Native and engineered promoters in natural product discovery

Journal:	*Natural Product Reports*
Manuscript ID	NP-HIG-01-2016-000002.R2
Article Type:	Review Article
Date Submitted by the Author:	06-Jul-2016
Complete List of Authors:	Myronovskyi, Maksym; Helmholtz-Zentrum fur Infektionsforschung GmbH Luzhetskyy, Andriy; Universität des Saarlandes, Dept. of Pharmaceutical Biotechnology

Native and engineered promoters in natural product discovery

Maksym Myronovskyi[a], Andriy Luzhetskyy[*ab]

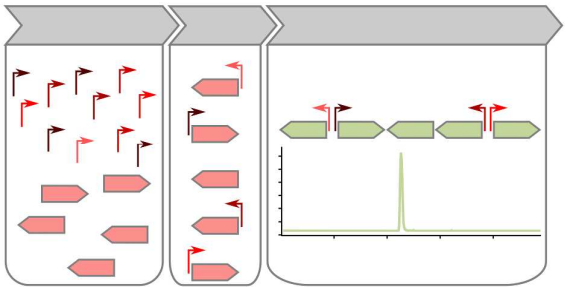
Covers the period up to 2016

Bacterial-based natural products have long represented a promising resource for the development of commercially relevant therapeutics, and more than two thirds of these products have been developed from members of the genus *Streptomyces*. The extensive sequencing of bacterial genomes suggests that the majority of gene clusters encoding natural products are silent and not expressed under standard laboratory conditions. However, these clusters can be activated through systematic exchanges between native transcriptionally silent promoters and transcriptionally active promoters. Therefore, the availability of well-studied constitutive and inducible promoters is of the utmost importance for identifying natural products encoded by silent gene clusters. This manuscript provides an overview of the promoter control elements for streptomycetes and examples of their successful application in refactoring the biosynthetic pathways of natural products.

[a]Helmholtz-Institute for Pharmaceutical Research Saarland, Saarland University Campus, Building E8 1, 66123 Saarbrücken, Germany

[b]Department of Pharmaceutical Biotechnology, Saarland University, 66123 Saarbrücken, Germany

E-mail: a.luzhetskyy@mx.uni-saarland.de



Transcriptional activation of biosynthetic gene clusters

Dr. Maksym Myronovskyi received his M.Sc. in Genetics and Biotechnology in 2007 and Ph.D in Molecular Genetics in 2012 from the Ivan Franko National University of Lviv. Subsequently he joined the Actinobacteria Metabolic Engineering Group at Helmholtz Institute for Pharmaceutical Research Saarland (HIPS). His main research interests lay in the area of discovery of new natural products and microbial host engineering.

Dr. Andriy Luzhetskyy gained a Ph.D. from the Ivan Franko National University of Lviv in 2004. At the same year he started his postdoctoral research in the Institute for Pharmaceutical Sciences at the Albert-Ludwigs University at Freiburg, where he founded his own junior research group in 2008. Since 2011, he heads a research group at the Helmholtz Institute for Pharmaceutical Research Saarland (HIPS). In 2015, Andriy Luzhetskyy is appointed as a Professor of Pharmaceutical Biotechnology at Saarland University. His research interests include synthetic biology and metabolic engineering of actinobacteria.

1. Introduction

Actinomycetes are a group of gram-positive soil bacteria that are widely known as producers of biologically active natural products. Large-scale genome mining projects have revealed that the potential of actinomycetes to produce natural compounds is much higher than expected, and most species contain more than twenty different biosynthetic gene clusters with the potential to encode novel natural products. However, few of these clusters are efficiently expressed, and the majority are inactive under common cultivation conditions[1–3]. Pathways that encode natural products are tightly regulated, and complex physiological and environmental signals are often required to elicit their expression[4–8]. The activation of cryptic gene clusters is challenging but extremely important in natural product discovery. Various experimental approaches are used to activate pathways, including the formulation of cell-culturing conditions, introduction of antibiotic resistance mutations, addition of environmental and physiological inducers and stress factors, expression or inhibition of regulatory genes, expression of clusters in heterologous strains, repression of competing secondary metabolite clusters, etc[6–11]. However, none of these approaches can be generally applied for cluster activation, which implies the presence of unique regulatory mechanisms that govern the expression of individual natural products. Under such circumstances, the development of a generally applicable strategy is highly desirable; therefore, a synthetic biology approach was developed to activate biosynthetic pathways (Fig. 1)[12]. Using synthetic regulatory elements that are not under cellular control allows for natural regulatory networks to be bypassed and forces the expression of the controlled genes. The construction and characterization of a comprehensive set of synthetic biology tools that includes various promoters, riboswitches, transcription terminators and ribosomal binding sites have been reported[13–17]. With regard to cluster activation, promoter elements are of indisputable importance because they are responsible for efficient transcription, which is the first stage of gene expression[18].

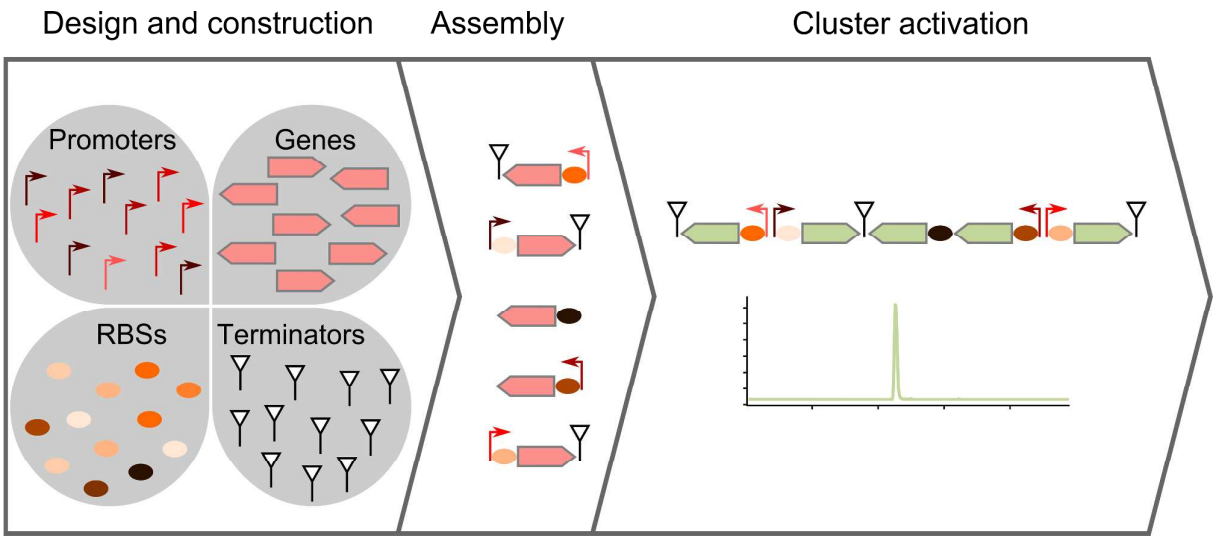


Fig. 1 Synthetic biology approach for silent cluster activation.

The aim of this report is to provide a comprehensive overview of the known promoter elements for streptomycetes, which belong to the most prolific producers of natural products. Several promoter engineering experiments are also discussed to demonstrate the historical evolution of cluster refactoring strategies.

2. Constitutive promoters

Constitutive promoters are widely used in expression studies of actinomycetes, including studies on the complementation of mutations, overexpression of genes, and heterologous expression of genes and gene clusters. Constitutive promoters with well-defined transcriptional activity can also be used to fine-tune the expression of modular elements in genetic circuits. The activities of certain constitutive promoters, however, are not constant and can vary significantly under particular cultivation conditions or growth stages. For example, constitutive promoters based on the widely used *ermE* lead to significantly stronger gene expression in the stationary growth phase than in the log phase, whereas the constitutive promoters based on *rpsL* increase the active transcription of genes in the earlier growth phases of actinomycetal cultures[19,20]. Considering that the majority of biosynthetic gene clusters reach their expression maximum in the late stationary phase, promoter actions must be considered when planning an experiment with heterologous gene cluster expression or mutation complementation. In this study, we provide a brief overview of the major classes of constitutive promoters that are widely utilized in actinomycetes genetics and examples of their application.

2.1 Promoters of the erythromycin-resistance gene *ermE*

The *ermE* gene of *Saccharopolyspora erythrea* encodes 23 S rRNA methyltransferase and confers resistance to the macrolide and lincosamide antibiotics erythromycin and lincomycin, respectively[21,22]. The promoter region of the *ermE* gene is frequently deployed to induce the overexpression of the genes in streptomycetes[23], and it presents a complicated structure. Two separate promoters, *ermEp1* and *ermEp2*, have been identified, and they initiate transcription at 1 and 72 bp from the translational start, respectively[21,24]. The -10 regions of *ermEp1* and *ermEp2* resemble a sequence of the consensus prokaryotic promoter, whereas the -35 region, which is also recognizable, shows much greater variability[21]. The spacing between the -10 and -35 regions is close to the optimal 17 bp distance observed in *Escherichia coli* in the case of *ermEp2*, although the spacing of *ermEp1* is significantly less[21]. Trinucleotide TGG deletion within the -35 region of *ermEp1* has been reported to increase the level of promoter activity by approximately five fold; however, subsequent studies detected only slight differences in activity between the wild-type and mutated variants of this promoter[24–26]. Considering all potential variability, five different types of *ermE* promoter can be distinguished: *ermEp*, the wild-type promoter region containing two individual promoters; *ermEp*, the* upregulated mutant of *ermEp* containing *ermEp1* with trinucleotide deletion (*ermEp1**); and *ermEp1*, *ermEp1** and *ermEp2*, which are individual single promoters (Table S1).

