Systems level identification and kinetic modelling of SigQ-mediated control of germination in *Streptomyces coelicolor*

K. Smidova[1,2], J. Bobek[2,3], A. Zikova[1], M. Cerny[1], M. Schwarz[1], and J. Vohradsky[1]*

[1] Institute of Microbiology, Academy of Sciences of the Czech Republic, Laboratory of Bioinformatics, Vídeňská 1083, 142 20 Prague 4, Czech Republic.
[2] Institute of Immunology and Microbiology, First Faculty of Medicine, Charles University in Prague, Studničkova 7, 128 00 Praha 2, Czech Republic.
3 Chemistry Department, Faculty of Science, J. E. Purkinje University, 400 96 Ústí nad Labem, Czech Republic.
*corresponding author
e-mail:
klara.novotna@biomed.cas.cz
Jan.Bobek@lf1.cuni.cz
zikova@biomed.cas.cz
martin.cerny@biomed.cas.cz
marek.schwarz@biomed.cas.cz
vohr@biomed.cas.cz

ABSTRACT

The bacterial genus of *Streptomyces* undergoes a complex cell cycle that includes morphological and physiological differentiations. The first part of the cycle, spore germination, can be viewed as a transition from a dormant stage into metabolically active forms. Opposed to the further steps of the development, germination is still insufficiently studied; when only little is known about the gene expression accompanying the process, nothing is known how the expression is directed. The most potent regulators of gene expression are sigma factors, the separable subunits of the RNA polymerase. Our previous study revealed a set of sigma factors required during germination. However, the synthesis of SigQ accelerated to the highest level of all sigma factors. Here, using a chromosomal epitope-tag and the ChIP-seq method, we focused on identification of the SigQ-regulon. Exploiting a gene expression kinetic model, we found a surprising feature that the SigQ binding not only initiates expression of one gene set but as well may somehow prevent other genes from the transcription. Our data show that SigQ especially directs the nitrogen and carbohydrate metabolism, cell envelopes reconstruction, and is also required for the osmotic stress responses coinciding the germination.

INTRODUCTION

The developmental program of complex bacteria is directed by a number of transcription regulators that respond to the stage of growth or to changes in the environment. Differential transcription regulators direct developmental transitions by switching on or off sets of responding genes. The major control under the transcription patterns is provided by means of a repertoire of sigma factors (σ). The sigma factor is a separable subunit of RNA polymerase (RNAP) and its selective binding to appropriate promoter initiates transcription. The number of sigma factors encoded by bacterial genome is proportional to developmental complexity of the organism. A high degree of the developmental complexity in bacterial kingdom is represented by the genus of *Streptomyces*. *Streptomyces* are mycelial bacteria that inhabit competitive environments, mostly soil. While struggling for nutrient sources, they evolved huge capacities to produce various compounds, including antibiotics, for which they are highly valued {Chater, 1993 #1468}. A widely used model of streptomycetes is *S. coelicolor* whose genome is best annotated {Hopwood, 1995 #1470}. A deeply studied part of this organism's cell cycle is the aerial mycelium development and the process of sporulation due to the connection of these developmental stages with the production of antibiotics (for reviews see {Flardh, 2009 #1471}{Yague, 2013 #1280}). Under a nutrient depletion, the vegetative mycelium differentiates into unbranched aerial hyphae. The aerial hyphae are being covered by chaplins (hydrophobic aerial proteins) and the SapB peptide {Kodani, 2004 #1417}. This covering enables the hyphae to grow into the air {Claessen, 2006 #1281;Gebbink, 2005 #1418}. The development completes by a sporulation that encompasses a septation of the aerial hyphae, followed by a maturation of unicellular spores that enter a dormant state. This process requires a complex reconstruction of cells. The septa are formed within formerly multinucleoid hyphae and chromosomes are segregated to assemble chains of haploid spores. The spore wall thickens, due to the movement of the peptidoglycan synthesis from the apical tips to lateral spore walls, resembling the elongasome of the rod-shaped bacteria {Sigle, 2015 #1472}. Spores are sheathed, as the chaplin filaments are assembled by the rodlin proteins into so called rodlet envelope {Claessen, 2004 #1419;Claessen, 2002 #1420}. In the last step, cell hydrolases separate individual metabolically faint spores from the chain {Haiser, 2009 #1241}.

From this point of view, a subsequent germination takes place in an opposite way; it appears in a moisture environment and proceeds in a nutrient sufficiency. The germination could be characterized as an awakening from the dormancy into a vegetative growth. The compact spore rodlet sheath is destroyed and the formerly highly cross-linked peptidoglycan is modified by cellular hydrolases {Haiser, 2009 #1241}. This, together with a high osmotic pressure allows water influx that ensures renaturation of aggregated proteins back into their active conformation {Bobek, 2004 #1242}. Spores thus swell until the emergence of first germ tubes {Rioseras, 2016 #1453}. During the swelling, cellular metabolism increases during the first 150 minutes {Bobek, 2014 #1355} and is being adjusted in order to adequately respond to the nutrient composition or generally to the environmental sources {Bobek, 2004 #1422}. Protein chaperones are intensively synthesized to protect aggregated proteins and transcriptional and translational system is rapidly re-activated {Bobek, 2004 #1242}. First DNA replication proceeds afterwards, followed by the germ tube appearance {Ensign, 1978 #1464;Hardisson, 1978 #1436;Susstrunk, 1998 #936}. Nevertheless, despite the mentioned details and probably due to the fact that the streptomycete germination is not visibly connected with a

production of any valuable compound, the molecular control that governs the process is still poorly understood.

Our previous investigations described gene expression during germination. We observed that known global regulators are not expressed at high levels, but the expression timing was definite {Bobek, 2014 #1355}. Therefore, based on these measured data, we created a model of gene expression changes, which revealed fast changes in the expression of several sigma factors (HrdD, SigB, SigE, SigH, SigI, SigJ, SigR, SigW, and SigQ) suggesting for their importance in germination {Bobek, 2014 #1423}. Specifically, our model suggested a high activity for sigma factors HrdD, SigR, SigE, SigD, and SigH during the initial 30 mins of germination. Nevertheless, the absolutely highest induction (18x) revealed SigQ {Bobek, 2014 #1355}. SigQ (SCO4908) is an extracytoplasmic sigma factor (ECF) divergently transcribed from afsQ1/afsQ2 gene cluster and its only described role is to antagonize antibiotic production {Shu, 2009 #1293}. In this study we focused on the SigQ regulon. For this, we developed an epitope-tag insertional mutagenesis approach which allowed us to subject the tagged SigQ protein to ChIP-Seq analysis, as described in methods. The ChIP-seq results were complemented with a kinetic analysis of expression of the genes and operons found to be under the SigQ control, using previously measured time series of gene expression during germination {Strakova, 2012 #1232;Strakova, 2013 #1357}.

MATERIALS AND METHODS

Strains and growth conditions

Strains used in our work are *S. coelicolor* M145, *E. coli* K-12 MG1655 {Datsenko, 2000 #1320} and derivatives from GM2929 {MacNeil, 1992 #1323}. *E. coli* BW25113/pIJ790 has λ Red recombination system under the control of arabinose inducible promoter and this strain was used to propagate S. coelicolor cosmid. *E. coli* ET12567/pUZ8002 is methylation-deficient strain for intergeneric conjugation with *S. coelicolor*. For the preparation of the epitope tagged mutant strain, *S. coelicolor* was cultivated on solid agar plates with MS medium (2% (w/v) mannitol, 2% (w/v) soya flour, 2% (w/v) bacterial agar in tap water) or DNA medium (2,3% (w/v) Difco nutrient agar){Kieser, 2000 #1231}. Apramycin (50 ug/ml), chloramphenicol (25 ug/ml), kanamycin (50 ug/ml) or nalidixic acid (25 ug/ml) was added to the media when needed. The list of genetic material used is given in Table 1.

