

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/138641> holds various files of this Leiden University dissertation.

Author: Du, C.

Title: Multi-omics studies of the control of growth and antibiotic production of streptomyces

Issue Date: 2020-12-09

Analysis of the background-reduced antibiotic production host *Streptomyces coelicolor* M1152 using quantitative proteomics

Chao Du¹, Dino van Dissel², Alexander Wentzel², Gilles P. van Wezel¹

1. Microbial Biotechnology, Institute of Biology, Leiden University, Leiden, The Netherlands

2. Department of Biotechnology and Nanomedicine, SINTEF Industry, Trondheim, Norway

Sulheim, S., Kumelj, T., van Dissel, D., Salehzadeh-Yazdi, A., Du, C., van Wezel, G.P., Nieselt, K., Almaas, E., Wentzel, A., and Kerkhoven, E.J. (2020) Enzyme-constrained models and omics analysis of *Streptomyces coelicolor* reveal metabolic changes that enhance heterologous production. *iScience*: 101525.

DOI: 10.1016/j.isci.2020.101525

Abstract

Streptomycetes are major producers of bioactive natural products. Genome mining has uncovered a wealth of biosynthetic gene clusters (BGCs) for bioactive molecules, yet many of the clusters are silent under laboratory conditions. An important strategy to explore these potentials is to express BGCs in an optimised heterologous production host, but the effects of the genomic changes on growth and production are not well understood. Here, using quantitative proteomics, we have analysed the differences in global protein profiles between *Streptomyces coelicolor* M145 and its derivative strain M1152, which is a widely used platform for heterologous gene expression. M1152 has four major natural BGCs removed, and in addition a point mutation in *rpoB*. Our data show that the response to phosphate depletion is deregulated in M1152, with delayed up-regulation of several members of the PhoP regulon. Level of proteins belonging to the GlnR regulon was reduced in M1152. GlnR and PhoP are connected regulons that globally control natural product biosynthesis. Biosynthetic proteins for the osmoprotectant hydroxyectoine were over-represented in strain M1152, while the level of γ -butyrolactone (GBL) synthase ScbA and its GBL-responsive regulator ScbR was strongly induced. Taken together, this work reveals that M1152 has undergone significant changes in the global regulatory networks that control antibiotic production. This provides novel insights into the consequences of strain-optimization approaches and will provide guidance for future efforts in this important field of research.

Introduction

Natural products are the primary source of antibiotics, antitumor compounds, food preservatives, and many more (Newman & Cragg, 2016, Hopwood, 2007). Human society has benefited enormously from the introduction of new antibiotics, allowing the treatment of infections with microbial pathogens. Recently, however, the dramatic decline of new antibiotics discovery combined with a rapid rise of multidrug resistant infections impose huge new threats to human health (Kolter & van Wezel, 2016, Cooper & Shlaes, 2011, Baltz, 2008). This antibiotic crisis had led to the urgent need of novel antibiotics (Baltz, 2007, Zhu *et al.*, 2014a). Most antibiotics are produced by members of the Actinobacteria, and the majority of those by streptomycetes (Barka *et al.*, 2016). Since the first complete genome of the model organism *S. coelicolor* was published, it is becoming increasingly clear that *Streptomyces* genomes harbour a very large repository of biosynthetic gene clusters (BGCs) that might potentially produce new compounds (Bentley *et al.*, 2002, Medema & Fischbach, 2015, Ziemert *et al.*, 2016).

One major effort in natural product discovery is to find, study, and manipulate the biosynthetic gene clusters (BGCs) of the producer organisms. This includes engineering the expression of biosynthetic enzymes and increasing the precursor and intermediates pool, thereby optimizing the production of the natural products which are already produced by native strains (Komatsu *et al.*, 2010, Myronovskiy & Luzhetskyy, 2019). However, the activation of the cryptic biosynthetic potential remains difficult. This is among others due to the fact that we do not yet understand the ecological conditions under which the biosynthetic gene clusters are activated in the natural habitat (van der Meij *et al.*, 2018). To overcome these limitations, expression in heterologous production hosts is becoming indispensable (Nepal & Wang, 2019, Gomez-Escribano & Bibb, 2011). One advantage of heterologous expression is that we can choose genetically well understood strains that can be cultured robustly and are genetically tractable. One of the most important organisms for heterologous expression is *Streptomyces*. Heterologous expression in *Streptomyces* is now becoming a popular tool for novel antibiotics discovery and production (Myronovskiy & Luzhetskyy, 2019, Ahmed *et al.*, 2020, Gomez-Escribano & Bibb, 2011). Heterologous expression can also be used to determine the boundaries of the BGC (Komatsu *et al.*, 2013, Liu *et al.*, 2018), to create artificial natural products by combining different BGCs (Park *et al.*, 2011), or for characterizing the functions of individual genes (Waldman *et al.*, 2015).

Streptomyces coelicolor, as the most studied *Streptomyces* model organism (Bentley *et al.*, 2002, Hoskisson & van Wezel, 2019), is a producer of many different natural products, including five antibiotics, namely actinorhodin (ACT), undecylprodigiosin (RED), calcium-dependent antibiotic (CDA), coelimycin (CPK) and the plasmid-encoded methylenomycin (MMY). Antibiotic production is tightly