2.2 *SF14P* promoter from phage *I19*

Promoterless aminoglycoside-O-acetyltransferase I gene (*aacC1*) has been used as a reporter to screen for DNA fragments with promoter activity within the genome of phage *I19* isolated from *Streptomyces ghanaensis*. Four promoter regions were identified, and *SF14P* was the strongest. S1 nuclease mapping revealed two transcriptional start sites in *SF14P* that corresponded to two individual phage promoters, *14-Ip* and *14-IIp*, the latter of which appeared to be stronger. The putative -10 regions of *14-Ip* and *14-IIp* (ATCAAT and TACAATC, respectively) overlapped, whereas their putative -35 regions (TTGATG and TTGACC) were adjacent. The -10 and -35 regions of *14-Ip* and *14-IIp* resemble consensus sequences of vegetatively expressed promoters. The results of in vitro assays showed that despite different spacing between the -10 and -35 regions, both promoters are recognized by the major housekeeping sigma factor HrdB[27]. In semi-quantitative antibiotic resistance assays, *SF14P* (Table S1) has shown 2-fold higher activity than *ermE** (Fig. 2); however, the results of another study indicated that the strength of *SF14P* and *ermE** is comparable[19,27].

2.3 Synthetic promoter library recognized by the major vegetative sigma factor HrdB

Streptomyces species contain a great variety of genes encoding sigma factors, with the number varying from 35 in *S. albus* to over 60 in *S. coelicolor* and *S. avermitilis*[1,2,28]. Because of their complex developmental cycle and varying environmental conditions, certain sigma factors undergo complicated regulation and are active at only specific points in the lifecycle or under specific conditions [29–34]. To bypass this regulatory complexity, consensus sequence of the promoters that are recognized by the vegetative sigma factor HrdB and expressed throughout the entire lifecycle has been used to construct synthetic promoter library[14,35,36]. The hexamers TTGACN (where N is A, T, C, or G) and TASVDT (where S is G or C, V is G, A, or C and D is A, T, or G) corresponding to the -35 and -10 consensus sequences were preserved during this construction, while a 17 bp spacer region between the -10 and -35 sequences, as well as a 10 bp region upstream of the -35 sequence and a 5 bp region downstream the -10 sequence were totally randomized. The constructed library was cloned in front of the neomycin resistance gene *aphII* and analysed in *S. lividans*. According to the resistance level to neomycin and the RT-PCR analysis, 38 individual synthetic promoters were classified into three groups: weak (*C2-6—A5-24*), medium (*C4-15—C4-14*) and strong (*C4-1—B4-8*) (Table S2). A 12-fold difference in the promoter strength was detected between the weakest A2-1 and the strongest A1-14 synthetic promoters; however, even the latter was slightly weaker than the *ermEp** promoter (Fig. 2)[14]. The selected promoters were also checked in other *Streptomyces* hosts, including *S. coelicolor* and *S. ambofaciens*, and they showed similar activities to those observed in *S. lividans*[14].

2.4 Engineered derivatives of *kasOP* promoter

The *kasO* gene encodes the SARP transcriptional activator of a type I polyketide synthase cluster responsible for coelimycin production in *S. coelicolor*[19,37]. The promoter of this gene (*kasOP*) is regulated by the γ-butyrolactone receptor ScbR and the pseudo γ-butyrolactone receptor ScbR2, which bind to OA and OB sites within the promoter sequence[38–40]. The core promoter sequence of *kasOP* is similar to the consensus sequence of the promoters recognized by the housekeeping sigma factor HrdB, which is highly expressed during growth[38,41]. The role of HrdB in the transcription of *kasOP* was experimentally demonstrated in the heterologous environment of *E. coli*[19]. Attempts to remove or abolish ScbR- and ScbR2-binding sites have been undertaken to circumvent the host regulation of *kasOP*. After removing the OB binding site of ScbR2 by trimming the 5'-end of the promoter region, the transcriptional strength of the resulting $kasOP_3$ increased nearly 40 times compared with that of the original promoter. To remove the ScbR-binding site OA, which overlaps the -10 and -35 regions of the promoter, the $kasOP_3$ sequence between these core regions was randomized. Four individual mutant promoters, $kasOP_{314}$, $kasOP_{361}$, $kasOP_{382}$ and $kasOP_{3154}$ (Table S3), were identified. All these promoters showed much higher activities than the native promoter, and they also lost the ability to bind ScbR. The $kasOP_{361}$ promoter (renamed *kasOP**) exhibited the highest activity and was further analysed. Compared with the *ermE** and *SF14P* promoters, *kasOP** ($kasOP_{361}$) exhibited higher activity in both biological and real-time qPCR assays. The results of the real-time qPCR assay indicated that *kasOP** ($kasOP_{361}$) was from two- to five-fold more active than *ermE** and *SF14P* depending on the time point and showed that *ermE** and *SF14P* had comparable activity (Fig. 2)[19].

2.5 Second-stage derivatives of the synthetic *kasOP** promoter

The superfolder green fluorescent protein (sfGFP) and a flow cytometry method optimized for streptomycetes were applied, and the results demonstrated that the engineered *kasOP** promoter is 20-fold more active than *ermEp** in *Streptomyces venezuelae*[42]. Therefore, the *kasOP** promoter was used for further construction of two randomly mutated promoter libraries. The first library contained *kasOP**

derivatives with randomized nucleotides downstream from the -10 sequence, and the second contained *kasOP** derivatives with a mutated spacer region between the -10 and -35 sequences. The analysis of 180 promoters indicated that the activity of six promoters was stronger than that of *kasOP** (Fig. 2), and 44 promoters from this library (*SP1 – SP44*) with activities ranging from 0.95 to 187.5% compared with that of *kasOP** were sequenced (Table S4)[43].

2.6 *ermEp1*-based synthetic promoter library

When the *ermEp1*-based promoter library was created, several synthetic promoters were available but only generally characterized. Therefore, the authors pursued the objective of generating a comprehensively characterized library of synthetic promoters[14,26,44]. The common approach based on the randomization of sequences surrounding the -10 and -35 promoter consensus regions was used to construct synthetic derivatives of the well-characterized *ermEp1* promoter[26,45–48]. Altogether, 56 synthetic promoters (*D4 - 21*) were constructed, and they had transcriptional activities ranging from 2% to 319% relative to the parental *ermEp1* (100%) (Table S5) (Fig. 2). The transcriptional activity of all constructed promoters was evaluated quantitatively using the *gusA* reporter gene. For a more comprehensive analysis, the single representatives of the weak, middle and strong promoter groups (82-, 57- and 21-promoter, respectively) were analysed using RNA-Seq and assessed relative to two additional reporter genes: green fluorescent protein *gfp* and neomycin-3'-phosphotransferase II (*aphII*). Regardless of the method used to assess the transcriptional activity, a strong correlation among the relative promoter strengths was always observed. Selected promoters that were originally characterized in *S. lividans* TK24 were also tested for their activity in *S. albus* J1074, as well as in the more distantly related *Saccharothrix espanaensis* DSM 44229 and *Salinispora tropica* CNB-440. All synthetic promoters exhibited similar strengths to those observed in *S. lividans*[26].

2.7 Promoter regions of glyceraldehyde-3-phosphate dehydrogenase and 30S ribosomal protein S12 from Actinobacteria

A plug-and-play strategy for refactoring secondary metabolite clusters requires a set of constitutive or inducible promoters. The promoter regions of 23 housekeeping genes originating from *Streptomyces griseus* were selected, and the expression levels of the corresponding genes were analysed by real-time PCR[3]. Two genes, *gapdh* and *rpsL*, which encode glyceraldehyde-3-phosphate dehydrogenase and 30S ribosomal protein S12, were transcribed at substantially higher levels than the other genes[49]. The *xylE* activity assay showed that the activities of their promoter regions, *Pgapdh* and *PrpsL*, were much stronger than that of the *ermEp** promoter[49,50]. Encouraged by the activities of *Pgapdh* and *PrpsL*, their counterparts from other Actinobacteria species were analysed. Sequences of the *Pgapdh* and *PrpsL* promoters from non-*Streptomyces* genera were shown to be highly diversified, whereas the coding regions of the corresponding genes were highly conserved. Altogether, 36 promoter regions from 18 distinct Actinobacteria were cloned upstream of *xylE* to quantitatively assess their transcriptional activity. Thirteen of these regions were active in *S. lividans*, and *Pgapdh(EL)*, *PrpsL(CF)* and *PrpsL(XC)* from *Eggerthella lenta*, *Cellulomonas flavigena* and *Xylanimonas cellulosilytica* had more than 10-fold higher activity than *ermEp** (Table S6) (Fig. 2)[49].

2.8 Panel of strong constitutive promoter regions from *Streptomyces albus*

To expand the set of strong constitutive promoters and identify exceptionally strong promoters, a transcriptome-guided survey of highly expressed genes within the chromosome of the widely used host strain *S. albus* J1074 was performed[20,51]. Thirty-two genes that were highly expressed under different cultivation conditions and at different time points were identified, and 20 of these genes resided upstream of the 30S ribosomal proteins. The identified intergenic regions with high promoter activity were

