Chapter 3

Application of systems biology methods for the identification of novel natural products in *Streptomyces* species

Chao Du, Changsheng Wu, Jacob Gubbens, Koji Ichinose, Young Hae Choi, and Gilles P. van Wezel

Adapted from:

Wu, C., Du, C., Gubbens, J., Choi, Y.H., and van Wezel, G.P. (2015) Metabolomics-driven discovery of a prenylated isatin antibiotic produced by *Streptomyces* species MBT28. J. Nat. Prod. 78: 2355-2363. DOI: 10.1021/acs.jnatprod.5b00276

Wu, C., Du, C., Ichinose, K., Choi, Y.H., and van Wezel, G.P. (2017) Discovery of C-Glycosylpyranonaphthoquinones in *Streptomyces* sp. MBT76 by a combined NMR-based metabolomics and bioinformatics workflow. J. Nat. Prod. 80: 269-277. DOI: 10.1021/acs.jnatprod.6b00478

Abstract

Actinomycetes are a major source of antimicrobials, anticancer compounds, and other medically important natural products. Genome mining has revealed an extensive biosynthetic potential of Actinomycetes. Major challenges in the screening of these microorganisms are to activate the expression of cryptic biosynthetic gene clusters and the development of technologies for efficient dereplication of known molecules. Here we applied a systems biology approach consisting of NMR-based metabolomics, proteomics, bioinformatics, and targeted activation of cryptic gene clusters as an effective pipeline for bioactive microbial natural products. We achieved the identification of a previously unidentified isatin-type antibiotic produced by *Streptomyces* sp. MBT28, and novel C-glycosyl-pyranonaphthoquinones in *Streptomyces roseifaciens* MBT76 (NCCB 100637, DSM 106196). Genome mining and proteomics study on production fluctuated mutants helped in 7-prenyl isatin and related isatin type antibiotics characterization. The *qin* gene cluster from *S. roseifaciens* was highlighted by genome mining as a potential producer of novel glycosylated pyranonaphthoquinones. An approach consisting of combined NMR-based metabolomics and regulator over-expression was applied to activate the biosynthetic gene cluster and identify novel C-glycosyl-pyranonaphthoquinones.

Introduction

Discovery and development of antibiotics to fight bacterial diseases is one of the greatest triumphs of modern medicine. Bacterial infections were once conquered by the use of antibiotics. However, the increasing antimicrobial resistance now again form a major threat to human health (WHO, 2014). In this battle, bioactive natural products (NPs) provide us prolific weapon sources in the battle against drug resistant bacteria. As producers of approximately two-thirds of all known antibiotics and many other medically relevant natural products, actinomycetes are a major source of bioactive NPs (Barka *et al.*, 2016, Hopwood, 2007). Whole genome sequencing revealed that actinomycetes harbour numerous silent and hence likely untapped biosynthetic gene clusters (BGCs) that may not yet be associated with known metabolites. Indeed, even the extensively studied model organism *Streptomyces coelicolor* was shown to possess a far greater potential than anticipated (Bentley *et al.*, 2002). Furthermore, bioinformatics tools developed specifically for mining genome sequences for the identification of biosynthetic gene clusters allow the prediction of the chemical output on the basis of accumulated biosynthetic knowledge (Medema *et al.*, 2015, Fedorova *et al.*, 2012). These developments in genome mining mark the start of a new era of genomics-based drug discovery, with the potential of greatly expanding the chemical space of bioactive natural products. A bottleneck is that many of the biosynthetic pathways uncovered by genome sequencing are in a dormant state under routine laboratory conditions, generally referred to as cryptic or silent gene clusters, and specific approaches to activate their expression are required (Rutledge & Challis, 2015, Zhu *et al.*, 2014a).

One way of activation cryptic BGCs utilizes different culturing conditions, allowing fluctuation of the production of possibly cryptic antibiotics, followed by the metabolic profiling-based identification of the bioactivity of interest. Systems biology methods can then step-in and reveal the genetic source of the bioactive compounds and its production mechanism. One challenge in this research path lies in finding the appropriate chemical triggers or ecological cues to elicit the production of cryptic antibiotics (Zhu *et al.*, 2014a, Yoon & Nodwell, 2014). Strategies that have been employed include heterologous expression (Schümann & Hertweck, 2006), the use of chemical elicitors (Zhu *et al.*, 2014b, Craney *et al.*, 2012), and inducing antibiotic resistance (Ochi *et al.*, 2014, Hosaka *et al.*, 2009). Selection of streptomycin- or rifampicin-resistant mutants, typically caused by mutations in *rpsL* (ribosomal protein S12) or *rpoB* (RNA polymerase β-subunit), respectively, enhances the production of antibiotics (Hosaka *et al.*, 2009, Wang *et al.*, 2008, Tamehiro *et al.*, 2003) and may also lead to the production of novel antibiotics (Fu *et al.*, 2014). Another method to find and activate the production of cryptic antibiotics is through direct genome mining and genetic manipulation of interesting BGCs. In bacteria, biosynthetic genes are often clustered in large gene