controlled, and correlates to the onset of development (Bibb, 2005, van der Heul *et al.*, 2018). Nutrient availability is thereby an important factor, and coordinated lysis of the vegetative or substrate mycelium precedes the onset of morphological and chemical differentiation (Chater & Losick, 1997, Manteca *et al.*, 2005, Tenconi *et al.*, 2018). Upon nutrient limitation, *S. coelicolor* responds with a flux change of energy and intermediate metabolites which initiate secondary metabolism (van Keulen & Dyson, 2014). This is a tightly controlled process that resulted in major metabolic change from primary to secondary metabolism, so this process is often referred as metabolic switch. The regulation of the metabolic switch involves both global and pathway-specific regulators that form complex regulatory networks (Rigali *et al.*, 2008, Romero *et al.*, 2014, Urem *et al.*, 2016). Depletion of inorganic phosphate (Pi) is one of the most studied inducer of major metabolic switch (Martín, 2004, Nieselt *et al.*, 2010). The primary response to Pi depletion is mediated through the two-component system PhoR-PhoP (Sola-Landa *et al.*, 2003, Sola-Landa *et al.*, 2005). PhoP acts as a global regulator of various metabolomic processes, including secondary metabolite biosynthesis (Martín *et al.*, 2011, Sola-Landa *et al.*, 2003, Santos-Beneit *et al.*, 2009). Another important global regulator that related with nutrient depletion is GlnR, which controls nitrogen uptake and metabolism related genes (Fink *et al.*, 2002, Tiffert *et al.*, 2008, Amin *et al.*, 2012, Wang & Zhao, 2009). Interestingly, there is significant cross-talk between the P- and N-responses in *Streptomyces*, due to partially overlapping regulons of GlnR and PhoP (Sola-Landa *et al.*, 2012, Santos-Beneit *et al.*, 2012). Novel systems biology approaches of multi-dimensional, multi-omics global analyses are required to help us understand the regulatory networks in a more comprehensive way (Hwang *et al.*, 2014, Rokem *et al.*, 2007). Another gene that has great regulatory impact on antibiotic production in *Streptomyces* is *rpoB*. It encodes an RNA polymerase β -subunit known to respond to stringent response elicitor guanosine 5'-diphosphate 3'-diphosphate (ppGpp), which causes rapid cessation of RNA synthesis and other cellular reactions (Cashel, 1996, Xu *et al.*, 2002). It has been suggested that the *rpoB*[C1298T] mutation functions by mimicking the activated form in promoting the onset of secondary metabolism in *Streptomyces* (Hu *et al.*, 2002).

In order to explore the full potential of *S. coelicolor* as a superhost for the production of antibiotics or other valuable natural product, we here analysed the background-reduced host *S. coelicolor* M1152 with its parent, *S. coelicolor* M145. As an important production host (Gomez-Escribano & Bibb, 2014, Nepal & Wang, 2019), *S. coelicolor* M1152 was generated by removal of four major BGCs, namely the polyketides ACT, RED and CPK, and the non-ribosomal peptide antibiotic CDA (Gomez-Escribano & Bibb, 2011). In addition, mutations were introduced in *rpoB*[C1298T] which boosts antibiotic production (Hu *et al.*, 2002).

To obtain more insights into how the metabolic switch redirects primary and secondary metabolism, a proteomics study was done as part of the systematic rebuilding of Actinomycetes for natural product formation (SYSTERACT) project using a multi-Omics approach, and applying the optimised cultivation system published previously (Wentzel *et al.*, 2012a). Here, we compared the protein profiles of *S. coelicolor* M1152 and its parent M145 using samples from batch fermenters obtained from the SYSTERACT project. For this, a new advanced proteomics pipeline was applied, featuring label-free and TOP3 quantification techniques. The differences between the two strains and the impact of the new insights on the use of heterologous production strains are discussed.

Results and Discussion

Comparison of the growth of background-reduced expression host *S. coelicolor* M1152 with its parent M145

We aimed to investigate the changes in global protein profiles in *S. coelicolor* M1152 as compared to its parent *S. coelicolor* M145. *S. coelicolor* M145 and M1152 were cultivated in bioreactors in SSBM-P media. SSBM was designed with two carbon sources, D-glucose and L-glutamate, to ensure sufficient biomass before the onset of secondary metabolism. Use of L-glutamate as the sole nitrogen source leads to a well-defined transition phase, which allows antibiotic production after sufficient biomass accumulation (Wentzel *et al.*, 2012a). This established system was tested for high-resolution time-course analyses on the levels of transcriptomics (Nieselt *et al.*, 2010, Alam *et al.*, 2010), proteomics (Thomas *et al.*, 2012), and metabolomics (Wentzel *et al.*, 2012b).

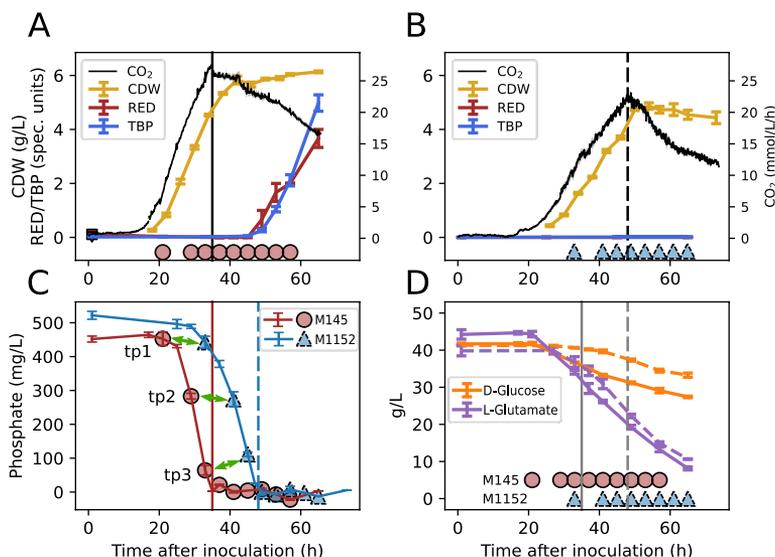


Figure 1. Batch cultivation data of *S. coelicolor* M145 and M1152. A) M145 and B) M1152 were grown on SSBM-P medium in benchtop bioreactors in triplicate. Time points of collecting proteomics and transcriptomics samples are indicated by purple triangle on the x axis line. CDW, cell dry weight; RED, undecylprodigiosin; TBP, total blue pigments / actinorhodins. Error bars for CDW, RED, TBP and greyed area for CO₂ generation represent the standard error of three biological replicates. C) Phosphate (Pi) concentration measured during the same batch culture. The first 3 sampling time points (tp1 to tp3) are indicated by green arrows. D) Concentration of the two carbon sources during fermentation of the two strains. Circle and triangles in all subplots indicate the time points of collecting proteomics and transcriptomics samples for *S. coelicolor* M145 and M1152, respectively. Vertical lines indicate the time when phosphate in the medium has been depleted for respective fermentation experiments.