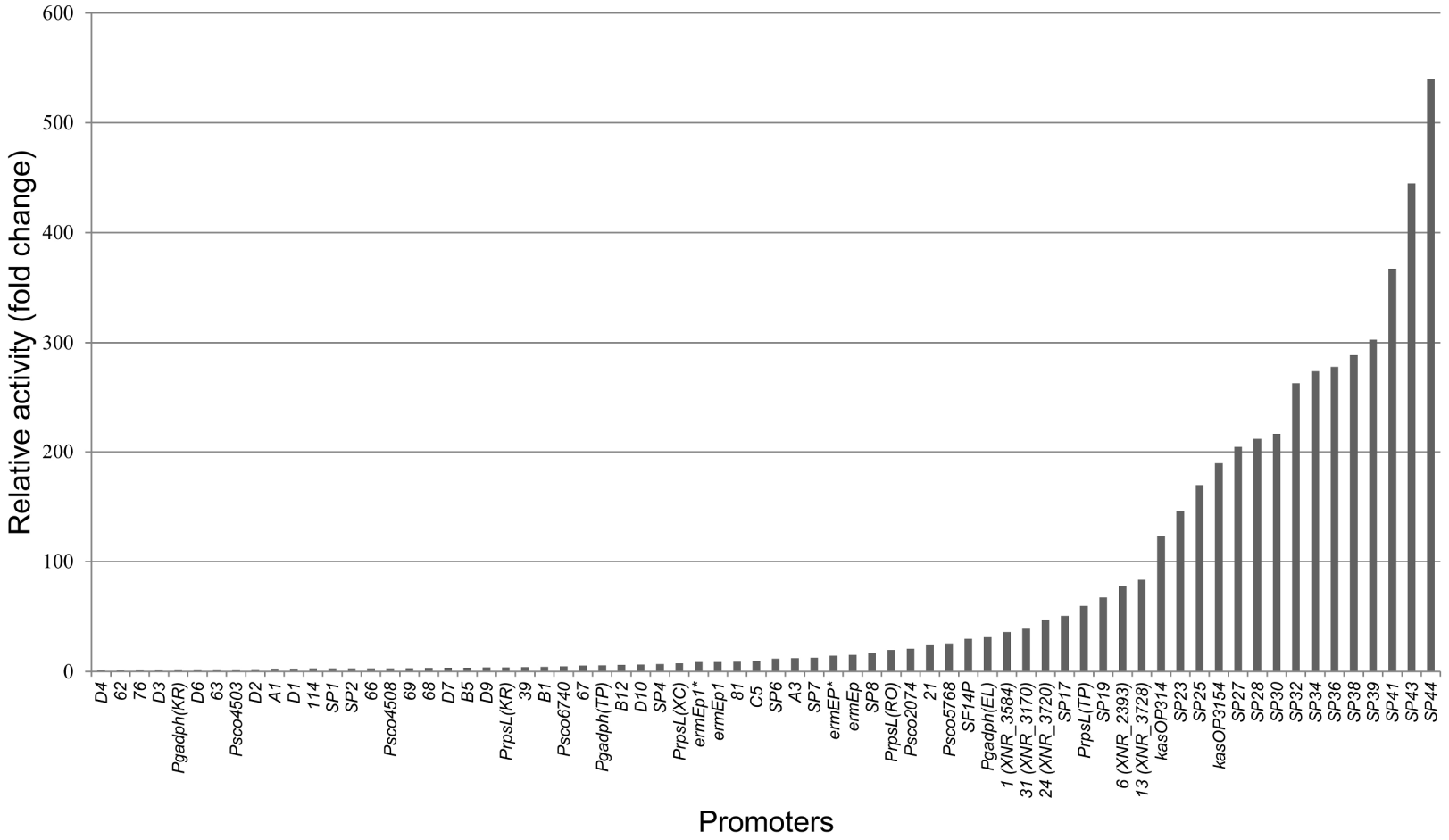


Fig. 2 Relative activities of the promoter elements from selected promoter libraries.

quantitatively analysed using the *xylE* reporter gene. Ten individual promoter regions (*1 – 31*) with unprecedented high transcriptional activities ranging from 200% up to 1300% of the activity of *ermEp** were discovered during this analysis (Table S7) (Fig. 2). Subsequent time-course studies also demonstrated the constitutively high level of activity in these regions. At the time that this manuscript was written, the minimal promoter sequences had not been identified within the discovered promoter regions. The size of these regions ranged from 247 bp to 695 bp; therefore, the possibility of multiple promoters and ribosomal binding sites residing within the identified promoter regions cannot be excluded. With regard to transcriptional activity, the identified promoter regions appear to be among the strongest ever reported in *Streptomyces*[20].

2.9 Library of synthetic promoters based on the *actII orf4* promoter of *Streptomyces coelicolor*

The *actII orf4* gene encodes a DNA-binding protein that positively regulates the transcription of the actinorhodin biosynthetic genes in *S. coelicolor*[4,52]. To modulate actinorhodin production, a synthetic promoter library was developed by randomizing sequences around the -10 and -35 consensus regions of the *actII orf4* promoter. Cloned in front of the native gene (*actII orf4*), the constructed promoters were analysed in *S. coelicolor* by the level of actinorhodin production. The *actII orf4* gene, which is under the control of the native promoter and *ermEp**, was used as a reference. The single promoter *ScoSPL20* confers a slightly higher actinorhodin production level than *ermEp** promoter, and it was identified after screening 10,000 colonies. The increased production corresponds to a 2.8-fold increase compared with the wild-type *actII orf4* promoter. An additional ten synthetic promoters (*ScoSPL7 – ScoSPL185*) that provide higher and lower actinorhodin productivity than the native promoter were also analysed (Table S8). The transcriptional activity of the selected eleven promoters was quantitatively assessed using the xyloglucanase reporter gene *xeg*. Although the results of actinorhodin production studies and the reporter assay showed a general correlation, the activity fold-changes among the individual promoters were not sustainable[53]. In addition, it was unclear whether the affinity of the AtrA transcriptional regulator known to bind directly to the native *actII orf4* promoter was affected with the use of synthetic derivatives[53,54].

2.10 Panel of constitutive promoter regions from *Streptomyces coelicolor*

Over the course of the transcriptome-guided screening, 166 promoter regions of *S. coelicolor* that presented stable transcriptional activity in the five transcriptome datasets obtained according to different cultivation conditions and time points were identified. This identified set of promoter regions was rationally preselected by discarding the promoters that presented altered activity in front of different genomic mutations, the promoters that were sensitive to external stress, the promoters that were involved in secondary metabolism pathways or operon structures, etc. The remaining 166 promoter regions covered 90% of the range of relative promoter strengths observed in the *S. coelicolor* M145 transcriptome[23,55]. Eight individual promoter regions of different strengths were selected from the identified set and used in the experimental tests (Table S9). The results of the GFP fluorescence assay were consistent with the results of the RT-qPCR, which implied the constitutive activity of the selected promoter regions. The eight selected promoter regions were originally characterized in *S. coelicolor* M1146 and experimentally evaluated in *S. venezuelae* WVR2006 and *S. albus* J1074[56–58]. In both strains, the results were consistent with those obtained in *S. coelicolor*, thereby indicating the applicability of the identified promoter regions for various *Streptomyces* strains. The strongest promoter region identified in this study, *Psco5768*, was approximately two-fold more active than *ermEp**, whereas *Psco5768* was approximately 5.5-fold more active than the weakest promoter region (*Psco4503*) (Fig. 2). Transcriptional start points and the minimal promoter sequences were not identified within the studied promoter regions; therefore, the latter might contain

multiple promoters and 5'-untranslated regions, which have the potential to affect the translation of the controlled genes[55].

3. Inducible promoters

Inducible promoters are valuable tools for designing and studying genetic circuits in basic research and for developing metabolically engineered strains, especially strains that produce highly toxic compounds, such as antibiotics. The strong advantage of inducible promoters compared with constitutive promoters is their ability to turn on gene or gene cluster expression to the desired level at a specific time point. However, the influence of inducer molecules on the global gene expression level (which can be dramatic, e.g., thiostrepton) must be considered when using inducible expression systems to avoid misleading results. Additionally, inducing gene expression in large-scale fermentations remains an unsolved issue. A great number of well-characterized inducible systems have been developed and utilized for *E. coli*[59,60]. Unfortunately, most of these systems cannot be directly applied for actinomycetes and therefore must be reconstituted and rewired. In the past decade, several new versatile and highly inducible promoters have been developed to control gene and gene cluster expression in actinomycetes. Below, we summarize the widely used and best-characterized inducible systems for actinomycetes and discuss their limitations and advantages.

3.1 Thiostrepton-inducible promoter *PtipA*

Routine utilization of thiostrepton for the selection of plasmids led to the discovery of unexpected biological effects. For instance, thiostrepton conferred increased resistance to *S. lividans* to heterogeneous antibiotics and induced the overexpression of several proteins[61,62]. Two proteins, TipAL and TipAS, are different in-frame translation products of the same gene, *tipA*. TipAL protein belongs to the MerR family of transcriptional regulators and consists of a conserved N-terminal helix-turn-helix DNA-binding motif and a C-terminal drug-recognition domain, whereas TipAS lacks the N-terminal helix-turn-helix motif and is translated as the C-terminal region of TipAL at a considerable molar excess[62]. Both TipAL and TipAS bind covalently to thiostrepton and structurally related thiopeptides with equimolar stoichiometry. The dehydroalanine tail in the structures of thiopeptide antibiotics is shown to react with the C-terminal cysteine residue of TipA proteins; therefore, it is crucial for complex formation[62,63]. In complex with thiostrepton, the TipAL protein activates transcription of its own gene by binding to its promoter *PtipA* (Table S10) in the form of a dimer[64,65]. DNAseI footprinting experiments indicate that TipAL binds the inverted repeat sequence located in *pTipA*, which is -13 to -36 bp upstream of the transcriptional start site[64,65]. Because it lacks a C-terminal region, the TipAS protein cannot to bind DNA, although it binds thiostrepton[62]. Because of the autogenous transcriptional activation of the *tipA* gene by thiopeptides, *PtipA* has been widely used in *Streptomyces* genetics as an inducible promoter[23,66]. The results of a Northern blot assay indicate that the ratio of *PtipA* transcriptional activity under induced conditions relative to uninduced conditions exceeds 200; however, the results of the DNA-binding assays detected only a 10-fold increase of TipAL binding to *PtipA*[61,64]. Highly efficient transcription of the genes under the control of PtipA and yields of up to 25% of the total extracellular proteins have been reported[67]. Although *PtipA* has been shown to provide reliable and controllable gene transcription in a number of cases, it also presents a number of drawbacks, including the following: the promoter is dependent on the presence of the TipAL protein and presents a considerable level of uninduced transcription and the use of *ptipA* often requires the presence of a resistance gene because of the high activity of thiostrepton, even at low concentrations[44,61,68,69].