clusters, with core biosynthetic enzymes and those patriated in decorating the products. In addition, the controlling regulators as well as enzymes related with the resistance are also encoded in the gene cluster. The number of functional enzymes in BGCs being revealed is increasing fast, which made genome mining tools including antiSMASH (Blin *et al.*, 2017), PRISM (Skinnider *et al.*, 2016) and many others capable of more and more accurate BGC discovery (Du & van Wezel, 2018). New biochemical potentials normally lie in the new combinations of decorating enzymes found by genome mining, studying the new potential in genome mining results becomes a good method in finding BGC of interest. When a potentially new BGC is located, activating its activity becomes essential again. This can be achieved by fluctuating the growth environments, as is discussed above. But now we can activate these cryptic BGCs by direct manipulating the regulators located in and out of the gene cluster. Generally, in actinomycetes, the transcription of genes encoding the biosynthetic machinery for secondary metabolites involves multiple regulatory cascades and networks (van Wezel & McDowall, 2011). The regulatory signals are transmitted through global regulatory networks and ultimately transmitted to the pathway-specific regulatory genes that control the expression of the biosynthetic genes (Bibb, 2005). Molecular biology strategies, including manipulating the pleiotropic regulators and pathway-specific regulators to circumvent the regulatory can possibly trigger the biosynthesis of the corresponding natural products (Rutledge & Challis, 2015, Abdelmohsen *et al.*, 2015).

Once differential expression of a bioactivity is achieved in a producing organism, e.g., by varying culturing conditions or by making mutants, the next challenge is to rapidly establish whether the natural product is sufficiently novel to warrant full structure elucidation. Metabolomics approaches globally identify the low molecular weight metabolites in biological samples. Chemical profiling techniques using LC-MS or NMR followed by multivariate data analysis allow scientists to compare and detect molecules that are differentially synthesized between diverse biological samples, thereby effectively narrowing down the search for the sought-after biomarker and avoiding chemical redundancy in an early stage (Wu *et al.*, 2015b, Krug & Muller, 2014). At the same time, the more readily available bioinformatic tools allows prediction of the metabolite produced by the biosynthetic gene clusters, and this knowledge can be directly related to the ^{1}H NMR spectrum by examining the expected chemical shift and/or splitting pattern of typical protons in the predicted molecular motifs. Together with proteomics and/or transcriptomics technologies applied, it is possible to bridge the gap between bioinformatics-driven gene cluster analysis and experimental NPs discovery (Wu *et al.*, 2016a, Du & van Wezel, 2018).

In this chapter, we first describe the use of a proteomics-based genome mining approach in combination with multivariate data analysis correlated bioactivity of

NMR peaks, facilitating the characterization of the previously undescribed antibiotic 7-prenylisatin and its possible BGC. In a workflow of using NMR-based metabolomics and bioinformatics to identify novel pyranonaphtoquinones in *S. roseifaciens*, we made direct changes in the genome to consecutively express the pathway specific activator of a cryptic type II PKS gene cluster (designated *qin*). This resulted in the activation of the production of a family of pyranonaphtoquinones with intriguing chemical architecture.

Results and discussion

Streptomycin resistant mutant of *Streptomyces* sp. MBT28 with elevated bioactivity caused by 7-prenylisatin production

Streptomyces sp. MBT28 is a strain of our collection of actinomycetes from remote mountain soils (Zhu *et al.*, 2014b). In this collection, *Streptomyces* sp. MBT28 was identified as a producer of various antibiotics, one of which was identified as borrelidin (Zhu *et al.*, 2014b). However, other secondary metabolites were produced in very low quantities. In an attempt to enhance antibiotic production, two streptomycin resistant (Str^R) mutants are selected from SFM agar plates containing 10 $\mu g \cdot mL^{-1}$ streptomycin. One mutant with a white phenotype (*Streptomyces* sp. MBT28-91) shows strong bioactivity and another with a wildtype phenotype (*Streptomyces* sp. MBT28-30) was chosen (Figure 1). They both have less enhanced bioactivity. We checked the streptomycin-resistant hot spot gene *rpsL* and *rsmG* in the two Str^R strain by sequencing, only *Streptomyces* sp. MBT28-30 carried one mutation on *rsmG* gene.

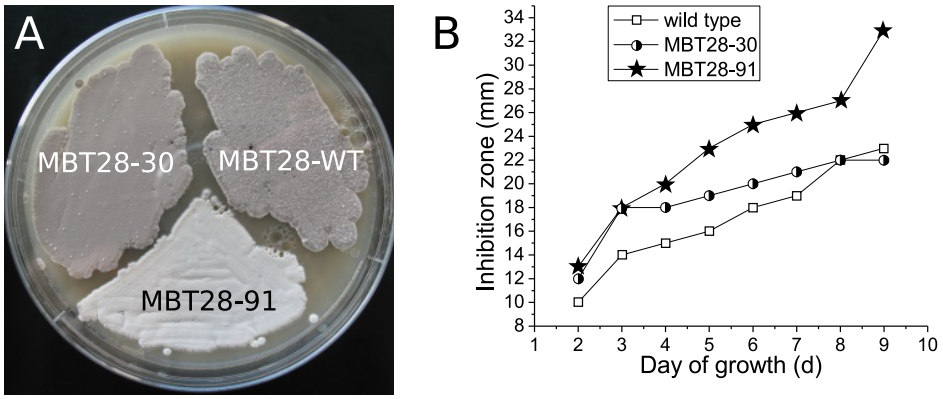

Figure 1. Antibiotic activity of *Streptomyces* sp. MBT28 and its streptomycin-resistant mutants. Morphological appearance of *Streptomyces* sp. MBT28 (WT) and its Str^R mutants *Streptomyces* sp. MBT28-30 and *Streptomyces* sp. MBT28-91 grown on SFM agar for 5 days at 30°C (A), and time course of inhibition activity against *Bacillus subtilis* (B). For activity test, strains were grown on MM agar supplemented with 0.5% mannitol and 1% glycerol, samples taken at 24 h intervals (*X* axis) were overlaid with *B. subtilis* to give inhibition zone (*Y* axis). MBT-28-91 showed much better activity than those of *Streptomyces* sp. MBT28 and *Streptomyces* sp. MBT28-30.

