Regulators Influencing the Heterologous Production of Liponucleoside Antibiotics in *Streptomyces coelicolor* M512

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Sarah Wilcken

aus Hattingen

Tübingen

2024

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	17.05.2024
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Leonard Kaysser
2. Berichterstatter/-in:	Prof. Dr. Harald Groß

Table of contents

Abbreviations

3-MG	3-methylglutaryl
°C	degree celsius
β-ME	β-mercaptoethanol
µ	micro
aa	amino acid(s)
Amp	ampicillin
Apra	apramycin
APS	ammoniumpersulfate
BGC	biosynthetic gene cluster
BLI	biolayer interferometry
bp	base pair
BSA	bovine serum albumin
CA	California
cAMP	cyclic adenosine-mono-phosphate
Carb	carbenicillin
Cas	CRISPR associated
CCR	carbon catabolite repression
cDNA	complementary DNA
Cml	chloramphenicol
CoA	coenzyme A
Cpz	caprazamycin
CRISPR	clustered regularly interspaced short palindromic repeats
CRP	cAMP receptor protein
CSR	cluster-situated regulator
CV	column volume
d	day(s)
dNTP	deoxyribonucleoside 5'-triphosphate
Da	Dalton
DACA	DNA affinity capturing assay
dH_2O	destilled water
DNA	deoxyribonucleic acid
ds	double-strand(ed)
DTT	dithiothreitol
e.g.	*exempli gratia*

EIC	extracted ion chromatogram
EMSA	electromobility shift assay
et al.	*et alii* (and others)
Fig.	figure
fwd	forward
g	gram
gDNA	genomic DNA
GlcNAc	*N*-acetylglucosamine
h	hour
HCl	hydrochloric acid
HF	high fidelity
his	histidine
HPLC	high performance liquid chromatography
i.e.	*id est*
IPTG	isopropyl β-D-1-thiogalactopyranoside
Kan	kanamycin
k	kilo
L	liter
Lpm	liposidomycin
M	molar
m	milli/meter
MA	Massachusetts
MCS	multiple cloning site
Mg	magnesium
MIC	minimal inhibitory concentration
min	minute
MO	Missouri
MS	mass spectrometry
MurNAc	*N*-acetylmuramic acid
MW	molecular weight
m/z	mass-to-charge ratio
n	nano
NaCl	sodium chloride
NaOH	sodium hydroxide
NC	North Carolina
Ndx	nalidixic acid
OD_{600}	optical density measured at 600 nm wavelength

oriT	origin of transfer
PA	Pennsylvania
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
pH	pondus hydrogenii
PKS	polyketide synthase
PMSF	phenylmethylsulfonylfluorid
psi	pound-force per square inch
qPCR	quantitative polymerase chain reaction
^{R}	resistant
rev	reverse
RNA	ribonucleic acid
ROK	repressor, open-reading-frame, kinase
RT	real time/room temperature
Rt	retention time
s	second
S	Svedberg
SEC	size-exclusion chromatography
sgRNA	single guide ribonucleic acid
SDS	sodium dodecyl sulphate
ss	single-strand(ed)
Tet	tetracycline
TFR	TetR family regulator
Theo	theophylline
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
TX	Texas
USA	United States of America
UV	ultraviolet
V	Volt
WHO	world health organization
WI	Wisconsin
WT	wild-type
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

List of tables

List of figures

1 Summary

Heterologous expression of biosynthetic gene clusters is a routinely applied approach in natural product research and facilitates studies on the genes involved in the biosynthesis of the resulting compound and possible strategies for yield improvements. However, global regulation of the heterologously expressed gene cluster or potential metabolic adaptations within the host are largely disregarded.

A previous study suggested an interplay between the heterologously expressed novobiocin gene cluster and precursor supply from primary metabolism via regulators encoded by the host (Bekiesch *et al*., 2016b). The present study likewise aimed for the identification and characterization of regulators, that interact with the caprazamycin and liposidomycin biosynthetic gene clusters, heterologously expressed in *Streptomyces coelicolor* M512.

Liposidomycins and caprazamycins are naturally produced by several *Streptomyces* strains and were discovered in the mid-1980s and early 2000s, respectively (Isono *et al*., 1985; Igarashi *et al*., 2003). Both liponucleoside antibiotics convey strong antimycobacterial bioactivity and share high similarities regarding their structural scaffold, thereby relying on the same building blocks for biosynthesis. Furthermore, the identification of the underlying biosynthetic gene clusters revealed a highly similar genetic arrangement and sequence homologies (Kaysser *et al*., 2009; Kaysser *et al*., 2010a). Therefore, similar regulatory responses in the heterologous strain *Streptomyces coelicolor* M512 harboring either the entire caprazamycin (cpzLK09) or liposidomycin (lpmLK01) gene cluster were hypothesized.

In the present study, DNA-affinity capturing assays, coupled with subsequent mass-spectrometry, were used to trap and identify host-derived regulators on selected intergenic regions of the caprazamycin and liposidomycin gene clusters. The intergenic fragments were verified as promoter regions by RNA-sequencing and a negative control (promoter region of vegetative sigma factor *hrdB*) was included to assess binding specificity.

In total, 1906 and 1540 proteins were identified at the gene cluster-derived promoter regions from *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01, respectively. Among these, only three proteins bound specifically to the corresponding promoter regions in both clusters, indicating differences in the regulatory interaction with the host. Based either on their specific binding pattern or a known involvement in precursor supply regulation, five putative regulators were selected for further characterization. Among these, overexpression of two genes

resulted in significantly increased caprazamycin levels. Overexpression of the cAMP-receptor protein Sco3571, a regulator of the leucine-isovalerate-utilization pathway, supplying precursors for liponucleoside synthesis (Gao *et al*., 2012; Bär *et al*., 2022), elevated caprazamycin and liposidomycin levels by 3,2- and 2,8-fold, respectively. The second protein, that was identified as a regulator of heterologous caprazamycin biosynthesis in this work is Sco4385, belonging to the TetR-family of regulators. Although this protein bound specifically to a corresponding promoter pair from both gene clusters in the DNA-affinity capturing assays, gene overexpression increased only caprazamycin (3-fold production enhancement) and not liposidomycin production. Therefore, the binding location of Sco4385 within the promoter regions was determined by biolayer-interferometry measurements. Interestingly, the position of the main binding site was found to be different: while in the selected region of the caprazamycin gene cluster, Sco4385 binds within the coding region of the upstream gene, it binds the center of the intergenic region derived from the liposidomycin gene cluster. This finding potentially explains the differential effects on liponucleoside production in both heterologous strains. Moreover, the identification of binding sites allowed the determination of a consensus recognition sequence, which was validated experimentally using biolayer-interferometry as well. However, a genome-wide screening for putative regulatory target genes of Sco4385 in *S. coelicolor* did not reveal an obvious connection to liponucleoside biosynthesis.

The last part of this work focused on the characterization of the cluster-situated regulators Cpz9 and LpmG, from the caprazamycin and liposidomycin biosynthetic gene cluster, respectively. Thus, the according genes, encoding for putative AraC-type activators, were individually overexpressed in the heterologous producer strains. While *cpz9* overexpression yielded slightly enhanced (1,35-fold) productivity of caprazamycins in *S. coelicolor* M512/cpzLK09, *lpmG* overexpression in *S. coelicolor* M512/lpmLK01 led to a strong 3,5-fold overproduction of liposidomycins. Therefore, low *lpmG* expression was identified as a plausible limiting factor for liposidomycin biosynthesis in the heterologous host.

Zusammenfassung

Die heterologe Expression von Biosynthese-Genclustern ist ein häufig angewandtes Vorgehen in der Naturstoffforschung und erleichtert Untersuchungen der involvierten Gene an der Biosynthese des resultierenden Naturstoffs, sowie die Entwicklung möglicher Strategien zur Produktionssteigerung. Die Auswirkungen globaler Regulatoren des Wirtsstammes auf die Expression eingebrachter Gencluster, sowie die Anpassungen im Stoffwechsel des Wirtsstammes finden jedoch größtenteils kaum Beachtung.

Eine vorangegangene Studie deutete auf ein regulatorisches Zusammenspiel zwischen der heterologen Expression der Biosynthesegene von Novobiocin auf der einen Seite, und der Bereitstellung von biosynthetischen Vorstufen aus dem Primärmetabolismus des Wirtes auf der anderen Seite hin (Bekiesch *et al*., 2016b). Ziel der vorliegenden Arbeit war es daher, Regulatoren des heterologen Wirtes, *Streptomyces coelicolor* M512, zu identifizieren, die mit den eingebrachten Genclustern für die beiden Liponucleosid-Antibiotika Caprazamycin und Liposidomycin interagieren und deren Produktion beeinflussen.

Liposidomycine und Caprazamycine werden von verschiedenen *Streptomyces* Stämmen produziert und wurden Mitte der 1980er bzw. Anfang der 2000er Jahre entdeckt (Isono *et al*., 1985; Igarashi *et al*., 2003). Beide Liponucleosid-Antibiotika zeigen eine gute Bioaktivität gegen Mykobakterien, bedingt durch ihre ausgeprägten chemisch strukturellen Ähnlichkeiten. Dementsprechend ähneln sich ihre Biosynthesewege, die auf den gleichen biosynthetischen Vorstufen beruhen, sowie die genetische Organisation der entsprechenden Gencluster, die untereinander eine sehr hohe Sequenzhomologie aufweisen (Kaysser *et al*., 2009; Kaysser *et al*., 2010a). Somit wurde für die heterologen Stämme *S. coelicolor* M512, mit vollständigem Caprazamycin (cpzLK09) oder Liposidomycin (lpmLK01) Gencluster, auch ein vergleichbares Zusammenspiel zwischen Regulatoren des Wirtes und Promoter-Regionen der eingebrachten Gencluster vermutet.

In der vorliegenden Arbeit wurden *DNA-affinity-capturing assays* mit nachfolgender Massenspektrometrie genutzt, um Regulatoren des Wirtsstammes zu identifizieren, die an ausgewählte intergene Bereiche des Caprazamycin- und Liposidomycin-Genclusters binden. Die untersuchten intergenen Bereiche wurden als Promotor-Regionen mittels RNA-Sequenzierung verifiziert. Als Negativkontrolle diente die Promotor-Region des vegetativen Sigma Faktors *hrdB*, um die Bindungsspezifität beurteilen zu können.

Insgesamt wurden jeweils 1906 und 1540 Proteine an den Promotor-Regionen der Gencluster aus *S. coelicolor* M512/cpzLK09 und *S. coelicolor* M512/lpmLK01 identifiziert. Von diesen banden nur drei Proteine spezifisch an zueinander gehörige Promotor-Regionen aus beiden Genclustern, was auf Unterschiede hinsichtlich der regulatorischen Wechselwirkung mit dem Wirtsstamm hindeutete. Basierend auf einer spezifischen Bindung, oder aber einer bereits bekannten Beteiligung an der Regulation zur Bereitstellung von Vorläufermolekülen, wurden fünf mutmaßliche Regulatoren für eine weitere Charakterisierung ausgewählt. Von diesen führte die Überexpression von zwei Genen zu signifikant erhöhter Caprazamycin Produktion. Überexpression des cAMP-Rezeptor Proteins Sco3571, ein Regulator des Leucin-Isovalerat-Abbauwegs, der Vorstufen für die Liponucleosid-Biosynthese bereitstellt, erhöhte Caprazamycin und Liposidomycin Produktionsmengen um jeweils das 3,2- und 2,8-fache. Der zweite Regulator, der die heterologe Caprazamycin-Produktion beeinflusste, war Sco4385 aus der Familie der TetR Regulatoren. Obwohl dieses Protein im *DNA-affiity-capturing assay* spezifisch an einer zueinander korrespondierenden Promotor-Region in beiden Genclustern gebunden hatte, steigerte dessen Gen-Überexpression nur die Caprazamycin (3-fache Produktionssteigerung) und nicht die Liposidomycin-Produktion. Deshalb wurde die Bindestelle von Sco4385 in den Promotor-Regionen durch *biolayer-interferometry* Messungen genauer bestimmt. Die Positionen der primären Bindestelle sind interessanterweise in beiden Genclustern unterschiedlich: während Sco4385 in der ausgewählten Region des Caprazamycin Genclusters im codierenden Bereich des Strang-aufwärts liegenden Gens *cpz9* bindet, bindet es in der Mitte der intergenen Region zwischen *lpmG* und *lpmH* aus dem Liposidomycin Gencluster. Dies könnte die unterschiedlichen Effekte auf die Liponucleosid Produktion in beiden heterologen Stämmen erklären. Zusätzlich bot die Identifikation der Bindestellen die Möglichkeit, eine Konsensus-Erkennungssequenz zu bestimmen, die experimentell ebenfalls mit *biolayer-interferometry* verifiziert wurde. Eine Genom-weite Suche nach möglichen regulatorischen Zielsequenzen von Sco4385 in *S. coelicolor*, erbrachte jedoch keine offensichtliche Verbindung zur Liponucleosid Biosynthese oder deren Vorläufermolekülen.

Im letzten Teil dieser Arbeit wurde der Fokus auf die Charakterisierung der Cluster-spezifischen Regulatoren des Caprazamycin und Liposidomycin Biosynthese-Genclusters, Cpz9 und LpmG, gelegt. Dafür wurden die entsprechenden Gene, die vermutlich für AraC-ähnliche Aktivatoren codieren, einzeln in den heterologen Produzentenstämmen überexprimiert. Während die *cpz9* Überexpression in *S. coelicolor* M512/cpzLK09 nur eine leichte Erhöhung (1,35-fach) der Produktivität

von Caprazamycinen erbrachte, führte die Überexpression von *lpmG* in *S. coelicolor* M512/lpmLK01 zu einer starken, 3,5-fachen Überproduktion von Liposidomycinen. Hierdurch konnte die geringe Expression von *lpmG* als möglicher limitierender Faktor für die Liposidomycin Produktion im heterologen Expressionsstamm identifiziert werden.

2 Introduction

2.1 Nucleoside antibiotics

2.1.1 The importance of nucleoside antibiotics research

Nucleoside- and nucleotide- derived compounds constitute a large group of microbial natural products. Due to the numerous fundamental roles played by these building blocks in cellular physiology and metabolism (Shiraishi and Kuzuyama, 2019), the biological functions of nucleoside antibiotics are extremely diverse. Representatives of this compound class and derivatives thereof may, among others, convey antitumor (Mini *et al*., 2006), insecticidal (Liu *et al*., 2012), immunosuppressive (Robak *et al*., 2006), antifungal (Nix *et al*., 2009), antiviral (Wang *et al*., 2010) or antibacterial (Isono *et al*., 1985; Igarashi *et al*., 2005; Xie *et al*., 2010) activity. Several of them counteract menacing germs and are therefore promising leads in drug development (Salomon *et al*., 2013; Thomson and Lamont, 2019; Larwood, 2020). Thus, the identification, biosynthesis, yield improvement and chemical derivatization of natural nucleoside antibiotics are important research interests to help tackle the worldwide threat of emerging antibiotic-resistant pathogens (Murray *et al*., 2022). In the following paragraph, some nucleoside-based compounds with industrial potency will be introduced.

An example for the successful application of nucleoside natural products is the gougerotin compound family, representing the sub-group of cytidine-derived nucleoside antibiotics. This family comprises gougerotin, bagougeramine and ningnanmycin, that all fulfill diverse biological functions, but are mainly studied for their antiviral activity (Iwasaki, 1962; Takahashi *et al*., 1986). As already mentioned, gougerotin is build up from cytosine, which, together with 4-amino-4-deoxyglucuronamide, forms the nucleoside skeleton, that is coupled to a sarcosyl-D-serine dipeptide (Fox *et al*., 1968). Gougerotin was found to be produced by different *Streptomyces* strains (Iwasaki, 1962; Ikeuchi *et al*., 1972; Arai *et al*., 1974) and heterologous expression of the gougerotin biosynthetic gene cluster (BGC) from *S. graminearus* in *S. coelicolor* allowed studies on its biosynthesis and regulation (Niu *et al*., 2013). The mode of action of gougerotin is to disturb interaction of tRNAs with the 23 S rRNA of the ribosome and thereby block translation (Kirillov *et al*., 1999). A commercially successful member of the gouterotin family is ningnanmycin, which was first isolated from *S. noursei var. xichangensis* (Guxi *et al*., 1995). Ningnanmycin, a stereoisomer of gougerotin, is used in agricultural applications in China to protect industrially relevant plants, such as tomatoes or tobacco, from viral infections (Wang

et al., 2023). By direct binding of the coating protein discs, ningnanmycin inhibits the assembly of the viral coat, which therefore looses its pathogenicity (Li *et al*., 2017). This example illustrates how nucleoside-derived natural products can be successfully exploited on an industrial scale.

A nucleoside antibiotic with high potential for commercial medical application in both humans and animals is nikkomycin Z. It was first isolated from the culture broth of *Streptomyces tendae* Tü 901 (Dähn *et al*., 1976), from which the underlying biosynthetic gene cluster (BGC) was identified more than two decades later (Bruntner *et al*., 1999). The potent chitin synthase inhibitor confers bioactivity against clinically relevant fungi, such as *Candida albicans*, *Aspergillus* spp. and *Coccidioides* spp.. *In vitro* studies reported significant chitin synthesis inhibition of nikkomycin Z at concentrations of 0,5 µg/mL (Larwood, 2020). Moreover, in an *in vivo* model with naturally infected dogs suffering from coccidioidal pneumonia, oral administration of nikkomycin Z led to significantly improved clinicopathological parameters (Shubitz *et al*., 2013). However, low production and challenging yield improvements of nikkomycin Z in *S. tendae* pose a problem for industrial scale manufacturing (Stenland *et al*., 2013), which is an often encountered obstacle in natural product drug development. Hence, basic research on these compounds is of utmost importance to be able to exploit their full potential.

Besides these compounds, studied for their antiviral and antifungal bioactivity, there is a variety of nucleoside antibiotics and their synthetically modified analogues, that act against bacteria and exhibit promising pharmaceutical features (Serpi *et al*., 2016). More specifically, the compound family of uridine-based nucleoside antibiotics will be introduced in the next paragraph.

