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., *Molecular microbiology*, 32 (1999) 41-50.
- [45] H.L. Carter, 3rd, 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. M  rquez, M.J. Chamberlin, Cloning, sequencing, and disruption of the *Bacillus subtilis* sigma 28 gene., *Journal of Bacteriology*, 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, *Journal of bacteriology*, 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, e. al., The *Bacillus subtilis* sigma-x protein is an extracytoplasmic function sigma factor contributing to survival at high temperature, *J. Bacteriol.* 179:2915-2921, 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, *Molecular Systems Biology*, 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.