NMR spectroscopy, particularly ^1H NMR, commonly used for the analysis of industrial natural products, is suited for such analyses as it is fast, reproducible, and benefits from a relatively easy sample preparation (Kim *et al.*, 2010). Multivariate data analysis (MDA) were used to reduce the dimensionality of

multivariate dataset and thus to discriminate samples. For this purpose, *Streptomyces* sp. MBT28, *Streptomyces* sp. MBT28-30, *Streptomyces* sp. MBT28-91 were grown on MM agar plates supplemented with 0.5% (w/v) mannitol and 1% (w/v) glycerol as carbon sources. Ten biological replicates were performed for each sample. ^{1}H NMR was done, and the results show distinct difference of antibiotic-overproducing mutant *Streptomyces* sp. MBT28-91. Multivariate data analysis of NMR signals with bioactivity shows one clear signal related with the increased bioactivity in *Streptomyces* sp. MBT28-91. NMR-guided fractionation (Grkovic *et al.*, 2014) tracing the characteristic signal resulted in the identification of 7-prenylisatin (Figure 2). The NMR-based metabolomics and compound identification are described in detail in (Wu *et al.*, 2015a).

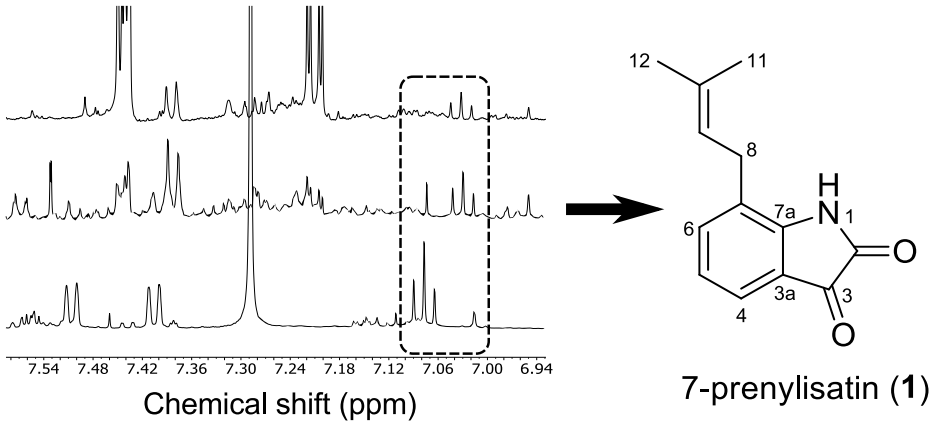

Figure 2. NMR-guided purification of 7-prenylisatin from *Streptomyces* sp. MBT28-91. Characteristic proton signals δ 7.03 (t, J = 7.2 Hz) that correlated to the high bioactivity are boxed, which were used as tracking signals for chromatographic separation (dash boxed). Chemical shift divergence was due to the different deuterated solvents used (methanol-d4 for top and middle rows, and $CDCl_3$ for bottom row).

Isatins exhibit a variety of biological activities such as antibacterial, antifungal, and anticancer properties (Vine *et al.*, 2009). However, only a few isatin-type compounds of microbial origin have been identified (Sobolevskaya *et al.*, 2009, Breinholt *et al.*, 1996, Graefe & Radics, 1986). 5-prenylisatin from the fungus *Chaetomium globosum* was reported to have antifungal activity (Breinholt *et al.*, 1996), while 6-prenylisatin was the first example of an isatin derivative discovered in a streptomycete, namely *Streptomyces albus*, which showed antimicrobial activity against Gram-positive bacteria, with an MIC against *Bacillus subtilis* of 20 $\mu g \cdot mL^{-1}$ (Graefe & Radics, 1986). Purified 7-prenylisatin was tested for its efficacy against *B. subtilis* using an MTP-based MIC test (See Materials and Methods section). This showed that 7-prenylisatin acted as an antibiotic with an MIC of around 25 $\mu g \cdot mL^{-1}$ against *B. subtilis*, which is similar to that of 6-prenylisatin. No

antimicrobial activity was observed against the Gram-negative *Escherichia coli* K12.