Table 1: Strains, plasmids, cosmids and oligonucleotides used in this study

Strains	Genotype/comments	Reference/source
S. coelicolor		
M145 WT	SCP1- SCP2-	Bentley et al., 2002{Bentley, 2002 #745}
M145 sigQ-HA	HA tagged *sigQ* mutant strain:: *apr oriT* cassette	This study
E. coli		
ET12567	dam, dcm, hsdM, hsdS, hsdR, cat, tet	Flett et al. 1997 {Flett, 1997 #1322}

		MacNeil et al., 1992 {MacNeil, 1992 #1323}
BW25113	K-12 derivative; Δ(*araD-araB*)567 Δ *lacZ4787*(::*rrnB-4*) *lacIp-4000*(*lacI^Q*), l-*rpoS369*(*Am*) *rph-1* K-12 derivative; Δ(*araD-araB*)567 Δ *lacZ4787*(::*rrnB-4*) *lacIp-4000*(*lacI^Q*), l-*rpoS369*(*Am*) *rph-1*	Datsenko and Wanner, 2000 {Datsenko, 2000 #1320}
Plasmids/cosmids		
pIJ773	P1-FRT-oriT-aac(3)IV-FRT-P2 (plasmid template for amplification of the *apr oriT* cassette for 'Redirect' PCR-targeting)	Gust *et al.*, 2003 {Gust, 2003 #1416}
pIJ790	λ-RED (gam, bet, exo), cat, araC, rep101^{ts}	Gust *et al.* 2003 {Gust, 2003 #1416}
pUZ8002	tra, neo, RP4	Paget et al., 1999 {Paget, 1999 #1248}
2StK8	carb, kan	cosmid library University of Wales Swansea
2StK8 sigQ-HA	carb, kan, HA tagged sigQ mutant strain:: *apr oriT* cassette	This study
Oligonucleotides		
SigQ_HAtag_left	GGCCGTGACCAGGAGGACGCCTTCGGCGAACTGGTCGCCTACCCGTACGACGTCCCGGACTACGCCTGAGAAGTTCCCGCCAGCCTCGC	
SigQ_HAtag_right	AGCCCCGCCCCCTCGGGCCGGTCCCCCCGGCCCGTGCCTGGAATAGGAACTTATGAGC	
SigQ_CTRL_left	GCTGGTCCTGCGCTACTACG	
SigQ_CTRL_right	AGTCCCGCTCCCTTTCCAC	

Construction of epitope tagged mutant strains

To avoid the construction of a deletion strain supplemented by the epitope-tagged gene on a separate plasmid, we modified the mutagenesis procedure {Gust, 2004 #1465} in order to insert the epitope tag directly to the gene within its native site on the chromosome. Primers containing the HA tag were chemically synthesized. The nucleic acid sequence of the HA tag (YPYDVPDYA) was optimized for the codon usage in *S. coelicolor* (TAC CCG TAC GAT GTG CCG GAT TAC GCG). A gene cassette containing FRT flanking regions, apramycin resistance marker and oriT was amplified from plasmid pIJ773 as described {Gust, 2004 #1465}, cut by EcoRI and HindIII restriction enzymes and used as a PCR template. The PCR was performed to a final volume of 50 ul containing 2,5 U iProof DNA polymerase (Bio-Rad), 5 μl GC buffer, 2,5 μl 100% DMSO, 1 μl 10mM dNTPs (50 uM each), 0,5 μl template DNA (50 ng pIJ773 cut by EcoRI and HindIII, gel purified) and 0,5 ul each primer for mutagenesis (50 pmol each, SigQ_HAtag_left, SigQ_HAtag_right). PCR cycle conditions were 10 cycles with 45 sec denaturation at

98°C, 45 sec annealing at 50°C, 90 sec extension at 72°C, 15 cycles with 45 sec denaturation at 98°C, 45 sec annealing at 50°C, 90 sec extension at 72°C. PCR product was then gel purified.

Purified PCR product was then electroporated into *E. coli* BW25113/pIJ790 containing *S. coelicolor* cosmid 2StK8. The cells were then cultivated at 37°C for 1 hour in 1 ml LB. The culture was centrifuged 15 sec, 10.000 g and spread onto LB agar with apramycin (50 µg/ml). The cosmid with inserted cassette was then transformed into methylation-deficient *E. coli* ET12567/pUZ8002 and the resulting strain was conjugated with *S. coelicolor* M145 {Flett, 1997 #1322}. Final mutants were selected on MS medium containing apramycin, kanamycin and nalidixic acid {Kieser, 2000 #1231}. Double cross-over exconjugants were selected. Chromosomal DNA was then isolated and cassette integration into the chromosome was confirmed by PCR and sequencing.

Spore germination
The details regarding *S. coelicolor* M145 spore cultivation and growth were published in our previous work {Strakova, 2013 #1357}. Briefly, matured spores were harvested after 14 days of growth in 28°C on solid agar plates (0.4% (w/v) yeast extract, 1% (w/v) malt extract, 0.4% (w/v) glucose, 2.5% (w/v) bacterial agar, pH 7.2) overlaid with cellophane discs. For the germination, spores were activated by mechanical disruption of the outer coat and 10 min heat shocked in 50°C to boost synchronicity. The germination course proceeded in liquid AM medium (20 amino acids at 0.2 mM, 20 mM KH_2PO4, 30 mM Na_2HPO4, 2% (w/v) glucose, 0.05% (w/v) $MgCl_2$, 0.5 mM $CaCl_2$, 7 mM KCl, and a mixture of nucleic acids, each 0.01% (w/v)) at 37°C. Spores were cultivated to the 30th min when the *sigQ* reaches the highest expression level during germination, as was found in our previous work {Strakova, 2013 #1357}.

Chromatin immunoprecipitation
The epitope-tagged streptomycete spores were grown in 80 ml AM medium at 37°C. After 30 min of cultivation, spores were crosslinked by adding formaldehyde (final 1%) at room temperature for 10 min. Then 125 mM glycine was added and culture was incubated at room temperature for other 5 min. Spores were centrifuged at 6000 g, 5 min at 4°C, washed three times with 50 ml ice cold phosphate-buffered saline (pH 7,4) and freezed in liquid nitrogen. The washed spores were resuspended in 2 ml lysis buffer (50 mM Tris-HCl, pH8; 150 mM sodium chloride, 1% Triton X100, 0,5% sodium deoxycholate, 0,1% SDS, 1mM PMSF, protease inhibitors (Roche, 4ul/ml) with zirconium

beads (1g/1.5ml tube) and disrupted in FastPrep-24 machine (Biomedicals) at 4°C, speed 5,5, 12x45 sec. The lysate was then centrifuged at 14000g for 10 min in 4°C. 20 ul protein A/G PLUS-agarose (Santa Cruz) was pre-equilibrated in RIPA buffer (Santa Cruz) 10 ul (2 ug) of mouse monoclonal anti-HA antibody (clone F-7, Santa Cruz) was added to the pre-equilibrated Protein A/G PLUS-agarose and incubated at 4°C for 3 hours on rotator. For the non-specific control and also for the sample, 1 ug IgG mouse serum (Sigma) was added to the cell extract. 20 ul protein A/G PLUS agarose was added to the non-specific control and 20 ul protein A/G PLUS agarose with bound anti-HA antibody was added to the sample. Both were incubated overnight at 4°C on rotator. Sample was treated three times with Lysis buffer RIPA (Santa Cruz), four-times with wash buffer (100 mM Tris-HCl, pH8,5; 0,5 M LiCl, 1% Triton X100, 1% sodium deoxycholate, twice with Lysis buffer RIPA (Santa Cruz) and twice with TE buffer (10 mM Tris-HCl, pH8,0; 1 mM EDTA). Sample was resuspended in 30 ul elution buffer (10 mM Tris-Cl, pH 8,0; 10 mM EDTA, and 1% SDS) and incubated for 15 min at 65°C. Then NaCl (200 mM final) was added to the sample and incubated for 30 min at 65 °C for reverse cross-linking. Beads were removed by centrifugation at 14.000 g for 5 min. Proteins were removed from the sample by incubation with 10 ug of Proteinase K (Roche) for 3 hours at 65 °C. The sample was then purified by NucleoSpin gDNA Clean-up (MACHEREY-NAGEL). The DNA size was checked by Bioanalyzer system (Agilent) if the fragments were between 75-1000 bp with median size around 250 bp. If not, than additional sonication was included. Samples and controls were prepared in triplicates.