3.2 *PnitA*-NitR inducible expression system

In the actinomycetes *Rhodococcus rhodochrous* J1, which is used in the industrial production of acrylamide, nitrilase is strongly induced by isovaleronitrile or ε-caprolactam[70]. Nitrilase catalyses the cleavage of nitriles to the corresponding acids and ammonia. Particularly in *R. rhodochrous* J1, induced nitrilase corresponds to 35% of all soluble proteins, implying the presence of a strong promoter in front of the nitrilase gene *nitA*. Sequencing and functional analyses of the *nitA* gene and its flanking regions showed that the transcriptional activity of the *PnitA* promoter (Table S10) was dependent on the *nitR* gene located downstream from *nitA*, which encoded the positive transcriptional regulator homologous to XylS and AraC[70,71]. In complex with ε-caprolactam, NitR (accession number BAA11038.1) most likely binds to an inverted repeat sequence within *PnitA* and induces the transcription of *nitA* and *nitR* genes[70,71]. Compared with the toxic isovaleronitrile, ε-caprolactam has been shown to have no effect on the growth and morphology of *Streptomyces* species and therefore has been used as an inducer. The pSH19 vector for inducible gene expression in streptomycetes was constructed based on *PnitA* and *nitR*. The level of gene expression using the nitrilase system was shown to be dose dependent[71]. Because of the high transcriptional activity of *PnitA* combined with the high copy number replication origin of pSH19, the use of this system resulted in a number of successful protein expressions that accounted for up to 40% of the soluble proteins in each case[71].

3.3 Synthetic tetracycline-inducible promoter *tcp830*

To expand the Streptomyces genetic toolkit for controlled gene expression, several tetracycline-inducible promoters were constructed using the tetracycline-inducible repressor-operator interaction from *E. coli* transposon Tn*10*. The promoter elements from *ermEp1* together with two or three operators (*tetO1* and *tetO2*) were used for the construction. The tests of different promoters indicated that *tcp830* (Table S10) showed the biggest difference in expression level between the induced and uninduced conditions, and when induced, *tcp830* was among the strongest promoters. The transcriptional activity of *tcp830* was dose-dependent, and varying levels of induction were observed after the addition 1-100 ng/ml of the inducer. Anhydrotetracycline is a more active inducer than tetracycline and also has a higher minimum inhibitory concentration. The full induction of the *tcp830* promoter was observed in *Streptomyces coelicolor* at a 100 ng/ml concentration of anhydrotetracycline, and detrimental effects on the growth rate were not observed. The *tcp830* promoter was initially studied with the cognate gene encoding the TetR-repressor adapted for use in streptomycetes but later it was used separately according to the functions of the indigenous TetR homologue (SCO0253) present in the chromosome of a number of *Streptomyces* strains. However, the high uninduced transcriptional activity of the *tcp830* promoter in several studied strains and the lack of increased activity after induction implies that the TetR homologue is not contained in certain *Streptomyces* strains or the homologue is functionally different from that in *E. coli*[44].

3.4 Glycerol-inducible system for gene expression

The pathway for glycerol catabolism in *Streptomyces coelicolor* is determined by the *glyCABX* operon, which can be induced by glycerol and repressed by glucose[72,73]. Interestingly, the induction and catabolite repression of the operon are accomplished by the single negative transcriptional autoregulator GylR[74]. The *glyR* gene (accession number X14188) is located immediately upstream from the *glyCABX* gene and undergoes GylR autoregulation. However, because the *glyR* gene is also induced by glycerol, it is only weakly repressed by glucose[74]. Two promoters, *glyP1* and *glyP2* (Table S10), were identified in front of the operon and most likely serve as binding sites for GylR. When grown on arabinose as the sole carbon source, the expression of the *glyCABX* operon can be increased by 35-fold by the addition of glycerol[75]. These features of *glyR* together with those of *glyP1* and *glyP2* enable their use for controlled gene expression.

Although several *gyl* vectors for glycerol-inducible gene expression have been constructed, this system has not been extensively used in *Streptomyces* genetics[23,76,77].

3.5 Synthetic resorcinol-inducible promoter

For a long time, the inducible gene expression systems for streptomycetes shared one common drawback: a leaky phenotype in the absence of an inducer[44,71,75]. The development of a tight inducible gene expression system was attempted by constructing resorcinol and cumate-inducible promoters[78]. RolR, a member of the TetR-family of transcriptional repressors, is involved in regulating the genes responsible for resorcinol catabolism in *Corynebacterium glutamicum*. Binding of RolR to its cognate operator *rolO* is hindered by the presence of the aromatic compounds resorcinol and hydroxyquinol[79,80]. A new resorcinol-inducible promoter for streptomycetes was constructed by fusing the *rolO* operator with the previously constructed synthetic promoter *A3*, which yielded the *PA3-rolO* promoter (Table S10)[26,78]. In the presence of a codon-optimized *rolR* gene (accession number KJ775861) under the control of the strong synthetic promoter *21*, the *PA3-rolO* promoter could be induced by the presence of resorcinol but not by the presence of the structurally related 1,2,4-benzenetriol. The promoter has a low level of basal expression and reaches a 33-fold induction ratio, which is comparable to the activity of the native *A3* promoter. The promoter induction was shown to be dose dependent and reached its maximum value at a resorcinol concentration of 40 μM. No negative effects on the *S. albus* strain were detected with resorcinol concentrations up to 50 μM; however, high concentrations of approximately 100 μM were shown to affect its growth[78].

3.6 Synthetic cumate-inducible promoter

The construction of the cumate-inducible system was similar to that of the resorcinol-inducible system. The CymR transcriptional regulator also belongs to the TetR family and is involved in controlling the degradation of cumate and cymene in *Pseudomonas putida*[81]. CymR binding to its cognate *cmt* operator is abrogated by the presence of cumate. The cumate-inducible promoter *P21-cmt* (Table S10) was constructed by fusing the *cmt* operator with the synthetic *21* promoter, which is approximately twice as strong as the *A3* promoter. In the background of the codon-optimized *cymR* gene (accession number KJ775862), the *P21-cmt* promoter showed negligible basal expression under uninduced conditions, whereas the addition of 50 μM cumate completely released the repression of transcriptional activity. The induction of the *P21-cmt* promoter was also dose dependent and reached its maximum expression level at a cumate concentration of 30—100 μM. Compared with resorcinol, cumate was shown to have no adverse effects on the growth of the *S. albus* strain. According to the results of the kinetics studies, an increase of the transcriptional activity was first observed 4 h after induction, whereas its maximum value was reached between 6 and 12 h after induction. A 50% reduction of the *P21-cmt* promoter's activity could be detected within the first 9 h after inducer removal, whereas a 70% reduction was reached after approximately 24 h. It is not clear whether *P21-cmt* stays in the transcriptionally active state for such a long period of time after inductor removal or whether transcription cessation is masked by the stability of the β-glucuronidase enzyme, whose gene *gusA* was used as a reporter in this study[78].

4 Heterologous promoters in antibiotic production

4.1 Expression of the jadomycin gene cluster from the *ermEp** promoter

Expression of the engineered jadomycin gene cluster from the *ermEp** promoter is one of the first examples of biosynthetic pathway refactoring using heterologous promoters[82]. Jadomycins produced by

Streptomyces venezuelae ISP5230 belong to the group of polyketide-derived angucycline antibiotics with broad-spectrum cytotoxic activities[83,84]. The jadomycin biosynthetic cluster spans 28.2 kb and contains 28 genes. All 22 structural genes predicted for jadomycin biosynthesis are oriented towards the same direction[85,86]. The presence of six regulatory genes implies the complex regulation of jadomycin production. Wild-type cultures produce only a negligible amount of jadomycins under standard cultivation conditions, whereas these antibiotics are greatly induced under the effect of stress factors, such as heat shock, ethanol treatment and phage infection[87]. To improve the production of jadomycins, the regulatory mechanisms of all biosynthetic genes were inhibited by deleting four regulatory genes (*jadW2*, *jadW3*, *jadR2* and *jadR1*), and control of the biosynthetic genes was then conferred to a single *ermEp** promoter. The resulting strain constitutively produced jadomycin B at the level of 84.3 mg L^{-1}, and the stress factors did not further increase the production level but rather decreased it by affecting the growth rate of the strain. After induction with ethanol, the wild-type strain produced less jadomycin than the engineered strain without induction (50.5 mg L^{-1}), although the specific productivities, such as micrograms of antibiotic per milligram of wet biomass, were similar for both strains[82]. Further, an increase in jadomycin production was achieved eight years later after the identification of strong constitutive promoters in the genome sequence of *S. coelicolor M145*. Replacing *ermEp** with the strongest identified promoter, *Psco5768*, caused a two-fold increase of jadomycin B production[55].

4.2 Overexpression of the novobiocin biosynthetic cluster from a single inducible *tcp830* promoter

Another example of whole-cluster transcription from a single promoter is the heterologous overexpression of a novobiocin cluster[88]. The aminocoumarin antibiotic novobiocin, which is produced by *Streptomyces spheroides*, is a potent inhibitor of the bacterial DNA gyrase[89,90]. The novobiocin biosynthetic cluster spans 23.4 kb and consists of 20 genes, including genes involved in antibiotic biosynthesis (*novH* to *novW*), regulation (*novE* and *novG*) and resistance (*gyrB*)[90]. Notably, all 20 genes are arranged in the same orientation and transcribed as a single polycistronic mRNA[91]. Because the natural producer of novobiocin was refractory to genetic manipulation, the biosynthetic cluster was expressed heterologously in *S. coelicolor* M512[92,93]. To increase the antibiotic production level in the heterologous host, two pathway-specific regulatory genes that are essential for transcription of the novobiocin cluster, *novE* and *novG*, were deleted and the inducible tetracycline promoter *tcp830* was inserted in front of the cluster[44,88]. After induction with anhydrotetracycline, the *S. coelicolor* strain harbouring the engineered novobiocin cluster produced 74 mg L^{-1} of antibiotic, which corresponded to a 3.4-fold increase in production compared with the strain with the native cluster. Under the uninduced cultivation condition, novobiocin was produced at a much lower level of 3 mg L^{-1}. Moreover, the single *tcp830* promoter was shown to be sufficient to transcribe 16 biosynthetic genes spanning an 18 kb region and the insertion of additional copies of *tcp830* within the cluster did not further increase the antibiotic production level[88].