Identification of the biosynthetic gene cluster for 7-prenylisatin

To characterize the 7-prenylisatin biosynthetic pathway, the genome of *Streptomyces* sp. MBT28 was sequenced, and the resulting draft genome sequence was annotated as described before (Girard *et al.*, 2014). The genes responsible for forming the 1*H*-indole-2,3-dione scaffold (isatin) have not previously been identified in bacteria. However, the prenyltransferase (PTase) catalysing the initial prenylation of the indole nucleus has been well characterized. Genes for indole PTases have been found in actinomycetes in two different genomic contexts (Ozaki *et al.*, 2013): Type A in combination with a tryptophanase, and Type B in combination with a flavin-dependent monooxygenase (FMO). For instance, in *Streptomyces sp*. SN-593, IptA (type A) installs a dimethylallyl group on the C-6 position of the tryptophan indole core, while in *Streptomyces coelicolor* A3(2), a Type B PTase prenylates tryptophan at C-5 (Takahashi *et al.*, 2010). Interestingly, the *Streptomyces* sp. MBT28 genome contained both types of gene clusters. However, considering that a tryptophanase is required for cleavage of the side chain (Figure 3A), the type A cluster was the most likely candidate for the bioassembly of 7-prenylisatin.

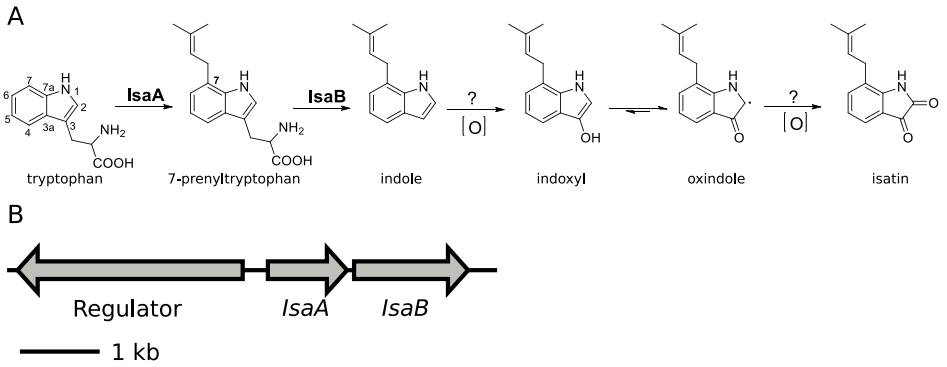

Figure 3. Proposed biosynthetic pathway of 7-prenylisatin (A) and its gene cluster (B).

To validate that 7-prenylisatin is indeed synthesized by a Type A indole PTase, we applied the *natural product proteomining*. This research pipeline makes use of the fact that the expression of biosynthetic proteins correlates closely with the bioactivity of interest. Then quantitative proteomics might possibly show the correlated protein levels and are therefore correlated to changes in the production of the bioactive compound (Gubbens *et al.*, 2014). For this, protein extracts of *Streptomyces* sp. MBT28 and its mutants *Streptomyces* sp. MBT28-30 and

Streptomyces sp. MBT28-91 were obtained from cultures grown on MM agar plates, followed by stable isotope dimethyl labelling, and subsequent LC-MS/MS analysis. Relative quantification of protein levels between the three strains were obtained from the result. In total, 1,382 proteins were identified, and the expression levels of 1,038 proteins could be obtained. Based on the intensity-dependent significance-B value (Cox & Mann, 2008), 177 of the quantified proteins were significantly enhanced or reduced ($p < 0.05$) in at least one comparison (Table S1). Among the proteins that were significantly enhanced in strain *Streptomyces* sp. MBT28-91 compare to the wild-type strain, we found the two key enzymes of the Type A indole PTase gene cluster, namely the aromatic prenyltransferase (IsaA) and the tryptophanase (IsaB) (Table 1 and Figure 3B), were all significantly upregulated in *Streptomyces* sp. MBT28-91. While in *Streptomyces* sp. MBT28-30, the two enzymes were upregulated only in a small scale. The change of IsaB is still significant but the fold change is in a much lower scale compared to *Streptomyces* sp. MBT28-91. As *Streptomyces* sp. MBT28-91 has much higher 7-prenylisatin production than *Streptomyces* sp. MBT28-30, and the production of the later strain is slightly higher than wild type, a strong correlation was found between the amount of 7-prenylisatin production and the expression level of the Type A indole PTase gene cluster. In contrast, none of the proteins encoded by the Type B cluster were detected, which indicates a possibility that this cluster was not expressed under these conditions. Although the non-detection of Type B indole PTase gene cluster in proteomics is not a definitive evidence of zero presence, but taken together the bioinformatics analysis, the biosynthetic pathway and the strong correlation on Type A cluster with 7-prenylisatin production, it is suggested that the Type A gene cluster is responsible for the biosynthesis of 7-prenylisatin.

Table 1. Protein level differences (in fold change mutant/wt) of the Type A indole PTase gene cluster as determined by quantitative proteomics.

Protein	MBT28-91/WT (2log)	Significance-B	MBT28-30/WT (2log)	Significance-B	Description*
Isa Regulator	ND†	ND	ND	ND	Transcriptional regulator
IsaA	3.0	0.023	0.9	0.116	Aromatic prenyltransferase
IsaB	2.8	0.031	1.6	0.040	Tryptophanase

* based on BLAST homology searches

† Not detected

Thus, a biosynthetic route for 7-prenylisatin was proposed based on the existing literature for related isatin compounds (Figure 3A). Three connected enzymatic reactions are required for the biosynthesis of 7-prenylisatin from tryptophan, namely a prenylation followed by carbon bond cleavage and an oxidation reaction. It is yet unclear which gene(s) are responsible for the oxidation of indole into isatin

backbone, as no such genes were found within or in close proximity of the *isa* gene cluster (Table S1). Three proteins with an oxidation-related PQQ domain were all found significantly up-regulated (with p-value < 0.05) in the overproducing strain as compared to the parental strain. Further analysis is required to establish whether one of these enzymes may execute the oxidation reaction to generate the final isatin moiety.