Western blotting
Streptomycete cells were cultivated 48 hours in 2YT medium at 30°C . When needed, 30 minutes prior to the termination of cultivation, the medium was supplemented by 5M NaCl or 30mM $(NH_4)_2SO_4$ to induce *sigQ* expression. For the protein extracts, cells were resuspended in lysis buffer with zirconium beads and disrupted in FastPrep-24 machine (Biomedicals) at 4°C, speed 5,5, 12x45 sec. The lysate was then centrifuged at 14.000g for 20 min at 4°C. Concentration of proteins was measured using Pierce BCA Protein assay kit (Thermo Scientific). Load buffer (1:4) was then added to the lysates or immunoprecipitated samples and boiled for 4 min. Samples were separated on SDS-PAGE (NuPAGE Bis-Tris Mini Gels, MES SDS Running buffer (Novex), 200 V, 35 min) and then transferred to the PVDF membrane (blotting buffer 48 mM Tris, 39 mM glycine, 0.0375 % SDS, 20 % methanol, 360 mA, 30 min). The anti-HA antibody conjugated to peroxidase (High Affinity Anti-HA-Peroxidase, 3F10, Roche) was used to probe HA-tagged sigma factor and detected by chemiluminescence using SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific).

Sequence analysis
Next generation sequencing was performed on the Illumina MiSeq platform in a paired-end 2x350 bp run. Illumina reads were trimmed by removing adapters and leading/trailing low quality or N bases with Trimmomatic (v 0.36) {Bolger, 2014 #1409}. Quality control of the reads were checked using FastQC and Qualimap (v2.2){Okonechnikov, 2016 #1411}. Filtered reads were mapped to the

6

S.coelicolor reference genome (GenBank: NC_003888) with Gsnap program {Wu, 2010 #1335}. The SAM files from three replicates and three controls were assembled and analyzed using MACS v1.4 to detect read enriched regions in the genome {Feng, 2012 #1321}. Enriched regions were selected for further study based on *p*-value≤0.05. For the visualization of the reads we used Tablet (1.15.09.1){Milne, 2010 #1412}.

DNA microarray and data processing

Gene expression data were processed as in our previous paper {Strakova, 2013 #1357}. The data preprocessing steps are repeated here to make clear how the values used for the analysis in this article were obtained. RNA quality control and gene expression levels were performed by Oxford Gene Technology (Oxford, UK) using Agilent DNA microarrays covering the entire S. coelicolor genome and the standard Bacterial RNA amplification protocol for two-channel assays by OGT.

The data were normalized using LOWESS and filtered for background and flag information (from Agilent documentation) in the GeneSpring software to obtain genes that were expressed significantly above background and to avoid side effects of possible cross hybridization. These methods reduced the number of entities on a single array from 43888 to 25312, which finally represented the outcome for 7115 genes out of 7825. The data discussed in this publication have been deposited in the NCBI Gene Expression Omnibus and are accessible using the GEO Series accession number GSE44415 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE 44415).

Array normalization

The experiment included 37 arrays from 13 distinct time points during S. coelicolor germination. The arrays shared a common reference in the red channel (Cy5), which consisted of a mixture of RNA samples from all examined time points. The distributions of Log2Ratio values (Log2Ratio = log2 (Sample (Cy3)/Reference (Cy5))) for all samples were scattered around a common mean and all had similar variance. Therefore, the distributions for each array were centered so that the medians and the median absolute deviations of all the array distributions were equal. To eliminate array outliers, we filtered out the 0.02 quantile of the least and the most intensive Log2Ratio values. Normalized Log2Ratios were exponentiated to return the values to the original. The outliers among gene replicates at individual time points were filtered using the Q-test (for 3-9 inputs) and the Pierce test (for > 10 inputs).

Kinetic model of gene expression

We build upon our previous work on an ordinary differential equation (ODE) model for gene regulation, inspired by the neural network formalism {Vohradsky, 2001 #571}. In this model the synthesis of new mRNA for a gene *z* controlled by set of *m* regulators $y_1,..,y_m$ (genes or any other regulatory influence) is determined by sigmoid function *f((t))* of the regulatory input $\rho(t) = \sum_{j=1..m} w_j y_j(t) + b$ $\rho(t) = \sum_{j=1..m} w_j y_j(t) + b$. Here w_j is the relative weight of regulator y_j, *b* is bias (inversely related to the regulatory influence that saturates the synthesis of the mRNA). In our case, *f* is the logistic function *f(x) = 1/(1 + e⁻ˣ)*. The transcript level of *z* is then governed by the ODE:

$$\frac{dz}{dt} = k_1 f(\rho) - k_2 z \tag{1}$$

where k_1 is related to the maximal level of mRNA synthesis and k_2 represents the decay rate of the mRNA. Both k_1 and k_2 must be positive. The complete set of parameters for this model is $= [k_1, k_2, b, w_1, , w_m]$. Given N samples from a time series of gene expression taken at time points $t_1, , t_N$ the inference task can be formalized as finding $\hat{\beta}\;\hat{\beta}$ that minimizes squared error with regularization:

$$\hat{\beta} = \operatorname*{argmin}_{\beta}\left[\sum_{i=1}^{N}\left(\hat{z}_\beta(t_i) - z(t_i)\right)^2 + r(\beta)\right] \tag{2}$$

Here z is the observed expression profile, $\hat{z}_\beta\;\hat{z}_\beta$ the solution to (Eq. 1) given the parameter values and the observed expression of $y_1,..,y_m$, and $r(\ )$ is the regularization term. The regularization term represents a prior probability distribution over that gives preference to biologically interpretable values for . The regularization is discussed in more detail in the supplementary material (supplementary file 1). Assuming Gaussian noise in the expression data, (Eq. 2) is the MAP estimate of .

Our approach is similar to the Inferelator algorithm{Bonneau, 2006 #1407}, although there are important differences: the Inferelator does not model decay (k_2) – it assumes decay is always one. Further, Inferelator minimizes the error of the predicted derivative of the expression profile, while we minimize the prediction error for the actual expression profile.

Since the expression data is noisy, we have smoothed the data prior to computation with a piecewise cubic spline with 6 knots (the best number of knots was determined empirically). By smoothing we get more robust results with respect to low frequency phenomena, but sacrifice our ability to discover high-frequency changes and regulations (oscillations with frequency comparable to the measurement interval are mostly suppressed). Further our experiments with fitting raw data or tight interpolations of the data (e.g. B-spline with knots at all measurement points) have had little success in fitting even the profiles that were highly correlated, due to the amplified noise in the data. Approaches to GRN inference using smoothed data have been described in the literature {Berrones, 2016 #1353}.

Further advantage of smoothing is that it lets us subsample the fitted curve at arbitrary resolution. We have subsampled the profiles at 1-minute time steps (360 points) which allowed us to integrate Eq. 1 accurately with the computationally cheap Euler method, making evaluation of the error function fast and easy to implement in OpenCL.

We minimized Eq. 2 by simulated annealing. For each gene and candidate regulator set we execute 128 runs with different initial parameter values. To increase speed, the inference algorithm was implemented for parallel execution on both CPU and GPU architectures with OpenCL. On a mid-tier GPU (NVidia GTX 960), the system found optimal parameters for 300 gene-regulator pairs in 11 seconds. The inference algorithm source with all data and scripts necessary to reproduce the results in this paper can be obtained upon request. The algorithm can be invoked from Matlab and R, Cytoscape integration is being developed.

For the task at hand, we have first tested single regulator models for all genes in the SigQ regulon. Since SigQ has three major peaks in the collected data, and it is plausible that those peaks correspond to different phases of germination with possibly different regulatory interactions. We have therefore also tried to fit the single regulator models on subset of the data containing only one or two of the peaks.