4.3 Activation of silent polycyclic tetramate macrolactam biosynthetic clusters from *Streptomyces griseus* and *Streptomyces albus*

Polycyclic tetramate macrolactams (PTMs) are a widely distributed class of natural products with important biological activities that range from antibacterial and antifungal to cytotoxic. Representative examples of PTMs are dihydromaltophilin, maltophilin, cylindramide, ikarugamycin, alteramide discodermide and frontalamide[94–100]. Genome-mining efforts have revealed strikingly conserved PTM-type gene clusters in the genomes of phylogenetically diverse bacteria, ranging from Proteobacteria to actinomycetes: *Streptomyces sp*. SPB78, *Streptomyces sp*. SPB74, *Streptomyces albus* J1074, *Streptomyces flavogriseus*, *Streptomyces griseus*, *Streptomyces roseosporus*, *Streptomyces sp*. ActE, *Streptomyces sp*. Act-1, *Streptomyces*

clavuligerus, *Streptomyces sp*. Mg1, *Salinispora arenicola*, *Saccharophagus degradans*, and *Lysobacter enzymogenes*[100]. Two conservative genes, *ftdA* and *ftdB*, encoding putative desaturase and hybrid PKS-NRPS, respectively, compose the most upstream flank of the PTM clusters and are often clustered with another four genes, *ftdC-F*, which are likely involved in antibiotic biosynthesis because of an apparent operon-like arrangement[100]. Two silent PTM clusters from *S. griseus* and *S. albus* were activated through a plug-and-play promoter insertion strategy[9,101]. The entire silent PTM cluster from *S. griseus* was reconstructed by inserting six constitutive promoters upstream of the biosynthetic genes. The activated cluster, which was constructed using the assembler method, led to the production of two PTM compounds in the heterologous host *S. lividans*, and one of these compounds has not been previously reported in the literature[101,102]. A homologous PTM cluster from *S. albus* was activated by independently inserting the single *ermEp** promoter upstream of the *sshg_05712* and *sshg_05713* genes, which correspond to *ftdA* and *ftdB,* respectively. As a result of these manipulations, two *S. albus* strains were obtained, and they produced two novel members of the PTM-family: 6-epi-alteramides A and B[9].

4.4 Activation of indigoidine synthetase

A small silent NRPS cluster consisting of a single gene encoding the putative indigoidine synthetase was identified in the genome sequence of *S. albus* J1074. Similar clusters were also identified in other *Streptomyces* strains, such as *S. lavendulae*, *S. aureofaciens*, and *S. chromofuscus*, and in *Erwinia chrysanthemi*[103–106]. Similar to the case of the alteramide cluster from *S. albus*, the strong constitutive promoter *ermEp** was inserted upstream of the *sshg_00313* gene that encodes indigoidine synthetase. The resulting *S. albus* strain produced a blue pigment, which implied the successful activation of the cluster[9].

4.5 Refactoring of the silent spectinabilin gene cluster

Spectinabilin is a rare nitroaryl-substituted polyketide from *Streptomyces spectabilis* and *Streptomyces orinoci* that exhibits antimalarial and antiviral activities[107,108]. Despite similar organization between the spectinabilin gene clusters from two strains, they appear to be differently regulated[109]. Heterologous expression experiments further confirm this differential regulation because only the *spn* cluster from *Streptomyces spectabilis* produced spectinabilin in *S. lividans*. A real-time PCR analysis showed that the *nor* cluster from *Streptomyces orinoci* is expressed at extremely low levels in a heterologous environment; therefore, it was used as a model pathway for scaffold design in refactoring gene clusters[49]. Nine strong constitutive promoters were used to drive the expression of the *nor* genes except for *norD* and *norG*, which encode the transcriptional repressor and the first enzyme in the spectinabilin biosynthetic pathway, respectively. The control of *norG* was conferred to the hyperinducible promoter *PnitA*[71]. After its assembly into the scaffold, the refactored biosynthetic pathway was successfully activated and produced spectinabilin in *S. lividans* with a titre of 105 μg L^{-1} [49].

4.6 Refactoring and resuscitation of the silent lazarimide gene cluster

A set of cassettes containing synthetic promoters recognized by the major vegetative sigma factor HrdB were constructed for the yeast homologous recombination-mediated activation of silent natural product biosynthetic gene clusters[14,110]. As a proof-of-concept experiment, a silent environmentally derived lazarimide (*Lzr*) gene cluster was activated through promoter engineering. The *Lzr* gene cluster was identified by a PCR screen of environmental cosmid libraries, and it resembles the clusters that encode cladoniamide and BE-54017, which belong to the tryptophan dimer class of structurally and functionally diverse natural products[111]. Tailoring enzymes, which are not used in the biosynthesis of any known indolotryptoline, were encoded by the genes inside the *Lzr* cluster, which indicates that this cluster should

encode for a novel indolotryptoline congener. Three bidirectional (*P1*, *P2* and *P3*) promoters and one unidirectional (*P4*) promoter were predicted to drive the expression of the biosynthetic genes. In a series of single cassette insertions, three bidirectional *lzr* promoters (*P1*, *P2* and *P3*) were replaced with the synthetic promoter cassettes. As expected, the *P1* and the *P1*+*P2* replacement constructs produced chromopyrrolic and indolocarbazole intermediates in *S. albus*, whereas the *P1*+*P2*+*P3*-replaced cluster produced an indolocarbazole intermediate instead of the expected indolotryptoline intermediate. A detailed examination of the *lzrX1* gene, which is responsible for the missing reaction step, revealed that a single base deletion led to the nonfunctional gene. Replacement of the mutated *lzrX1* gene with its functional full-length homologue *abeX1* (from the BE-54017 cluster) led to the production of new indolocarbazole (lazarimide C)- and indolotryptoline (lazarimide B)-based compounds [110,112]. The fully reengineered *Lzr* gene cluster, which included a replaced *P4* promoter, produced one additional major metabolite, lazarimide A, which differs from cladoniamide and BE-54017 by its halogenation pattern and the oxidation of the flipped indole moiety[110,112,113].

5. Outlook and conclusions

The scientific interest in actinomycetes and streptomycetes in particular is based on their ability to produce a great variety of natural products, which frequently possess commercially relevant properties. The development of genetic tools for streptomycetes in recent decades, including DNA-manipulation techniques and methods for studying and controlling gene expression, served a purpose of natural product discovery and facilitation of their biosynthetic study. Similarly, a number of promoter libraries for streptomycetes have been constructed to resolve the main problem associated with natural product discovery: the activation of silent gene clusters, which are believed to constitute a potentially valuable but unexplored source of natural products[114].

As demonstrated in this review, the currently available promoter toolkit disposes an extensive variety of constitutive and inducible elements. The transcriptional activity of these promoters spans an approximately 1,000-fold range, thereby allowing the precise fine tuning of gene expression levels. Many of these promoters, however, are not precisely defined and instead are presented as intergenic promoter regions that span several hundreds of base pairs. Because the refactoring of silent gene clusters relies primarily on DNA recombineering and DNA assembly techniques that are to some extent dependent on PCR, the size of certain identified promoter regions (+100 bp) limits their use for cluster activation because they cannot be easily introduced in the sequence of a single primer and require much more comprehensive PCR-based approaches. Considering the number, transcriptional strength and robustness of the reported promoter regions, it would be highly desirable to delineate minimal promoters within their sequences and exclude unessential DNA regions that may potentially affect the expression of the controlled genes. In the current state it is unlikely that the reported promoter regions will find their application in cluster activation experiments and will be most probably discarded. The available libraries of minimal promoters that consist of several dozen base pairs with defined transcription start points are not affected by the abovementioned problem and offer unprecedented flexibility in the design of cluster activation experiments.

One important but frequently overlooked problem related to cluster activation is balancing gene expression. The number of silent multioperon clusters activated in streptomycetes through comprehensive multiple-promoter insertions is currently small to negligible. Thus, cluster activation requires further investigations, which will likely occur within the next few years. Currently, the crucial role of gene balancing for successful pathway activation is indicated by the heterologous expression of metabolic pathways and their engineering in different bacterial hosts. The most prominent examples of gene balancing are the

engineering of terpenoid pathways that lead to the production of artemisinin and taxadiene and the functional optimization of nitrogen fixation (*nif*) gene cluster by combinatorial design and assembly in *E. coli* [115–117]. During optimization of the *nif* gene cluster by varying promoters, ribosome binding sites, gene order, gene orientation and operon occupancy, 122 variants of the complete pathway were constructed. The most productive, fully synthetic cluster, v2.1, which recovers 57% of the wild-type activity, shows the transcriptional profile of the single genes much more similar to the wild type than the other less productive and less balanced clusters, though with higher transcription levels. Interestingly while the transcripts of the v2.1 cluster remained 2- to 100-fold higher than in wild type, the protein expression in this refactored cluster was nearly identical to that in the wild-type cluster implying significance of the RBS design and optimization during pathway refactoring[117]. The other associated studies indicated that the elimination of all bottlenecks within a pathway and maximization of the optimal yield of product requires simultaneous tuning of all the pathway's enzyme expression levels. The optimal enzyme expression level in its turn is commonly some intermediate value in between very low and very high expression[115,116,118,119]. In this respect, the construction of numerous promoter libraries that span a wide range of transcription initiation activities (Fig. 2) was an inevitable prerequisite for efficient cluster activation studies.

Even if the minimal promoters of all possible strengths are provided, cluster activation can still be a challenging task, especially when it should be performed in a high-throughput manner. The main reason for this is an imperfect annotation of open reading frames of silent gene clusters. It is not a trivial task to predict gene start codons, ribosome binding sites and promoters, although the exact identification of transcriptional start points may be a key to successful cluster activation. If we assume that silent clusters are fully functional and their lack of expression is caused by our inability to reproduce the growth conditions or environmental signals experienced by bacteria in their natural habitats, then the gene and ribosome binding sites should be functional because the lack of expression is caused by only a lack of promoter activity. Thus, to activate the silent cluster, we only have to substitute native promoters that are inactive under our cultivation conditions with active promoters that are readily available. Therefore, the only remaining task is to define the exact sequences in the silent cluster that should be substituted with active promoters from the numerous constructed libraries or identify the exact positions where these promoters should be integrated. Unfortunately, current bioinformatics tools cannot define transcriptional start points in front of transcriptional units, and promoter sequences can be inserted into only arbitrary sites upstream from transcriptional units within the cluster. Such an approach cannot guarantee that the natural transcriptional organization of the cluster or the 5'-untranslated regions (UTRs) will be achieved. The crucial importance of UTRs for effective gene expression has been reported in numerous studies; therefore, changing the length and sequence of the UTRs by improper promoter insertions can severely affect the expression of the genes within refactored clusters[120–122]. Even several nucleotide changes within the UTR can have a dramatic effect on translation efficiency through the formation of unfavourable mRNA secondary structures or sequestration of ribosome binding sites. Thus, in the worst case scenario, the affected genes will remain translationally inactive regardless of how efficiently they are transcribed during promoter shuffling studies. The development of bioinformatics instruments capable of predicting promoters and transcription start points, similar to those developed for predicting translational rates of mRNA, would greatly facilitate cluster activation-based drug discovery[17,123]. In the case of streptomycetes, establishing such a bioinformatics instrument is an especially complicated task because of the presence of numerous sigma factors[1–3]. The most straightforward method of resolving this problem will likely involve developing an in vitro transcription system for streptomycetes that is similar to the system developed for corynebacteria[124]. The 5'-enriched sequencing of transcriptomes obtained with in vitro reactions of mono

sigma factor RNA polymerase holoenzymes will reveal the promoter specificity of each sigma factor and define the transcription start points of the respective promoters[125].