Samples were taken in 4 h intervals during exponential growth, around the metabolic switch and during the stationary phase (when natural product formation takes place) (Figure 1). Growth of M145 reached stationary phase at around 42 h with a cell dry weight (CDW) of 5.8 g·L⁻¹, while for M1152, the stationary phase was reached at around 50 h with a CDW of only 4.7 g·L⁻¹. This alteration in the growth characteristics is consistent with the literature (Gomez-Escribano & Bibb, 2011). The online analysis showed maximum CO₂ production after 35 h for M145 and after 47 h for M1152 (Figure 1A, B). This is consistent with the timing of phosphate depletion in both strains (Figure 1C). Shortly after phosphate depletion, both ACT and RED were produced by *S. coelicolor* M145, indicating that the metabolic switch had induced antibiotic production (Figure 1A). D-glucose and L-glutamate were consumed concomitantly during fermentation, with glutamate consumed slightly faster than glucose (Figure 1D). Both remained in excess throughout the cultivation. The consumption of glucose and glutamate by M1152 were all reduced, with glucose consumption reduced more.

Proteomics analysis reveals major differences between *S. coelicolor* M145 and M1152

To investigate the changes in more detail, samples from the fermentations with M145 and M1152 were compared by proteomics. 27 mycelial samples were collected from *S. coelicolor* M145 corresponding to triplicate samples from 9 time points, and 24 mycelial samples were collected from *S. coelicolor* M1152 corresponding to triplicate samples from 8 time points (Figure 1). In total, 3,278 proteins were identified combining all samples from *S. coelicolor* M145 and M1152 fermentations. After the removal of proteins with low quantities (see Materials and Methods section for details on the criteria), 1,879 proteins were reliably quantified in at least one time point (Table S1). These same samples were also analysed by RNA-seq (Sulheim *et al.*, 2020), allowing comparison of proteome and transcriptome data.

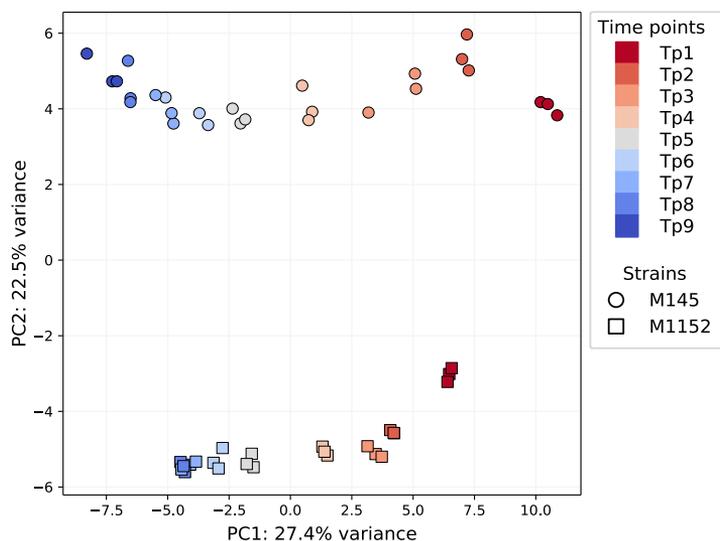


Figure 2. Principal component analysis (PCA) of proteomics data from all samples. The variance stabilizing transformed (VST) was used. The proteins from genes deleted in M1152 were excluded. Aligned sampling time points are coloured coded from red to blue, different strains are distinguished by the shape of data point.

5

As *S. coelicolor* M1152 grows slower than its parent M145 (Figure 1), data were aligned to represent a comparative growth phase. The metabolic switch occurred between the third and fourth sampling points for both strains (Figure 1C). This allowed sequential alignment of the samples from the two strains, named tp(n). To understand the overall differences between different samples, variance stabilizing transformation (VST) of the normalized TOP3 quantification data was done to reduce the systematical variance especially for the low abundant proteins (Love *et al.*, 2014). This normalization was recently reviewed to be excellent for label-free proteomics data (Välikangas *et al.*, 2016). General assessments of data transformation are shown in Figures S1 and S2. To investigate the coherence of the collected protein concentrations between the different samples a principal component analysis (PCA) performed on the transformed dataset (Figure 2). All proteins encoded by the four BGCs that had been deleted from M1152 were omitted from the PCA analysis. The PCA plot shows a clear time course migration of samples on principal component 1 (PC1) between samples from the same strain. The continuity of samples from one strain indicates a good stability of our data. Moreover, there was a very clear separation between *S. coelicolor* M145 and M1152 samples on principal component 2 (PC2), which represents 22.5% of the variance (Figure 2), indicating that the genome modifications in M1152 led to a profound overall protein level alteration. However, like for M145, M1152 samples also shifted to one side of PC1 in a time-dependent manner. To investigate which

proteins caused the general shift in the protein profiles, protein levels were compared between *S. coelicolor* M1152 and M145 for the aligned time points, whereby the raw quantification value with fold change ≥ 2 or ≤ 0.5 and with T-test p -value < 0.05 were considered differentially expressed proteins (DEPs). In total, there were 408 DEPs in at least one time point comparing the two strains, representing 21.7% of quantifiable proteome in this experiment (Table S2).

Response of the PhoP regulon to Pi depletion is delayed in strain M1152

Since the metabolic switch was induced by phosphate depletion (Figure 1), we investigated the change in protein profiles during phosphate depletion. The same criteria for DEP were used comparing selected timepoints before (tp2) and after (tp6) Pi depletion within each strain. The distribution of variance in these samples was comparable (Figure S3). In total, 108 DEPs were found in M145 and 47 in M1152 (Table S3). The reduced number of total DEPs in the BGC-deleted background suggests that secondary metabolism may act as a modulator of the phosphate response in *S. coelicolor*. Indeed, many PhoP regulon proteins were up-regulated after Pi depletion (Figure 3, purple dots). Especially for strain M1152, the PhoP regulon proteins were dominant (Fisher's exact test p -value < 0.001) among the DEPs.