2.1.2 MraY-inhibiting uridine-derived nucleoside antibiotics

The class of uridine-based nucleoside antibiotics can be divided into several sub-groups (McErlean *et al*., 2021). Herein, based on their common ability to inhibit the bacterial phospho-*N*-acetylmuramyl (MurNAc)-pentapeptide translocase MraY (translocase I), five of these antibiotic families will be highlighted: uridyl peptides, capuramycins, tunicamycins, muraymycins and liponucleosides. Translocase I is an interesting unexploited target of uridine-based nucleoside antibiotics and will be introduced in detail in section 2.2.2.

The first group, called uridyl peptide antibiotics, encompasses sansanmycins, pacidamycins, mureidomycins and napsamycins. They all share highly selective activity against the Gram-negative pathogen *Pseudomonas aeruginosa* (McErlean *et*

al., 2021), which was declared a major global health threat by the WHO (Tacconelli *et al*., 2018). The highest bioactivity is conferred by mureidomycin C with minimal inhibitory concentration (MIC) values between 0,4 µg/mL and 1,5 µg/mL against various *P. aeruginosa* strains (Inukai *et al*., 1993). Several *Streptomyces* species are natural producers of uridyl peptide antibiotics and were the source for their first isolation: Sansanmycins were gained from *Streptomyces* sp. SS (Xie *et al*., 2007), pacidamycins from *S. coeruleorubidus* AB 1183F-64 (Karwowski *et al*., 1989), mureidomycins from *S. flavidovirens* SANK 60486 (Isono *et al*., 1989) and napsamycins from *Streptomyces* sp. HIL Y-82 (Chatterjee *et al*., 1994). Moreover, the biosynthetic gene clusters (BGCs) of napsamycins, pacidamycins and sansanmycins were identified (Zhang *et al*., 2010; Rackham *et al*., 2011; Kaysser *et al*., 2011; Li *et al*., 2013). On the structural level, these compounds are comprised of a 3´-deoxyuridine nucleoside moiety, which is covalently bound to a central *N*-methyl-2,3-diaminobutyric acid (*N*-methyl-DABA) residue within a tetra- or pentapeptide chain via a 4´,5´-enamide bond (McErlean *et al*., 2021). As mentioned before, the molecular target of these compounds is the phospho-MurNAc-pentapeptide translocase MraY, an essential enzyme in bacterial cell wall biosynthesis (Mashalidis *et al*., 2019).

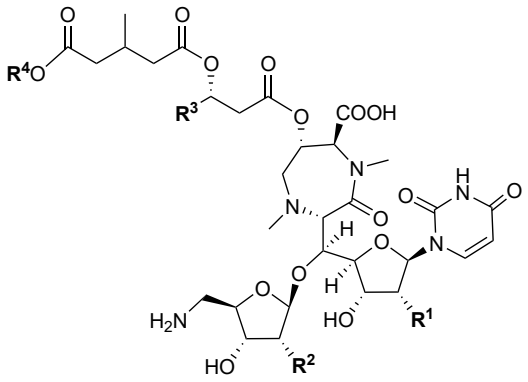
Uridyl glycosylpeptide antibiotics pose another MraY-inhibiting family, with capuramycin as their representative (derivatives of capuramycin called A-500359s were isolated later (Muramatsu *et al*., 2003a; Muramatsu *et al*., 2003b)). Capuramycin is characterized by a 5´-carbamoyluridine core, an aminocaprolactam ring and an unsaturated hexuronic acid. It was isolated from *S. griseus* 446-S3 and shows potency against *Streptococcus pneumoniae* and *Mycobacterium smegmatis* with MICs of 12,5 µg/mL and 3 µg/mL, respectively (Yamaguchi *et al*., 1986). The gene cluster encoding for the biosynthetic genes of A-500359s was identified in *S. griseus* SANK 60196 in 2009 (Funabashi *et al*., 2009).

Tunicamycins were first discovered in 1971 from *S. lysosuperificus* nov. sp., (Takatsuki *et al*., 1971). In the following decades, they were found to be produced by many other *Streptomyces* strains as well (Doroghazi *et al*., 2011) and the associated gene cluster was identified in *S. clavuligerus* ATCC 27064 (Chen *et al*., 2010; Wyszynski *et al*., 2010). Their structure is composed of an aminodialdose core, *N*-acetylglucosamine, uracil and variable fatty acid chains (Takatsuki *et al*., 1971). Besides targeting translocase I (Brandish *et al*., 1996), tunicamycins interfere with mammalian glycoprotein biosynthesis, which makes them less suitable as candidates for drug development compared to other nucleoside antibiotics (Almahayni *et al*., 2022).

In summary, 19 naturally occurring muraymycins were isolated from *Streptomyces* sp. NRRL 30471 (McDonald *et al*., 2002; Cui *et al*., 2018), which was also the source for the identification of the associated biosynthetic gene cluster (Cheng *et al*., 2011). These compounds are build up from a disaccharide core, composed of glycyluridine and 5-amino-5-deoxyribose, which they share with the structurally related liponucleosides (see 2.1.2.1). In contrast, muraymycins lack the diazepanone moiety and instead contain a pseudo-tripeptide consisting of L-epicapreomycidine, L-valine and L-leucine. Hydroxylation or acylation of the L-leucin moiety leads to the distinction of four different series (A-D) of muraymycins (Wiegmann *et al*., 2016). Muraymycin A1 shows bioactivity against a range of bacterial genera, such as *Staphylococci* (2-16 µg/mL MIC) and *Enterococci*, as well as *Escherichia coli* (<0,03 µg/mL MIC; McDonald *et al*., 2002).
The last sub-class are liponucleoside antibiotics, which will be introduced in the following paragraph.

2.1.2.1 Liponucleoside antibiotics

Members of the liponucleoside antibiotics are liposidomycins, caprazamycins, muraminomicins and A-90289s. These compounds share a common scaffold, consisting of a disaccharide core, which is composed of uridine and an aminoribose, a diazepanone ring and a fatty acid moiety. Attachment of the diazepanone ring to the disaccharide results in the formation of a central structure, the (+)-caprazol, which is further extended by addition of β-hydroxy fatty acids of different length and conformation. Both, the (+)-caprazol and the fatty acyl side chains are essential for bioactivity (Kimura *et al*., 1998; Hirano *et al*., 2008; Wiker *et al*., 2019a). However, attachment of a structurally unique 3-methylglutaryl (3-MG) moiety to the β-hydoxyl group of the fatty acid is not required for bioactivity. While muraminomicins carry a succinylated heptopyranose linked to the 3-MG moiety, this component is replaced by a permethylated L-rhamnose in caprazamycins and A-90289s. Liposidomycins, in contrast, completely lack this sugar moiety. Furthermore, muraminomicins are characterized by two 2-deoxy furanoses while the other representatives carry hydroxy- or sulfate-groups at these positions (McErlean *et al*., 2021). The highly similar chemical structures of caprazamycin E, liposidomycin C, muraminomicin F and A-90289 A are representatively depicted in Figure 1.



Liposidomycin C

R^1 = OH or OSO_3H

R^2 = OH

R^3 =

R^4 = H

Caprazamycin E

R^1 = OH or OSO_3H

R^2 = OH

R^3 =

R^4 =

Muraminomicin F

R^1 = H

R^2 = H

R^3 =

R^4 =

A-90289 A

R^1 = OSO_3H

R^2 = OH

R^3 =

R^4 =

Figure 1: Overview of structurally related liponucleoside antibiotics. The common scaffold of liposidomycins, caprazamycins, muraminomicins and A-90289s is shown and distinguishing moieties are depicted for one representative of each compound family.

Except for muraminomicins, which were isolated from *Streptosporangium* sp. SANK-60501 (Fujita *et al*., 2019), all other representatives were first isolated from *Streptomyces* species: liposidomycins from *Streptomyces griseosporus* (Isono *et al*., 1985), caprazamycins from *Streptomyces* sp. MK730-62F2 (Igarashi *et al*., 2003; Igarashi *et al*., 2005) and A-90289s from *Streptomyces* sp. SANK 60405 (Fujita *et al*., 2011).

Moreover, for all of them, the underlying biosynthetic gene clusters (BGCs) have been identified (Kaysser *et al*., 2009; Funabashi *et al*., 2010; Kaysser *et al*., 2010a; Chi *et al*., 2013), which allowed manipulation of the genes and thereby laid the basis for the analysis of their biosynthesis and regulation (see 2.3 and 2.4).

By inhibition of MraY, the liponucleosides convey modest to excellent bioactivity against a range of Gram-positive bacterial pathogens, such as *Streptococcus pyrogenes*, *Streptococcus pneumoniae*, *Enterococcus faecium*, *Staphylococcus aureus* and various species of *Mycobacteria*, most importantly *M. tuberculosis*.

A-90289 A, for example, is potent against *Staphylococcus aureus* ATCC 6538P (methicillin-sensitive strain) at a MIC of 8 µg/mL, against *S. aureus* 10925 (methicilin-resistant strain) at 16 µg/mL MIC and against *Streptococcus pyrogenes* at 4 µg/mL MIC (Fujita *et al*., 2011).

Another example, muraminomicin F, is able to impede growth of various clinical isolates, such as methicillin-resistant *S. aureus* at a MIC of 6,25 µg/mL and penicillin-resistant *S. pneumoniae* at MICs ranging from 3,13 µg/mL to 12,5 µg/mL (Fujita *et al*., 2019).

Especially *M. tuberculosis* is currently causing a resurgence of tuberculosis and an emerging threat to human health. In 2022, a world wide number of 7,5 million people newly got infected with tuberculosis and the disease caused an estimated 1,3 million deaths (Global tuberculosis report, 2023). Due to the increase of antibiotic resistant strains, potent antimycobacterial agents with new molecular targets are urgently needed (Verma *et al*., 2023).

Both caprazamycins and liposidomycins show strong bioactivity against *Mycobacteria*. In the original publication presenting liposidomycins, a MIC of 1,6 µg/mL against *M. phlei* was demonstrated (Isono *et al*., 1985). Caprazamycin B exhibits a MIC of 3,13 µg/mL against *M. tuberculosis* H37Rv and 6,25-12,5 µg/mL against both drug-susceptible and drug-resistant strains of *M. tuberculosis*, indicating no cross-resistance to commercially used antibiotics. At the same time, no significant toxicity was conferred in murine models or in geno- and cytotoxicity assays (Igarashi *et al*., 2003).

Due to their excellent biological properties, liponucleosides were subject of further developments after their discovery. Chemical modification of these naturally derived compounds led to the creation of new (semi)-synthetic derivatives with improved bioactivity and/or pharmacokinetic properties. Kagoshima *et al*. (2019) created semi-synthetic derivatives of muraminomicin Z_1, by attachment of a lipid moiety to the diazepanone ring. Two compounds showed high potency against methicillin-resistant *Staphylococcus aureus* (MRSA) *in vitro* (MIC_{50} of 1 µg/mL) and their therapeutic

efficiency, comparable to that of vancomycin, was validated in mouse-model experiments. Thus, these liponucleoside derivatives represent promising candidates for the development of clinically used drugs.

Other lipouridine antibiotics have been subjected to derivatization as well. The most advanced example is CPZEN-45, a semi-synthetic caprazene-based 4-butylanilide derivative of caprazamycins. This compound was identified among a plethora of other derivates to exhibit the best combination of excellent bioactivity and minimal hemolytic effects. The MIC of CPZEN-45 against *M. tuberculosis* H37Rv is 1.56 µg/mL (which is half the MIC of caprazamycin B) and 6.25 µg/mL against a multidrug-resistant strain of *M. tuberculosis* (Takahashi *et al*., 2013). Interestingly, the structural modifications came along with a target switch from MraY to the mycobacterial *N*-acetylglucosamine-1-phosphate transferase WecA, another essential enzyme in cell wall biosynthesis (Ishizaki *et al*., 2013; Huszár *et al*., 2017). Due to poor oral bioavailability, an administration of CPZEN-45 via inhalation is currently tested in clinical trails (Stewart *et al*., 2022; Kumar and Amrutha, 2023).

2.2 Mode of action of liponucleoside antibiotics

2.2.1 Bacterial cell wall biosynthesis

The cell wall constitutes a major feature that distinguishes bacteria from other forms of life. It fulfills essential tasks for bacterial survival, such as maintenance of the bacterial shape, protecting the cells from disruption by turgor and displays a barrier to external influences (Madigan *et al*., 2013). The main components of the bacterial cell wall is peptidoglycan. It is composed of alternating units of β-(1→4)-linked *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc), which are cross-linked by short peptide bridges (Sibinelli-Sousa *et al*., 2021). Therefore, the mechanical robustness of the peptidoglycan scaffold is strongly increased. The thickness of the peptidoglycan layer and composition of the peptide bridges linking the GlcNAc and MurNAc units, differ significantly between Gram-positive and Gram-negative bacteria. While the peptide bridges in Gram-negatives are mostly composed of L-alanine, D-iso-glutamic acid, *meso*-diaminopimelic acid and a D-alanine-D-alanine motif, they may contain D-iso-glutamine and a diamino acid residue (e.g. L-lysine and D-alanine) in their Gram-positive counterparts (Sibinelli-Sousa *et al*., 2021). However, in the latter group, a large variety in the composition of the peptide stems is observed (Schleifer and Kandler, 1972). Moreover, in Gram-negatives, peptidoglycan only accounts for 10 % of the cell wall (Madigan *et al*., 2013), but, in

contrast to Gram-positives, they possess an outer membrane in addition to the inner (cytoplasmic) membrane. The outer membrane is mainly build up from lipopolysaccharides and plays an important role in conferring toxicity and drug resistance (Zhou *et al*., 2022). In Gram-positives, a peptidoglycan layer that is several times thicker than in Gram-negatives, is installed on the cytoplasmic membrane (Madigan *et al*., 2013).

The biosynthesis of peptidoglycan can be spatially divided into three stages (Zhou *et al*., 2022) and is depicted in Figure 2.

In the cytoplasm, peptidoglycan biosynthesis starts with uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc), which is converted to the soluble cell wall precursor UDP-MurNAc-pentapeptide via the sequential activity of MurA to MurF (Sibinelli-Sousa *et al*., 2021). The second stage is membrane-associated and starts with the activity of the membrane-bound UDP-MurNAc-pentapeptide phosphotransferase MraY. MraY transfers phospho-MurNAc-pentapeptide from the newly synthesized UDP-MurNAc-pentapeptide to undecaprenyl phosphate (C_{55}-P), that is anchored within the cytoplasmic membrane (Ikeda *et al*., 1991; Boyle and Donachie, 1998). This reaction leads to the formation of lipid I, which is subsequently coupled with UDP-GlcNAc by translocase MurG to yield lipid II. Further strain-specific modification and branching of the peptide stem of lipid II is accomplished by activity of the enzyme complex MurT/GatD, MurM and MurN (Filipe *et al*., 2001; Münch *et al*., 2012). Finally, the modified lipid II is switched to the external side of the membrane by lipid II flippase MurJ, where the last steps of peptidoglycan biosynthesis take place. In stage three, conferred by class I penicillin binding proteins, transglycosylation between the glycan components of the lipid II units leads to polymerization and transpeptidation between their peptide stems, performed by class II penicillin binding proteins, results in cross-linking (Figure 2). The newly synthesized peptidoglycan is eventually released from the membrane and is thus available for incorporation into the cell wall.

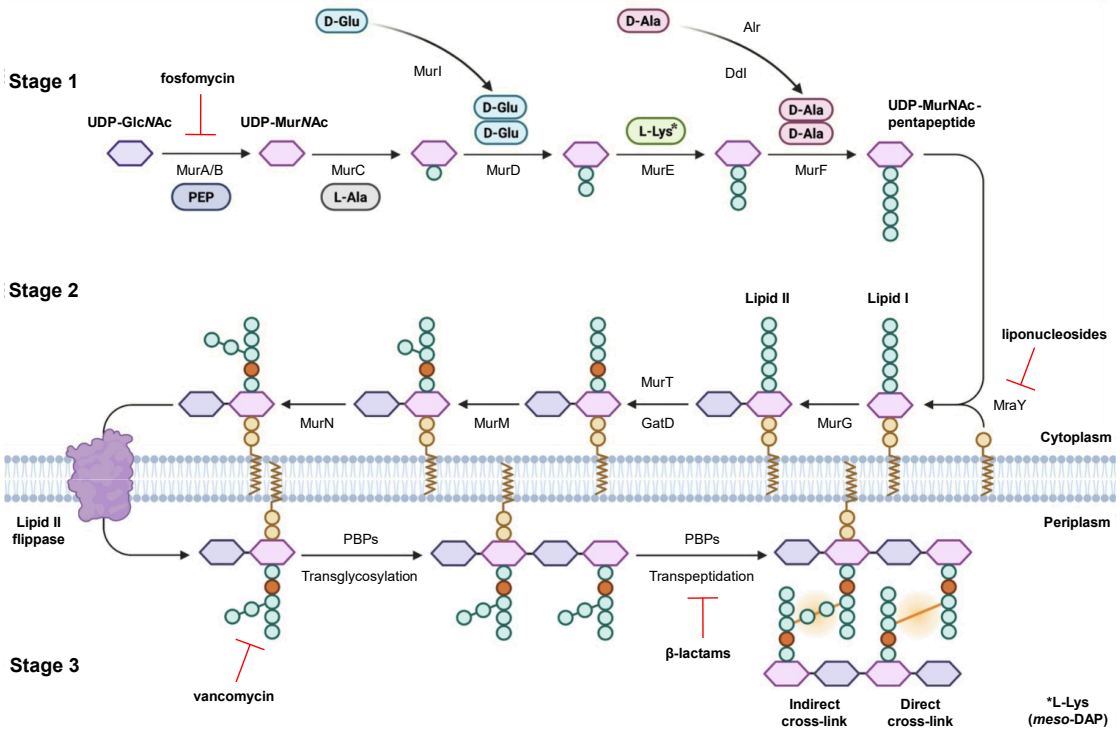


Figure 2: Schematic representation of the peptidoglycan biosynthetic pathway in Gram-negative bacteria. UDP: Uridine diphosphate; GlcNAc: N-acetylglucosamine; MurNAc: N-acetylmuramic acid; PEP: phosphoenol pyruvate; D-/L-Ala: D-/L-alanine; D-Glu: D-glutamic acid; L-Lys: L-lysine; PBPs: penicillin binding proteins; DAP: diaminopimelic acid. Adapted from Zhou *et al.*, 2022.