RESULTS

Epitope tagging and SigQ induction

The eligibility of the epitope tagging approach was proven by western blot (Figure 1.). The SigQ-HA strain was cultivated for 48h because the SigQ from germination was undetectable by this method. The original wild type strain lacking the HA-tag was used as a negative control and the immunoprecipitated SigQ-HA protein as a positive control. Cell-free protein extract from the SigQ-HA strain revealed a band of approximately 23,5 kDa (corresponding to 22,5 kDa of the SigQ protein + 1,07 kDa of the HA tag). To see whether expression of SigQ is induced by osmotic or nitrogen stress, the cultivation medium was for the last 30 minutes supplemented by 0.5M NaCl or 30mM $(NH_4)_2SO_4$. After this, the SigQ level highly increased in the presence of ammonium sulfate and even more increased in NaCl. These data suggest that the SigQ sigma factor is another alternative sigma factor of streptomycetes that controls cellular stress responses. As the osmotic stress is concomitant to germination, the suggested role of SigQ in the osmotic stress response may explain the 18x increase in its expression during germination.
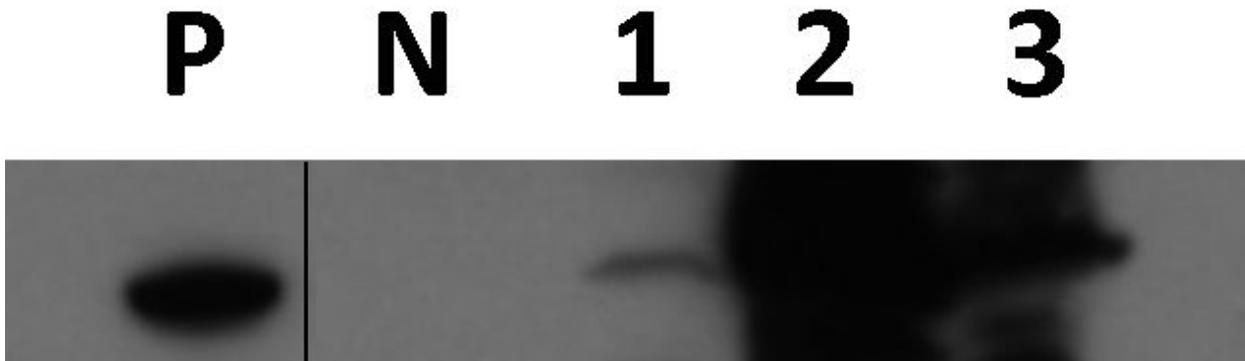
Figure 1. Control and SigQ-induction western blot. In order to prove eligibility of the method, western blot was performed using anti-HA antibody. Cells were cultivated 48h in 2YT medium. P - precipitated SigQ-HA protein as a positive control; N - protein extract from the wild type strain as a negative control; 1 - protein extract from the SigQ-HA strain; 2 - protein extract from the SigQ-HA strain induced by NaCl (0.5M final); 3 - protein extract from the SigQ-HA strain induced by (NH4)2SO4 (30mM final).

Identification of SigQ binding regions by ChIP-seq

Genome wide analysis of the direct interaction between sigma factor SigQ and its cognate DNA had not been described before. To study the SigQ binding on the *S. coelicolor* genome, we tagged SigQ by

HA tag. We constructed an HA-tagged SigQ strain using epitope-tagging mutagenesis. Resulting cells were cultivated in the defined AM medium. The reason was to keep consistency with our previous gene expression analyses of germinating spores that had been done in the same medium {Strakova, 2012 #1232;Strakova, 2013 #1357;Strakova, 2014 #1300}. Then, to determine *in vivo* SigQ binding regions we performed next-generation sequencing. Using a bioinformatics pipeline (see Materials and Methods), we detected 130 binding regions.

For the peak summits located in ≤300 bp upstream of an annotated start codon, SigQ was considered to regulate the respective downstream gene. Using this criterion, we identified 260 direct target genes. We also incorporated genes of putative operons predicted according to operon prediction approach as published in {Castro-Melchor, 2010 #1197}. This approach identified additional 66 genes. Altogether our dataset contained 326 genes.

Regulatory function of SigQ from kinetic modelling

Time series of gene expression for the genes identified as potential targets of SigQ by ChIP-seq were used together with the expression profile of SigQ for modeling of incident gene expression kinetics. This analysis revealed possibility of regulatory influence of SigQ on the identified target genes and also mode of its action (activation/repression). From the original number (326 genes), 25 expression curves had either very low response to the expression pattern of SigQ or were too flat; among them, 3 gene expression profiles were not identified in the expression time series dataset. These genes have been discarded from the analysis. Remaining 301 genes found by ChIP-seq and operon analysis were subjected to the kinetic modelling.

Target gene expression profile clustering

Expression profiles of the selected genes were normalized to have zero mean and the same standard deviation and the normalized profiles were clustered (see Figure 2) using k-means scheme for 10 clusters (number of clusters was determined empirically).
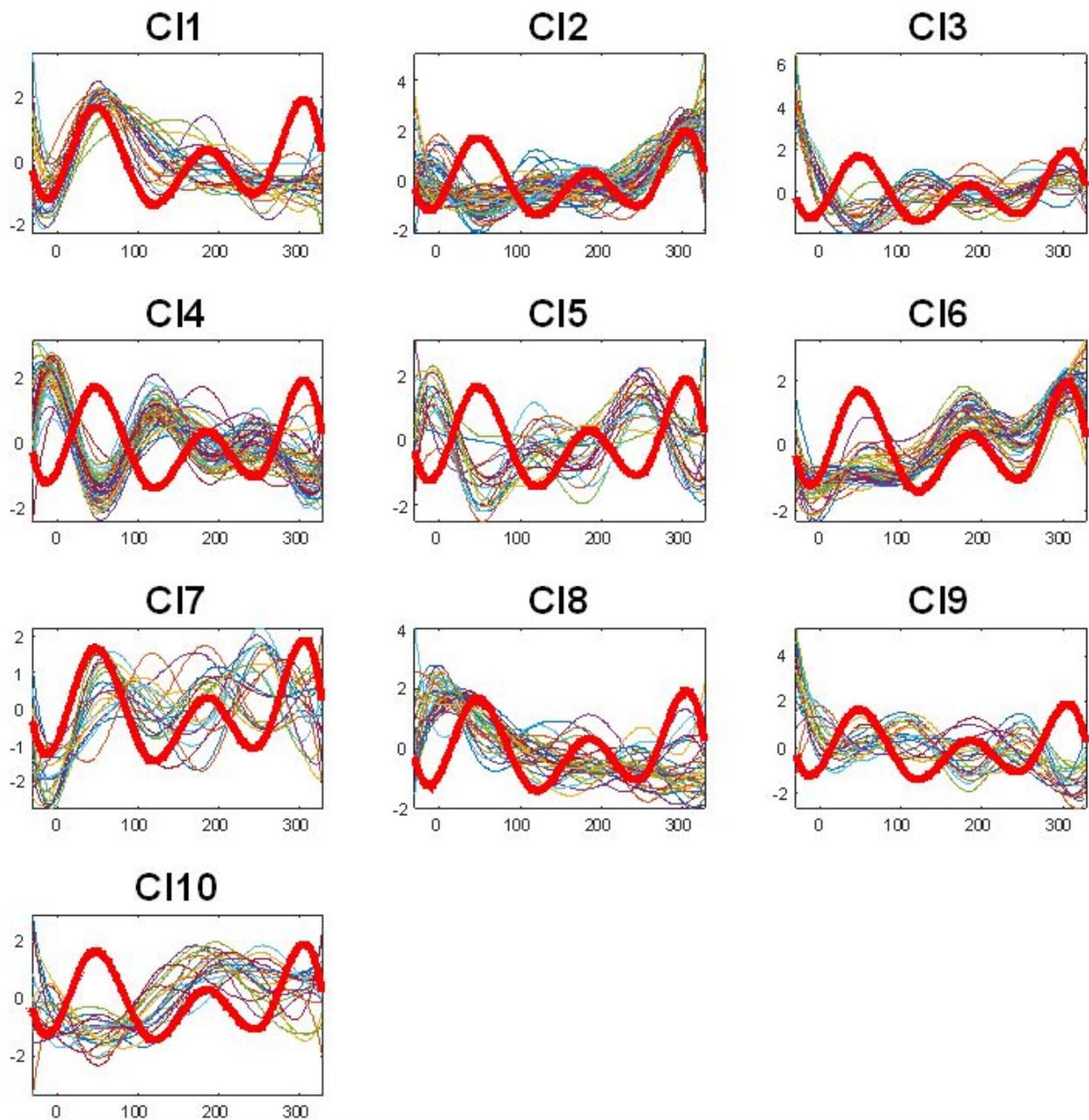
Figure 2. Gene expression profiles of the SigQ regulon clustered according to similarity of their expression profiles with superimposed profile of *sigQ* (red). For the clustering purposes, the profiles were smoothed with a piecewise cubic spline with 6 knots and normalized to have zero mean and the same standard deviation.

The clustering showed two principal trends; 1. profiles that were at some interval correlated with the expression profile of *sigQ* (CL1,6,10) and, 2. profiles inversely correlated with the *sigQ* profile (CL3,4,5,9), suggesting that, for a large part of the SigQ regulon, SigQ acts as a repressor. Figure 2 also shows that some of the profiles were correlated (positively or negatively) only in a certain interval, e.g

CL6,10 was positively correlated with *sigQ* profile after ca 120 min., or CL3 was negatively correlated with the *sigQ* profile consistently during the first ca 150 minutes. Some, as CL8, were not correlated with *sigQ* at all. This observation led us to dissect the expression profiles into three parts spanning roughly in the range <120 min., 120-250 min., and >250 min.