Identification and activation of the cryptic type II PKS gene cluster (*qin*) for glycosylated pyranonaphthoquinones

Streptomyces roseifaciens (van der Aart *et al.*, 2019), previously named *Streptomyces* sp. MBT76, which originates from the Qinling mountains in China, was identified as a prolific producer of antibiotics, including those with efficacy against multiple Gram-positive and Gram-negative multi-drug resistant pathogens (Zhu *et al.*, 2014b). Further detailed metabolic characterization of the strain identified many natural products often with interesting chemistry, including isocoumarins, prodiginines, acetyltryptamine, and fervenulin, among others (Wu *et al.*, 2016c). To unravel the biosynthetic potentials of *S. roseifaciens*, it was subjected to Illumina/Solexa whole genomic sequencing, and the genome was assembled in 13 contigs, with a total genome size of 8.64 Mb. In total, 7974 coding sequences (CDS) were predicted using the GeneMark algorithm (Lukashin & Borodovsky, 1998). Analysis of the contigs by antiSMASH (Blin *et al.*, 2013) presented a possible 55 putative biosynthetic gene clusters specifying secondary metabolites, 22 of which encoding polyketide synthases (PKS).

One 41 kb biosynthetic gene cluster for a type II PKS, designated *qin* was of particular interest considering its potential to specify pyranonaphthoquinones. These molecules represent a well-studied family of aromatic polyketides with highly complex chemical architecture and pronounced bioactivities (Oja *et al.*, 2015, Metsä-Ketelä *et al.*, 2013), including the representative members actinorhodin (Okamoto *et al.*, 2009), medermycin (Ichinose *et al.*, 2003), and granaticin (Ichinose *et al.*, 1998). In the *qin* gene cluster, besides the central PKS genes that are responsible for the biosynthesis of the pyranonaphthoquinone backbone, the presence of genes for the deoxyaminosugar D-forosamine strongly suggested that the end product should be glycosylated. This genetic organization was similar to the clusters for the synthesis of the glycosylated pyranonaphthoquinone medermycin in *Streptomyces* sp. AM-7161 (*med*) (Ichinose *et al.*, 2003) and granaticin in *Streptomyces violaceoruber* Tü22 (*gra*) (Ichinose *et al.*, 1998). Coclustering of glycosylation-associated biosynthetic genes with those for the aglycones is typical of microbial genomes, which facilitates matching the biosynthetic gene cluster to the corresponding NPs (Kersten *et al.*, 2013). In comparison, *qin*-ORF29 encoding an NADPH-dependent FMN reductase was absent in either the *med* or *gra* biosynthetic gene clusters, and glycosylation with a D-forosamine is unprecedented in the pyranonaphthoquinone

family. This promoted an investigation into the potentially novel product(s) of the *qin* gene cluster. Despite our previous detailed chemical investigations of *S. roseifaciens* (Wu *et al.*, 2016c, Zhu *et al.*, 2014b), the corresponding molecules had not been identified, suggesting that the gene cluster may be cryptic under the many different growth conditions that had been tested.

The *qin* gene cluster contains two putative regulatory genes, namely the SARP-family transcriptional regulatory gene *qin*-ORF11, and the TetR-family regulatory gene *qin*-ORF35. Many pathway-specific activators for secondary metabolite production in streptomycetes belong to the SARP family (Bibb, 2005), while TetR-family regulators often act as repressors (Ramos *et al.*, 2005). As an example, in the type II PKS gene cluster *aur1* of *Streptomyces aureofaciens* CCM 3239, which specifies the angucycline-family auricin (Novakova *et al.*, 2005), *aur1P* and *aur1PR3* encode SARP-family activators (Novakova *et al.*, 2011), while the *tetR*-type *aur1R* encodes a negative regulator (Novakova *et al.*, 2010). To activate the *qin* gene cluster, we opted to over-express the likely pathway-specific activator gene *qin*-ORF11. For this, the gene was amplified by PCR and inserted behind the *ermE* promoter (Bierman *et al.*, 1992) in the conjugative and integrative vector pSET152. The resulting plasmid was then introduced to *S. roseifaciens* via conjugation and integrated into the chromosome to create the recombinant derivative *S. roseifaciens*-1.

Proposed biosynthetic pathway for qinimycins based on bioinformatics analysis

Ex-conjugant *S. roseifaciens*-1 and its parent *S. roseifaciens* were cultivated in parallel in modified NMMP liquid media (Wu *et al.*, 2016c). Extracts of the fermentation were analysed by ^{1}H NMR profiling (Kim *et al.*, 2010). Some metabolites usually produced by the wild-type strain, such as 1*H*-pyrrole-2-carboxamide, acetyltryptamine, fervenulin and 2-hydroxy-3-methoxy-benzamide (Wu *et al.*, 2016c), was aborted in *S. roseifaciens*-1. While the production of 6,8-dihydroxy-3-methyl-isocoumarin was not affected by ORF11 over-expression. Conversely, the ^{1}H NMR spectra of *S. roseifaciens*-1 indicates the presence of an α,β-unsaturated ketone aryl moiety typical of pyranonaphthoquinones (Wu *et al.*, 2017). These evidences indicated that indeed the *qin* gene cluster was actively expressed in *S. roseifaciens*-1. In light of the substantial change in the metabolome of *S. roseifaciens* and *S. roseifaciens*-1, we hypothesized that *qin*-ORF11 may control multiple biosynthetic gene clusters for specialized metabolites (Wu *et al.*, 2017).