Due to the fact, that a cell wall and thus peptidoglycan biosynthesis is only present in bacterial cells and not in mammalian ones, almost all essential steps are potential targets for antibiotic activity. One of these targets, the glycosyl transferase MraY will be highlighted in the following section.

2.2.2 The translocase MraY

As illustrated in the previous section, translocase I, or MraY, is one of the central enzymes involved in peptidoglycan biosynthesis. Being located within the cytoplasmic membrane, it catalyzes the reversible and Mg^{2+}-dependent transfer of UDP-MurNAc-pentapeptide to the lipid carrier C_{55}-P, yielding lipid I and UMP. A *mraY* deletion mutant of *Escherichia coli* is no longer able to withstand the osmotic pressure, which eventually results in cell lysis (Boyle and Donachie, 1998). Therefore, MraY is an interesting target in the search for antibiotics with a new mode of action. As set out in paragraph 2.1.2, several uridine-based nucleoside antibiotics are highly potent MraY inhibitors and were extensively subjected to structure-activity-relationship (SAR) studies. Since the first successful crystallization of MraY from *Aquifex aeolicus* in 2013 (Chung *et al.*, 2013; PDB 4J72), several further MraY crystal structures were solved

in complex with various nucleoside antibiotics, such as tunicamycin, capuramycin, carbacaprazamycin (PDB 6OYH), 3´-hydroxymureidomycin A and muraymycin D2 (Chung *et al*., 2016; Hakulinen *et al*., 2017; Mashalidis *et al*., 2019). The tertiary structure of MraY consists of 10 trans-membrane helices, a periplasmic hairpin and helix, five cytoplasmatic loops and an interfacial helix. The N- as well as the C-terminus of the enzyme are located in the periplasm (Chung *et al*., 2013).

The mode of action of nucleoside antibiotics targeting MraY has been addressed in numerous studies and IC_{50} were determined *in vitro* for most of them (Wiegmann *et al*., 2016; Bugg and Kerr, 2019).

Best IC_{50} values were conferred by capuramycins, that inhibit translocase I at an IC_{50} of 0,017 µg/mL (Muramatsu *et al*., 2003b), muraymycins at 0,027 µg/mL and both mureidomycin A and liposidomycin C at a similar value of 0,03 µg/mL (Kimura *et al*., 1989; Wiegmann *et al*., 2016). Moreover, while liposidomycin C inhibits total peptidoglycan biosynthesis at an IC_{50} of 0,038 µg/mL, it was determined to be 12 µg/mL for tunicamycin, based on a comparative study (Kimura *et al*., 1989). Another study showed that liposidomycin B is a non-competitive inhibitor of the UDP-MurNAc-pentapeptide substrate, but competitively inhibits interaction of the lipid carrier with MraY. It is furthermore characterized by slow-binding inhibition (K_i = 80 ± 15 nM), which implies reversible and fast formation of the enzyme-inhibitor complex (Brandish *et al*., 1996).

2.3 Biosynthesis of caprazamycins and liposidomycins

The previously introduced liponucleoside antibiotics caprazamycins and liposidomycins were subject of the present study. Especially the biosynthesis of caprazamycins has been studied quite extensively since the identification of the associated BGC in 2009 and is understood to some extend. The caprazamycin BGC was initially identified in *S*. sp. MK730-62F2 spanning 42,3 kbp and comprising designated genes *cpz1-cpz34* (Kaysser *et al*., 2009). Following studies on the cluster boundaries revealed, that genes *cpz4-cpz31* (33,7 kbp) are sufficient for the production of sulfated caprazamycin-aglycons and a minimal BGC from *cpz9-cpz31* (28,1 kbp) for unsulfated caprazamycin-aglycons (Kaysser *et al*., 2010a). Besides, the liposidomycin gene cluster comprises 25 open reading frames (orfs), designated as *lpmA-lpmY*, thereby spanning 29,7 kbp and was first discovered in *S*. sp. SN-1061M (Kaysser *et al*., 2010a). The caprazamycin and liposidomycin BGCs share an almost identical genetic organization with functionally corresponding genes being arranged in the same order and orientation. Every essential gene for biosynthesis, regulation,

resistance and transport has a corresponding counterpart in the respective other gene cluster (Table S 1). Moreover, intergenic regions larger than 100 bp are located at corresponding positions within both clusters and show moderate to high sequence identity on DNA level (Figure 3).

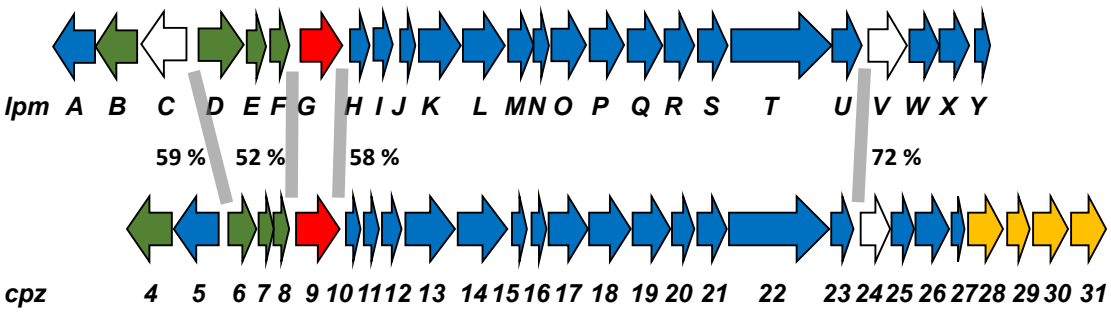


Figure 3: Comparison of the genetic arrangement of the caprazamycin (*cpz*) and liposidomycin (*lpm*) BGC. Corresponding intergenic regions of both clusters are marked with grey bars and their DNA sequence identity is indicated in percentages. Color code: biosynthesis and transfer of the sulphate group (green), unknown function (white), cluster-situated regulators (red), biosynthesis, resistance and transport (blue), transfer and methylation of the rhamnose moiety (yellow).

Due to the high structural and genetic similarities of caprazamycins, liposidomycins, A-90289s and muraminomycins (see 2.1.2.1), they share widely common biosynthetic pathways. Therefore, some biosynthetic steps have been functionally assigned *in vitro* with equivalent enzymes of the related pathways, which allowed deduction to caprazamycin and liposidomycin biosynthesis. The biosynthetic pathway of both compounds starts with the synthesis of the disaccharide core, composed of glycyluridine and 5-amino-5-deoxyribose, from uridine-mono-phosphate (UMP) via the sequential activity of dioxygenase Cpz15, aldolase Cpz14, glycosyltransferase Cpz17 and aminotransferase Cpz13 (Chi *et al.*, 2011; Barnard-Britson *et al.*, 2012; Cui *et al.*, 2020; Zadeh *et al.*, 2022). In a branching pathway, uridine 5´-aldehyde, the product of Cpz15 (Yang *et al.*, 2011), serves as the substrate for aminotransferase Cpz18, catalyzing the transamination to 5´-amino-5´-deoxyuridine. This intermediate is further converted by a pyrimidine phosphorylase (Cpz19) and a nucleotidyltransferase (Cpz16) to UDP-5-amino-5-deoxyribose (Chi *et al.*, 2011). This product putatively serves as the new substrate for Cpz17, merging both pathway branches. The resulting, aforementioned disaccharide core is the shared feature of muraymycins and liponucleosides and for the latter, the central intermediate in (+)-caprazol synthesis. Although no experimental data are present about the cyclization and *N*-methylation of the diazepanone ring yet, it is speculated, that after β-

hydroxylation by Cpz10, methylation is carried out by the methyltransferases Cpz11 and Cpz26 (Wiker *et al*., 2019a). Subsequently, Cpz27 phosphorylates the intermediate (Zadeh *et al*., 2022) and ring closure is carried out in an unknown mechanism. The resulting (+)-caprazol is finally equipped with a fatty acid moiety by the lipase Cpz23 (Zadeh *et al*., 2022), yielding hydroxyacylcaprazols. Subsequently, a 3-methylglutaryl (3-MG) moiety is added by the carboxyesterase Cpz21 (Kaysser *et al*., 2009), which results in caprazamycin-aglycon formation. Recently, it was shown that the hydroxy-methylglutaryl-CoA synthase Cpz5, the acyl-CoA synthase Cpz20 and the acyl dehydrogenase Cpz25 are involved in a cluster-encoded pathway for the production of 3-methylglutaryl-CoA, the precursor of 3-MG (Bär *et al*., 2022). In dependence on the length of the fatty acid chain incorporated by Cpz23/LpmU, caprazamycins A-G and 11 different liposidomycins can be distinguished. Attachment and methylation of a L-rhamnose moiety in native caprazamycins (Figure 1) is facilitated by Cpz28-Cpz31, resulting in the final natural product (Kaysser *et al*., 2010b).

As indicated before, Cpz4, Cpz6, Cpz7 and Cpz8 are involved in the production of sulfated caprazamycins (Tang *et al*., 2013). The type III polyketide synthase Cpz6 facilitates synthesis of a triketide pyrone, which is then sulfated by the phosphoadenosine-phosphosulfat- (PAPS-) dependent sulfotransferase Cpz8. The resulting sulfate donor molecule is utilized by Cpz4, which finally transfers the sulfate group to the hydroxyacylcaprazols (Wiker *et al*., 2019b). All these enzymes have functional equivalents in the liposidomycin BGC (LpmB, LpmD, LpmE and LpmF; see Table S 1), in which the degree of sulfation is much higher.

Further gene products of the caprazamycin cluster are the ABC transporter Cpz22, the phosphotransferase Cpz12, both putatively involved in self-resistance, hypothetical protein Cpz24 and the AraC-type regulator Cpz9 (see 2.4).

2.3.1 Heterologous expression of the biosynthetic gene clusters (BGCs)

Heterologous expression of putative natural product BGCs is a routinely applied method to enhance production yields, link a compound to its genetic origin or to facilitate molecular manipulations thereof. For their identification, the caprazamycin and liposidomycin BGC were heterologously expressed in the well-studied model strain *Streptomyces coelicolor* M512 (Kaysser *et al*., 2009; Kaysser *et al*., 2010a). Screening for compounds synthesized by the resulting strains *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01, caprazamycin-aglyca and liposidomycins were found to be produced, respectively. Due to the fact that the genes

for the production of the L-rhamnose moiety, incorporated in native caprazamycins, are encoded elsewhere within the genome, this sugar part is lacking in heterologously produced caprazamycins (caprazamycin-aglyca). This implies, that the host does not possess the genetic capacity to supply the required biosynthetic building blocks (Kaysser *et al*., 2010b). All caprazamycin-aglyca and liposidomycins from the heterologous producer strains, that were subject of the present study, are shown in Figure 4.

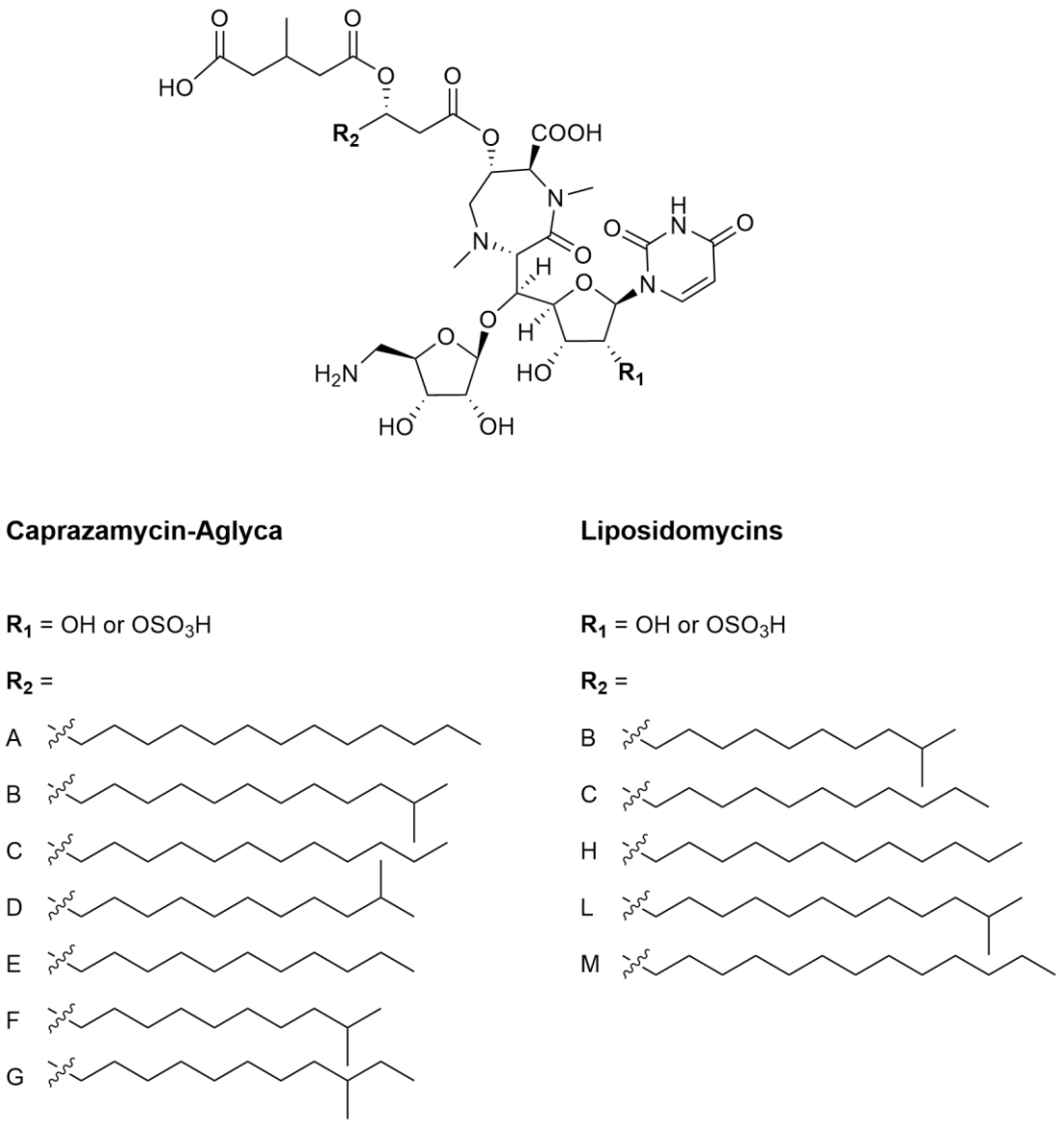


Figure 4: Common scaffold and distinguishing fatty acid chains of caprazamycin-aglyca and liposidomycins. Caprazamycin-aglyca A-G and liposidomycins B, C, H, L and M are distinguished based on their fatty acid chains. Only selected compound representatives, encompassing a saturated fatty acid side chain, are shown, as these were subject of the present study. However, in the native producer strains, unsaturated fatty acids were also found to be incorporated (Siebenberg *et al*., 2011).

The structural difference between natively and heterologously synthesized caprazamycins emphasizes the strong reliability on the hosts primary metabolism and how much both processes are intertwined. The complex chemical structures of caprazamycins and liposidomycins require precursor supply from various metabolic pathways, such as lipid and sugar metabolism and rely on deployment of SAM (S-adenosyl-methionine), UMP and PAPS (phosphoadenosine-phosphosulfat) by the host. As explained in the previous paragraph, caprazamycins and liposidomycins share astonishing similarity in their BGCs and biosynthesis. Therefore, they are well-suited subjects for a comparative investigation on the hosts response to newly incorporated heterologous gene clusters. The major way how processes can be adapted and synchronized in cells is by gene regulation.

2.4 Regulation of the caprazamycin and liposidomycin BGC

Regulation of natural product biosynthesis, especially in *Streptomyces*, is highly complex and stongly connected with cellular physiology and morphological differentiation of the producing organism. In order to prevent unnecessary loss of energy and resources, secondary metabolism is, in most cases, regulated on many levels and by multi-layered regulatory cascades (van der Heul *et al.*, 2018). The most common form of regulation is conferred on the transcriptional level mediated by regulators, which are also called transcription factors (Browning and Busby, 2004; Cuthbertson and Nodwell, 2013). Other regulatory mechanisms may include, for instance, post-translational modifications, regulatory small RNAs, DNA methylation and supercoiling. Regulators are DNA-binding proteins, that occur in a high structural or functional diversity and alter transcription of either single genes or entire operons. Depending on their effect, regulators may be classified as activators (enhancing transcription), repressors (suppressing transcription) or proteins of dual function (Romero-Rodríguez *et al.*, 2015).

Generally speaking, regulators can be divided into three groups (Niu and Tan, 2015). Global regulators are the masters of regulatory proteins, that are involved in control of other regulators, cell physiology, life cycle, primary and secondary metabolism. Pleiotropic regulatory genes are spread all over the genome, like global regulators, and influence several genes or operons, but with less severe impact on cellular processes and development. The last class is the most exploited regulator group in natural product research. So-called cluster-situated regulators (CSRs) are encoded within the BGC itself. These proteins are usually highly specific for the affiliated gene

cluster and their activity often has a drastic impact on the production of the final metabolite.

In case of the liponucleoside antibiotics caprazamycin and liposidomycin, almost nothing is known about their gene cluster regulation in both the native producers (*S*. sp. MK730-62F2 and *S*. sp. SN-1061M, respectively) and the heterologous host *S. coelicolor* M512. By BLAST analysis, Cpz9 and LpmG were proposed as the cluster-situated putative regulators of the caprazamycin and liposidomycin BGC, respectively (Kaysser *et al*., 2009; Kaysser *et al*., 2010a). These putative AraC-type regulators share 77 % sequence similarity on amino-acid level (by BLASTp sequence alignment), but were not further characterized. AraC-type regulators are typically transcriptional activators (Martin and Rosner, 2001) and were first described in *Escherichia coli* to be involved in L-arabinose metabolism (Sheppard and Englesberg, 1966).