Each individual profile, or profile segment (as defined above), were then modeled according to the scheme described in Methods, searching for parameters of the model for whole profile and for the profiles in the three above mentioned intervals. Results are summarized in supplementary File 2. Modeling results as given in supplementary File 2, were classified into four categories "Full" – regulated gene expression profile was successfully modeled with SigQ as regulator in whole measured interval, "Weak" – modeling was successful in only a part of the interval, "No" – SigQ could not model the given profile, and "Inconsistent" – the modeling results were inconsistent, i.e. their segments were modeled with SigQ as both positive and negative regulator, therefore such gene expression profiles were considered as not controlled by SigQ. Excluding those profiles that could not be modeled, we got altogether 118 genes that could be negatively regulated by SigQ (while 44 were in category Weak), and 94 that could be positively regulated (58 were in category Weak). 89 gene expression profiles could not be modeled by SigQ (categories No and Inconsistent). This results suggest that SigQ acts in majority of cases as repressor, when the ratio of successfully fitted positively/negatively regulated genes was 74/36, i.e double of negatively regulated genes. Supplementary file 2 also contains parameters of the model that allow to simulate expression kinetics under the SigQ control.

2-regulator model

For exploratory purposes, the remaining genes that were not fitted at all or were fitted only on a subset of the data were then tested for interactions with alternative sigma factors that could be active during germination {Strakova, 2014 #1300} (158 genes, 44 sigma factors, 6951 combinations). We tested both 2-regulator models (SigQ + alternative regulator) and regulation solely by the alternative regulator. The fits were then annotated by computer to find plausible regulations. A 2-regulator model was deemed plausible only if it provided significantly better fit than the corresponding model with the other regulator only. Of the 6591 combinations, 869 resulted in a plausible fit without SigQ and further 1173 gave plausible fit for SigQ + alternative regulator. The large number of plausible fits is probably a consequence of the large number of sigma factors tested together with the limited number of data points and we expect them to be mostly false positives. We are however confident that the ratio of false negatives is minimal. The list of plausible fits with alternative regulators is given in the supplementary file 4.

Functional classification of the SigQ regulon

Genes of the SigQ regulon were further characterized by classification into functional groups according to The Sanger Institute database or KEGG database (http://www.genome.jp/kegg/pathway.html){Kanehisa, 2000 #1121} (supplementary file 3). We identified SigQ bound regulatory regions, that included ribosomal proteins, dehydrogenases, ABC transporters, aminotransferases, ATP/GTP binding proteins, cold shock proteins, components of

cytochrome, helicases, elongation proteins, transcriptional regulators, hydrolases, hypothetical proteins, chaperonins, integral membrane transport proteins, lipoproteins, oxidoreductases, regulatory proteins, sigma factors, sugar transporters, amino acid synthases, two component system proteins, transmembrane proteins.

Tables of genes and their functional and pathway annotations, found to be controlled by SigQ, were too large to be placed into the text of the manuscript and were thus placed to supplementary materials (supplementary file 3). A summary of the tables results is given below.

Genes positively regulated by SigQ

We identified 90 genes falling into positive regulation pattern, where the SigQ played a role as an activator of gene expression. Most of the genes in this group were formed by unknown (18 genes) and not classified genes (6 genes). Second largest group was formed by genes from group 4.1.0 Periplasmic/exported/lipoproteins (17 genes), namely hypothetical proteins (SCO2067, 2095, 2096, 2347, 2622, 2960, 2978, 4471, 4472, 4474, 5540, 5623, 5628, and 5751) and alkaline phosphatase (SCO2068), small membrane protein (SCO1634) and lipoprotein (SCO3107).

Third largest group was formed by genes of group 3.1.0 Amino acid biosynthesis (11 genes) formed by glutamine synthetase (SCO2241, 2198), glutamate synthase (SCO2025) and threonine synthase (SCO5355) followed by imidazole glycerol phosphate synthase genes (SCO2048,2051), imidazole glycerol phosphate dehydratase (SCO2052). The rest formed the genes: phosphoribosyl isomerase A (SCO2050), histidinol-phosphate aminotransferase (SCO2053), diaminopimelate decarboxylase (SCO5353) and homoserine dehydrogenase (SCO5354).

Nine genes were assigned to group 2.2.0 Macromolecule synthesis, modification, formed by peptidyl-prolyl cis-trans isomerase (SCO1638), leucyl-tRNA synthetase (SCO2571), DEAD/DEAH box helicase (SCO3732), SCO4318, elongation factor Tu (SCO4662), tRNA (guanine-N(1)-)-methyltransferase (SCO5594), elongation factor Ts (SCO5625), ribosome recycling factor (SCO5627), and peptidyl-prolyl cis-trans isomerase (SCO7510).

Four genes were assigned to the group 3.2.0 Biosynthesis of cofactors, carriers, namely molybdenum cofactor biosynthesis protein A (SCO1821), cytochrome biogenesis-like protein (SCO4473, 4475) and polyprenyl diphosphate synthase (SCO4583). Next group was formed by genes of 3.5.0 Energy metabolism, carbon with 4 genes - dihydrolipoamide dehydrogenase (SCO2180), dihydrolipoamide succinyltransferase (SCO2181), type II citrate synthase (SCO2736) and 6-phosphogluconate dehydrogenase (SCO6658).

Group 1.5.0 Transport/binding proteins comprised 3 genes – 2 ABC transporters (SCO3947 and SCO5392) and transport system kinase (SCO5400). Three genes were involved in groups 3.4.0 Degradation of small molecules (rhamnose kinase SCO814, acetyl-CoA acetyltransferase (SCO5399) and tagatose 6-phosphate kinase (SCO5848). Group 4.2.0 Ribosome constituents comprised 3 genes - 30S ribosomal protein S16 (SCO5591), 30S ribosomal protein S2 (SCO5624) and 16S rRNA-processing protein RimM (SCO5593).

Two genes were in group 3.8.0 Secondary metabolism (subtilisin-like protease SCO1824, AgaS protein SCO5849) and in group 6.5.0. Others, including regulatory proteins, regulatory protein SCO2094 and regulatory protein SCO6169 were present.

Remaining functional groups constituted only one gene - 1.4.0 Protection responses (penicillin acylase SCO3184), 1.6.0 Adaptation (acetyltransferase SCO1864), 1.8.1 Differentiation/sporulation (SCO1772), 3.3.0 Central intermediary metabolism (uridylate kinase SCO5626), 3.6.0   Fatty acid biosynthesis (3-hydroxybutyryl-CoA dehydrogenase SCO1591), 5.1.0 Laterally acquirred elements (SCO5902), 6.2.0 RNA polymerase core enzyme binding (SCO3548) and 1.3.1 Chaperones (SCO4761).

Metabolic pathways positively regulated by SigQ

Only 29 genes of 90 positively regulated genes were identified within the genes annotated in KEGG pathways database. We identified two major positively regulated pathways – amino acid metabolism and carbohydrate metabolism.

Amino acid metabolism

The largest metabolic pathway positively regulated by SigQ was the amino acid metabolism (28 genes). It is probably caused by the cultivation in amino acid rich medium. From this finding we can deduce, that one of the main functions of SigQ during spore germination might be a positive regulation of amino acid metabolism, mainly metabolism of histidine, glycine, serine, threonine.

Carbohydrate metabolism

The second largest pathway potentially controlled by SigQ (21 genes) was carbohydrate metabolism. Mainly glyoxylate and dicarboxylate metabolism (6 genes), citrate cycle and propanoate metabolism (3 genes) and others like pyruvate metabolism, butanoate metabolism, glycolysis/gluconeogenesis, pentose phosphate pathway, pentose and glucuronate interconversion, fructose and mannose metabolism, galactose metabolism.