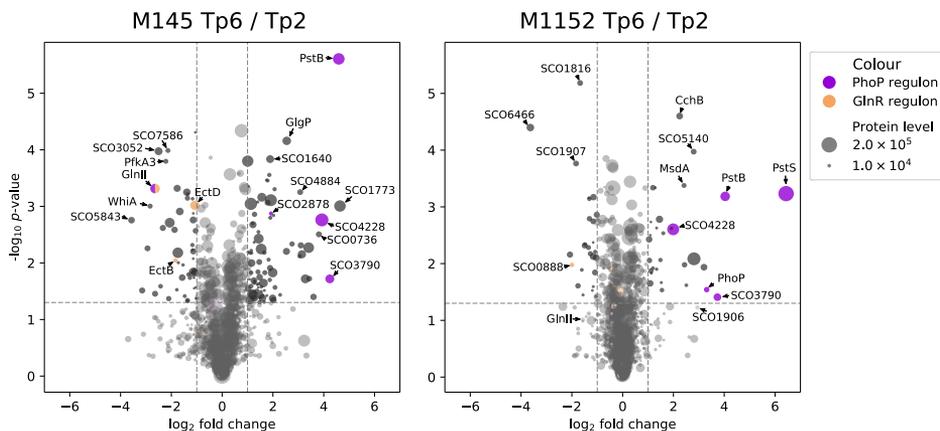


Figure 3. Volcano plot showing comparisons between representative time points before (tp2) and after (tp6) phosphate depletion. The time points were chosen to represent different developmental state. Proteins known to be regulated by PhoP and GlnR are coloured purple and orange, respectively. The protein names of outstanding points are marked with an arrow. The higher average protein level of the test and control samples is transformed to the size of each dot. Proteins not significantly changed (p -value ≥ 0.05 or $0.5 < \text{fold change} < 2$) are greyed out as transparent dots. Dashed lines showing the boundaries of significance threshold.

To investigate further how PhoP regulon proteins respond to Pi depletion and the differences of the response between two strains, a time-course protein level curve was made for all known PhoP regulon proteins that were quantified (Figure 4). At the transcriptome level, all genes belonging to the PhoP regulon were shown to be up-regulated immediately after Pi depletion in both strains (Sulheim *et al.*, 2020). This is consistent with previous microarray data based on the same system (Nieselt *et al.*, 2010). As expected, also many proteins that belong to the PhoP regulon were up-regulated directly after Pi depletion. However, some proteins were unchanged or even down-regulated after Pi depletion (Table 1). This delay in the regulation of protein levels may be explained by the effect of post-translational control (Vogel & Marcotte, 2012).

PhoP itself (Figure 4C) showed a delayed response in both strains, whereby the response to Pi depletion was delayed by one sampling interval (4 h). The pattern of protein levels of the PhoP regulon can be categorized into three subcategories (Table 1), namely those that were up-regulated in both M145 and M1152 at the time of Pi depletion (group A), those that showed delayed up-regulation in M1152 (group B), and those that did not show any up-regulation (group C). Group A includes proteins that were up-regulated in both *S. coelicolor* M145 and M1152. This group includes the phosphate transport proteins PstB, PstC, PstS, phosphate scavenging protein SCO3790, and SCO4878 that relates to phosphate storage (Martín *et al.*, 2012) (Figure 4B). The second category consists of proteins that were up-regulated in strain M145 immediately upon Pi depletion but showed delayed or reduced response in M1152 (Figure 4D). Four of these proteins are related to phosphate scavenging. The delayed up-regulation of several proteins of the PhoP regulon and of PhoP itself, in the phase immediately following the Pi depletion, may be one of the factors that causes slow growth. Perhaps this difference may contribute to the slower growth of *S. coelicolor* M1152.

Table 1. Response pattern of PhoP regulon proteins*

	Protein ID (name)	Response pattern	Functions related with phosphate depletion	Annotation
PhoP	SCO4230 (PhoP)	Delayed up-regulation (both)	Signal transduction	Pho response regulator
	SCO3790	Both up-regulated at Pi depletion	Phosphate scavenging	Putative phosphatase
Group A	SCO4139 (PstB)	Both up-regulated at Pi depletion	Phosphate transport	Phosphate ABC transport system ATP-binding protein
	SCO4141 (PstC)	Both up-regulated at Pi depletion	Phosphate transport	Phosphate ABC transport system permease protein
	SCO4142 (PstS)	Both up-regulated at Pi depletion	Phosphate transport	Secreted phosphate-binding protein
	SCO4878	Both up-regulated at Pi depletion	Phosphate storage	Putative glycosyltransferase
	SCO2878	Up-regulated in M145, but no change in M1152	Unknown	Hypothetical protein
	SCO1196	Delayed up-regulation in M1152	Phosphate scavenging	Putative secreted protein, PLC-like phosphodiesterase motif
	SCO1906	Delayed up-regulation in M1152	Phosphate scavenging	Putative secreted phosphatase
Group B	SCO1968 (GlpQ2)	Delayed up-regulation in M1152	Phosphate scavenging	Secreted glycerophosphodiester phosphodiesterase
	SCO2286 (PhoA)	Delayed up-regulation in M1152	Phosphate scavenging	Secreted alkaline phosphatase
	SCO4228 (PhoU)	Delayed up-regulation in M1152	Signal transduction	Putative modulator of the Pho system
	SCO4159 (GlnR)	Detected only in aligned time point 1	Signal transduction	Transcriptional regulatory protein
	SCO4879	Up-regulated but with fluctuation	Phosphate storage	Conserved hypothetical protein
	SCO2210 (GlnII)	Down-regulated	Cross regulation with GlnR regulon	Glutamine synthetase
	SCO0033	No response	Unknown	Putative secreted neuraminidase
Group C	SCO1565 (GlpQ1)	No response	Phosphate scavenging	Secreted glycerophosphodiester phosphodiesterase
	SCO2068 (PhoD)	No response	Phosphate scavenging	Secreted phospholipase D
	SCO2198 (GlnA)	No response	Cross regulation with GlnR regulon	Glutamine synthetase I
	SCO2262	No response	Unknown	Putative oxidoreductase
	SCO4229 (PhoR)	No response	Signal transduction	Pho histidine kinase
	SCO4261	No response (ND [†] in M1152)	Unknown	Putative response regulator
	SCO5583 (AmtB)	No response (ND in M1152)	Cross regulation with GlnR regulon	Ammonium transporter

* PhoP regulon gene list adapted from (Allenby *et al.*, 2012), with extension of the proteins belonging to the same regulon: SCO4139 and SCO4141 from SCO4142 regulon, SCO4230 from SCO4228 regulon. Only proteins identified in the proteomics experiment are listed.