In order to fulfill their physiological role, transcription factors recognize and bind (mostly) specific DNA-sequence patterns, which function as transcriptional promoters for RNA-polymerase binding (Lee *et al*., 2012). In case of the caprazamycin and liposidomycin BGC, there are four intergenic regions that are large enough to serve as putative promoters, respectively (Figure 3).

Regulation of the BGCs from the closely related muraminomicins and A-90289s remains elusive to this day as well, although putative cluster-situated regulators, both likewise probably belonging to the AraC-type family, have been identified (Funabashi *et al*., 2010; Chi *et al*., 2013).

2.5 Identification of regulators by DNA-affinity capturing assay

Due to the fundamental impacts of gene regulation on all cellular processes, it requires a high degree of fine-tuning, that adapts and evolves over time. An in-depth understanding of the regulation of a (biosynthetic) pathway is highly desirable for its full exploitation, but is also challenging because of its complexity.

The identification of regulatory proteins on a certain target DNA region is often the first step required for deeper regulatory studies. This aim can be addressed by performance of DNA-affinity-capturing assays (DACAs) coupled with highly sensitive mass-spectrometry (MS) analysis. Combining DACAs with subsequent semi-quantitative nano-liquid-chromatography-mass-spectrometry (nLC-MS/MS) without a pre-separation step of the captured proteins was first described by Park *et al*. (2005), which strongly enhanced the practicability of this method. Beforehand, eluted proteins from the DACA had to be prepared by electrophoretic separation, protein bands were

excised and identified using matrix-assisted-laser-desorption/ionization-time-of-flight (MALDI-TOF; Rey *et al*., 2003; Engels *et al*., 2004).

DACAs are pull-down assays, based on the (specific) interaction between a DNA-binding transcription factor and a given DNA sequence. It allows the identification of proteins binding to an immobilized DNA sequence at a fixed time point (Figure 5).

In the past years, DACAs with subsequent analysis by mass-spectrometry were successfully applied in several studies, either to verify binding of a known regulator to a certain promoter or to identify new regulators (Qu *et al*., 2015; Kim *et al*., 2022; Liu *et al*., 2023). By the implementation of DACAs, our working group was previously able to shed light on the regulation of the heterologously expressed novobiocin BGC (Bekiesch *et al*., 2016a; Bekiesch *et al*., 2016b) and the native tacrolimus BGC in *S. tsukubaensis* (Bakker, 2021). Furthermore, this sensitive and powerful method was also central in the present study.

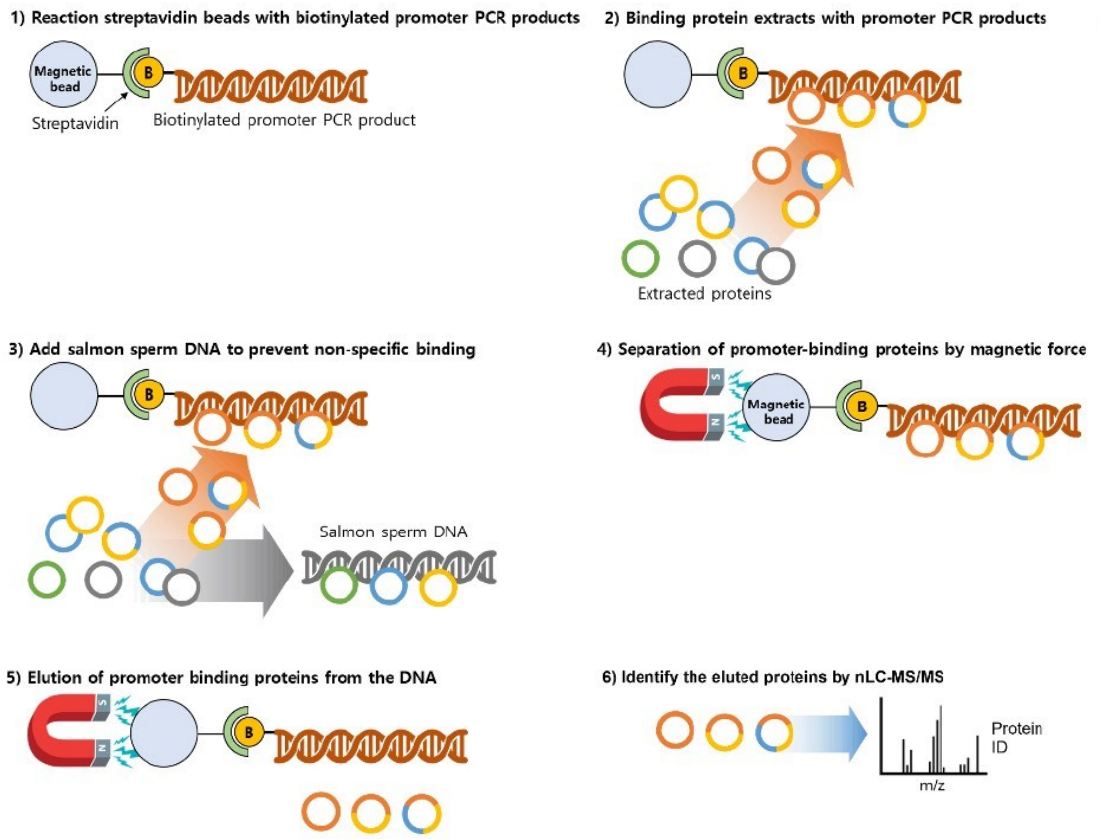


Figure 5: General workflow of a DNA-affinity-capturing assay (DACA). Amplified and biotinylated promoter regions are immobilized via streptavidin-covered magnetic beads. Addition of the whole-cell protein extracts leads to binding of proteins to the DNA. Unspecifically and only weakly binding proteins are removed by addition of unspecific salmon sperm DNA and rigorous washing of the beads, facilitated by magnetic separation. Specifically binding proteins are finally eluted using the salt-out strategy. Eluted and precipitated proteins are identified by semi-quantitative mass-spectrometry. Adapted from Kim *et al.*, 2022.

2.6 Aims of the study

The overall aim of this study was a comparative investigation of the regulation of the heterologously expressed caprazamycin and liposidomycin biosynthetic gene clusters. Therefore, the heterologous caprazamycin and liposidomycin producer strains *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01 were selected. The following objectives were addressed:

1. **RNA-sequencing of the heterologous caprazamycin and liposidomycin producer strains for determination of the operon structure and promoter sites**

In order to track and quantify transcription of the genes from the caprazamycin and liposidomycin BGC, total RNA should be isolated from the heterologous producer strains *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01 at two time points. Additionally, for an advanced whole-genome based analysis, strain *S. coelicolor* M512, without an inserted cluster, was included. In cooperation with the NGS competence center Tübingen (NCCT), the RNA should be further processed and subjected to RNA-sequencing. Resulting raw data should be pre-processed in cooperation with Theresa Harbig from the working group of Prof. Kay Nieselt (Interfaculty institute for bioinformatics and medical informatics, University of Tübingen).

2. **Identification of regulators from *S. coelicolor* M512 binding to the heterologously expressed caprazamycin and/or liposidomycin BGC**

For the identification of regulators, that bind to the caprazamycin and/or liposidomycin gene cluster during heterologous expression in *S. coelicolor* M512, DACAs should be conducted. Putative promoter regions, should be applied as bait DNA-sequences and incubated with the whole-cell protein extract of *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01. In order to isolate sufficient amounts of protein, large-scale cultivations of both heterologous producer strains in 20-L fermenters are conducted in cooperation with Andreas Kulik at the Institute of Biology, University of Tübingen. During cultivation, growth, liponucleoside production and transcription of the biosynthetic genes should be monitored to determine optimal time points for sample withdrawals for the subsequent DACA experiments. Proteins captured by DACA should be identified by nLC-MS/MS in cooperation with Dr. Mirita Franz-Wachtel (working group of Prof. Macek, proteome center Tübingen). Detected

proteins will be compared between the two heterologous producer strains and candidate regulators will be selected for further characterization.

3. Overexpression and deletion of selected, putative regulatory genes

Based on the DACA results, candidate regulatory genes should be selected for overexpression and deletion in *S. coelicolor* M512/cpzLK09 and/or *S. coelicolor* M512/lpmLK01. For overexpression, the respective genes should be cloned into a replicative plasmid harboring a strong, constitutive promoter, while for scar- and markerless gene deletion, a plasmid-based CRISPR-Cas9 approach should be exploited. The generated mutant strains should be cultivated, liponucleosides should be extracted and subjected to semi-quantitative LC-MS.

4. DNA-protein binding studies with selected regulators using biolayer interferometry

The proteins encoded by genes, that conferred a significant impact on liponucleoside production levels upon overexpression, should be subjected to *in vitro* binding assays. Thus, the regulatory proteins should be recombinantly overexpressed in *E. coli* and purified by affinity chromatography. A protocol for protein-DNA interaction measurements by biolayer interferometry first needs to be established. Subdivision of the promoter region, at which the protein was found to bind in the DACA, should allow verification of the interaction and definition of binding sites. After experimental validation of the DNA sequences as regulator binding sites, a consensus sequence can be deduced. Finally, the consensus sequence may be utilized to scan the *S. coelicolor* M512 genome for further target genes of the respective regulators.

5. Overexpression of the cluster-situated regulators in *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01

The last part of the present study comprises the overexpression of the cluster-situated regulator genes *cpz9* and *lpmG* in both *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01. As the strength of *cpz9* and *lpmG* expression in the heterologous producer strains is unknown, overexpression of the, probably activating CSRs may lead to a significant enhancement in liponucleoside antibiotic production. After their generation, overexpression mutants should be cultivated, extracted and measured by LC-MS. By comparison with the control strains, containing the empty vector, the sole impact of the regulator gene overexpression will be assessed.

3 Materials and Methods

3.1 Materials

3.1.1 Devices

Table 1: Devices used in this study.

Manufacturer	Device
Agilent Technologies, Santa Clara, CA, USA	Agilent HPLC-MS system components: • 1200 autosampler with thermostat • 1200 binary pump • 1200 column thermastat • 1200 Diodearraydetector • LC/MSD Ultra Trap System XCT 6330
Bio-Rad Laboratories Inc., Hercules, CA, USA	Electroporation apparatus Micro Pulser™ PCR thermocycler MyiQ2
Biozym Scientific GmbH, Hessisch Oldendorf, Germany	Mini-Centrifuge Color Sprout Plus
Büchi Deutschland, Essen, Germany	Rotary evaporator components: • Heating Bath B-491 • Recirculating Chiller F-108 • Rotavapor R-210 • Vacuum Controller V-855 • Vacuum Pump V-700
Eppendorf, Hamburg, Germany	SpeedVac Vacuum Concentrator plus Centrifuge 5920 R Centrifuge 5810 R Centrifuge 5424
FortéBiosciences Inc., Dallas, TX, USA	Octet K2 system
GE Healthcare Technologies Inc., Chicago, Illinois, USA	ÄKTApurifier components: • Box-900 • Fraction Collector Frac-920 • Pump P-900 • UPC-900 Monitor
Giovanola Frères SA, Monthey, Switzerland	Giovanola b20 fermenter (20 L volume)
Heidolph Instruments GmbH & Co.KG, Schwabach, Germany	Magnetic stirrer Heidolph MR Hei-Standard
KERN & SOHN GmbH, Balingen, Germany	Scale Kern 572
MP Biomedicals, Santa Ana, CA, USA	MP FastPrep-24™

Manufacturer	Device
New Brunswick Scientific, Eppendorf, Hamburg, Germany	Incubator Innova® 44 Incubator Shaker Series
Pharmacia Biotech, Uppsala, Sweden	Electrophoresis power supply EPS 300 Spectrophotometer Novaspec II
Sartorius AG, Göttingen, Germany	Incubator Certomat HK Scale Sartorius MC210 P ISO9001
Scientific Industries Inc., Bohemia, New York, USA	Vortex mixer Genie 2, VF2
SHARP, Sakai, Japan	Microwave Sharp Inventer
Sigma-Aldrich®, St. Louis, MO, USA	pH-meter HANNA Instruments pH 211
SLM Instruments Inc.	French pressure cell press Aminco FA-078
Systec GmbH, Linden, Germany	Autoclave Systec DE-23 Autoclave Systec VX-150
Tecan Trading AG, Männedorf, Switzerland	Tecan Reader Infinite F200 PRO
ThermoFisher Scientific, Waltham, MA, USA	Centrifuge Multifuge 1 S-R 75002005 G Sorvall Heraeus Clean bench Heraeus Herasafe Incubator Heraeus Function Line B12
Veolia, Paris, France	MilliQ system Purelab flex
Vilber Lourmat, Collégien, France	Gel documentation system Vilber E-Box VX5
VWR International, Radnor, PA, USA	Centrifuge MicroStar 17R NanoDrop 1000 PCR Cycler peqSTAR Thermo Shaker TS 100

3.1.2 Equipment and consumables

Table 2: Equipment and consumables used in this study.

Manufacturer	Equipment/Consumable
Becton, Dickinson and Company, Franklin Lanes, New Jersey, USA	Syringes (20 mL, 10 mL)
Bemis®, Neenah, WI, USA	Parafilm
BIOplastics BV, Landgraaf, Netherlands	Opti-Seal Optical Sealing Sheet
Bio-Rad Laboratories Inc., Hercules, CA, USA	SDS Gel Electrophoresis Chamber

Manufacturer	Equipment/Consumable
BRAND Scientific GmbH, Wertheim, Germany	Microtiterplate 96-well, transparent
Büchi Deutschland, Essen, Germany	Round-bottom and pear-shaped glass flask
Cytiva, Marlborough, MA, USA	SEC column Superdex 16/600 75 pg SEC column Superdex 200 10/300 GL
Eppendorf, Hamburg, Germany	Eppendorf® Safe-Lock microcentrifuge tubes (2 mL, 1,5 mL) Pipette tips (10000 µL, 1000 µL, 200 µL, 20 µL, 2 µL) Pipettes Discovery (10000 µL, 1000 µL, 200 µL, 20 µL, 2 µL)
FortéBiosciences Inc., Dallas, TX, USA	SA Streptavidin Dip-and-Read Biosensors
Greiner bio-one GmbH, Kremsmünster, Austria	Centrifuge tubes (50 mL, 15 mL) Culture dishes 94 mm x 16 mm
LMS Consult GmbH & Co.KG, Brigachtal, Germany	Bottletop Vacuum Filter Corning® 500 mL 0,22 µm pore size
Macherey-Nagel, Düren, Germany	pH paper
Merck KGaA, Darmstadt, Germany	Amicon® Ultra-4 (10.000 Da) and -15 Centrifugal Filter Units
MP Biomedicals, Santa Ana, CA, USA	Lysing matrix B tubes (2 mL)
Phenomenex Inc., Torrance, CA, USA	HPLC column Luna Omega Polar C18 (5 µm, 150 x 2,1 mm)
Qiagen, Hilden, Germany	Ni-NTA Agarose (Suspension)
Sarstedt AG & Co. KG, Nümbrecht, Germany	ELISA 96-well plate black PCR Tubes Multiply®-µStrip Pro 8-strip PCR Plate 96-wells Pipette filter tips (10000 µL, 1000 µL, 200 µL, 20 µL, 2 µL)
Sartorius AG, Göttingen, Germany	Sterile filters (0,2 µm, 0,45 µm)
ThermoFisher Scientific, Waltham, MA, USA	Dynabeads™ M-280 Streptavidin (Invitrogen)
VWR International, Radnor, PA, USA	Electroporation cuvettes Peqlab, 2 mm gap
WICOM, Heppenheim, Germany	LC-MS caps 11 mm aluminium crimp LC-MS CRIMPSNAP-Vial 11 mm, polypropene, 500 µL, canonical bottom LC-MS CRIMPSNAP glass vials, 11 mm

3.1.3 Chemicals

Table 3: Chemicals and media components used in this study.

Manufacturer	Chemical/Media component
AppliChem GmbH, Darmstadt, Germany	Iron(III) chloride hexahydrate ($FeCl_3 \cdot 6\ H_2O$)
Becton, Dickinson and Company, Franklin Lanes, New Jersey, USA	Agar Soluble starch Soytone Trypticase soy broth Tryptone Yeast extract
Bio-Rad Laboratories Inc., Hercules, CA, USA	Tetramethylethylendiamin (TEMED)
Carl Roth GmbH & Co.KG, Karlsruhe, Germany	Acetonitrile Acrylamide (30 %) Ammoniumpersulfate (APS) Bromophenol blue Carbenicillin Di-potassium hydrogen phosphate D-maltose monohydrate Ethylene-diamine-tetra acetic acid (EDTA) Formic acid Glucose Glycerol Isopropyl β-D-1-thiogalactopyranoside (IPTG) Kanamycin Magnesium chloride hexahydrate Mannitol MOPS Nalidixic acid Pepstatin A Phenol/Chloroform/Isoamylalcohol 25:24:1 (*v/v*) Phosphoric acid 85 % Sodium dodecyl sulphate (SDS) Trichloroacetic acid Triton X-100 X-Gal Zinc chloride ($ZnCl_2$)
Genaxxon Bioscience GmbH, Ulm, Germany	Agarose dNTPs
Macherey-Nagel, Düren, Germany	RNase-free H_2O
Merck KGaA, Darmstadt, Germany	Imidazole

Manufacturer	Chemical/Media component
Sigma-Aldrich®, St. Louis, MO, USA	β-Mercaptoethanol Acetic acid Ammonium molybdate tetrahydrate Apramycin Benzamidin Betaine Bovine serum albumin (BSA) Chloroform Chloramphenicol Coomassie brilliant blue R-250 Copper(II) chloride dehydrate ($CuCl_2 \cdot 2\ H_2O$) Dimethyl sulfoxide (DMSO) Dithiothreitol (DTT) Ethanol HPLC-grade Hydrochloric acid Manganese(II) chloride tetrahydrate ($MnCl_2 \cdot 4\ H_2O$) Methanol LC/MS grade Phenylmethylsulfonylfluorid (PMSF) Potassium acetate Sodium acetate Sodium chloride Sodium hydroxide Sodium tetraborate decahydrate ($Na_2B_4O_7 \cdot 10\ H_20$) Sucrose Trizma base TWEEN-20
SOBO Naturkost, Köln, Germany	Soy flour
Takara Bio Inc., Kusatsu, Japan	RNase-off
ThermoFisher Scientific, Waltham, MA, USA	Butan-1-ol Isopropanol Potassium hydrogen phosphate Theophylline Salmon sperm DNA (Invitrogen)

3.1.4 Enzymes and kits

Table 4: Enzymes and kits used in this study.