Other positively regulated metabolic pathways were pathways of energy metabolism, signal transduction biosynthesis of secondary metabolites lipid metabolism, metabolism of cofactors and vitamins ribosome, xenobiotics biodegradation and metabolism membrane transport metabolism of terpenoids and polyketides and in small numbers folding, sorting and degradation pathway, metabolism of other amino acids and nucleotide metabolism

Genes negatively regulated by SigQ

We found 110 genes negatively regulated by SigQ at least at one of the three intervals defined in Methods, where SigQ played probably a role as a repressor of gene expression. Altogether 46 genes fell to the non-annotated groups of unknown function or not classified (0.0.0 and 7.0.0),

Largest annotated group was the group 4.1.0. Periplasmic/exported/lipoproteins (22 genes), most of them formed by hypothetical proteins (SCO1590, 1823, 2199, 2348, 2454, 2572, 2621, 4002, 4128, 4129, 4511, 4515, 4610, 4995, 5204, 5461, 5823, 6573, 7191, 7224), neuraminidase (SCO0033) and D-alanine--D-lactate ligase (SCO3595).

17 genes were assigned to the group 1.5.0 Transport/binding, including genes of ABC transporter (SCO2258), ABC transporter ATP-binding protein (SCO4963, 6062), ABC transporter permease (SCO6063, 6064), integral membrane and transmembrane transport proteins (SCO1822, 2979, 4964,

5827), integral membrane permease (SCO0067) and sugar transport proteins (SCO1056, 1058, 6229). Remaining were L-lactate permease (SCO3029), solute binding protein (SCO6569), ammonium transporter (SCO5583) and BCCT family transporter (SCO6739).

Five genes were assigned to group 6.3.0. Defined families, namely LuxR family- (SCO6334), GntR family- (SCO3264), MarR family- (SCO7727) and TetR family transcriptional regulators (SCO1699, SCO1702). Three genes fell into groups 2.1.0. Macromolecular degradation (protease SCO2920, D-alanine:D-alanine dipeptidase SCO3596, and heat shock protein HtpX SCO4609) and 3.5.0. Energy metabolism (cytochrome oxidase subunit II SCO3946, D-lactate dehydrogenase SCO3594, and glyceraldehyde 3-phosphate dehydrogenase SCO7511).

Two genes were found in the groups, 3.1.0 Amino acid biosynthesis (methylmalonic acid semialdehyde dehydrogenase SCO2726, and cysteine synthase SCO0992), 6.1.0 Two component system (two-component sensor SCO5824 and two-component sensor kinase SCO7089), 6.2.0 RNA polymerase core enzyme binding (ECF sigma factors SCO4996 and sigma factor SCO7192), 6.5.0 Others (nitrogen regulatory protein P-II SCO 5584 and DNA binding protein SCO6571), 3.8.0 Secondary metabolism (transcriptional regulator cdaR SCO3217, that was previously found to be regulated by SigQ (33)) and hydrolase SCO0878.

Single genes were assigned to groups 2.2.0 Macromolecule synthesis, modification (helicase SCO3550, 1.3.1 Chaperones (SCO5210), 3.2.0 Biosynthesis of cofactors,carriers (NAD(+) synthase (glutamine-hydrolysing), SCO2238), 3.3.0 Central intermediary metabolism (membrane-bound proton-translocating pyrophosphatase, SCO3547).

Metabolic pathways negatively regulated by SigQ

19 genes out of 110 negatively regulated genes were assigned to KEGG pathways. The major negatively regulated pathways were membrane transport, signal transduction, carbohydrate metabolism and energy metabolism.

Membrane transport

Membrane transport occupied around 5 genes. There were genes coded for ABC transporters - SCO2258 (ABC transporter), SCO2979 (integral membrane transport protein), SCO6062 (ABC transporter ATP-binding protein), SCO6063 (ABC transporter permease), SCO6569 (solute binding protein).

Signal transduction

The second most influenced pathway was signal transduction with the same percentage of incident genes as the membrane transport (5 genes). This group comprised genes from two-component system pathway, including genes SCO3594 (*dldh*, *vanH*, D-lactate dehydrogenase), SCO3595 (*ddlA2*, *vanA*, D-alanine-D-lactate dehydrogenase), SCO3596 (*vanX*, D-alanine:D-alanine dipeptidase), coming from one operon. Then, genes SCO3946 (*cydB*, cytochrome oxidase subunit II) and SCO5584 (*glnK*, nitrogen regulatory protein P-II).

Carbohydrate metabolism

Four genes from the carbohydrate metabolism pathway were negatively regulated by SigQ - SCO0462 (glyceraldehyde 3-phosphate dehydrogenase) from the glycolysis/gluconeogenesis, SCO0951

(methylmalonic acid semialdehyde dehydrogenase) falling into propanoate metabolism group, SCO1031 (methylmalonic acid semialdehyde dehydrogenase) and SCO1022 (hypothetical protein) from the inositol phosphate pathway.

Energy metabolism

From the energy metabolism group, we detected four genes as repressed by SigQ. SCO1096 (membrane-bound proton-translocating pyrophosphatase) and SCO1086 (cytochrome oxidase subunit II) from the oxidative phosphorylation pathway, SCO1106 (oxidoreductase) assigned to methane metabolism and SCO1188 (cysteine synthase) from sulfur metabolism.

Other SigQ-negatively regulated metabolic pathways were amino acid metabolism, drug resistance, metabolism of cofactors and vitamins, Biosynthesis of other secondary metabolites, Cell motility, Metabolism of other amino acids Xenobiotics biodegradation and metabolism

Identification of consensus SigQ binding site

To discover potential binding sites, we have hand curated MACS identified peaks and filtered out those which were either whole located in gene or where located more than 300 bp upstream of annotated genome element. This restricted our dataset from 130 peaks identified by MACS to 114 peaks. Sequences of these peaks, were passed to RSAT (www.rsat.eu) peak-motifs web server and analyzed.

We have obtained 25 possible binding sites from which we have selected one which best matched our criteria: reasonable amount of peaks with at least one predicted site, appeared to be centrally enriched, and binding information was carried by TA nucleotides in the GC rich organism (Figure 3, for this motif there were also hits in DBTBS for LmrA (negative regulator) and in regulonDB for NanR (dual-regulator)).
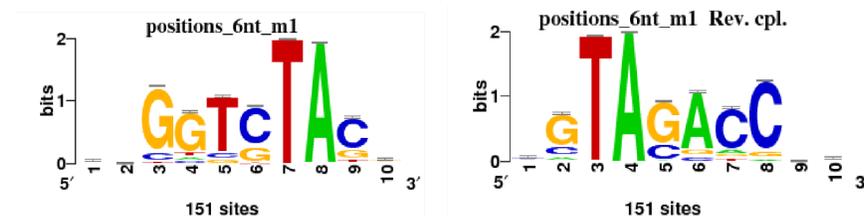
Figure 3 Potential SigQ binding site motifs.

We found consensus SigQ-binding motif for 16 genes listed in Table2.

Table2. List of genes with consensus sequence binding motif. p-value - ChIP-seq peak identification p-value.

SigQ binding site	Distance (nts)	Gene	Gene product	Regulation type	Fold enrichment	pvalue
CTGATCTACA	439	SCO0068	hydrolase	-	2,74	0,0245
GAGTTCTACG	140	SCO0814	rhamnose kinase	+	2,73	0,0240
TAGTTCTACG	37	SCO0878	hydrolase	-	4,23	0,0286
ATGTTCTACG	68	SCO2622	hypothetical protein	+	3,07	0,0306
ATGTTCTACG	558	SCO2621	hypothetical protein	-	3,78	0,0306
GTGGTGTACA	412	SCO4442	hypothetical protein	-	3,18	0,0015
GAGTTCTACC	207	SCO5399	acetyl-CoA acetyltransferase	+	2,88	0,0182
AAGATCTACG	30	SCO5625	tsf - elongation factor Ts	+	3,1	0,0174
TACGTCTACA	506	SCO5823	hypothetical protein	-	2,97	0,0366
CAGGTCTACC	260	SCO5848	tagatose 6-phosphate kinase	+	3,06	0,0402
ACGGTGTACT	130	SCO6170	oxidoreductase subunit	-	4,48	0,0237
ACGGTGTACT	178	SCO6169	regulatory protein	+	4,48	0,0237
CGGGTGTACC	505	SCO6658	6-phosphogluconate dehydroge	+	2,84	0,0063
CCGGTCTACC	450	SCO6886	hypothetical protein	-	3,36	0,0086
CTGATCTACA	215	SCO7088	two-component system respons	no profile available	3,76	0,0097
TTGGTCTAGA	148	SCO7224	hypothetical protein	-	3,32	0,0295

DISCUSSION

The method that allowed us to characterize the SigQ regulon is based on the epitope tagging approach accompanied by ChIP-seq analysis. The approach differs from standard epitope tagging methods which are usually based on a preparation of deletion mutant and its complementation by a tagged gene carried on a plasmid. Despite the fact that the plasmid-based methods are ineligible in the cases of indispensable genes, it is uncertain whether the gene expression from the plasmid is equal to that from the wild-type genome position. The insertion-based method used here is novel within the field of streptomycetes in that point of view that the epitope tag was delivered directly to the gene of interest within the chromosome. Similar approach was previously, but independently, proposed by {Kim, 2012 #1443}. In our work, a 27 bp sequence coding the HA-tag was inserted within the *sigQ* gene in its native genomic position. The mutagenesis procedure followed the REDIRECT technology protocol (http://streptomyces.org.uk/redirect/protocol_V1_4.pdf) with several modifications (see Methods) leading to an addition of the HA-tag to the inspected protein. Our preliminary western blot results that identified such a tagged protein approved the competence of the method (Figure 1). As the tag amino acid sequence was very short (9 amino acids), no influence on the protein function and folding had been expected and none has been revealed.