To identify the metabolic product(s) of the *qin* gene cluster, the crude extract of *S. roseifaciens*-1 was separated by semi-preparative HPLC-UV chromatography. In the ^{1}H NMR spectroscopy and further UHPLC-UV-ToF-HRMS analysis of the resulting fractions, we confirmed the presence of unknown compounds $C_{23}H_{27}NO_8$

(**1**), $C_{23}H_{29}NO_9$ (**2**), and $C_{24}H_{31}NO_9$ (**3**) with a relative abundance of 6:2:1 (Figure 4), which we designated as qinimycins.

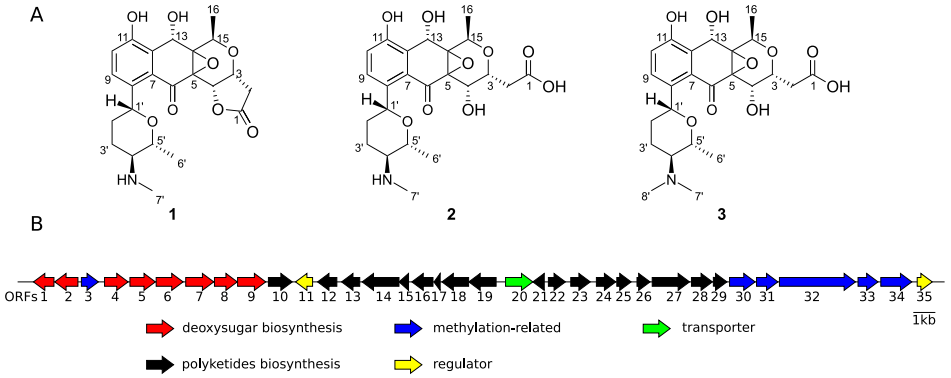

Figure 4. Secondary metabolites produced by *S. roseifaciens*-1. (A) Organization of the qin locus in *S. roseifaciens*. (B) Genes for the minimal PKS (presented in black) are similar to those for biosynthesis of the pyranonaphthoquinone kalafungin, while the eight genes in red encode enzymes for production of deoxyaminosugar D-forosamine.

The qinimycins have intriguing chemical features, representing a new branch within the extensively studied pyranonaphthoquinone family of natural products. A detailed discussion of the biosynthetic pathway and chemical novelties of qinimycins is provided in Wu *et al.* (2017).

Table 2. Antimicrobial activity of qinimycins.

Compound	Inhibition zone (mm)			
	Bacillus subtilis	*Escherichia coli*	*Staphylococcus aureus*	*Pseudomonas aeruginosa*
Qinimycins (Fr1)	15	0	20	0
AMP	23	20	10	7
APRA	10	7	7	15
NC	0	0	0	0

For Qinimycins, 25 µL of HPLC fragment 1 (Fr1) was spotted as a 2 mg·mL^{-1} solution in methanol; For AMP and APRA, 5 µL was spotted as a 1 mg·mL^{-1} solution in miliQ water. AMP, ampicillin; APRA, apramycin; NC, negative control (methanol)

The bioactivity of qinimycins was tested using agar diffusion assays. Growth inhibition was seen against the Gram-positive bacteria *B. subtilis* 168 and *Staphylococcus aureus* CECT976, but not for Gram-negative *E. coli* JM109 or *Pseudomonas aeruginosa* PAO1, thus establishing bioactivity against Gram-positive but not Gram-negative bacteria (Table 2). However, assessment of the MIC of the mixture of molecules revealed minimal inhibition concentrations (MICs) against *B. subtilis* and *S. aureus* of 50 µg·mL^{-1} and 100 µg·mL^{-1}, respectively. These

high values are indicative of very limited bioactivity of the qinimycins under the tested conditions. A biosynthetic model for qinimycins is proposed based on the functional assignments from sequence analysis and the known metabolites that are produced, see detailed discussion in Wu *et al.* (2017).