Manufacturer	Kit/Enzyme
AppliChem GmbH, Darmstadt, Germany	Lysozyme

Manufacturer	Kit/Enzyme
Bio-Rad Laboratories Inc., Hercules, CA, USA	iCycler External Well Solution iScript cDNA Synthesis Kit
Carl Roth GmbH & Co.KG, Karlsruhe, Germany	Proteinase K
In-house production	Taq and Pfu polymerase
Macherey-Nagel, Düren, Germany	NucleoSpinRNA clean up kit rDNase Set RNase A
New England Biolabs, Ipswich, MA, USA	*Bam*HI-HF *Eco*RI-HF *Eco*RV-HF NEBuilder® HiFi DNA Assembly Mix *Hind*III-HF *Nco*I-HF Phusion High Fidelity Polymerase Q5 High-Fidelity DNA Polymerase rSAP (shrimp alkaline phosphatase) *Sna*BI *Spe*I-HF *Stu*I T4-DNA Ligase
Qiagen, Hilden, Germany	QIAquick Gel Extraction Kit QIAquick PCR Purification Kit QuantiNova SYBR Green PCR Kit RNAprotect® Bacteria Reagent RNeasy Plus Mini Kit

3.1.5 Dyes and markers for DNA and protein gel electrophoresis

Table 5: Gel electrophoresis supplies used in this study.

Manufacturer	Dye/Marker
New England Biolabs, Ipswich, MA, USA	2-Log DNA Ladder
Sigma-Aldrich®, St. Louis, MO, USA	Thiazole orange dye
ThermoFisher Scientific, Waltham, MA, USA	DNA Gel Loading Dye GeneRuler 1 kb DNA Ladder PAGE Ruler Plus Prestained Protein Ladder
VWR International, Radnor, PA, USA	peqGREEN

3.1.6 Software

Table 6: Software used in this study.

Manufacturer	Software
Bruker Daltonik GmbH Life Sciences, Bremen, Germany	DataAnalysis, 6300 Series Ion Trap LC/MS Software, Version 6.1
Bio-Rad Laboratories Inc., Hercules, CA, USA	MyiQ2 real-time PCR detection system, Version 2.1
DNA Star, Madison, WI, USA	SeqMan Pro, Lasergene 7, Version 7.1.0
FortéBiosciences Inc., Dallas, TX, USA	Octet Data Acquisition 11.0.0.64
GraphPad Software, Boston, MA, USA	Graphpad Prism Version 8.3.0
GSL Biotech LLC, San Diego, CA, USA	SnapGene Viewer, Educational Free Version
National Institute of Health, Bethesda, Maryland, USA	ImageJ, Java 1.8.0.112
PerkinElmer, Waltham, MA, USA	ChemDraw Professional Suite Version 8
Sci Ed Software, Westminster, CO, USA	Clone Manager Professional Suite, Version 8
University of North Carolina, NC, USA	Integrated Genome Browser (IGB), Version 9.1.10

3.1.7 Bacterial strains

Table 7: Bacterial strains used in this study.

Organism	Strain	Genotype	Reference
E. coli	BL21 (DE3)	strain B, *F^{-}, ompT, gal, dcm, lon, hsdS$_B$(r$_B^-$m$_B^-$), λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]), [malB$^+$]$_{K-12}$(λ^S)*	Studier *et al*., 1990
E. coli	ET12567	Strain defective in DNA methylation *dam-13::Tn9, dcm-6, hsdM, hsdR, recF143, zjj201::Tn10, galK2, gal T22, ara14, lacY1, xyl5, leuB6, thi1, tonA31, rpsL136, hisG4, tsx78, mtli, glnV44*; Tet^R, Cml^R	MacNeil *et al*., 1992
E. coli	Top10	K-12 strain, F$^-$, *mcrA,* Δ(*mrr-hsdRMS-mcrBC*),	Invitrogen, Life Technologies,

Organism	Strain	Genotype	Reference
		ϕ80*lacZ*ΔM15, Δ*lacX74, recA1, araD139,* Δ(*araleu*)*7697, galU, galK, rpsL*, *str, endA1, nupG*	Cergy-Pontoise
E. coli	XL1 blue	*recA1, endA1,gyrA96 (nal^{R}), thi-1, hsdR17(rk^{-} mk^{+}), supE44, relA1, lac [F´, proAB+, lacIqZ_M15, Tn10, (*Tet^{R}*)]*	Stratagene Inc., La Jolla, CA, USA
S. coelicolor	M512	SCP1^{-}, SCP2^{-}, *ΔactII-ORF4*, *ΔredD*	Floriano and Bibb, 1996
S. coelicolor	M512/cpzLK09	*S. coelicolor* M512 strain containing the caprazamycin gene cluster; Kan^{R}	Kaysser *et al*., 2009
S. coelicolor	M512/lpmLK01	*S. coelicolor* M512 strain containing the liposidomycin gene cluster; Kan^{R}	Kaysser *et al*., 2010a
S. coelicolor	M1154/cpzLK09	*S. coelicolor* M1154 strain containing the caprazamycin gene cluster; Kan^{R}	Bär *et al*., 2022
S. coelicolor	M1154/cpzLK09 Δ*cpz5*	*S. coelicolor* M1154 strain containing the caprazamycin gene cluster with deletion of *cpz5*; Kan^{R}	Bär *et al.*, 2022
S. coelicolor	M1154/cpzLK09 Δ*liu*	*S. coelicolor* M1154 strain containing the caprazamycin gene cluster and the deletion of the complete *liu*-pathway (*sco2776-sco2779*); Kan^{R}	Bär *et al.*, 2022
S. coelicolor	M1154/cpzLK09 Δ*cpz5* Δ*liu*	*S. coelicolor* M1154 strain containing the caprazamycin gene cluster with deletion of *cpz5* and deletion of the complete *liu*-pathway (*sco2776-sco2779*); Kan^{R}	Bär *et al.*, 2022
S. coelicolor	M512/cpzLK09 Δ*sco3571*	*S. coelicolor* M512/cpzLK09 with marker- and scarless deletion of *sco3571*	This study

Organism	Strain	Genotype	Reference
S. coelicolor	M512/cpzLK09 Δ*sco4385*	*S. coelicolor* M512/cpzLK09 with marker- and scarless deletion of *sco4385*	This study
S. coelicolor	M512/lpmLK01 Δ*sco3571*	*S. coelicolor* M512/lpmLK01 with marker- and scarless deletion of *sco3571*	This study
S. coelicolor	M512/lpmLK01 Δ*sco4385*	*S. coelicolor* M512/lpmLK01 with marker- and scarless deletion of *sco4385*	This study

3.1.8 Plasmids

Table 8: Plasmids used in this study.

Plasmid name	Resistance and characteristics	Reference
pBlueScript II SK (+)	$Carb^R$ / Amp^R Cloning vector	Stratagene Inc., La Jolla, CA, USA
pCRISPR-TT	$Apra^R$ Carrying the CRISPR-Cas9 system; Cas9 under control of a theophylline dependent riboswitch; temperature sensitive replication	Prof. Dr. Marta Mendes, IBMC, University of Porto, Portugal
pET28a	Kan^R Overexpression vector for proteins carrying a C- or N-terminal His-tag	Sigma-Aldrich®, St. Louis, MO, USA
pHis8	Kan^R Overexpression vector for proteins carrying a N-terminal 8x His-tag	Jez *et al*., 2000
pMVM3100	$Apra^R$ Integrative vector for φBT1 *attB* integration site	Prof. Dr. Marta Mendes, IBMC, University of Porto, Portugal
pUWL-apra-oriT	$Apra^R$ Carrying *ermE** promoter for gene overexpression; replicative vector	Erb *et al*., 2009
pUZ8002	Kan^R Carrying *tra* genes; for biparental conjugation	Paget *et al*., 1999
pSW4	$Apra^R$	This study

Plasmid name	Resistance and characteristics	Reference
	pUWL-apra-oriT with *sco2987* (MarR) cloned into *Hind*III and *Spe*I sites	
pSW5	Apra^R pUWL-apra-oriT with *sco3571* (CRP) cloned into *Hind*III and *Spe*I sites	This study
pSW6	Apra^R pUWL-apra-oriT with *sco4385* (TetR) cloned into *Hind*III and *Spe*I sites	This study
pSW7	Apra^R pUWL-apra-oriT with *sco5956* (TetR) cloned into *Hind*III and *Spe*I sites	This study
pSW8	Apra^R pUWL-apra-oriT with *cpz9* cloned into *Hind*III and *Spe*I sites	This study
pSW9	Apra^R pUWL-apra-oriT with *lpmG* cloned into *Hind*III and *Spe*I sites	This study
pSW10	Apra^R pUWL-apra-oriT with *sco7543* (ROK/CRP) cloned into *Hind*III and *Spe*I sites	This study
pSW11	Apra^R pCRISPR-TT with spacer 1 for *sco4385* in *Nco*I and *Sna*BI sites	This study
pSW12	Apra^R pCRISPR-TT with spacer 2 for *sco4385* in *Nco*I and *Sna*BI sites	This study
pSW13	Carb^R pBlueScript II SK (+) with homology domains for *sco4385* in *Eco*RV site	This study
pSW14	Carb^R pBlueScript II SK (+) with homology domains for *sco3571* in *Eco*RV site	This study
pSW15	Apra^R pCRISPR-TT with spacer 1 for *sco3571* in *Nco*I and *Sna*BI sites	This study
pSW16	Apra^R pCRISPR-TT with spacer 2 for *sco3571* in *Nco*I and *Sna*BI sites	This study
pSW17	Apra^R pCRISPR-TT with spacer 1 for *sco4385* in *Nco*I and *Sna*BI sites and homology domains in *Stu*I site	This study
pSW18	Apra^R	This study

Plasmid name	Resistance and characteristics	Reference
	pCRISPR-TT with spacer 2 for *sco4385* in *Nco*I and *Sna*BI sites and homology domains in *Stu*I site	
pSW19	Apra^R pCRISPR-TT with spacer 1 for *sco3571* in *Nco*I and *Sna*BI sites and homology domains in *Stu*I site	This study
pSW20	Apra^R pCRISPR-TT with spacer 2 for *sco3571* in *Nco*I and *Sna*BI sites and homology domains in *Stu*I site	This study
pSW21	Kan^R pHis8 with *sco4385* (TetR) cloned into *Eco*RI and *Hind*III sites	This study
pSW22	Kan^R pHis8 with *lpmG* cloned into *Eco*RI and *Hind*III sites	This study
pSW23	Kan^R pET28a with *lpmG* cloned into *Eco*RI and *Hind*III sites (N- and C-terminal his-tag)	This study
pSW24	Apra^R pMVM3100 with *sco3571* under *ermE** promoter control (pSW5 as template) cloned into *Spe*I and *Bam*HI sites	Master thesis Felix Emmelmann

3.1.9 Oligonucleotides (primers)

Table 9: Oligonucleotides used in this study.

Name	Sequence (5´ → 3´)	Application
lpmDp_fwd	CGATGTCGCTGTTCGCCTG	544 bp promoter fragment upstream of *lpmD* incl. 23 bp DAC biotin linker
lpmDp_rev	GAGGAGTCGTCGATGTGGAGACCCCTGTGTGGTGACCTTGTGACC	
lpmGp_fwd_neu	GGATGGGAGCCGTTGAGCACG	520 bp promoter fragment upstream of *lpmG* incl. 23 bp DAC biotin linker
lpmGp_rev_neu	GAGGAGTCGTCGATGTGGAGACCCTTCACCGATGCGCCGCACC	
lpmHp_fwd	CCTGCGCAACGGTTTTGCC	

Name	Sequence (5´ → 3´)	Application
lpmHp_rev	GAGGAGTCGTCGATGTGGAGACCGTGCTGGAGGGCGAGTTGCC	369 bp promoter fragment upstream of *lpmH* incl. 23 bp DAC biotin linker
lpmVp_fwd	CCGACGGCATCCATCTGACG	455 bp promoter fragment upstream of *lpmV* incl. 23 bp DAC biotin linker
lpmVp_rev	GAGGAGTCGTCGATGTGGAGACCGACTCGTACCACCGCAACACG	
cpz6p_fwd	GAGGAGTCGTCGATGTGGAGACCCAGGAAGAACTCGGTTGTGC	424 bp promoter fragment upstream of *cpz6* incl. 23 bp DAC biotin linker
cpz6p_rev	GCTGTTGTGACCTTGTTGC	
cpz9p_fwd	GAGGAGTCGTCGATGTGGAGACCGCTGGGAGCCATTGTGCGC	358 bp promoter fragment upstream of *cpz9* incl. 23 bp DAC biotin linker
cpz9p_rev	CGGCCACCCTCCTGTCTTTGC	
cpz10p_fwd	GAGGAGTCGTCGATGTGGAGACCGGGAAATCGGCCTGCGTAACG	365 bp promoter fragment upstream of *cpz10* incl. 23 bp DAC biotin linker
cpz10p_rev	CGGATTCGCGGAGCTTGTCG	
cpz24p_fwd	GAGGAGTCGTCGATGTGGAGACCCCGACGGCATCCATCTGACG	431 bp promoter fragment upstream of *cpz24* incl. 23 bp DAC biotin linker
cpz24p_rev	CACGTCCCGCTGTCTGTCC	
hrdBp_fwd	GTCAACTTCTGACCGTCCAC	559 bp promoter fragment upstream of *hrdB* incl. 23 bp DAC biotin linker
hrdBp_rev	GAGGAGTCGTCGATGTGGAGACCAATGAGCGCCATGACAGAG	
DAC Biotin	Biotin-GAGGAGTCGTCGATGTGGAGACC	Linker sequence for attachment of the biotin-tag to amplified promoter fragments
qPCRcpz9_fwd	GATCCGAGGATTCATCGACC	qPCR primers for detection of *cpz9* transcripts (105 bp)
qPCRcpz9_rev	TTGAACAGCCGTTGCAGAC	
qPCRcpz10_fwd	CTTCCGTTCAAGCCCATTC	qPCR primers for detection of *cpz10* transcripts (156 bp)
qPCRcpz10_rev	AATGTGCAGCACCTTGTC	

Name	Sequence (5´ → 3´)	Application
qPCRlpmG_fwd	TGAAGTCAGACATCAAGACCC	qPCR primers for detection of *lpmG* transcripts (102 bp)
qPCRlpmG_rev	ACCGCACTGAACTTGATCC	
qPCRlpmH_fwd	AGCCTGTTGGAAAAGCATC	qPCR primers for detection of *lpmH* transcripts (121 bp)
qPCRlpmH_rev	CCGAGGAAATCAACGACTG	
qPCRhrdB_fwd	TGACGCTGATGGTCAGTGC	qPCR primers for detection of *hrdB* transcripts (124 bp)
qPCRhrdB_rev	GTCGCCTTCCTGCTGGTC	
hrdB_test_fwd	TGACGCTGATGGTCAGTGC	Testprimers for tracing gDNA contaminations after RNA isolation (1066 bp)
hrdB_test_rev	ATGAGGTCACCGAACTCGC	
sco7543_HindIII_fwd	aaaAAGCTTTTGACCGGCATGCAGATGCC	Amplification of *sco7543* and attachment of *Hind*III and *Spe*I restriction site for cloning (1283 bp)
sco7543_SpeI_rev	aaaACTAGTTCAAGGACGGACATGGGCGG	
sco4385_HindIII_fwd	aaaAAGCTTGTGAACGCGGCCGACCGCACG	Amplification of *sco4385* and attachment of *Hind*III and *Spe*I restriction site for cloning (621 bp)
sco4385_SpeI_rev	aaaACTAGTTCACTGTGCCAGCGCGGCGTCC	
sco5956_HindIII_fwd	aaaAAGCTTGTGACGGCACCTGCCACGGCC	Amplification of *sco5956* and attachment of *Hind*III and *Spe*I restriction site for cloning (825 bp)
sco5956_SpeI_rev	aaaACTAGTCTACTCGGCCGGGACGGACTTCG	
sco2987_HindIII_fwd	aaaAAGCTTATGACCACGCCCGCTACCGAG	Amplification of *sco2987* and attachment of *Hind*III and *Spe*I restriction site for cloning (519 bp)
sco2987_SpeI_rev	aaaACTAGTTCAGGACGGTTCGGGTGCCG	
sco3571_HindIII_fwd	aaaAAGCTTGTGGACGACGTTCTGCGGCG	Amplification of *sco3571* and attachment of *Hind*III and *Spe*I restriction site for cloning (687 bp)
sco3571_SpeI_rev	aaaACTAGTTCAGCGGGAGCGCTTGGCC	
cpz9_HindIII_fwd	aaaAAGCTTGTGATCTTCCAGGCGTCACCGAC	Amplification of *cpz9* and attachment of *Hind*III and *Spe*I restriction site for cloning (1059 bp)
cpz9_SpeI_rev	aaaACTAGTCTATTGACTGATCGCGCCCCACC	
lpmG_HindIII_fwd	aaaAAGCTTGTGTGGAAACGGGGGTTTGG	Amplification of *lpmG* and attachment of *Hind*III and *Spe*I restriction site for cloning (1023 bp)
lpmG_SpeI_rev	aaaACTAGTTTAGCCGACCGCGTCCCG	