The SigQ-binding sites throughout the genome were shown by the ChIP-Seq approach. Based on the set parameters we defined which binding site corresponded with classical promoters (i.e. those localized in intergenic regions upstream from relevant ORF). The sequencing data, however, did not elucidate the strand on which the transcription may proceed and whether the control is possible also from the point of view of gene expression kinetics. Therefore, we created a gene expression control model which exploited our previous expression data from germination {Strakova, 2012 #1232}. The created model parameters were optimized with SigQ-gene expression profile as regulator and expression profiles of regulated genes. From the model arose an interesting evidence that SigQ (probably as well as other sigma factors) functions not only as a gene expression activator but surprisingly its binding may also repress the transcription of the incident gene, as could be seen from the inverse expression profiles (Figure 2) and, specifically, by the computational modeling. A possible

explanation for this phenomenon is that the promoter-bound SigQ-RNAP complex requires for the transcription initiation a specific activator that binds to an upstream enhancer, as suggested in {Rappas, 2007 #1451}.

The putative sigma factor SigQ belongs to ECF σs (Group 4), which are the most divergent group of alternative sigma factors broadly distributed across bacteria (except for *Mycoplasma*) {Lonetto, 1994 #1447}. The role of ECF σs is supposed to maintain transcriptional responses to environmental signals. To achieve these responses in required values, the expression of bacterial ECF regulators is often controlled by adjacently encoded two-component systems (in the case of *Streptomyces* by CseB-CseC for SigE {Hong, 2002 #1288} or AfsQ1-AfsQ2 for SigQ {Shu, 2009 #1293}), and their activity by specific anti-sigma factors {Campagne, 2015 #1426}. Whereas an average bacterial genome encodes about six different ECF σs, *Streptomyces coelicolor* possesses more than 30 ECF σs. Such a high number per a single genome allows cells to orchestrate their complex developmental programs and stress responses. An exemplar model where the ECF σs coordinated functions are realized is the streptomycete spore germination, accompanied with an outset of metabolic activities and distinct morphological changes. Our previous investigation revealed about 10 different sigma factors that are known to control various stress responses and whose striking expression changes suggest for their importance during germination {Bobek, 2014 #1355}. From these data we suggested that the process of germination evokes both osmotic (due to the water influx) and oxidative stress responses. The SigQ is one of the early expressed regulators whose expression is accelerated in highest degree during the ignition of germination. Consistently, we found here that the intracellular SigQ level is highly increased by the osmotic stress (0.5 NaCl). Based on these facts we suggested that SigQ is one of the major regulators of the germination process.

The sigQ directed nitrogen metabolism

The *sigQ* gene is adjacent to the divergently oriented *afsQ1-Q2* (SCO4907, SCO4906, resp.) region and, as shown previously, its expression is under a positive control of the AfsQ1-Q2 two component system {Shu, 2009 #1293}. In our germination course, the expression of both a*fsQ1* and a*fsQ2* was hardly detected from the beginning (corresponding with s*igQ* peak1) and then their transcripts were slowly but constitutively increasing (Figure 4).
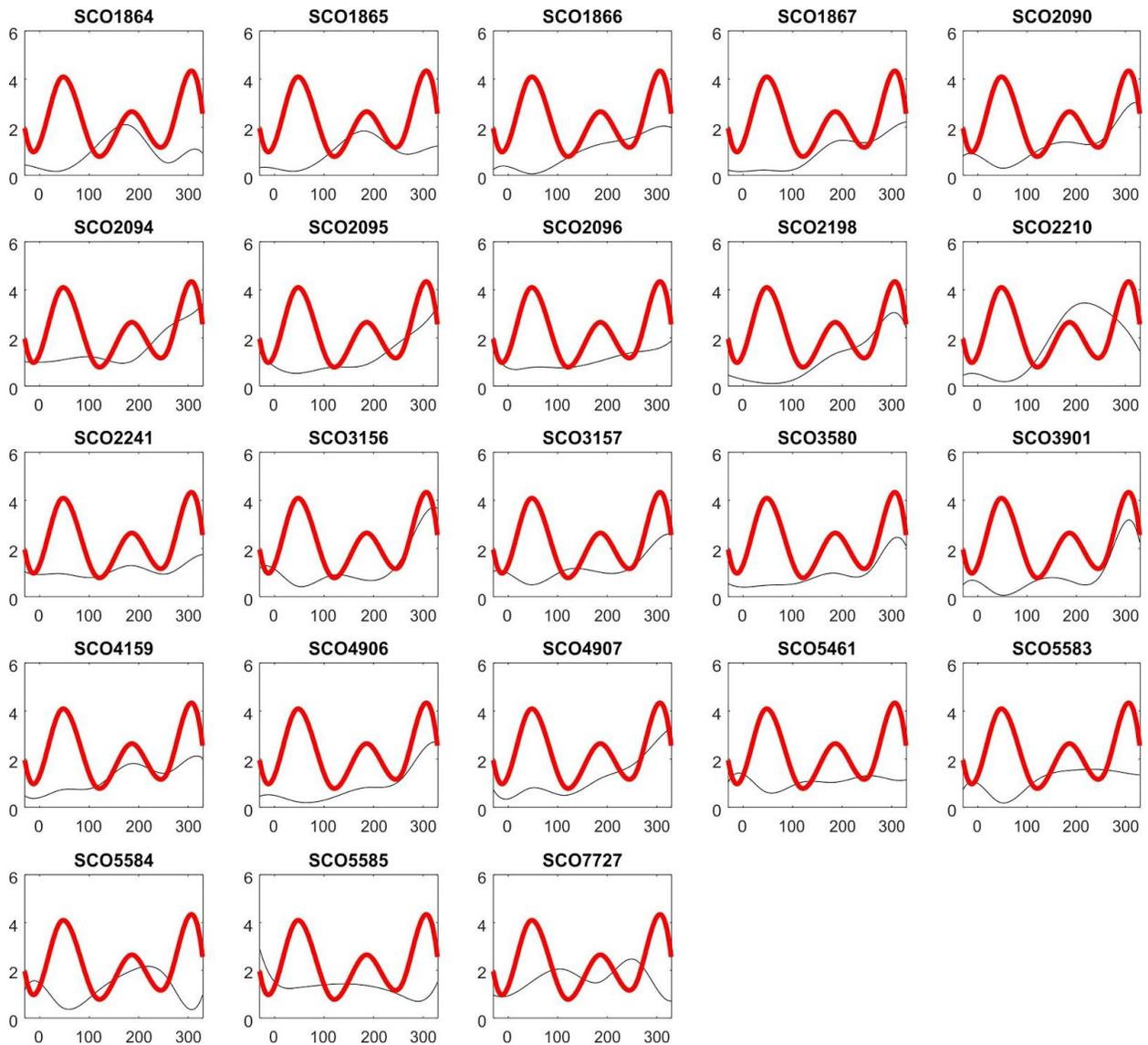
Figure 4. Expression profiles of genes mentioned in discussion with the *sigQ* profile (red thick) superimposed.

