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

Du, C.

Chapter 7

General discussion and future perspectives

Actinobacteria are well known for their complex life cycle and the ability to produce bioactive natural products (Barka *et al.*, 2016, Hopwood, 2007, Bérdy, 2005). Due to the extensive screening of these bacteria in the 20th century, the frequency of finding truly novel antibiotics has decreased dramatically. Therefore, new systematic approaches should deliver the goods in terms of finding the novel drugs needed to counteract the problems associated with the emerging antibiotic resistance. Given that conventional NP screening methods identify primarily those compounds that are produced at higher concentrations under laboratory conditions, new methods are required to uncover the bioactive compounds that are specified by cryptic BGCs (Baltz, 2007). From a metabolomics perspective, this means that we need to improve the efficiency in dereplication (Wu *et al.*, 2015b, Krug & Muller, 2014). To allow us to identify the NPs produced from BGCs that are cryptic under laboratory conditions, we need to discover ways to activate them. Mutagenesis, changing culturing conditions and mimicking environmental factors are three main methods to achieve undirected activation of BGCs. The mutagenesis method is exemplified in **Chapter 3**, where streptomycin-resistant mutants were generated, and strains with enhanced or reduced antibiotic activity were selected for analysis. The sought-after BGC was then identified by correlating the production of the bioactive metabolite using LC-MS with the expression of BGCs as seen by proteomics. Alternatively, multiple small molecules mimicking environmental factors may be applied to activate antibiotic production, such as shown for the promising NP producer *S. roseifaciens* in **Chapter 4**. This study revealed eliciting strategies that may be applied in a systematic way.

Systems biology is a key element in NP research because it allows us to understand the biological system as a whole, which takes advantage of the huge technological advances that were missing in the past (Butcher *et al.*, 2004). The recent advances in 'omics technologies, especially transcriptomics, metabolomics, and proteomics, has made systems-level investigation more achievable. Of these technologies, proteomics is important for the study of NP biosynthesis as it allows scientists to determine the expression levels of the metabolic enzymes (Du & van Wezel, 2018). As shown in **Chapter 4**, statistical analysis revealed that jasmonic acid, N-acetylglucosamine, chitosan and benzoic acid induced a surprisingly similar response at whole proteome level in *S. roseifaciens*. However, this did not lead to a similar response in the antimicrobial activity, indicating that this activity might relate to the changes in a few key enzymes, small differences in the growth conditions between experiments, or the effects of elicitors on growth rate or morphology. Second, gene ontology enrichment analysis gives finer detail of the samples. The results showed that hydroxycoumarin, which is a good elicitor of antimicrobial activity, probably affects energy metabolism in *S. roseifaciens*. Finally, high-resolution proteomics studies on the short-term responses to JA identified a gene cluster (*jar*) that relates to the antimicrobial activity of *S.*

roseifaciens, though it does not directly correlate to the biosynthesis of the bioactive metabolites.

Expressing BGCs to obtain more insights into the metabolites they specify is not always straight forward, especially for those from non-culturable organisms or identified in metagenomic studies. In order to access this part of the NP repository, heterologous expression of BGCs in an optimised host is an important method (Nepal & Wang, 2019, Gomez-Escribano & Bibb, 2011). It has been shown to be effective in finding novel antibiotics (Du *et al.*, 2013), determining the boundaries of the BGC (Komatsu *et al.*, 2013, Liu *et al.*, 2018), creating artificial natural products by combining different BGCs (Park *et al.*, 2011), and characterizing functions of individual genes (Waldman *et al.*, 2015). However, choosing a suitable host for targeted expression of BGCs remains difficult. This fact, together with other technical difficulties such as transfer of large gene fragments and the requirement of biosynthetic repertoire from the natural producer that may be lacking in the new host, slows down the application of heterologous production platforms. Furthermore, what is the basis to decide which BGC to prioritise, if we do not know what the cognate NP is? Gene synthesis is still relatively expensive and hence high-throughput gene synthesis is not commonplace. In attempts to solve these issues, an background-reduced strain *S. coelicolor* M1152 was made to maximise production and at the same time reduce the complexity of extracts (Gomez-Escribano & Bibb, 2011). However, the optimization of this strain is mainly based on experience, while systems level details about the effect of the optimization process was missing so far. Therefore, we performed a time-series multi-omics study in a fully defined fermentation system and analysed the samples via high resolution proteomics (**Chapter 5**). Significant differences were observed in the global protein profiles between the background-reduced strain and its wild-type parent, including many members of the regulon of the global transcription factors PhoP, GlnR, and ScbR. We also found upregulation of ectoine biosynthetic enzymes, which may reflect increased osmotic stress in *S. coelicolor* M1152. These findings will be beneficial for further rational design of this strain or provide guidance to the optimization of other promising host strains.