Name	Sequence (5´ → 3´)	Application
pUWL_test_fwd	ACGCCTGGTCGATGTCGGAC	Test- and sequencing primers for cloning into the MCS of pUWL-apra-oriT vector
pUWL_new_rev	GAGCGAGGAAGCGGAAGAGC	
pCRISPR-TT_sco3571_fwd1	CATGCCATGG**GACCACGTAGAGGCGGTCTC**GTTTTAGAGCTAGAAATAGC*	Amplification of the sgRNA with insertion of spacer 1 for *sco3571* (123 bp)
pCRISPR-TT_sco3571_fwd2	CATGCCATGG**TCGGGGCGGACGTTCAGCCA**GTTTTAGAGCTAGAAATAGC*	Amplification of the sgRNA with insertion of spacer 2 for *sco3571* (123 bp)
pCRISPR-TT_sco4385_fwd1	CATGCCATGG**CAACGCGATGGAGCGTTCCT**GTTTTAGAGCTAGAAATAGC*	Amplification of the sgRNA with insertion of spacer 1 for *sco4385* (123 bp)
pCRISPR-TT_sco4385_fwd2	CATGCCATGG**CAGGCCCTGGACGGTTTCCC**GTTTTAGAGCTAGAAATAGC*	Amplification of the sgRNA with insertion of spacer 2 for *sco4385* (123 bp)
pCRISPR-TT_rev	ACGCCTACGTAAAAAAAGCACCGACTCGGTGCC	Universal reverse primer for amplification of the sgRNAs
pCRISPR_test_fwd	CGAATTGTACGCGGTCGATC	Test- and sequencing primers for cloning the sgRNAs into the pCRISPR-TT backbone
pCRISPR_test_rev	GACCCCCCATTCAAGAACAG	
sco3571_HD_upstream_fwd	CCCCCGGGCTGCAGGAATTCGATATCGTCGGGTACTTGGCGAAGAGG	Amplification of upstream homology domain of *sco3571* (942 bp) and assembly in pBlueScript SK II
sco3571_HD_upstream_rev	CCCGGGGAGACCCCAGGGGGGAGTTCTCTCCTTGTCGACCGG	
sco3571_HD_downstream_fwd	GGTCGACAAGGAGAGAACTCCCCCCTGGGGTCTCCCCGG	Amplification of downstream homology domain of *sco3571* (935 bp) and assembly in pBlueScript SK II
sco3571_HD_downstream_rev	TCGACGGTATCGATAAGCTTGATATCCGGCGTCCTGTCGGGACCG	
sco4385_HD_upstream_fwd	CCCCCGGGCTGCAGGAATTCGATATCCTTCCAGGTGGCCCCCGCCGAA	Amplification of upstream homology domain of *sco4385* (1074 bp) and assembly in pBlueScript SK II
sco4385_HD_upstream_rev	GCCCCGGTGCCGGACGCCGGGGCGCCGCTCACACCACCCATT	

Name	Sequence (5´ → 3´)	Application
sco4385_HD_ down_fwd_neu	CCGGCGTCCGGCACCGG G	Amplification of downstream homology domain of *sco4385* (1079 bp) and assembly in pBlueScript SK II
sco4385_HD_ down_rev_neu	TCGACGGTATCGATAAGCT TGATATCGCGCGCGAGCT GGACGCG	
pSET152_test_ fwd	ACGCCAGGGTTTTCCCAG TCAC	Test- and sequencing primers for cloning into the MCS of pBlueScript SK II vector
pSET152_test_ rev	AGCTGGCACGACAGGTTT CCC	
pCRISPR-TT_test_Stu_fwd	GATCCACCAGAGCATCAC CG	Test- and sequencing primers for cloning the homology domains into *Stu*I restriction site of pCRISPR-TT-spacer
pCRISPR-TT_test_Stu_rev	GTCGACGCGCTGTTCTCT CG	
sco4385_KO_ test_new_fwd	ATCGACGAGGAGGGCTGG CTG	Test- and sequencing primers for validation of the deletion of *sco4385*
sco4385_KO_ test_new_rev	CATGGTGGTCTCCACGGC GGA	
sco3571_KO_ test_fwd	CCGGAAGAAGCCGGTCGG AC	Test- and sequencing primers for validation of the deletion of *sco3571*
sco3571_KO_ test_rev	CAAGCACGTCCCGAGCAC GG	
sco4385_EcoRI_ fwd	aaaGAATTCGTGAACGCGG CCGACCGCACG	Amplification of *sco4385* and attachment of *Eco*RI and *Hind*III restriction site for cloning (621 bp)
sco4385_HindIII_r ev	aaaAAGCTTTCACTGTGCC AGCGCGGCGTCC	
lpmG_EcoRI_fwd	aaaGAATTCGTGTGGAAAC GGGGGTTTGG	Amplification of *lpmG* and attachment of *Eco*RI and *Hind*III restriction site for cloning (1023 bp)
lpmG_HindIII_rev	aaaAAGCTTTTAGCCGACC GCGTCCCG	
lpmG_HindIII_ Ctag_rev	aaaAAGCTTGCCGACCGCG TCCCGCCG	Alternative reverse primer for cloning of LpmG with a C-terminal his- tag
T7_fwd	TAATACGACTCACTATAGG G	Standard sequencing primers
T7term_rev	CTAGTTATTGCTCAGCGGT	
pMVM_test_fwd	GCCTTGAAATCGTTAGTTA G	Test- and sequencing primers for cloning and integration of pMVM3100
pMVM_test_rev	CGCAATTAATGTGAGTTAG C	

Restriction sites are underlined

* Sequences to be inserted are bold

Table 10: Oligonucleotides used for BLI measurements.

Name	Sequence (5´ → 3´)
ReDCaT linker	Biotin-GCAGGAGGACGTAGGGTAGG
PhrdB_NC_fwd	CTGTGCATCTCCCCGGCCCGCCCGCACCGTCGGCCCATTCCCAAGCCGGT
PhrdB_NC_rev	ACCGGCTTGGGAATGGGCCGACGGTGCGGGCGGGCCGGGGAGATGCACAGcctaccctacgtcctcctgc
Pcpz10_1_fwd	GGGAAATCGGCCTGCGTAACGGTTTCGCCCAGGCGTCCGACTTCAGCCGG
Pcpz10_1_rev	CCGGCTGAAGTCGGACGCCTGGGCGAAACCGTTACGCAGGCCGATTTCCCcctaccctacgtcctcctgc
Pcpz10_2_fwd	CGCCCAGGCGTCCGACTTCAGCCGGGTGTTTCGTGGCCGCTACGGTGTCC
Pcpz10_2_rev	GGACACCGTAGCGGCCACGAAACACCCGGCTGAAGTCGGACGCCTGGGCGcctaccctacgtcctcctgc
Pcpz10_3_fwd	GTGTTTCGTGGCCGCTACGGTGTCCCGCCGGGTAAATTCCGCGACGACTG
Pcpz10_3_rev	CAGTCGTCGCGGAATTTACCCGGCGGGACACCGTAGCGGCCACGAAACACcctaccctacgtcctcctgc
Pcpz10_4_fwd	CGCCGGGTAAATTCCGCGACGACTGGTTCCGGTGGGGCGCGATCAGTCAA
Pcpz10_4_rev	TTGACTGATCGCGCCCCACCGGAACCAGTCGTCGCGGAATTTACCCGGCGcctaccctacgtcctcctgc
Pcpz10_5_fwd	GTTCCGGTGGGGCGCGATCAGTCAATAGCCGGTCGCATGAGGACAATGAC
Pcpz10_5_rev	GTCATTGTCCTCATGCGACCGGCTATTGACTGATCGCGCCCCACCGGAACcctaccctacgtcctcctgc
Pcpz10_6_fwd	TAGCCGGTCGCATGAGGACAATGACAGAGCCCGAAGTGACCGGGAAAATA
Pcpz10_6_rev	TATTTTCCCGGTCACTTCGGGCTCTGTCATTGTCCTCATGCGACCGGCTAcctaccctacgtcctcctgc
Pcpz10_7_fwd	AGAGCCCGAAGTGACCGGGAAAATAGTGTGAGACAGGGAGCGTTCGACCT
Pcpz10_7_rev	AGGTCGAACGCTCCCTGTCTCACACTATTTTCCCGGTCACTTCGGGCTCTcctaccctacgtcctcctgc
Pcpz10_8_fwd	GTGTGAGACAGGGAGCGTTCGACCTCGCTCGCCCTTGAGTCCCCAGGTGG
Pcpz10_8_rev	CCACCTGGGGACTCAAGGGCGAGCGAGGTCGAACGCTCCCTGTCTCACACcctaccctacgtcctcctgc
Pcpz10_9_fwd	CGCTCGCCCTTGAGTCCCCAGGTGGCTCGTCCATTCCTGCCAGGACATCA
Pcpz10_9_rev	TGATGTCCTGGCAGGAATGGACGAGCCACCTGGGGACTCAAGGGCGAGCGcctaccctacgtcctcctgc
Pcpz10_10_fwd	CTCGTCCATTCCTGCCAGGACATCAACGGCGGAGCAGAGAAGGCGCGCCG

Name	Sequence (5´ → 3´)
Pcpz10_10_rev	CGGCGCGCCTTCTCTGCTCCGCCGTTGATGTCCTGGCAGGAATGGACGAGcctaccctacgtcctcctgc
Pcpz10_11_fwd	ACGGCGGAGCAGAGAAGGCGCGCCGCTGTGCCCGGCCGTCGAAAGGCTTG
Pcpz10_11_rev	CAAGCCTTTCGACGGCCGGGCACAGCGGCGCGCCTTCTCTGCTCCGCCGTcctaccctacgtcctcctgc
Pcpz10_12_fwd	CTGTGCCCGGCCGTCGAAAGGCTTGCTTCGTGACAGCACTCACGTCCAGG
Pcpz10_12_rev	CCTGGACGTGAGTGCTGTCACGAAGCAAGCCTTTCGACGGCCGGGCACAGcctaccctacgtcctcctgc
Pcpz10_13_fwd	CTTCGTGACAGCACTCACGTCCAGGACCGAACTCGACATCGACCCCGACA
Pcpz10_13_rev	TGTCGGGGTCGATGTCGAGTTCGGTCCTGGACGTGAGTGCTGTCACGAAGcctaccctacgtcctcctgc
Pcpz10_14_fwd	ACCGAACTCGACATCGACCCCGACAAGCTCCGCGAATCCGTCGTGGAGTT
Pcpz10_14_rev	AACTCCACGACGGATTCGCGGAGCTTGTCGGGGTCGATGTCGAGTTCGGTcctaccctacgtcctcctgc
PlpmH_1_fwd	CCTGCGCAACGGTTTTGCCCAGCCCTCCGACTTCAGTCGGGTATTCCGTG
PlpmH_1_rev	CACGGAATACCCGACTGAAGTCGGAGGGCTGGGCAAAACCGTTGCGCAGGcctaccctacgtcctcctgc
PlpmH_2_fwd	TCCGACTTCAGTCGGGTATTCCGTGCCAATTACGGCATACCGCCGGGCAA
PlpmH_2_rev	TTGCCCGGCGGTATGCCGTAATTGGCACGGAATACCCGACTGAAGTCGGAcctaccctacgtcctcctgc
PlpmH_3_fwd	CCAATTACGGCATACCGCCGGGCAAGTTTCGGGACGACTGGTTCCGGCGG
PlpmH_3_rev	CCGCCGGAACCAGTCGTCCCGAAACTTGCCCGGCGGTATGCCGTAATTGGcctaccctacgtcctcctgc
PlpmH_4_fwd	GTTTCGGGACGACTGGTTCCGGCGGGACGCGGTCGGCTAACCAGGTGGTC
PlpmH_4_rev	GACCACCTGGTTAGCCGACCGCGTCCCGCCGGAACCAGTCGTCCCGAAACcctaccctacgtcctcctgc
PlpmH_5_fwd	GACGCGGTCGGCTAACCAGGTGGTCGCTTCACGGCAATGACACATCCTGC
PlpmH_5_rev	GCAGGATGTGTCATTGCCGTGAAGCGACCACCTGGTTAGCCGACCGCGTCcctaccctacgtcctcctgc
PlpmH_6_fwd	GCTTCACGGCAATGACACATCCTGCGGTGGCCGGAAGAATGAAGTTCAGG
PlpmH_6_rev	CCTGAACTTCATTCTTCCGGCCACCGCAGGATGTGTCATTGCCGTGAAGCcctaccctacgtcctcctgc
PlpmH_7_fwd	GGTGGCCGGAAGAATGAAGTTCAGGCAGGAAAGCATGAACCTCGTTCGCC

Name	**Sequence (5´ → 3´)**
PlpmH_7_rev	GGCGAACGAGGTTCATGCTTTCCTGCCTGAACTTCATTC TTCCGGCCACCcctaccctacgtcctcctgc
PlpmH_8_fwd	CAGGAAAGCATGAACCTCGTTCGCCGTTGCCGGACTTG GCTGATCCGTTC
PlpmH_8_rev	GAACGGATCAGCCAAGTCCGGCAACGGCGAACGAGGTT CATGCTTTCCTGcctaccctacgtcctcctgc
PlpmH_9_fwd	GTTGCCGGACTTGGCTGATCCGTTCCGCAGGGGGAATT ACTCCGTGACAG
PlpmH_9_rev	CTGTCACGGAGTAATTCCCCCTGCGGAACGGATCAGCC AAGTCCGGCAACcctaccctacgtcctcctgc
PlpmH_10_fwd	CGCAGGGGGAATTACTCCGTGACAGTACTGACGTCCAG GACCGTGCTCGA
PlpmH_10_rev	TCGAGCACGGTCCTGGACGTCAGTACTGTCACGGAGTA ATTCCCCCTGCGcctaccctacgtcctcctgc
PlpmH_11_fwd	TACTGACGTCCAGGACCGTGCTCGACATCGACCCGGTC AGGCTCCGCGAA
PlpmH_11_rev	TTCGCGGAGCCTGACCGGGTCGATGTCGAGCACGGTCC TGGACGTCAGTAcctaccctacgtcctcctgc
PlpmH_12_fwd	CATCGACCCGGTCAGGCTCCGCGAATCCGTGGCAAGCC TGTTGGAAAAGC
PlpmH_12_rev	GCTTTTCCAACAGGCTTGCCACGGATTCGCGGAGCCTG ACCGGGTCGATGcctaccctacgtcctcctgc
PlpmH_13_fwd	TCCGTGGCAAGCCTGTTGGAAAAGCATCCGTTGGTATTC GAGGGCACACG
PlpmH_13_rev	CGTGTGCCCTCGAATACCAACGGATGCTTTTCCAACAGG CTTGCCACGGAcctaccctacgtcctcctgc
PlpmH_14_fwd	ATCCGTTGGTATTCGAGGGCACACGGCAACTCGCCCTC CAGCACCGGTCG
PlpmH_14_rev	CGACCGGTGCTGGAGGGCGAGTTGCCGTGTGCCCTCGA ATACCAACGGATcctaccctacgtcctcctgc
Pcpz10_1_ Konsensus_fwd	ACGCCTGGGCGAAACCGTTACGC
Pcpz10_1_ Konsensus_rev	GCGTAACGGTTTCGCCCAGGCGTcctaccctacgtcctcctgc
PlpmH_1_ Konsensus_fwd	AGGGCTGGGCAAAACCGTTGCGC
PlpmH_1_ Konsensus_rev	GCGCAACGGTTTTGCCCAGCCCTcctaccctacgtcctcctgc
PlpmH_7/8_ Konsensus_fwd	AGGAAAGCATGAACCTCGTTCGC
PlpmH_7/8_ Konsensus_rev	GCGAACGAGGTTCATGCTTTCCTcctaccctacgtcctcctgc

3.1.10 Culture media

3.1.10.1 Media for cultivation of *E. coli*

Unless otherwise indicated media were sterilized by autoclaving at 121 °C for 20 min. Heat sensitive substances such as antibiotics were added after cooling down the media to approximately 60 °C.

LB (Luria-Bertani) medium (Sambrook and Russell, 1989)

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
ddH_2O	Ad 1 L

To obtain solid medium agar was added in a final concentration of 1,5 %.

TB (terrific broth) medium

Tryptone	12 g
Yeast extract	24 g
Glycerol	4 mL

Components were dissolved in 900 mL ddH_2O and autoclaved.

K_2HPO_4	12,54 g
KH_2PO_4	2,31 g

Components were dissolved in 100 mL ddH_2O, autoclaved and cooled down to RT. Prior to inoculation it was added to TB medium under sterile conditions.

3.1.10.2 Media for cultivation of *Streptomyces*

TSB medium (Kieser *et al*., 2000)

Trypticase soy broth	30 g
ddH_2O	Ad 1 L

To obtain solid medium (TSA), agar was added in a final concentration of 2 %.

Production medium (P-medium) (Fronko *et al*., 2001)

Soytone	10 g
Soluble starch	10 g
D-maltose	20 g
Trace elements solution	5 mL
ddH_2O	Ad 1 L

Before autoclaving, a pH of 6,7 was adjusted.

Trace elements solution

$FeCl_3 \cdot 6\ H_2O$	200 mg
$ZnCl_2$	10 mg
$CuCl_2 \cdot 2\ H_2O$	10 mg
$MnCl_2 \cdot H_2O$	10 mg
$Na_2B_4O_6 \cdot 10\ H_2O$	10 mg
$(NH_4)_6Mo_7O_{24} \cdot 4\ H_2O$	10 mg
ddH_2O	Ad 1 L

2x- YT medium (Kieser *et al*., 2000)

Tryptone	16 g
Yeast extract	10 g
NaCl	5 g
ddH_2O	Ad 100 mL

MS agar (Kieser *et al*., 2000)

Mannitol	20 g
Soy flour	20 g
Agar	20 g

Mannitol was dissolved in 1 L of tap water and aliquots of 100 mL were poured into 300 mL flasks each containing 2 g agar and 2 g soy flour. MS agar was autoclaved twice (115 °C, 2 bar, 15 min) with gentle shaking between the two runs.