† Not detected

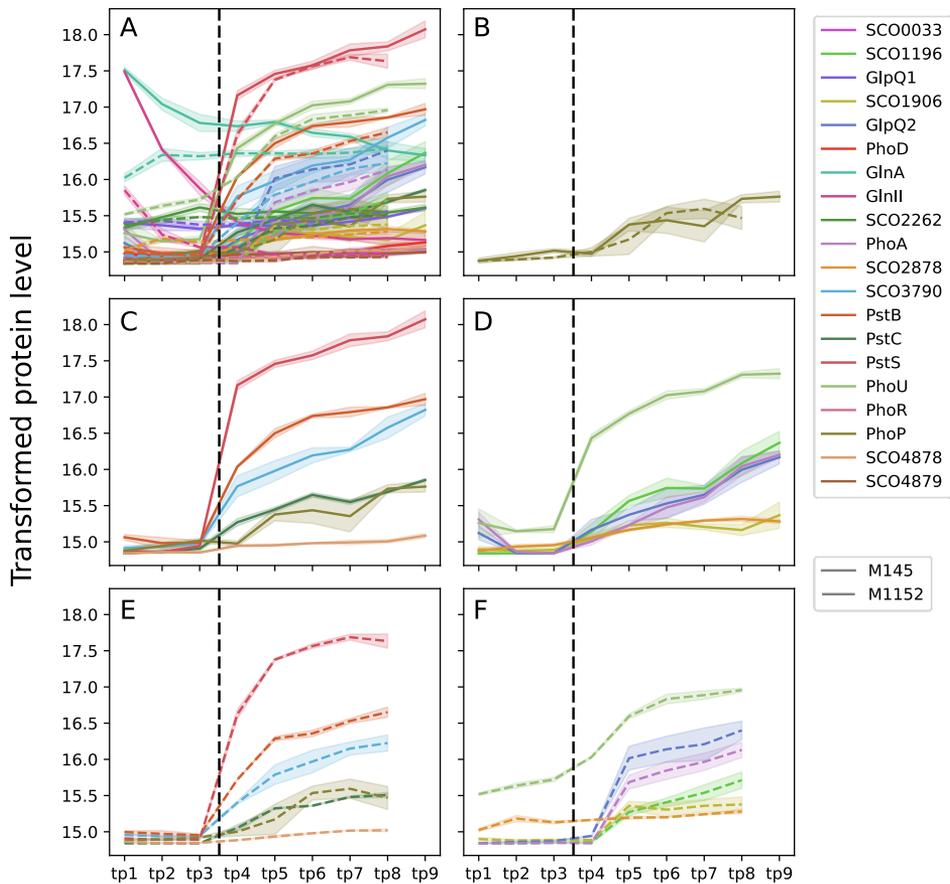


Figure 4. Level of detected PhoP regulon proteins in aligned time points. A), all detected PhoP regulon proteins, including some GlnR regulon proteins that proved to be co-regulated by PhoP. All other subplots are showing a subset of proteins in figure A). Protein levels in *S. coelicolor* M145 are shown in solid lines, strain M1152 are shown in dashed lines. B) PhoP. C) and E) PhoP regulon proteins regulated similarly in both strains (group A in Table 1). D) and F), PhoP regulon proteins differently regulated in the two strains (Group B in Table 1). For all plots, variance stabilizing transformed (VST) data was used for better demonstration of low abundant proteins (Love *et al.*, 2014). Black vertical dashed lines indicating the time of phosphate depletion. Shaded area indicating the standard error of triplicates, colour coding is shared across all subplots. PhoP regulon was adapted from (Allenby *et al.*, 2012), see Table 1 for detailed annotation.

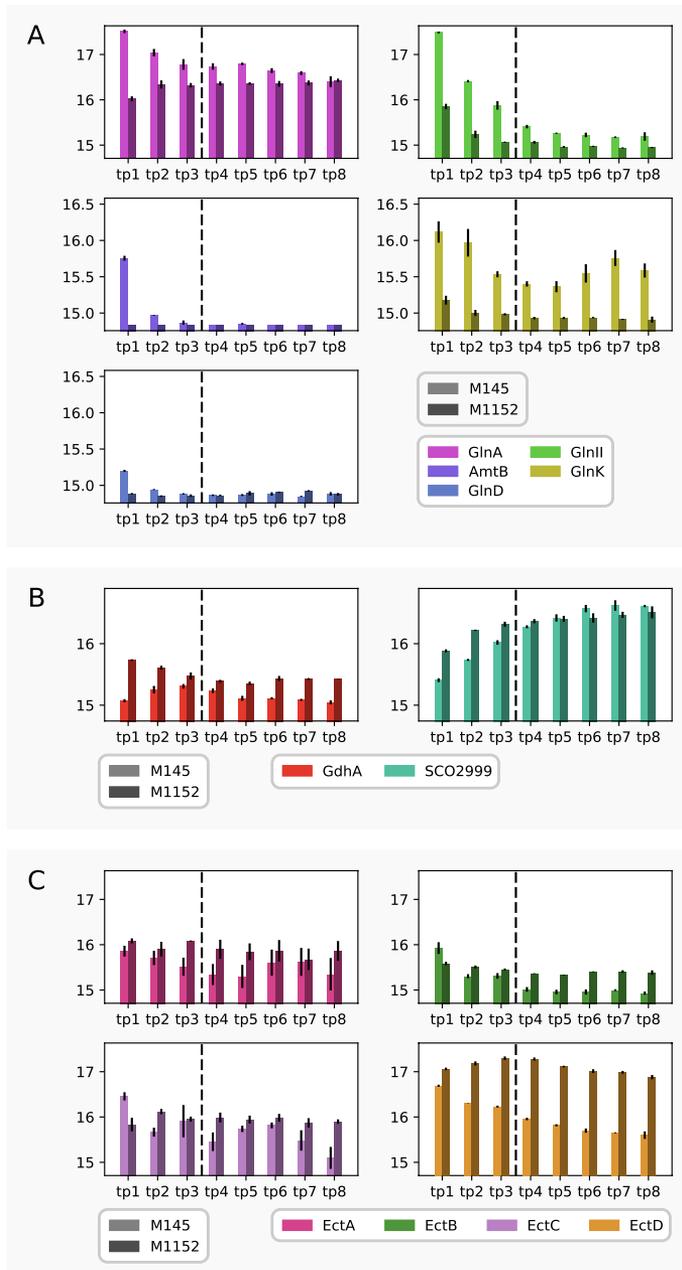


Figure 5. Protein profile of significantly changed GlnR regulon proteins. (A), two glutamate dehydrogenase (B), and Ect proteins (C). All protein levels were plotted as bar plots of the VST transformed data, note that quantification level in zero is transformed to around 15. Protein levels in strain M1152 are shown with darkened colours next to that of strain M145 at the same aligned time point (tp). Error bar indicates the standard error of the triplicates. Vertical dashed lines indicate the time of phosphate depletion.