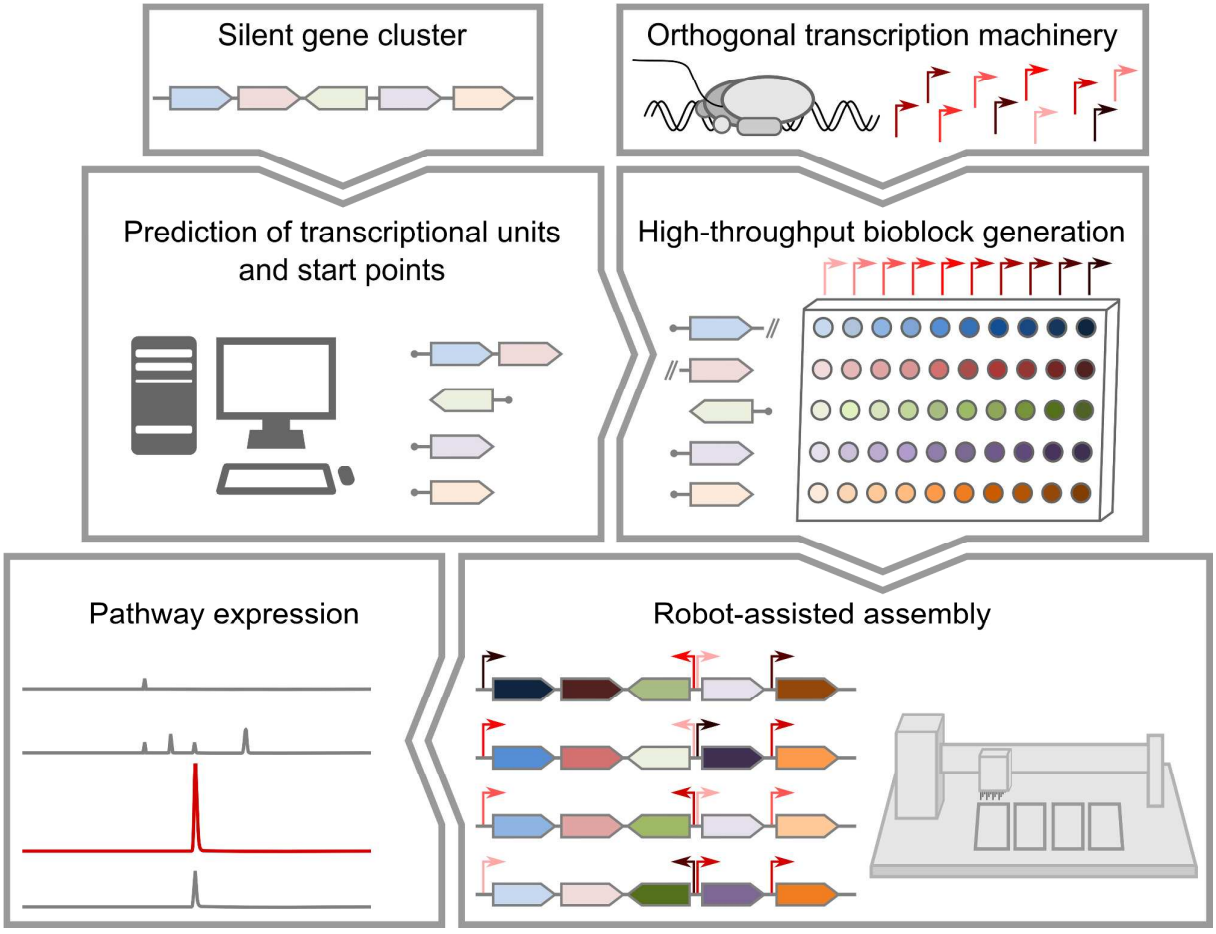


Fig. 3 High-end workflow allowing for the high-throughput activation of silent gene clusters.

Another issue for discussion is the nature of the promoters in the constructed libraries and their behaviour during cluster activation. As previously mentioned, all promoter elements described in this review can be divided into two major categories: promoter regions and minimal promoters. Promoter regions are relatively large chromosomal regions containing a native promoter or promoters and possibly other genetic elements. Minimal promoters encompass defined minimal sequences of native promoters and their derivatives, which are frequently regarded as synthetic promoters. During the generation of these synthetic promoters, the -10 and -35 core sequences of the parent promoters are preserved and only the adjacent sequences are changed. From a genetic perspective, these promoters can be regarded as mutant promoters because they are still recognized by the same RNA polymerase holoenzyme or the same sigma factor[126]. Therefore, promoter regions and minimal promoters constitute the underlying mechanisms of global host regulation, and the activity of these promoters can be changed, especially under stress conditions[61,127]. Since the activation of silent clusters in the same host can be problematic, especially when the strain is refractory to genetic manipulations or is uncultivable under standard laboratory conditions, the refactored and activated clusters are primarily expressed in genetically amenable heterologous hosts that

have never previously encountered these pathways and their corresponding compounds. The production of these foreign compounds can drastically change the transcription pattern of the encoding genes and indirectly alter the genes in the activated cluster[127]. The use of orthogonal RNA polymerase and the cognate orthogonal promoter library, which includes promoters that are completely decoupled from the host regulatory network, can help avoid these potential issues. Construction of the host-decoupled and autonomously self-regulated system for the expression of T7 polymerase was reported for *E. coli* DH10B, *Pseudomonas putida* KT2440 and *Bacillus subtilis* 168[128]. It is worth mentioning that studies have attempted to express orthogonal T7 polymerase in *Streptomyces lividans*; however, the yield of the target protein was lower than that with conventional noninducible expression systems[129].

In conclusion, the natural product discovery field is on the verge of a new qualitative and quantitative level of cluster activation technology (Fig. 3). An extensive promoter toolkit that spans a 1,000-fold activity range is available, and the number of sequenced genomes is steadily growing, which provides the perfect starting conditions for developing and improving cluster activation platforms. Constantly improving cluster prediction and annotation tools allow for the rapid identification of potent candidates for activation[130–132]. Breakthrough technologies such as Red/ET- and TAR-recombineering approaches allow for rapid cluster cloning without requiring the construction of genomic libraries and further cluster refactoring[133,134]. For complex clusters, numerous DNA assembly technologies, such as yeast-assisted assembly, Gibson assembly and Golden Gate Cloning, allow for separate cluster modules to be engineered and rapidly assembled into the complete pathway[133,135–137]. These technologies also empower a robotic-based mix-and-match approach for high-throughput balancing of the engineered pathways. Several genome-minimized *Streptomyces* strains with improved host properties have been made available for refactored gene cluster expression[56,138,139]. The simplified metabolite backgrounds of these engineered host strains advance the detection limits of expressed compounds, and the improved precursor supply positively affects the production yields of these compounds.

6. Acknowledgements

This work was supported by the European Commission under the 7th Framework Programme through the “Collaborative Project” action, “STREPSYNTH” grant No. 613877 and through the ERC starting grant EXPLOGEN No. 281623 to AL.

7. References

1 S. D. Bentley, K. F. Chater, A.-M. Cerdeño-Tárraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C.-H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M.-A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill and D. A. Hopwood, *Nature*, 2002, **417**, 141–147.

2 H. Ikeda, J. Ishikawa, A. Hanamoto, M. Shinose, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori and S. Omura, *Nat. Biotechnol.*, 2003, **21**, 526–531.

3 Y. Ohnishi, J. Ishikawa, H. Hara, H. Suzuki, M. Ikenoya, H. Ikeda, A. Yamashita, M. Hattori and S. Horinouchi, *J. Bacteriol.*, 2008, **190**, 4050–4060.

4 M. J. Bibb, *Curr. Opin. Microbiol.*, 2005, **8**, 208–215.

5 M. Bibb and A. Hesketh, *Methods Enzymol.*, 2009, **458**, 93–116.

6 S. Rigali, F. Titgemeyer, S. Barends, S. Mulder, A. W. Thomae, D. A. Hopwood and G. P. van Wezel, *EMBO Rep.*, 2008, **9**, 670–675.

7 A. Craney, C. Ozimok, S. M. Pimentel-Elardo, A. Capretta and J. R. Nodwell, *Chem. Biol.*, 2012, **19**, 1020–1027.

8 Y. Tanaka, T. Hosaka and K. Ochi, *J. Antibiot. (Tokyo)*, 2010, **63**, 477–481.

9 C. Olano, I. García, A. González, M. Rodriguez, D. Rozas, J. Rubio, M. Sánchez-Hidalgo, A. F. Braña, C. Méndez and J. A. Salas, *Microb. Biotechnol.*, 2014, **7**, 242–256.

10 H. B. Bode, B. Bethe, R. Höfs and A. Zeeck, *Chembiochem Eur. J. Chem. Biol.*, 2002, **3**, 619–627.

11 Y. Tanaka, K. Kasahara, Y. Hirose, K. Murakami, R. Kugimiya and K. Ochi, *J. Bacteriol.*, 2013, **195**, 2959–2970.

12 M. H. Medema, R. Breitling and E. Takano, *Methods Enzymol.*, 2011, **497**, 485–502.

13 J. Huff, A. Czyz, R. Landick and M. Niederweis, *Gene*, 2010, **468**, 8–19.

14 N. Seghezzi, P. Amar, B. Koebmann, P. R. Jensen and M.-J. Virolle, *Appl. Microbiol. Biotechnol.*, 2011, **90**, 615–623.