3.1.11 Solutions and buffers

3.1.11.1 Antibiotics and inductors

Antibiotics (except chloramphenicol and tetracycline) were sterilized by filter sterilization (pore size 0,22 µm) and stored at -20 °C. Theophylline served as the inductor of *cas9* endonuclease expression in CRISPR-Cas9 deletion experiments and was stored at RT under light protection.

Table 11: Antibiotics and inductors used in this study.

Antibiotic/inductor	Stock concentration [mg/mL]	Working concentration [µg/mL]	Dissolved in
Apramycin	50	50	H_2O

Antibiotic/inductor	Stock concentration [mg/mL]	Working concentration [µg/mL]	Dissolved in
Carbenicillin	100	100-200	H_2O
Chloramphenicol	25	12,5-25	EtOH
Kanamycin	50	50	H_2O
Nalidixic acid	25	25	0,15 M NaOH
Tetracycline	25	12,5	70 % EtOH
Theophylline	160 mM	10 mM	0,1 M NaOH

3.1.11.2 Solutions for protein isolation and purification

Table 12: Solutions used for protein isolation and processing.

Component	Stock concentration	Working concentration	Dissolved in	Storage
Benzamidin	-	1 mM	Directly added into buffer	Prepared freshly for each use
DTT	1 M	1 mM	ddH_2O	-20°C
Pepstatin A	1 mM	1 µM	MeOH:Acetic acid (9:1)	-20°C
PMSF	100 mM	1 mM	EtOH	-20°C

3.1.11.3 Solutions for blue/white screening of *E. coli*

Table 13: Solutions for blue/white screeing of *E. coli*.

Component	Stock concentration	Working concentration	Dissolved in	Storage
IPTG	100 mM	1 mM	ddH_2O (filter sterilized)	-20°C
X-Gal	20 mg/mL	0,2 mg/mL	DMSO (autosterile)	-20°C (light protected)

3.1.11.4 Solutions for alkaline lysis

<u>Solution I</u>

50 mM Glucose

10 mM EDTA, pH 8

25 mM Tris-HCl, pH 8

100 µg/mL RNase, added freshly before usage

Solution II

0,2 M	NaOH
1 %	SDS

(prepared freshly before use)

Solution III

3 M	Potassium acetate, pH 5,2

(stored on ice before use)

3.1.11.5 Buffers for gDNA isolation from *Streptomyces*

TSE buffer

10 %	Sucrose
25 mM	EDTA
25 mM	Tris-HCl

The buffer was autoclaved for long-term storage.

3.1.11.6 Buffers for DNA-affinity capturing assay

If not otherwise stated all buffers were stored at 4 °C.

Binding and Washing (B+W) Buffer, pH 7,5

10 mM	Tris-HCl, pH 7,5
2 M	NaCl

The buffer was autoclaved for long-term storage.

TGED-Buffer, pH 7,5

20 mM	Tris-HCl, pH 7,5
1 mM	EDTA
100 mM	NaCl
0,01 %	Triton X-100
10 %	Glycerol
1 mM	DTT (add from stock solution right before use)

The buffer was adjusted to pH 7,5 and autoclaved.

Elutionbuffer

See TEGED buffer with 2 M NaCl

3.1.11.7 Buffers for agarose gel electrophoresis

50x TAE buffer, pH 7,8

2 M Tris-HCl

0,05 M Disodium EDTA

1 M Acetic acid

The buffer was autoclaved, stored at RT and diluted to its working concentration (1x) prior to use.

10x MOPS buffer

0,4 M MOPS, pH 7,0

0,1 M Sodium acetate

0,01 M EDTA

This buffer was especially used for gel electrophoresis of RNA samples; autoclaved twice with cooling down between the two runs. The buffer was stored at RT.

3.1.11.8 Buffers for protein isolation, purification and quantification

Table 14: Buffers used for protein isolation and purification.

Component	Lysis buffer	Running buffer	Elution buffer
Tris-HCl, pH 7,5	50 mM	50 mM	50 mM
NaCl	500 mM	150 mM	150 mM
Glycerol	10 % (*v/v*)	10 % (*v/v*)	10 % (*v/v*)
DTT	-	5 mM	-
ß-ME	10 mM	-	-
TWEEN-20	1 %	-	-
Imidazole	-	20 mM	100-250 mM

Buffers were autoclaved and stored at 4°C. DTT, β-ME and TWEEN-20 were added right before use.

Bradford reagent

100 mg/L Coomassie Brilliant Blue R-250

50 mL/L Ethanol 95 %

100 mL/L H_3PO_4 85 %

The components were mixed in ddH_2O and passaged through filter paper until the solution had a brown color. The solution was stored at 4 °C.

3.1.11.9 Buffers for SDS-PAGE

4x Sample buffer (50 mL)

12 mL	0,24 M Tris-HCl (pH6,8)
4 g	SDS
25 mL	Glycerol (80 %)
13 mL	β-mercaptoethanol
Spatula tip	Bromophenol blue

The buffer was stored at -20 °C.

10x Running buffer

30 g	Trizma base
144 g	Glycerol
10 g	SDS
Ad 1 L	ddH_2O

The buffer was stored at 4 °C and diluted to working concentration before use.

Coomassie staining solution

5 %	Acetic acid
10 %	Ethanol
1 %	Coomassie Brilliant blue R-250

Destaining solution

5 %	Acetic acid
10 %	Ethanol

3.1.11.10 Buffers for SEC and BLI

Running buffer for SEC

50 mM	Tris-HCl, pH 7,5
250 mM	NaCl

The buffer was sterilized using bottletop filters (0,22 µm pore size) and stored at 4°C.

Binding buffer for BLI

50 mM	Tris-HCl, pH 7,5
150 mM	NaCl

The buffer was autoclaved and kept at 4°C for long-term storage.

3.1.11.11 Further solutions

Magnesium chloride

2,5 M Magnesium chloride hexahydrate

The solution was autoclaved and stored at RT.

3.2 Methods

3.2.1 Microbiological methods

3.2.1.1 Cultivation and long-term storage of *E. coli*

Cultivation of *E. coli* was routinely carried out in liquid LB medium or on LB agar plates at 37 °C. Overnight cultures were either inoculated from plates or from glycerol stocks into test tubes containing 4 mL of medium supplemented with antibiotics, if applicable. Standard conditions for cultivation were 37 °C and 200 rpm. For long-term storage of *E. coli* strains, 4 mL overnight culture were mixed thoroughly with 1,5 mL 80 % glycerol and 1-2 mL of this suspension were transferred into an Eppendorf tube. Glycerol stocks were stored at -80 °C.

3.2.1.2 Cultivation and long-term storage of *Streptomyces*

Streptomyces were routinely maintained on MS plates at 30 °C. For cultivation, pre-cultures were inoculated from single colonies from plates or spore stocks into 50 mL of TSB medium and main cultures, composed of 70 mL of P-medium, were inoculated with 1 mL of the pre-culture. Both pre- and main cultures were cultivated in baffled 300 mL Erlenmeyer flasks containing a stainless steel spiral and incubated at 30 °C and 200 rpm. In the following protocols, these conditions are referred to as standard conditions for *Streptomyces*.

For long-term storage of *Streptomyces* strains, spore stocks were prepared. To allow sufficient sporulation, the strains were confluently streaked on MS agar plates and incubated at 30 °C for approximately one week. The plates were then flooded with 4 mL of ddH_2O and spores were scraped off using a sterile cotton bud. The suspension was pipetted into a falcon tube and vortexed thoroughly. Passing the suspension through a disposable syringe filled with sterile cotton wool leaded to a separation of spores and mycelium which could not pass the cotton wool. The spores were pelleted by centrifugation at 4 °C and 4.000 xg for 12 min and the supernatant was discarded. Finally, the resulting spore pellet was dissolved in 0,6-1,5 mL 20 % glycerol and stored at -80 °C.

3.2.1.3 Preparation of chemically competent *E. coli* cells

Chemically competent cells of *E. coli* were achieved from an overnight culture grown under standard conditions. It was used to inoculate 100 mL LB medium in a 300 mL Erlenmeyer flask. Cells were grown under the same conditions to an OD_{600} of 0,6 and were subsequently harvested by centrifugation at 4 °C and 3.000 xg for 20 min. Pelleted cells were resuspended in 7,5 mL ice-cold 0,1 M $CaCl_2$ and incubated on ice for another 30 min. Next, the cells were pelleted again by centrifugation at 4 °C and 3.000 xg for 20 min. Finally, the resulting cell pellet was gently resuspended in 2 mL ice-cold 0,1 M $CaCl_2$ containing 15 % glycerol. The suspension was splitted into 140 µL aliquots and stored at -80 °C.

3.2.1.4 Transformation of chemically competent *E. coli*

For chemical transformation of *E. coli*, aliquots of competent cells were thawed on ice for 5-10 min. Subsequently, 10-20 µL of a ligation reaction or 1-2 µL of plasmid DNA were added to the cells and the tubes were incubated on ice for 30 min. For DNA uptake, the cells were heat-shocked at 42 °C on a thermoshaker block for 90 s without agitation. The tubes were then directly moved back on ice where they were cooled down for 2 min before 500 µL of fresh LB medium were added. Subsequently, cells were allowed to grow at 37 °C with agitation (950 rpm) for 1-1,5 hours. Finally, 100 µL of the cell suspension were spread out on LB plates containing the selective antibiotic. The rest of the suspension was briefly centrifuged, most of the supernatant poured off and the remaining cells were plated on agar plates. Incubation of the plates was carried out at 37 °C over night.

3.2.1.5 Preparation of electrocompetent *E. coli*

For the preparation of electrocompetent *E. coli* cells 1 mL of a 4 mL overnight culture (containing antibiotics, if necessary) was used to inoculate 100 mL LB medium. The culture was incubated under standard conditions until an OD_{600} of 0,4-0,6 was reached. The cells were then harvested by centrifugation at 4°C, 4000 xg for 5 min and washed twice with 10 mL of ice-cold 10 % glycerol. After the last centrifugation step most of the supernatant was removed and the cells were resuspended in the residual volume of 10 % glycerol. Aliquots of 50 µL volume in 1,5 mL Eppendorf tubes were stored at -80 °C or directly used for electroporation.

3.2.1.6 Electroporation

For the transformation of *E. coli* with plasmid DNA, aliquots of electrocompetent cells were thawed on ice and 2-4 µL of DNA were added. This mixture was transferred to a pre-cooled 2 mm electroporation cuvette without generating any air bubbles within the suspension. The cells were pulsed at 200 Ω, 25 *µ*F and 2,5 kV and directly taken up in 500 µL of ice-cold LB medium. Afterwards the cells were transferred back to a 1,5 mL Eppendorf tube and allowed to regenerate at 37 °C on a thermoshaker block under constant shaking at 950 rpm for one hour. Following this incubation time cells were plated on LB agar containing the selective antibiotic and incubated at 37 °C over night.

3.2.1.7 Intergenic conjugation of DNA from *E. coli* to *Streptomyces*

In order to transfer DNA into *Streptomyces* strains, the DNA was passaged through the non-methylating *E. coli* strain ET12567. Competent *E. coli* ET12567 carrying the pUZ8002 plasmid were transformed with the DNA to be transferred (biparental conjugation). One colony was used to inoculate an overnight culture containing kanamycin, chloramphenicol and the selective antibiotic for the DNA to be transferred. 200 µL of this over night culture were transferred into 10 mL of fresh LB containing the same antibiotics and the culture was grown at 37 °C until an OD_{600} between 0,4 and 0,6 was reached. The cells were pelleted by centrifugation and washed twice with fresh LB to get rid of the antibiotics. The final cell pellet was resuspended in 1 mL of LB. In parallel to the centrifugation steps, 10 µL of the respective *Streptomyces* spores were added to 500 µL 2x- YT medium (for every conjugation, respectively). The spores were heat shocked at 50 °C for 10 min and then cooled down to room temperature for approximately 15 min to allow germination of the spores. Afterwards, 500 µL of the washed *E. coli* cells were mixed together with the heat shocked spores and spun down briefly. Most of the supernatant was poured off and the remaining cell-spore-pellet resuspended in the residual medium. This suspension was used to prepare a dilution series from 10^{-1} to 10^{-4} in a total of 100 µL of sterile water. The whole volume of each dilution was spread out on MS agar plates containing 10 mM $MgCl_2$ to promote sporulation. The plates were incubated at 30 °C for 16-20 hours. The next day, the plates were overlayed with 1 mL ddH_2O containing 0,5 mg nalidixic acid and the appropriate amount of selective antibiotic for positive exconjugants. The plates were well dried under the cleanbench and then incubation was continued at 30 °C. After the formation of single colonies, four of them were picked and confluently

streaked out on MS plates containing nalidixic acid and the selective antibiotic. These plates were incubated until the appearance of spores. Spore suspensions were made following the standard protocol (3.2.1.2). These spore stocks were used in a dilution series from 10^{-5} to 10^{-9} in 200 µL of sterile water. Again, the whole volume of each dilution step was plated out on MS plates containing nalidixic acid and the selective antibiotic. Plates were incubated under standard conditions until the appearance of single colonies. For each exconjugant, one single colony was streaked out confluently on a fresh MS agar plate and grown until sporulation. Theses plates were used to prepare the final spore suspensions for production cultures. The successful transfer or integration of the desired DNA was reviewed by colony-PCR and thus the spore stocks were verified.

3.2.1.8 Determination of cell dry weight from *Streptomyces*

For cell dry weight determination, one mL of the *Streptomyces* culture was centrifuged in a pre-weighted 1,5 mL Eppendorf tube at maximum speed for 4 minutes. The used Eppendorf tubes were pre-heated for at least 48 hous at 80 °C and then cooled down to room temperature before their empty weight was determined. After pelleting the cells, the supernatant was removed and the cells were washed with 1 mL ddH_2O to remove remaining components of the medium. After another round of centrifugation at maximum speed, the supernatant was removed again and the cells dried at 80 °C for at least 48 hours before the tubes were weighted again.

3.2.2 Methods of molecular biology

3.2.2.1 Isolation of genomic DNA from *Streptomyces*

gDNA was isolated from *Streptomyces* cultures grown in TSB medium for 2-3 days. The whole culture volume of 50 mL was harvested by centrifugation and the resulting cell pellet washed in 10 mL TSE buffer. The washed cells were then resuspended in 10 mL Solution I containing a spatula tip of lysozyme and a small amount of proteinase K and incubated at 37 °C for approximately one hour. In the next step, 1 mL of a 20 % SDS solution was added, mixed by gently inverting the tube and then incubated for another 30 min at 60 °C. After this time the cells should be lysed. One volume of phenol/chloroform/isoamyl alcohol (25:24:1) was added and mixed by inversion of the tube. The following centrifugation was carried out at 4000 xg for 15 min. In order to yield high quality DNA, the phenol/chloroform extraction was repeated. Finally, 0,1 volumes of sodium acetate (3 M) solution and 0,8 volumes of ice-cold isopropanol

were added, mixed thoroughly by vortexing and centrifuged at 4 °C at maximum speed for at least 30 min. The resulting DNA pellet was washed twice with 0,5 volumes of 70 % ethanol and air-dried for a couple of minutes. The DNA was dissolved in an appropriate volume of ddH_2O or Tris-HCl buffer (10 mM, pH 8).

3.2.2.2 Isolation of plasmid DNA from *E. coli*

Plasmid DNA was isolated from *E. coli* using the standard alkaline lysis protocol. 3 mL of an overnight culture were centrifuged at maximum speed for a couple of minutes and the supernatant was removed. Next, the pellet was resuspended in 250 µL Solution I containing RNase (100 µg/mL). After incubation at room temperature for 5 min, 250 µL Solution II was added and the tubes were inverted 10 times to mix the solutions well. After another incubation step on ice for 5 min, 250 µL Solution III was added, the tube gently inverted and then cooled on ice for 5 min. To remove the cell debris, a centrifugation step at 4 °C and 17.000 xg for 8 min was conducted and the clear supernatant transferred into a new Eppendorf tube. The supernatant was mixed with one volume of isopropanol and incubated on ice for 10 min before being centrifugated at 4 °C at 17.000 xg for 45 min. The resulting supernatant was discarded and the DNA pellet washed with 500 µL 70 % ethanol, centrifuged shortly again and finally air-dried. Plasmid DNA was dissolved in 60 µL of ddH_20 or Tris-HCl buffer (10 mM, pH 8).

3.2.2.3 DNA amplification by PCR

Standard DNA amplification by PCR was carried out using a mixture of Taq:Pfu (8:1) polymerases in a total reaction volume of 50 µL. Addition of Betaine was optional. The pipetting scheme was the following:

10x Taq reaction buffer	5 µL
Template DNA	1 µL
Fwd primer (10 µM)	1 µL
Rev primer (10 µM)	1 µL
Taq:Pfu polymerase (8:1)	1 µL
dNTPs (2,5 mM each)	4 µL
DMSO	2,5 µL
(Betaine	10 µL)
ddH_2O	Ad 50 µL

Standard PCR cycler conditions were set as described below:

Table 15: Cycler program of a standard PCR reaction.

Step	Temperature	Time	Repetitions
Hotstart	94 °C	30 s	1 x
Denaturation	94 °C	30 s	30 x
Annealing	52-72 °C	40 s	
Elongation	72 °C	Depending on amplicon length	
Final Elongation	72 °C	10 min	1 x
Cooling	4 °C	∞	-

After each PCR run the product was checked on an agarose gel. PCR reactions were then purified using the PCR purification kit (Qiagen) according to the manufacturer´s protocol.