Levels of the GlnR regulon proteins are reduced in *S. coelicolor* M1152

According to the offline measurement of L-glutamate (Figure 1), nitrogen was not depleted during the course of the fermentation. However, the cross-talk between the PhoP and GlnR regulons (Sola-Landa *et al.*, 2012, Santos-Beneit *et al.*, 2012) prompted investigation of the GlnR regulon during Pi depletion. 14 out of 21 reported GlnR regulon proteins were identified by our proteomics analysis. GlnII, GlnA, AmtB, GlnK, GlnD, EctD and EctB were down-regulated only in the parental strain M145 during fermentation (Figure 3 for GlnII, EctD and EctB, Figure 5A for GlnA, AmtB). None of these proteins was changed in M1152 after Pi depletion. Interestingly, GlnA showed up-regulation at tp2 in M1152, while it was down-regulated in M145. Previous studies have shown that *glnA* and *glnII* are differentially regulated and probably related to the control of development (Fink *et al.*, 2002, Reuther & Wohlleben, 2007). The different level of GlnA and other GlnR regulon proteins in strain M1152 indicates that their growth phase-dependent regulation has changed in M1152 as compared to M145. The *amtB-glnK-glnD* operon is regulated by both PhoP and GlnR (Rodríguez-García *et al.*, 2009, Wang *et al.*, 2012), but the levels of these proteins was unchanged upon Pi depletion.

5

Interestingly, the level of GlnK was significantly lower at all time points in M1152 as compared to M145 (Figure 5A, Table S2). GlnK is one of the nitrogen-sensory proteins that acts as a pleiotropic regulator for secondary metabolism and morphological differentiation (Waldvogel *et al.*, 2011, Perez-Redondo *et al.*, 2012). It is important to realise that the reduced level of GlnK in M1152 may have a major effect on the expression of BGCs that are regulated by this protein. GlnII and GlnK directly relate to nitrogen assimilation (Tiffert *et al.*, 2008). Their down-regulation in M1152 is an indication of reduced nitrogen assimilation. Conversely, the level of glutamate dehydrogenase (GdhA) was higher in M1152 at all time points as compared to M145. The protein profile changes along the time series could be separated by Pi depletion (Figure 5B). The second glutamate dehydrogenase encoded by the *S. coelicolor* genome, SCO2999, showed a five-fold increase from tp1 to tp6 in M145, corresponding to the onset of antibiotic production. In M1152 the increase was less pronounced, although SCO2999 levels were higher at early time points than in M145. Interestingly, the ammonia excretion was not significantly different between the two strains, which has been shown in the metabolism simulations build upon the same proteomics data (Sulheim *et al.*, 2020). This stable ammonia excretion can be explained by the observed balance between SCO2999 and GdhA. Combining these observations, we propose that the levels of the two glutamate dehydrogenases is inversely regulated in *S. coelicolor* and may compensate each other.

Another important GlnR-controlled process is that of the biosynthesis of ectoine and hydroxyectoine (Shao *et al.*, 2015, Bursy *et al.*, 2008). Ectoine compounds act as protectants against osmotic and heat stress (Bursy *et al.*, 2008), and the ectoine

BGC is also a likely target of control by GlnK (Waldvogel *et al.*, 2011). All Ect proteins (EctA to EctD) were detected at a higher level in M1152 than in M145 (Figure 5C, Table S2). The up-regulation of the biosynthesis of osmoprotectant compounds on stress resistance of M1152 needs to be investigated.