15 M. M. Rudolph, M.-P. Vockenhuber and B. Suess, *Methods Enzymol.*, 2015, **550**, 283–299.

16 M. M. Rudolph, M.-P. Vockenhuber and B. Suess, *Microbiol. Read. Engl.*, 2013, **159**, 1416–1422.

17 H. M. Salis, *Methods Enzymol.*, 2011, **498**, 19–42.

18 W. R. Strohl, *Nucleic Acids Res.*, 1992, **20**, 961–974.

19 W. Wang, X. Li, J. Wang, S. Xiang, X. Feng and K. Yang, *Appl. Environ. Microbiol.*, 2013, **79**, 4484–4492.

20 Y. Luo, L. Zhang, K. W. Barton and H. Zhao, *ACS Synth. Biol.*, 2015, **4**, 1001–1010.

21 M. J. Bibb, G. R. Janssen and J. M. Ward, *Gene*, 1985, **38**, 215–226.

22 H. Uchiyama and B. Weisblum, *Gene*, 1985, **38**, 103–110.

23 T. Kieser, *Practical Streptomyces Genetics*, John Innes Foundation, 2000.

24 M. J. Bibb, J. White, J. M. Ward and G. R. Janssen, *Mol. Microbiol.*, 1994, **14**, 533–545.

25 T. Schmitt-John and J. W. Engels, *Appl. Microbiol. Biotechnol.*, 1992, **36**, 493–498.

26 T. Siegl, B. Tokovenko, M. Myronovskyi and A. Luzhetskyy, *Metab. Eng.*, 2013, **19**, 98–106.

27 G. Labes, M. Bibb and W. Wohlleben, *Microbiol. Read. Engl.*, 1997, **143 (Pt 5)**, 1503–1512.

28 N. Zaburannyi, M. Rabyk, B. Ostash, V. Fedorenko and A. Luzhetskyy, *BMC Genomics*, 2014, **15**, 97.

29 K. Flärdh and M. J. Buttner, *Nat. Rev. Microbiol.*, 2009, **7**, 36–49.

30 M. J. Bibb and M. J. Buttner, *J. Bacteriol.*, 2003, **185**, 2338–2345.

31 K. F. Chater, *Curr. Opin. Microbiol.*, 2001, **4**, 667–673.

32 L. Potúcková, G. H. Kelemen, K. C. Findlay, M. A. Lonetto, M. J. Buttner and J. Kormanec, *Mol. Microbiol.*, 1995, **17**, 37–48.

33 E.-J. Lee, N. Karoonuthaisiri, H.-S. Kim, J.-H. Park, C.-J. Cha, C. M. Kao and J.-H. Roe, *Mol. Microbiol.*, 2005, **57**, 1252–1264.

34 M. S. Paget, V. Molle, G. Cohen, Y. Aharonowitz and M. J. Buttner, *Mol. Microbiol.*, 2001, **42**, 1007–1020.

35 W. R. Bourn and B. Babb, *Nucleic Acids Res.*, 1995, **23**, 3696–3703.

36 W. R. Strohl, *Nucleic Acids Res.*, 1992, **20**, 961–974.

37 J. P. Gomez-Escribano, L. Song, D. J. Fox, V. Yeo, M. J. Bibb and G. L. Challis, *Chem. Sci.*, 2012, **3**, 2716–2720.
38 E. Takano, H. Kinoshita, V. Mersinias, G. Bucca, G. Hotchkiss, T. Nihira, C. P. Smith, M. Bibb, W. Wohlleben and K. Chater, *Mol. Microbiol.*, 2005, **56**, 465–479.
39 G. Xu, J. Wang, L. Wang, X. Tian, H. Yang, K. Fan, K. Yang and H. Tan, *J. Biol. Chem.*, 2010, **285**, 27440–27448.
40 J. Wang, W. Wang, L. Wang, G. Zhang, K. Fan, H. Tan and K. Yang, *Mol. Microbiol.*, 2011, **82**, 236–250.
41 K. L. Brown, S. Wood and M. J. Buttner, *Mol. Microbiol.*, 1992, **6**, 1133–1139.
42 J.-D. Pédelacq, S. Cabantous, T. Tran, T. C. Terwilliger and G. S. Waldo, *Nat. Biotechnol.*, 2006, **24**, 79–88.
43 C. Bai, Y. Zhang, X. Zhao, Y. Hu, S. Xiang, J. Miao, C. Lou and L. Zhang, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 12181–12186.
44 A. Rodríguez-García, P. Combes, R. Pérez-Redondo, M. C. A. Smith and M. C. M. Smith, *Nucleic Acids Res.*, 2005, **33**, e87.
45 K. Hammer, I. Mijakovic and P. R. Jensen, *Trends Biotechnol.*, 2006, **24**, 53–55.
46 P. R. Jensen and K. Hammer, *Biotechnol. Bioeng.*, 1998, **58**, 191–195.
47 P. R. Jensen and K. Hammer, *Appl. Environ. Microbiol.*, 1998, **64**, 82–87.
48 C. Solem and P. R. Jensen, *Appl. Environ. Microbiol.*, 2002, **68**, 2397–2403.
49 Z. Shao, G. Rao, C. Li, Z. Abil, Y. Luo and H. Zhao, *ACS Synth. Biol.*, 2013, **2**, 662–669.
50 C. Ingram, M. Brawner, P. Youngman and J. Westpheling, *J. Bacteriol.*, 1989, **171**, 6617–6624.
51 R. H. Baltz, *J. Ind. Microbiol. Biotechnol.*, 2010, **37**, 759–772.
52 P. Arias, M. A. Fernández-Moreno and F. Malpartida, *J. Bacteriol.*, 1999, **181**, 6958–6968.
53 S. V. Sohoni, A. Fazio, C. T. Workman, I. Mijakovic and A. E. Lantz, *PloS One*, 2014, **9**, e99701.
54 G. C. Uguru, K. E. Stephens, J. A. Stead, J. E. Towle, S. Baumberg and K. J. McDowall, *Mol. Microbiol.*, 2005, **58**, 131–150.
55 S. Li, J. Wang, X. Li, S. Yin, W. Wang and K. Yang, *Microb. Cell Factories*, 2015, **14**, 172.
56 J. P. Gomez-Escribano and M. J. Bibb, *Microb. Biotechnol.*, 2011, **4**, 207–215.
57 K. Fan, G. Pan, X. Peng, J. Zheng, W. Gao, J. Wang, W. Wang, Y. Li and K. Yang, *Chem. Biol.*, 2012, **19**, 1381–1390.
58 K. F. Chater and L. C. Wilde, *J. Gen. Microbiol.*, 1980, **116**, 323–334.
59 K. Terpe, *Appl. Microbiol. Biotechnol.*, 2006, **72**, 211–222.
60 F. Saïda, *Curr. Protoc. Protein Sci. Editor. Board John E Coligan Al*, 2007, **Chapter 5**, Unit 5.19.
61 T. Murakami, T. G. Holt and C. J. Thompson, *J. Bacteriol.*, 1989, **171**, 1459–1466.
62 M. L. Chiu, M. Folcher, T. Katoh, A. M. Puglia, J. Vohradsky, B. S. Yun, H. Seto and C. J. Thompson, *J. Biol. Chem.*, 1999, **274**, 20578–20586.
63 M. L. Chiu, M. Folcher, P. Griffin, T. Holt, T. Klatt and C. J. Thompson, *Biochemistry (Mosc.)*, 1996, **35**, 2332–2341.
64 D. J. Holmes, J. L. Caso and C. J. Thompson, *EMBO J.*, 1993, **12**, 3183–3191.
65 M. L. Chiu, P. H. Viollier, T. Katoh, J. J. Ramsden and C. J. Thompson, *Biochemistry (Mosc.)*, 2001, **40**, 12950–12958.
66 E. Takano, J. White, C. J. Thompson and M. J. Bibb, *Gene*, 1995, **166**, 133–137.
67 M. Lara, L. Servín-González, M. Singh, C. Moreno, I. Cohen, M. Nimtz and C. Espitia, *Appl. Environ. Microbiol.*, 2004, **70**, 679–685.
68 J. Weaden and P. Dyson, *Microbiol. Read. Engl.*, 1998, **144 (Pt 7)**, 1963–1970.
69 N. Ali, P. R. Herron, M. C. Evans and P. J. Dyson, *Microbiol. Read. Engl.*, 2002, **148**, 381–390.
70 H. Komeda, Y. Hori, M. Kobayashi and S. Shimizu, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 10572–10577.
71 S. Herai, Y. Hashimoto, H. Higashibata, H. Maseda, H. Ikeda, S. Omura and M. Kobayashi, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 14031–14035.
72 C. P. Smith and K. F. Chater, *Mol. Gen. Genet. MGG*, 1988, **211**, 129–137.
73 C. P. Smith and K. F. Chater, *J. Mol. Biol.*, 1988, **204**, 569–580.
74 Z. Hindle and C. P. Smith, *Mol. Microbiol.*, 1994, **12**, 737–745.
75 E. T. Seno and K. F. Chater, *J. Gen. Microbiol.*, 1983, **129**, 1403–1413.
76 M. S. Paget, G. Hintermann and C. P. Smith, *Gene*, 1994, **146**, 105–110.
77 G. Bucca, A. M. Brassington, H. J. Schönfeld and C. P. Smith, *Mol. Microbiol.*, 2000, **38**, 1093–1103.