Conclusion

A major issue in harnessing the huge potential reservoir of natural products is the increasingly high frequency of re-discovery of known antibiotics used in routine screening (Wang *et al.*, 2008, Tamehiro *et al.*, 2003, Hu & Ochi, 2001, Hosoya *et al.*, 1998). This is known as the Replication Issue. One promising strategy is to ensure (i) fluctuations in the production level of the compound of interest, and (ii) perform metabolomics combined with multivariate data analysis to correlate bioactive molecules to the observed bioactivity, combined with (iii) correlate the changes in the transcriptome or proteome profiles to the bioactivity to identify candidate gene clusters (Wu *et al.*, 2015d). This strategy was successfully applied in elucidating bioactive compounds produced by *Streptomyces* sp. MBT28, resulted in the identification of the antibiotic 7-prenylisatin, which was a near silent antibiotic under routine growth conditions. To activate the production of 7-prenylisatin and elucidate its structure, streptomycin-resistant mutants were selected and subsequently analysed by NMR-based metabolomics (Nicholson & Lindon, 2008) to identify those NMR signals that correlated statistically to the enhanced bioactivity, and natural product proteomining (Gubbens *et al.*, 2014) to identify the corresponding biosynthetic proteins. Another efficient way of solving the Replication Issue and to streamline the discovery of novel molecules from actinomycetes is to combine genome mining, targeted genetic manipulation, and metabolomics analysis. Prior knowledge of NPs that can be expected based on bioinformatics thereby provides information on which biosynthetic cluster is more interesting in terms of discovering new natural products. This information also simplifies and reduces the time consumption of the chromatographic isolation process and structure determination of target compound(s). The identification of qinimycin compounds from *S. roseifaciens* provides proof of concept for this principle. Whereby activation of the cryptic *qin* type II PKS gene cluster by constitutive expression of its pathway-specific activator gene (*qin*-ORF11) followed by NMR-based metabolic profiling, identified novel glycosylated pyranonaphthoquinones. The elucidated biosynthetic pathway for the qinimycins in the genetically tractable *S. roseifaciens* offers new insights into the biosynthesis of this family of natural products.

Experimental section

Strains and culturing conditions

Streptomyces sp. MBT28 and *Streptomyces roseifaciens* MBT76 (NCCB 100637, DSM 106196) were obtained from Molecular Biotechnology culture collection, IBL, Leiden University. Spores were stored in 20% glycerol and maintained in -20°C. Soy flower agar medium (SFM) (Kieser *et al.*, 2000) was used to generate streptomycin-resistant (Str^R) mutants, grow over-expression conjugants of *S. roseifaciens*, and for preparing spore suspensions. Standard solid minimal medium (MM) (Kieser *et al.*, 2000) supplemented with both 0.5% (w/v) mannitol and 1% (w/v) glycerol was used for activity tests and to prepare samples for metabolomics and proteomics for *Streptomyces* sp. MBT28 experiments. *Bacillus subtilis* 168 was grown in Luria-Bertani broth (LB) at 37°C. *Escherichia coli* JM109 (Sambrook *et al.*, 1989) was used for routine cloning. *E. coli* ET12567 (MacNeil *et al.*, 1992) containing pUZ8002 (Kieser *et al.*, 2000) was used for introducing non-methylated DNA into *S. roseifaciens* by conjugation. The basal medium for *S. roseifaciens* growth was modified minimal liquid medium NMMP (Kieser *et al.*, 2000) without PEG6000 and containing 1% (w/v) glycerol and 0.5% (w/v) mannitol as the carbon sources, which was further supplemented with 0.8% (w/v) Bacto peptone (Wu *et al.*, 2016c). Tryptone soy broth with 10% (w/v) sucrose (TSBS) was used to grow *S. roseifaciens* mycelia as receptor for conjugation experiments. *Streptomyces* sp. MBT28 Str^R mutants were selected by plating 10 µl of spores containing 1×10^5 colony forming units (CFU) onto SFM agar plates containing streptomycin (10 $\mu g \cdot mL^{-1}$) at 30°C. Colonies typically developed after nine days. For chemical analysis of *S. roseifaciens* secondary metabolites, 50 mL modified NMMP was inoculated with 10^6 spores of *S. roseifaciens* in 250 mL Erlenmeyer flasks equipped with a spring. Cultures were incubated at 30°C with constant shaking at 220 rpm for 120 h.

Genome sequencing, assembly, and annotation

DNA was extracted from *Streptomyces* sp. MBT28 and *S. roseifaciens* as described previously (Kieser *et al.*, 2000). Genome sequencing and annotation was done essentially as described previously (Girard *et al.*, 2014). 100-nt paired-end reads were obtained and the quality of the short reads verified using FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Depending on quality, reads were trimmed to various lengths at both ends. Processed raw reads were subsequently used as input for the Velvet assembly algorithm. The resulting contigs were analysed using the GeneMark hmm algorithm with the *Streptomyces coelicolor* genome as model for ORF finding (Lukashin & Borodovsky, 1998). The genomes were additionally annotated using the RAST server with default options. BLASTP for putative function prediction and HMMER for protein-domain prediction, manually inspected for some and visualized using Artemis. The resulting genomes have been deposited at GenBank. The GenBank WGS project

accession number of *Streptomyces* sp. MBT28 is LARV00000000, for *S. roseifaciens* the accession number is LNBE00000000.

Antibiotic activity assays

Streptomyces sp. MBT28 was grown for six days on MM agar plates and overlaid with LB agar containing 100 µL of exponentially *B. subtilis* cells (OD_{600} of 0.4). The activity was initially assessed as the zone of growth inhibition after overnight incubation at 37°C. MIC values against *B. subtilis* were carried out as described before (Zhu *et al.*, 2014b). All MIC determinations were performed in triplicate. Ampicillin was used as positive control, and the solvent chloroform as the negative control.