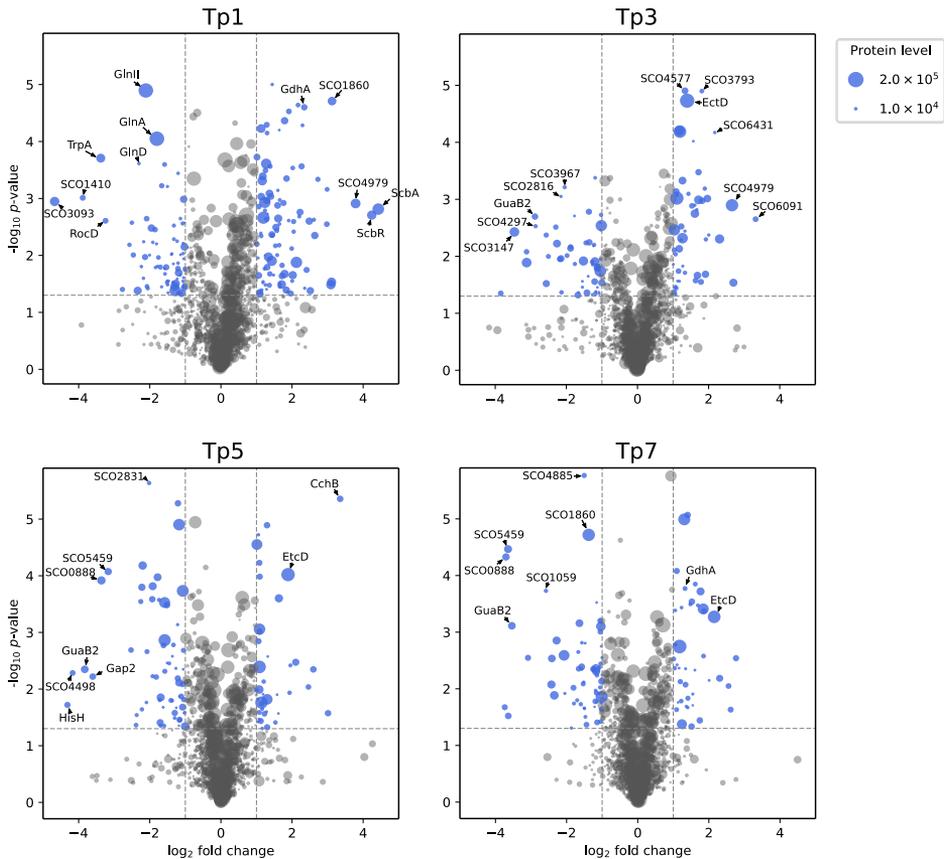


Figure 6. Volcano plots showing comparisons between *S. coelicolor* M1152 and M145 at representative aligned time points. The proteins from genes deleted in M1152 were excluded. Dashed lines showing the boundaries of significance threshold (p -value 0.05, \log_2 fold change ± 1). Protein names of outstanding points or those have been discussed in the text were marked with an arrow. The proteins are coloured blue as they met the significance threshold. The higher average protein level of the test and control samples is transformed to the size of each dot.

ScbR and ScbA are up-regulated in *S. coelicolor* M1152

To get a full view of the differences caused by the genome modifications in *S. coelicolor* M1152, the aligned corresponding samples were compared between

M1152 and M145 (Figure 6, Table S2). The most pronounced protein level differences were seen at tp1, corresponding to the time just before phosphate depletion, with 171 DEPs. The number of DEPs then dropped to values between 126 (time point 2) and 80 (time point 6). Interestingly, at tp1, ScbR and ScbA were among the most up-regulated proteins (Figure 6). These two proteins play a major role in γ -butyrolactone (GBL) mediated auto-regulation of antibiotic production and development (Takano *et al.*, 2001). ScbA synthesizes the GBL Scb1, while ScbR binds Scb1 and acts as a ScbA-responsive activator (Hsiao *et al.*, 2007). GBL-activated ScbR is a repressor of many developmental pathways including antibiotic production and development (Li *et al.*, 2015). The up-regulation of ScbR and ScbA is consistent with previous study showing that in AlaMM (minimal medium supplemented with alanine as nitrogen source), *S. coelicolor* M1152 produces more GBLs as compared to its parent strain M145 (Sidda *et al.*, 2016). It is known that ScbA is also controlled by ScbR2 (Li *et al.*, 2015), a gene that is lacking in M1152 because it is part of the *cpk* cluster that was deleted from its genome. Thus, the over-presence of ScbA is a likely result of *scbR2* deletion, which might relieve the suppression of *scbA* (Gomez-Escribano & Bibb, 2011, Li *et al.*, 2015). Further quantitative studies are needed to dissect the actual mechanism of ScbA, ScbR, ScbR2 and possible other participants. However, the up-regulation of ScbR should definitely be taken into account when expressing heterologous BGCs that are subject to control (repression) by ScbR.

Conclusions

In this work, the protein profiles in background-reduced host *S. coelicolor* M1152 - which lacks four major BGCs and has a mutation in *rpoB* - were compared to those of its parent *S. coelicolor* M145 in samples obtained from batch fermentations. We found that upon Pi depletion, part of the PhoP regulon was up-regulated, while the rest remained unchanged. Interestingly, half of the up-regulated proteins showed delayed response in M1152, which might result in reduced response in Pi depletion, and thus affects growth. Worth noticing is also the major differences shown for proteins of the GlnR regulon. Most of the GlnR regulon proteins were not affected by the Pi depletion induced by the metabolic switch. However, we noticed that some of these proteins showed significant differences between M1152 and M145. In particular, the down-regulation at early time points was not seen in M1152, indicating changes in growth phase-dependent control. We also observed significant reduction in the level of GlnK in M1152 as compared to that in M145, which may be an additional factor that may be relevant for heterologous expression of BGCs. The strong induction in ectoine biosynthesis in strain M1152 may reflect of elevated osmotic stress. Finally, the ScbR/ScbA regulatory system was highly up-regulated in M1152, likely as a result of the absence of *scbR2*, which alters the global control of BGC expression.

In summary, we have systematically compared the global protein profiles in *S. coelicolor* M145 and its derivative M1152 in a well-defined fermentation system. This provides comprehensive insights into the changes in growth and protein profiles that are caused by the genomic modifications in *S. coelicolor* M1152. Our proteomics data, connected to the extant metabolic and transcriptome data, provide detailed insights into the effects of rational strain design of *S. coelicolor* aimed at obtaining optimised platforms for the expression of heterologous BGCs. At the same time, these studies will also form useful reference datasets for future 'omics studies on the model organism *S. coelicolor*.

Material and methods

Strains, cultivation conditions, sampling procedures, and analyses of media components

Experiments were performed using *S. coelicolor* A3(2) M145 (Kieser *et al.*, 2000) and its derivative M1152 (Gomez-Escribano & Bibb, 2011). M1152 lacks the four major biosynthetic gene clusters for actinorhodin (ACT), undecylprodigiosin (RED), coelimycin (CPK), and calcium-dependent antibiotic (CDA), and carrying the pleiotropic, previously described antibiotic production enhancing mutation *rpoB*[C1298T] (Hu & Ochi, 2001). Both strains were kindly provided by Mervyn Bibb (John Innes Centre, Norwich, UK).