78 L. Horbal, V. Fedorenko and A. Luzhetskyy, *Appl. Microbiol. Biotechnol.*, 2014, **98**, 8641–8655.
79 Y. Huang, K.-X. Zhao, X.-H. Shen, M. T. Chaudhry, C.-Y. Jiang and S.-J. Liu, *Appl. Environ. Microbiol.*, 2006, **72**, 7238–7245.
80 T. Li, K. Zhao, Y. Huang, D. Li, C.-Y. Jiang, N. Zhou, Z. Fan and S.-J. Liu, *Appl. Environ. Microbiol.*, 2012, **78**, 6009–6016.
81 E. N. Choi, M. C. Cho, Y. Kim, C.-K. Kim and K. Lee, *Microbiol. Read. Engl.*, 2003, **149**, 795–805.
82 J.-T. Zheng, S.-L. Wang and K.-Q. Yang, *Appl. Microbiol. Biotechnol.*, 2007, **76**, 883–888.
83 J. L. Doull, S. W. Ayer, A. K. Singh and P. Thibault, *J. Antibiot. (Tokyo)*, 1993, **46**, 869–871.
84 J.-T. Zheng, U. Rix, L. Zhao, C. Mattingly, V. Adams, Q. Chen, J. Rohr and K.-Q. Yang, *J. Antibiot. (Tokyo)*, 2005, **58**, 405–408.
85 L. Han, K. Yang, E. Ramalingam, R. H. Mosher and L. C. Vining, *Microbiol. Read. Engl.*, 1994, **140 (Pt 12)**, 3379–3389.
86 L. Wang and L. C. Vining, *Microbiol. Read. Engl.*, 2003, **149**, 1991–2004.
87 J. L. Doull, A. K. Singh, M. Hoare and S. W. Ayer, *J. Ind. Microbiol.*, 1994, **13**, 120–125.
88 V. Dangel, L. Westrich, M. C. M. Smith, L. Heide and B. Gust, *Appl. Microbiol. Biotechnol.*, 2010, **87**, 261–269.
89 A. Maxwell and D. M. Lawson, *Curr. Top. Med. Chem.*, 2003, **3**, 283–303.
90 M. Steffensky, A. Mühlenweg, Z.-X. Wang, S.-M. Li and L. Heide, *Antimicrob. Agents Chemother.*, 2000, **44**, 1214–1222.
91 V. Dangel, J. Härle, C. Goerke, C. Wolz, B. Gust, J.-L. Pernodet and L. Heide, *Microbiol. Read. Engl.*, 2009, **155**, 4025–4035.
92 B. Floriano and M. Bibb, *Mol. Microbiol.*, 1996, **21**, 385–396.
93 A. S. Eustáquio, B. Gust, S.-M. Li, S. Pelzer, W. Wohlleben, K. F. Chater and L. Heide, *Chem. Biol.*, 2004, **11**, 1561–1572.
94 M. Jakobi, G. Winkelmann, D. Kaiser, C. Kempler, G. Jung, G. Berg and H. Bahl, *J. Antibiot. (Tokyo)*, 1996, **49**, 1101–1104.
95 K. Jomon, Y. Kuroda, M. Ajisaka and H. Sakai, *J. Antibiot. (Tokyo)*, 1972, **25**, 271–280.
96 P. R. Graupner, S. Thornburgh, J. T. Mathieson, E. L. Chapin, G. M. Kemmitt, J. M. Brown and C. E. Snipes, *J. Antibiot. (Tokyo)*, 1997, **50**, 1014–1019.
97 S. Kanazawa, N. Fusetani and S. Matsunaga, *Tetrahedron Lett.*, 1993, **34**, 1065–1068.
98 H. Shigemori, M. A. Bae, K. Yazawa, T. Sasaki and J. Kobayashi, *J. Org. Chem.*, 1992, **57**, 4317–4320.
99 S. P. Gunasekera, M. Gunasekera and P. McCarthy, *J. Org. Chem.*, 1991, **56**, 4830–4833.
100 J. A. V. Blodgett, D.-C. Oh, S. Cao, C. R. Currie, R. Kolter and J. Clardy, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 11692–11697.
101 Y. Luo, H. Huang, J. Liang, M. Wang, L. Lu, Z. Shao, R. E. Cobb and H. Zhao, *Nat. Commun.*, 2013, **4**, 2894.
102 Z. Shao, H. Zhao and H. Zhao, *Nucleic Acids Res.*, 2009, **37**, e16.
103 H. Takahashi, T. Kumagai, K. Kitani, M. Mori, Y. Matoba and M. Sugiyama, *J. Biol. Chem.*, 2007, **282**, 9073–9081.
104 R. Novakova, Z. Odnogova, P. Kutas, L. Feckova and J. Kormanec, *Folia Microbiol. (Praha)*, 2010, **55**, 119–125.
105 D. Yu, F. Xu, J. Valiente, S. Wang and J. Zhan, *J. Ind. Microbiol. Biotechnol.*, 2013, **40**, 159–168.
106 S. Reverchon, C. Rouanet, D. Expert and W. Nasser, *J. Bacteriol.*, 2002, **184**, 654–665.
107 M. Isaka, A. Jaturapat, J. Kramyu, M. Tanticharoen and Y. Thebtaranonth, *Antimicrob. Agents Chemother.*, 2002, **46**, 1112–1113.
108 K. Kakinuma, C. A. Hanson and K. L. Rinehart, *Tetrahedron*, 1976, **32**, 217–222.
109 Y. S. Choi, T. W. Johannes, M. Simurdiak, Z. Shao, H. Lu and H. Zhao, *Mol. Biosyst.*, 2010, **6**, 336–338.
110 D. Montiel, H.-S. Kang, F.-Y. Chang, Z. Charlop-Powers and S. F. Brady, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 8953–8958.
111 K. S. Ryan and C. L. Drennan, *Chem. Biol.*, 2009, **16**, 351–364.
112 F.-Y. Chang and S. F. Brady, *J. Am. Chem. Soc.*, 2011, **133**, 9996–9999.
113 K. S. Ryan, *PLoS ONE*, 2011, **6**, e23694.

114 P. J. Rutledge and G. L. Challis, *Nat. Rev. Microbiol.*, 2015, **13**, 509–523.
115 P. K. Ajikumar, W.-H. Xiao, K. E. J. Tyo, Y. Wang, F. Simeon, E. Leonard, O. Mucha, T. H. Phon, B. Pfeifer and G. Stephanopoulos, *Science*, 2010, **330**, 70–74.
116 J. R. Anthony, L. C. Anthony, F. Nowroozi, G. Kwon, J. D. Newman and J. D. Keasling, *Metab. Eng.*, 2009, **11**, 13–19.
117 M. J. Smanski, S. Bhatia, D. Zhao, Y. Park, L. B A Woodruff, G. Giannoukos, D. Ciulla, M. Busby, J. Calderon, R. Nicol, D. B. Gordon, D. Densmore and C. A. Voigt, *Nat. Biotechnol.*, 2014, **32**, 1241–1249.
118 I. Farasat, M. Kushwaha, J. Collens, M. Easterbrook, M. Guido and H. M. Salis, *Mol. Syst. Biol.*, 2014, **10**, 731.
119 C. Y. Ng, I. Farasat, C. D. Maranas and H. M. Salis, *Metab. Eng.*, 2015, **29**, 86–96.
120 H. M. Salis, E. A. Mirsky and C. A. Voigt, *Nat. Biotechnol.*, 2009, **27**, 946–950.
121 M. Teilhet, M. B. Rashid, A. Hawk, A. Al-Qahtani and K. Mensa-Wilmot, *Gene*, 1998, **222**, 91–97.
122 M. Mie, S. Shimizu, F. Takahashi and E. Kobatake, *Biochem. Biophys. Res. Commun.*, 2008, **373**, 48–52.
123 T. Tian and H. M. Salis, *Nucleic Acids Res.*, 2015, **43**, 7137–7151.
124 J. Holátko, R. Silar, A. Rabatinová, H. Sanderová, P. Halada, J. Nešvera, L. Krásný and M. Pátek, *Appl. Microbiol. Biotechnol.*, 2012, **96**, 521–529.
125 K. Pfeifer-Sancar, A. Mentz, C. Rückert and J. Kalinowski, *BMC Genomics*, 2013, **14**, 888.
126 A. Feklístov, B. D. Sharon, S. A. Darst and C. A. Gross, *Annu. Rev. Microbiol.*, 2014, **68**, 357–376.
127 A. Urban, S. Eckermann, B. Fast, S. Metzger, M. Gehling, K. Ziegelbauer, H. Rübsamen-Waigmann and C. Freiberg, *Appl. Environ. Microbiol.*, 2007, **73**, 6436–6443.
128 M. Kushwaha and H. M. Salis, *Nat. Commun.*, 2015, **6**, 7832.
129 F.-X. Lussier, F. Denis and F. Shareck, *Appl. Environ. Microbiol.*, 2010, **76**, 967–970.
130 M. H. Medema, K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano and R. Breitling, *Nucleic Acids Res.*, 2011, **39**, W339-346.
131 K. Blin, M. H. Medema, D. Kazempour, M. A. Fischbach, R. Breitling, E. Takano and T. Weber, *Nucleic Acids Res.*, 2013, **41**, W204-212.
132 T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, S. Y. Lee, M. A. Fischbach, R. Müller, W. Wohlleben, R. Breitling, E. Takano and M. H. Medema, *Nucleic Acids Res.*, 2015, **43**, W237-243.
133 N. Kouprina and V. Larionov, *Nat. Rev. Genet.*, 2006, **7**, 805–812.
134 X. Bian, F. Huang, F. A. Stewart, L. Xia, Y. Zhang and R. Müller, *Chembiochem Eur. J. Chem. Biol.*, 2012, **13**, 1946–1952.
135 C. Engler, R. Kandzia and S. Marillonnet, *PloS One*, 2008, **3**, e3647.
136 D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison and H. O. Smith, *Nat. Methods*, 2009, **6**, 343–345.
137 D. G. Gibson, G. A. Benders, K. C. Axelrod, J. Zaveri, M. A. Algire, M. Moodie, M. G. Montague, J. C. Venter, H. O. Smith and C. A. Hutchison, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 20404–20409.
138 M. Komatsu, T. Uchiyama, S. Omura, D. E. Cane and H. Ikeda, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 2646–2651.
139 M. Komatsu, K. Komatsu, H. Koiwai, Y. Yamada, I. Kozone, M. Izumikawa, J. Hashimoto, M. Takagi, S. Omura, K. Shin-ya, D. E. Cane and H. Ikeda, *ACS Synth. Biol.*, 2013, **2**, 384–396.