The first fully sequenced model organism in the major antibiotic producing phylum, Actinobacteria, is *S. coelicolor* (Bentley *et al.*, 2002). Its ability to produce many different antibiotics, in particular PKS and NRPS, suggests it is a logical choice as host for the heterologous expression of BGCs. *S. coelicolor* has been the model organism for research on streptomycetes since the 1960s (Hoskisson & van Wezel, 2019) and a large proportion of our understanding of the regulatory networks that control secondary metabolism is obtained from this species (Rigali *et al.*, 2018, Urem *et al.*, 2016). This regulatory network is also closely correlated with the development of *Streptomyces* (van Wezel & McDowall, 2011).

An interesting protein is SCO1839, which is member of a yet unstudied protein family that is unique to Actinobacteria. Previous studies on developmental genes of *S. coelicolor* suggested that SCO1839 is involved in the regulation of sporulation (Kim *et al.*, 2015). **Chapter 6** provides a detailed analysis on the DNA binding properties of SCO1839. We found that SCO1839 is a likely nucleoid associated protein (NAP) that binds to a very small DNA sequence centred around the palindrome GATC. As expected from the short motif, the binding site of SCO1839 is widespread on the chromosome. Besides SCO1839, many other small DNA binding proteins are encoded by the *S. coelicolor* genome that control development, as exemplified by BldC (Hunt *et al.*, 2005). Its mode of DNA binding appears to be distinct from that of other regulators, involving asymmetric head-to-tail oligomerization on direct repeats that results in dramatic DNA distortion (Schumacher *et al.*, 2018, Bush *et al.*, 2019). Unlike SCO1839, no significant binding motif has so far been proposed for BldC. However, just like the DNA binding sites of BldC vary significantly in length, the lengths of the SCO1839 target sequences also vary, as shown by the ChIP-Seq data. Thus, we need to elucidate the DNA-binding model for SCO1839 and compare it with that of BldC and other actinobacterial regulators. *Streptomyces* have many different NAPs whose functions have not yet been fully deciphered. These proteins not only provide structure to the nucleoid but also perform important roles in the regulation of various cellular processes. The study of SCO1839 and its specific binding motif has provided a new route to decode function of NAPs in *Streptomyces*. It also emphasised the importance of NAPs in participating development processes of *Streptomyces*, which are major regulatory factors in secondary metabolism.

In conclusion, this thesis describes the application of advanced proteomics and other 'omics technologies in NP research in *Streptomyces*, in particular as part of systems-wide studies. This includes the proteomining technology to connect BGCs to a bioactivity of interest, studies of proteome-level responses of *Streptomyces* to small molecules mimicking environmental signals, and analysis of the consequences of strain design approaches for the optimization of *S. coelicolor* as heterologous host for NP production. Additionally, a new family of small DNA binding proteins that might function in epigenetic regulation of antibiotic production was described, combining molecular biology and bioinformatics analysis. Nowadays, new high-throughput methods, and especially 'omics techniques, generate magnitudes more data than ever before. The challenge thereby lies in combining multi-dimensional datasets in such a way that we can answer major biological questions. I hope that this thesis has made a contribution to fulfil this daunting task. Systems-wide approaches are key to obtaining a more comprehensive understanding of the regulatory networks that control NP biosynthesis and the complex biology of *Streptomyces*.