The AfsQ1-Q2 system is only known to stimulate actinorhodin and undecylprodigiosin production {Ishizuka, 1992 #1440}, whereas SigQ antagonizes the antibiotic production and sporulation in a glutamate-based medium {Shu, 2009 #1293}. The glutamate presented in medium (as was also the case of the AM medium used here) is utilized by glutamine synthetase (GS) as a donor of nitrogen. As our data indicate, the nitrogen metabolism is maintained by the SigQ-directed transcription. *Streptomyces coelicolor* possesses two GS genes (*glnA* and *glnII*) {Rexer, 2006 #1452}. The *glnA* gene (SCO2198), here found as a member of the SigQ regulon, encodes the heat-stable GSI {Hillemann, 1993 #1438}. Based on our data, the transcription from the *glnA* gene increases during germination in a stepwise manner; then, during cell development, its expression and the GSI activity is constitutive, as was shown previously {Fink, 2002 #1433}. On contrary, the second gene, *glnII* (SCO2210), encodes a heat-labile GSII {Hillemann, 1993 #1438;Weisschuh, 2000 #1459}, which was described to be active

only on solid medium {Fink, 2002 #1433}. Nevertheless, its expression, which was not shown to be connected with SigQ, increased after 120 min of the germination course in liquid medium (Figure 4). Additionally, to these two GS enzymes mentioned, there are other three genes with sequence similarities to GSs, *glnA2*, *glnA3* and *glnA4*. The function of their protein products is unknown, however, none of them exhibits GS activity. Our results showed that, besides the *glnA*, the *glnA2* gene (SCO2241) also belonged to the SigQ regulon and, according to the model, the expression of both these genes was activated by this sigma factor (Figure 4). Cellular responses to nitrogen accessibility in medium are not well understood. Besides the GSI enzyme whose function is mediated by adenylyltransferase GlnE, another nitrogen regulating system is active in *Streptomyces*, encompassing AmtB ammonium transporter (SCO5583) and GlnK PII regulator (SCO5584), which is subjected to adenylylation by GlnD nucleotidyltransferase (SCO5585) under low nitrogen conditions {Hesketh, 2002 #1437}. The genes of these three proteins are adjacent, but according to our expression data (Figure 4), only AmtB and GlnK were under the control of SigQ during germination, whereas expression of the *glnD* was diminished. Although functional connection between the two nitrogen regulating systems, GSI and GlnK-PII, remains obscure, based on our data, both they belong to the SigQ-regulon.

One found member of the SigQ-regulon is the gene SCO5461 coding for mono-ADP-ribosyltransferase (ADPRTase). This is a transmembrane enzyme responsible for a transfer of ADP-ribose moiety to a specific amino acid chain of a protein. ADP-ribosylation is an important reversible regulatory mechanism affecting normal growth of cells. Deletion mutant has delayed or completely inhibited morphological differentiation in culture dependent conditions {Szirak, 2012 #1415}. In *S. griseus* it was shown that the ADP ribosylation of glutamine synthetase (GS) results in inactivation of this enzyme {Szirak, 2012 #1415}. Here, SigQ, acting as a repressor of the ADPRTase expression (Figure 4), probably allowed the GSI to remain active.

An important co-regulator of the nitrogen metabolism is OmpR-like protein GlnR (SCO4159) that is responsible for glutamine synthesis {Wray, 1988 #1462}. Based on our search, SigQ probably does not determine the GlnR synthesis (Figure 4); on the other hand, both transcriptional regulators share several promoter targets (*glnA*, *amtB*, SCO2960 and *ORF1863*) {Bursy, 2008 #1425}.

Adjacent to the *ORF1863* (encoding an arginine-rich hypothetical protein) but divergently orientated is *ectABCD* operon (SCO1864-SCO1867) {Bursy, 2008 #1425} which also belonged to the SigQ regulon, but were found to be regulated inconsistently by SigQ. The *ectA* gene encodes N-gamma-acetyltransferase (EctA) and the downstream genes encode L-2,4-diaminobutyric acid transaminase (EctB), ectoine synthase (EctC), and ectoine hydroxylase (EctD). The role of this enzyme set is to catalyze a biochemical reaction from L-aspartate-gamma-semialdehyde to (hydroxy-) ectoine {Kuhlmann, 2008 #1446}. The ectoine is a protective substance that helps cells to survive high salinity or temperature {Bursy, 2008 #1425;Kol, 2010 #1445}. During the germination course, transcription of the responsible genes peaks (*ectA-B*) at or accelerates (*ectC-D*) after 120 min, and their further expression more or less responds to the SigQ profile (Figure 4), suggesting that the ectoine biosynthesis takes place from this time. In the consequence that germination evokes an osmotic stress {Bobek, 2014 #1355}, the role of the produced compound is probably to restore and maintain appropriate volume of water in germinating spores. EctR1 is a MarR-like repressor of the *ectABCD* transcription, whose gene was detected upstream of the *ect* genes in 107 various bacteria but such an adjacent regulator has not been found in *Streptomyces* {Widderich, 2014 #1460}. Within the SigQ

regulon, we detected another MarR-like regulator (SCO7727), distantly located on the genome. Its expression profile (Figure 4) appeared antagonistic to the SigQ and partially also to the *ectABCD*. From this, one may speculate that the SCO7727 regulator might be involved in the *ect* transcription control. However, this we remain as an open question.

The SigQ-directed cell wall reconstruction and growth

The results mentioned above suggest that SigQ governs (among others) the nitrogen metabolism. Having this in mind, one may argue that such an enormous activation of the *sigQ* expression might be due to the cultivation of spores in the amino-acid rich medium. And indeed after ammonium stress, the SigQ protein was clearly detectable during the vegetative growth of *S. coelicolor* in SMM and R2YT media {Nieselt, 2010 #1405} or in 2YT medium (this work) whereas without the stress the SigQ protein was faintly detected on our western blots (Figure 1). On contrary, the salt (osmotic) stress, a concomitant event of germination, showed a similar effect. There are also some other aspects that suggest for the importance of SigQ during germination.

Spore germination involves reconstruction and growth of the cell wall. For the cell wall enlargement, a cleavage of the pentaglycine bridges of the peptidoglycan (PG) layer is required. For the process, penicillin binding proteins (PBP) with various enzymatic activities, such as D-alanine carboxypeptidases, peptidoglycan transpeptidases, and peptidoglycan endopeptidases, are essential {Spratt, 1977 #1457}. Several different genes encoding PBPs (SCO3156, SCO3157, SCO3580, SCO3901, and probably also the *ftsI* gene) were expressed under the control of SigQ during germination (Figure 4). In addition, there are LytM family proteins (endopeptidases containing a lysostaphin-domain) that cleave the pentaglycine cross-bridges, as found in staphylococcal PG {Browder, 1965 #1424;Cloud, 2004 #1431}. In *E. coli*, the peptidoglycan growth is facilitated by a LytM family protein YebA {Singh, 2012 #1456}. The probable streptomycete YebA homolog is encoded by the last member of the SCO2094-2096 operon (Figure 4). Within this gene region, we mapped two SigQ-binding sites, one preceding SCO2093 and second preceding the SCO2094. The region belongs to a wider cluster of the conserved division and cell wall (dcw) genes. The dcw cluster is highly conserved in many rod-shaped bacteria. In *E. coli* it is consisted of 16 genes involved in the cell division and synthesis of peptidoglycan precursors {Vicente, 1998 #1475}. Our data suggest that SigQ participates on the expression of dcw region in *Streptomyces*.

Conclusions

For this study, an epitope tag was inserted directly into the genome. This allowed the ChIP-Seq technique to proceed without construction of plasmid supplemented gene deletions. Using the method, the genome-wide binding of SigQ was analyzed. Combining epitope tagging based ChIp-seq analysis with kinetic modeling of gene expression we were able to identify genes and operons potentially controlled by ECF sigma factor SigQ, that was shown to be largely and divergently expressed during germination. Modeling of the gene expression time series using SigQ as regulator revealed different mode of its action on the genes under its control. Surprisingly we found that SigQ acted not only to initiate transcription but mostly as a repressor, which, in some parts or throughout the whole course of germination, suppressed expression of the genes whose promoters it bound. Consequent functional analysis using gene annotation and regulatory pathways databases discovered

SigQ to be mostly involved in nitrogen metabolism and vegetative cell wall reconstruction during germination. Kinetic models also allowed to optimize parameters of the model making possible to simulate kinetics of expression under other conditions defined by the expression profile of the principal regulator – SigQ, or other alternative regulators found by the computational modeling. Our strategy for further work is the identification of regulons of other sigma factors and creation of comprehensive network of gene expression under their control. We believe that such models will allow to simulate kinetic behavior of large molecular systems and will contribute to understanding of behavior of temporal control processes in the cell on a system level.