5

Triplicate cultivations of the strains were performed based on germinated spore inocula in 1.8 L phosphate-limited medium SSBM-P, applying all routines of the optimised submerged batch fermentation strategy for *S. coelicolor* established and described before (Wentzel *et al.*, 2012a). Spores were obtained by cultivation on soy flour-mannitol (SFM) agar plates (Kieser *et al.*, 2000), harvested and suspended in 20% (v/v) glycerol, following by storage at -80°C . 10^9 CFU of spores were germinated and inoculated into 50 mL $2 \times$ YT medium (van Wezel *et al.*, 2006). An even dispersion of the germinated spores was achieved by vortex mixing (30 s), ensuring comparable inocula among biological replicas. Each fermenter (1.8 L starting volume culture medium in 3 L Applikon fermenters) was inoculated with 4.5 mL germinated spore suspension (corresponding to 9×10^8 CFU). Phosphate-limited medium SSBM-P (Nieselt *et al.*, 2010) was used, containing Na-glutamate, $55.2 \text{ g}\cdot\text{L}^{-1}$; D-glucose, $40 \text{ g}\cdot\text{L}^{-1}$; MgSO_4 , 2.0 mM; phosphate, 4.6 mM; supplemented minimal medium trace element solution SMM-TE (Claessen *et al.*, 2003), $8 \text{ mL}\cdot\text{L}^{-1}$ and TMS1, $5.6 \text{ mL}\cdot\text{L}^{-1}$. TMS1 consisted of $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, $5 \text{ g}\cdot\text{L}^{-1}$; $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, $390 \text{ mg}\cdot\text{L}^{-1}$; $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, $440 \text{ mg}\cdot\text{L}^{-1}$; $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, $150 \text{ mg}\cdot\text{L}^{-1}$; $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, $10 \text{ mg}\cdot\text{L}^{-1}$; $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, $20 \text{ mg}\cdot\text{L}^{-1}$, and HCl, $50 \text{ mL}\cdot\text{L}^{-1}$. Clerol FBA 622 fermentation defoamer (Diamond Shamrock, Pennsylvania, U.S.) was added to the growth medium before inoculation. Throughout fermentations, pH 7.0 was maintained constant by automatic addition of 2 M HCl. Dissolved oxygen levels were maintained at a minimum of 50% by automatic adjustment of the stirrer speed (minimal agitation 325 rpm). The aeration rate was constant $0.5 \text{ L}\cdot(\text{L} \times \text{min})^{-1}$ sterile air. Dissolved oxygen, agitation speed and carbon dioxide evolution rate were measured and logged on-line, while samples for the determination of cell dry weight, levels of growth medium components and secondary metabolites concentrations, as well as for transcriptome and proteome analysis were withdrawn throughout the fermentation trials. For proteome analysis, 5 mL samples were taken at 21, 29, 33, 37, 41, 45, 49, 53 hours after inoculation of strain M145, and at 33, 41, 45, 49, 53, 57, 61, 65 hours after inoculation of strain M1152. Samples of bacteria culture were centrifuged ($3200 \times g$, 5 min, 4°C), and the resulting cell pellets frozen rapidly at -80°C until further processing.

Proteomics and data processing

Sample preparation and NanoUPLC-MS analysis

Samples were prepared essentially as described before (Gubbens *et al.*, 2012). Briefly, mycelia for proteome analysis were resuspended in 100 μL lysis buffer (4% SDS, 100 mM Tris-HCl pH 7.6, 50 mM EDTA) and disrupted by sonication. Total protein was precipitated using the chloroform-methanol method (Wessel & Flugge, 1984), and the proteins dissolved in 0.1% RapiGest SF surfactant (Waters, Massachusetts, U.S.) at 95°C. The protein concentration was measured at this stage using BCA method. Protein samples were then reduced by adding 5 mM DTT and incubate at 60°C for 30 min, followed by thiol group protection with 21.6 mM iodoacetamide incubation at room temperature in dark for 30 min. Then 0.1 μg trypsin (recombinant, proteomics grade, Roche, Bavaria, Germany) per 10 μg protein was added, and samples were digested at 37°C overnight. After digestion, trifluoroacetic acid was added to 0.5% and samples were incubated at 37°C for 30 min followed by centrifugation to degrade and remove RapiGest SF. Peptide solution containing 8 μg peptide was then cleaned and desalted using STAGE-Tipping (Rappsilber *et al.*, 2007). Briefly, 8 μg of peptide was loaded on a conditioned StageTip with 2 pieces of 1 mm diameter SDB-XC plug (Empore, Minnesota, U.S.), washed twice with 0.5% formic acid solution, and eluted with elution solution (80% acetonitrile, 0.5% formic acid). Acetonitrile was then evaporated in a SpeedVac. Final peptide concentration was adjusted to 40 $\text{ng}\cdot\mu\text{L}^{-1}$ using sample solution (3% acetonitrile, 0.5% formic acid) for analysis.

200 ng digested peptide was injected and analysed by reversed-phase liquid chromatography on a nanoAcquity UPLC system (Waters, Massachusetts, U.S.) equipped with HSS-T3 C18 1.8 μm , 75 μm \times 250 mm column (Waters, Massachusetts, U.S.). A gradient from 1% to 40% acetonitrile in 110 min (ending with a brief regeneration step to 90% for 3 min) was applied. [Glu¹]-fibrinopeptide B was used as lock mass compound and sampled every 30 s. Online MS/MS analysis was done using Synapt G2-Si HDMS mass spectrometer (Waters, Massachusetts, U.S.) with an UDMS^E method set up as described in (Distler *et al.*, 2014).

Quantification and statistical analysis

Raw data from all samples were first analysed using the vander software ProteinLynx Global SERVER (PLGS) version 3.0.3. Generally, mass spectrum data were generated using an MS^E processing parameter with charge 2 lock mass 785.8426, and default energy thresholds. For protein identification, default workflow parameters except an additional acetyl in N-terminal variable modification were used. Reference protein database was downloaded from GenBank with the accession number NC_003888.3. The resulted dataset was imported to ISOQuant version 1.8 (Distler *et al.*, 2014) for label-free quantification. Default high identification parameters were used in the quantification process.

TOP3 quantification was filtered to remove identifications meet these two criteria: 1. identified in lower than 70% of samples of each strain and 2. sum of TOP3 value less than 1×10^5 . Cleaned quantification data was further subjected to DESeq2 package version 1.22.2 (Love *et al.*, 2014) and PCA was conducted after variance stabilizing transformation (vst) of normalized data.

Proteome data availability

The proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol *et al.*, 2018) partner repository with the dataset identifier PXD013178 and 10.6019/PXD013178.

Acknowledgments

We would like to acknowledge Bogdan I. Florea of Leiden University, Leiden, The Netherlands, for running and for monitoring the proteome measurements, and the bio-organic synthesis group at Leiden University for providing the opportunity to use their instrumentation. This study was conducted in the frame of ERA-net for Applied Systems Biology (ERA-SysAPP) project SYSTERACT and the project INBioPharm of the Centre for Digital Live Norway