Antimicrobial activity of qinimycins was determined according to a disc diffusion method as described before (Wu *et al.*, 2015c, Wu *et al.*, 2016b). 25 µL of Fr1 (2 $mg{\cdot}mL^{-1}$ in methanol) was spotted onto paper discs (6 mm diameter) placed on agar plates containing a soft agar overlay with indicator bacteria. Indicator bacteria were *Bacillus subtilis* 168, *E. coli* ASD19, *Staphylococcus aureus* CECT976, or *Pseudomonas aeruginosa* PAO1. Ampicillin and apramycin were used as positive controls, whereby 5 µL was spotted of a 1 $mg{\cdot}mL^{-1}$ solution in miliQ water. The solvent methanol was used as the negative control. After incubation at 37°C for 18 h, growth inhibition zones (in mm) were recorded as antimicrobial activity. The MIC assay against *B. subtilis* and *S. aureus* were carried out in 96-well plate by serial double dilution method, as previously described (Zhu *et al.*, 2014b). All MIC determinations were performed in duplicate.

Primers and plasmid construction

PCRs were performed in a minicycler (MJ Research, Massachusetts, U.S.), using Phusion polymerase (Stratagene, California, U.S.) as described before (Colson *et al.*, 2007). Sequences of all oligonucleotides used in this study was listed in Table S2. For cloning, *rpsL* was amplified from *Streptomyces* sp. MBT28 genomic DNA using oligo nucleotides *rpsL*_For and *rpsL*_Rev; *rsmG* was amplified using oligonucleotides *rsmG*_For and *rsmG*_Rev. *qin*-ORF11 was amplified by PCR from *S. roseifaciens* genomic DNA as described before (Colson *et al.*, 2007) using primers: SC06_0044_F_EcoRI and SC06_0044_R_EcoRI_NdeI. The PCR product was subcloned as an *Eco*RI fragment and subsequently constructed as an *Nde*I-*Hin*dIII with constitutive *erm*E promoter fragment into pSET152 (Bierman *et al.*, 1992), which integrates at the genome φC31 sites. The final recombinant plasmid pCSW01 (pSET152/*erm*E/*qin*-ORF11) was made and *E. coli* ET12567/pUZ8002 was transformed with it. This transformant was then used to conjugate two-day old mycelia of *S. roseifaciens*. The ex-conjugants were confirmed by PCR using primers SC06_0044_SF and SC06_0044_SR.

Metabolomics sampling and preparation

Streptomyces sp. MBT28 was grown on MM agar plates supplemented with 0.5% mannitol and 1% glycerol. The agar with mycelia was cut into small pieces and extracted with ethyl acetate by soaking in the solvent overnight at room temperature. Samples were dried under reduced pressure at 40°C and re-dissolved in methanol. The solvent was then evaporated at room temperature under nitrogen gas flow. 50 mL cultures of *S. roseifaciens* or *S. roseifaciens*-1 were harvested by centrifugation at 4000 rpm for 10 min, and the supernatant extracted twice with 20 mL of ethyl acetate (EtOAc). The organic phase was washed with 30 mL of water and subsequently dried with 5 g of anhydrous Na_2SO_4. EtOAc was removed under vacuum at 38°C and the residue was dissolved in 2.0 mL of EtOAc. The solvent was then evaporated at room temperature under nitrogen gas. The extracts from both experiments were subsequently dipped into liquid nitrogen and lyophilized using a freeze dryer (Edwards Ltd., Crawley, England). Crude extracts were partitioned between methanol and n-hexane to remove lipids. Ten replicates were performed for *Streptomyces* sp. MBT28 samples and fifteen replicates were performed for *S. roseifaciens* samples.

Metabolomics, NMR guided separation, compound purification

For methods on NMR-based metabolomics and compound identification, we refer to the original publications by Wu, Du et al. (2015a, 2017)).

Quantitative proteomics analysis

To prepare samples for proteomics, strains were grown on MM agar plates overlaid with 0.1 μm polycarbonate filter papers prior to inoculation. After six days of incubation at 30 °C, mycelia were scraped from the filters. Protein extraction, in-solution digestion, C18 column dimethyl labelling, strong cation-exchange chromatography (SCX) fractionation, and LC-MS/MS analysis were performed as previously described (Gubbens *et al.*, 2014, Li *et al.*, 2013). Generally, mycelia were washed and then sonicated for disruption of cell wall. After removal of cell debris, protein concentration was measured using a Bradford assay, and 0.167 mg of protein was precipitated for each sample. Proteins were dissolved using RapiGest SF Surfactant (Waters, Massachusetts, U.S.) and treated with iodoacetamide before trypsin digestion. Peptides were dimethyl labelled on C-18 resin and the three samples mixed to obtain 0.5 mg of peptides for fractionation by SCX. 24 fractions were collected and subjected to LC-MS/MS analysis on an LTQ-orbitrap MS (Thermo, Pennsylvania, U.S.) (Li *et al.*, 2013). Data analysis was performed using MaxQuant 1.4.1.2 (Cox & Mann, 2008). Dimethyl labelling was selected as quantification method, carbamidomethylation of cysteine was selected as fixed modification, and oxidation of methionine was set as variable modification (Gubbens *et al.*, 2012). MS/MS spectra were searched against a database of translated coding sequences. Peptide and protein identification FDR were set to 0.01, and a minimum of two quantification events was specified to obtain a protein

quantification. To determine the significance of the found expression ratios, intensity-dependent B-significance values (Cox & Mann, 2008) were calculated.

Acknowledgments

We are grateful to H. Zhu for discussions. The work was supported by a grant from the Chinese Scholarship Council to C.S.W. and by VENI grant 11890 from The Netherlands Technology Foundation STW to J.G.