3.2.2.4 Colony-PCR

For colony-PCR of *Streptomyces*, 4 mL of LB medium containing the selective antibiotic, if applicable, were inoculated with 10 µL of a spore suspension in a test tube. These cultures were grown under standard conditions (see 3.2.1.2) for 3 days. 1,5 mL of these cultures were transferred to an Eppendorf tube and spun down at 20.240 xg (maximum speed) for 5 min. After discarding the supernatant completely, the pelleted mycelium was dissolved in 60 µL DMSO by pipetting up and down. Cells were heated up to 98 °C for 10 min and then cooled down to room temperature. 2,5 µL of the resulting supernatant were added to a PCR reaction mix with the following pipetting scheme:

10x Taq reaction buffer	5 µL
Supernatant from heated mycelia	2,5 µL
Fwd primer (10 µM)	1 µL
Rev primer (10 µM)	1 µL
Taq:Pfu polymerase (8:1)	1 µL
dNTPs (2,5 mM each)	4 µL
Betaine	10 µL
ddH_2O	25,5 µL

The PCR cycler program was analogous to the standard PCR program with the only difference that the hotstart step was elongated to 7 min.
For *E. coli* colony-PCR, the pre-treatment as described for *Streptomyces* was not necessary, but the colonies were directly picked from the plate and dissolved in the PCR reaction mix.

3.2.2.5 Agarose gel electrophoresis

Agarose gel electrophoresis was routinely carried out using 1 % agarose dissolved in 1x TAE buffer. The up-heated agarose was supplemented with either thiazole orange dye (1:10.000) or peq-green dye (1:1.000) and then filled into the gel chamber where it was allowed to solidify for 20 min. The samples were mixed with the appropriate amount of 6x loading dye and applied into the gel slots. The volumes applied to the slots varied between 5 and 20 µL depending on the application. For determination of fragment sizes, 6 µL of a DNA ladder (ThermoFisher 1 kb DNA ladder) was applied as standard. For DNA fragments smaller than 250 bp, a 3 % agarose gel was prepared and the 2-Log DNA ladder (NEB) was applied in addition to the standard marker. Agarose gels were run at 80-120 V for 45 to 120 min and analyzed using UV light of 312 nm wavelength.

3.2.2.6 Enzymatic digestion of DNA

Restriction enzymes were applied in classical cloning and in test digestions to verify the present DNA construct. rSAP alkaline phosphatase was only added when plasmids were digested for cloning. Standard restriction digestion reactions were made up from the following pipetting scheme:

10x CutSmart buffer	2 µL
Restriction enzyme 1	1 µL
Restriction enzyme 2	1 µL
(rSAP alkaline phosphatase	1 µL)
DNA	2-10 µL
ddH_2O	Ad 20 µL

The restriction digest was carried out at 37 °C for 1-2 h, depending on the applied enzymes. If possible, the enzymes were subsequently heat inactivated at 65 °C or 80 °C for 20 min or the reaction mix was directly purified using the PCR purification

kit (Qiagen). The success of the digestion was either checked on an agarose gel (in the case of test digestions) or digested fragments were directly applied in DNA ligation (for cloning).

3.2.2.7 Ligation of DNA

Ligation of DNA fragments with compatible ends was accomplished by T4 DNA ligase activity. Different molar ratios (typically 3:1; 3:2 and 3:6) of backbone and insert DNA were applied in the reaction mixtures. A standard ligation reaction had the following composition:

10x T4 DNA ligase buffer	2 µL
Backbone DNA	x µL [30 ng]
Insert DNA	x µL [depending on size]
T4 DNA ligase	1 µL
ddH_2O	Ad 20 µL

The reaction mixture was either incubated at room temperature for 2 h or at 16 °C over night. Subsequently, the reaction mixture was used to transform competent *E. coli* cells.

3.2.2.8 Sequencing

Sequencing was performed by Eurofins Genomics (Ebersberg, Germany). In case of purified PCR products, the samples were diluted to 5 ng/µL final concentration (for PCR products of 300-1.000 bp length) or 1 ng/µL final concentration (for PCR products of 150-300 bp length) in 15 µL volume. Plasmid DNA had a final concentration of 50-100 ng/µL in 15 µL volume. If applicable, 2 µL of sequencing primers (10 µM) were added. Sequencing results were analyzed using SeqMan software, Snapgene and BLAST.

3.2.2.9 Isolation and purification of RNA

For the isolation of high quality RNA, several attempts using different methodologies were conducted. The protocol that finally worked best is described here.
Due to the fact, that transcription should be monitored during production of liponucleosides, cultures were set up exactly like for the production of liponucleosides

(3.2.5.1). 5 mL of the grown cultures were harvested for RNA isolation. The pelleted cells were gently resuspended in 350 µL RLT buffer from the RNeasy mini kit (Qiagen), containing ß-mercaptoethanol (10 µL/mL RLT buffer) according to the manufacturers protocol. The cell suspension was transferred into a bead-beating tube and from there on kept on ice. Cell disruption was achieved by two runs in the FastPrep machine at 6,5 M/s for 30 s with a cooling step on ice for 2 min in between the two runs. Afterwards the tubes were centrifuged for 3 min at maximum speed at 4 °C to remove the cell debris. The supernatant containing all nucleic acids was transferred into a new tube and purification was achieved using the RNeasy mini kit (Qiagen) according to the manufacturers instructions. RNA was finally eluted from the column in 50 µL of RNase-free H_2O. Due to the fact that on-column DNase-digestion resulted in insufficient DNA removal, the DNase digestion step was conducted in solution, using the rDNase kit (Macherey & Nagel), following the manufacturers instructions. After DNase treatment, RNA was purified using the Macherey & Nagel RNA clean up kit and finally eluted in 60 µL of RNase-free H_2O. Integrity and amount of the isolated RNA was finally determined on an agarose gel using MOPS buffer (3.1.11.7). For quantification, the RNA samples were applied to Nanodrop measurement. The absence of remaining gDNA traces was verified by the absence of a PCR product using primers to amplify *hrdB* as the target gene (primer pair hrdB_test_fwd/ hrdB_test_rev). As long as the PCR reaction yielded a product, the respective RNA sample was repeatedly digested with DNase and the PCR was run again.

3.2.2.10 RT-qPCR

In order to determine the optimal time points from the 20-L fermentation at which the collected protein samples should be taken for the DACA experiment, RT-qPCR was conducted. Especially transcription of biosynthetic genes can serve as a good indicator for ideal sample recovery. In the DACAs, one protein sample reflecting low and one high biosynthetic activity should be used. Thus, specific qPCR primers were designed for the first biosynthetic genes within the caprazamycin and liposidomycin gene cluster, namely *cpz10* and *lpmH*. The RNA sigma factor gene *hrdB* functioned as a control reference. RNA was isolated from the fermenter samples for each timepoint and both heterologous producer strains.

Following up RNA isolation and quantification, reverse transcription of the RNA samples was carried out immediately due to the improved stability of cDNA in comparison to RNA.

From every RNA sample, 60 ng (for *S. coelicolor* M512/cpzLK09 strain) and 50 ng (for *S. coelicolor* M512/lpmLK01 strain) were reverse transcribed into cDNA using the iScript cDNA synthesis kit (Bio-Rad), following the manufacturers instructions. From this reaction mixture, 2 µL were applied in the following qPCR using the QuantiNova SYBR green PCR kit (Qiagen). To allow the fluorescence readout of the PCR cycler it was first calibrated using the exact same conditions (96-well plate, volumes, sealing method etc.) as for the qPCR itself. For the qPCR, the pipetting scheme and cycler program was set according the the manufacturers protocol. After each run, a melting curve over a temperature range between 55°C to 95 °C in 0,5 °C steps was measured to verify the identity of the amplified PCR products.
Analysis of the measured c(T) values was done via normalization against the expression of the control gene *hrdB,* applying the cycle threshold method (Livak and Schmittgen, 2001). Due to small sampling volumes and therefore insufficient RNA yield, data points are only present for the later time points of the fermentation.

3.2.2.11 Cloning of the pCRISPR-TT vectors

Genes encoding for regulators that showed a significant impact on liponucleoside production upon overexpression, were subjected to gene deletion experiments using the CRISPR-Cas9 system (*sco3571* and *sco4385*). The pCRISPR-TT plasmid carries all required components, harbors the *cas9* gene under the control of a theophylline inducible riboswitch and exhibits temperature sensitive replication. It is a derivative of the pCRISPR-Cas9 plasmid introduced by Tong *et al*., in 2015.
In order to equip the plasmid with all genetic information to be specific for the target gene, a definite spacer sequence (guiding the Cas9 endonuclease to the gene to be deleted) and the homology domains for the repair of the Cas9-induced double strand break (DSB) needed to be incorporated into the vector.
The spacer sequence is required to be a part of the sgRNA (single-guide RNA) and was cloned in front of the tracrRNA (trans-activating CRISPR RNA) to guide the Cas9 nuclease to the position where to induce the DSB. In order to find suitable spacer sequences within the target gene the online-software tool “CRISPY-web” (https://crispy.secondarymetabolites.org) was used. This tool identifies a protospacer adjacent motif (PAM) which consists of three nucleotides of the sequence NGG and which will be recognized by the Cas9 enzyme. The spacer sequence is defined by the 20 nucleotides downstream of the PAM. When searching for spacer sequences the CRISPY-web tool lists its hits by the number of off-target hits. Of course, this number should be as small as possible to reduce unwanted side effects of the Cas9 activity.

The two sequences with the lowest number of off-target hits were chosen for cloning, respectively, and specific primers were designed with the fwd- primer carrying the complete 20-nt sequence. This primer pair was used to amplify the complete sgRNA sequence from the pCRISPR-TT vector and to include restriction sites for *NcoI* and *SnaBI* by which the PCR product could be cloned into the vector, thereby inserting the additional 20-nt spacer sequence. Correct cloning was checked by PCR and sequencing of the PCR product. Verified constructs were assigned “pCRISPR-TT-spacer”.

The assembly of the homology arms was done in a subcloning step into the cloning vector pBlueScript SK II to allow blue/white screening of the *E. coli* transformants. Homology arms were defined as the ~ 1 kb genome region up- and downstream of the target gene. These regions were amplified by PCR and assembled with the backbone vector using Gibson assembly (see 3.2.2.12). Transformants were checked by colony-PCR and the subcloned construct was verified by sequencing. The assembled homology domain could be excised by *Eco*RV, purified from the agarose gel and cloned into the pCRISPR-TT-spacer construct via the *Stu*I restriction site. Final pCRISPR-TT constructs (pSW17-pSW20) harbor the specific spacer sequence and homology arms specific for each targeted gene.

3.2.2.12 Gibson assembly

For the assembly of the two homology arms carrying overhangs to the digested and dephosphorylated pBlueScript SK II backbone the HiFi DNA Assembly Master Mix (NEB) was used. The reaction setup had the following composition:

Upstream homology arm	x µL [0,2 pmol]
Downstream homology arm	x µL [0,2 pmol]
Backbone vector	x µL [0,1 pmol]
2x DNA assembly mix	15 µL
ddH_2O	Ad 30 µL

The components were mixed by pipetting up and down and incubated at 50 °C for 45 min. Subsequently, two times 15 µL of the reaction volume were used to transform chemocompetent *E. coli* XL1 blue cells which were plated out on LB-plates supplemented with the selective antibiotic, IPTG and X-Gal for blue/white screening.

3.2.2.13 *In-vivo* induction of *cas9* and plasmid curing

Following successful cloning of the final pCRISPR-TT constructs (pSW19 and pSW20 for deletion of *sco3571* and pSW17 and pSW18 for deletion of *sco4385*) the plasmids were conjugated into *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01 following the standard protocol (see 3.2.1.7). After the first round of spore dilution, four single colonies per plasmid were streaked confluently on MS-agar containing 10 mM $MgCl_2$, apramycin and nalidixic acid. These plates were used to prepare spore stocks which were diluted and plated out to obtain single colonies on TSA (solid TSB medium) plates containing 10 mM theophylline and apramycin. Addition of theophylline induces the transcription of the *cas9* gene (under the control of a theophylline inducible riboswitch) from the pCRISPR-TT vector. Emerging colonies were diluted, restreaked on MS-agar to obtain spore suspensions and checked for successful deletion of the target gene by colony-PCR (see 3.2.2.4). Primers were designed to bind within the genome outside of the homology arms, otherwise the primers could also bind within the plasmid which is still present within the cells. In case of a positive PCR result, the corresponding PCR product was checked by sequencing to validate the deletion on nucleotide level. Positive mutants were streaked to obtain spore suspensions. While for the deletion of *sco4385* three independent mutants could rapidly be generated, it was only possible to achieve two independent knockouts of *sco3571* in *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01, respectively. Although induction of *cas9* was carried out several times, most of the colonies yielded a mixture of the knockout and wildtype PCR product. Even mutants showing only a slight wildtype and a strong knockout band on the agarose gel could not be genetically purified by several rounds of dilution and restreaking.

Finally, spore suspensions of mutants containing the successful gene deletion were diluted and streaked on TSA plates which were incubated at 37 °C instead of 30 °C. Increasing the growth temperature prevents replication of the pCRISPR plasmid and thus curing the plasmid could be achieved after two to three rounds of incubation. The mutant strains were tested for the loss of the vector by streaking them on plates with and without apramycin simultaneously. In the case of one independent mutant of *S. coelicolor*/lpmLK01 Δ*sco4385*, *S. coelicolor*/lpmLK01 Δ*sco3571* and *S. coelicolor*/cpzLK09 Δ*sco3571*, respectively, plasmid curing was unsuccessful after five rounds of incubation, therefore, these mutants were cultivated still harboring the pCRISPR plasmid. At the end of the cultivation, the strains were tested positive for their apramycin resistance still showing the presence of the plasmid.

3.2.3 DNA-affinity capturing assay (DACA)

3.2.3.1 Amplification of promoter fragments

Promoters/intergenic regions within the caprazamycin and liposidomycin gene clusters that should be applied in the DACAs were identified by RNA-sequencing of the heterologous strains (see 4.1.2). These selected regions were amplified using primer pairs resulting in DNA amplicons of different lengths:

Caprazamycin cluster		Liposidomycin cluster	
P*cpz6*	424 bp	P*lpmD*	544 bp
P*cpz9*	358 bp	P*lpmG*	520 bp
P*cpz10*	365 bp	P*lpmH*	369 bp
P*cpz24*	431 bp	P*lpmV*	455 bp

Negative control	
P*hrdB*	559 bp

In a first PCR the promoter fragments were amplified from genomic DNA of the heterologous *Streptomyces* strains using a standard PCR procedure (see 3.2.2.3). The exact cycler conditions for each promoter fragment were the following:

Table 16: Cycler conditions for the amplification of promoter fragments used in the DACA.

Promoter	Annealing temp.	Elongation time	Promoter	Annealing temp.	Elongation time
P*cpz6*	62 °C	50 s	P*lpmD*	61 °C	50 s
P*cpz9*	63 °C	45 s	P*lpmG*	65 °C	50 s
P*cpz10*	62 °C	50 s	P*lpmH*	63 °C	30 s
P*cpz24*	61 °C	30 s	P*lpmV*	61 °C	40 s
P*hrdB*	62 °C	40 s			

One of the primers was equipped with a linker sequence complementary to the DAC-biotin primer for adding the biotin tag in the follow-up PCR. The PCR product of this first PCR was verified by sequencing and then used in the second PCR in which the primer carrying the linker was replaced by the DAC-biotin primer to yield large amounts of biotinylated promoter fragments. This second PCR had a total volume of 5 mL and was carried out in a 96-well plate with 50 µL for each reaction:

10x Taq reaction buffer	500 µL
Template DNA	15-30 µL (depending on DNA concentration)
Fwd primer (10 µM)	100 µL
Rev primer (10 µM)	100 µL
Taq:Pfu polymerase (8:1)	100 µL
dNTPs (2,5 mM each)	400 µL
DMSO	250 µL
(Betaine	1000 µL)
ddH_2O	Ad 5.000 µL

The resulting PCR product was purified using the Qiagen PCR purification kit according to the manufacturer´s protocol. Every 96-well plate yielded between 15 and 35 µg DNA. 120 µg of each promoter fragment were required for the DACA.

3.2.3.2 20-L fermentation of *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01

20-L fermentation of the two strains *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01 was carried out with the kind help of Andreas Kulik at the Institute of Biology, University of Tübingen.

Fermentation of both heterologous strains was carried out in parallel. For inoculation of the two 20-L stainless steel fermenters, 1 L pre-culture was required, respectively. Pre-cultures were set up in 20 baffled flasks, for each strain respectively, each containing 50 mL TSB medium which was inoculated with 40 µL of spore suspension. Pre-cultures were grown at 30 °C, 200 rpm for two days and were then pooled. Under sterile conditions this volume was used to inoculate the fermenters, each containing 19 L of P-medium and 4 mL Antifoam. The conditions were set to 1000 revolutions and the ventilation was 0,5 vvm (volume of air per volume of the fermenter per minute). Ventilation rate and pH value was monitored constantly during fermentation. Samples were taken at the following time points: 18 h, 24 h, 36 h, 42 h, 48 h, 54 h, 60 h, 72 h and 90 h. At each time point, samples were taken for protein extraction, cell dry weight determination, RNA isolation and extraction of liponucleosides to track the production of the antibiotics. The sampling volume ranged from 550 mL to 1060 mL depending on the cell density.

For protein extraction, cells were pelleted by centrifugation at 4 °C at 8000 rpm for 50-60 minutes and washed in TGED buffer. The suspension was then transferred to 50 mL falcons, centrifuged again at 4 °C at 5000 rpm for 25 minutes followed by the

removal of the supernatant. Resulting cell pellets were weighted and finally stored at -80 °C until protein extraction.
Cell dry weight was determined in technical duplicates according to the standard protocol (see 3.2.1.8).
For the extraction of liponucleosides from the fermenter samples, two times 20 mL of the culture was centrifuged and the supernatant used for liponucleoside extraction following the standard protocol (see 3.2.5.1).
Samples for RNA isolation needed to be processed right after being taken from the fermenter. At each time point, five times 0,5 mL of the culture were added to a 2 mL Eppendorf tube already containing 1 mL of RNAprotect bacteria reagent (Qiagen). After vortexing for 5 s, the mixture was incubated at RT for five minutes and then centrifuged for 10 min at 5000 xg. The supernatant was removed completely and the cell pellet containing stabilized RNA was stored at -80 °C until RNA isolation (see 3.2.2.9).

3.2.3.3 Protein isolation from fermenter samples

Cell pellets achieved by fermentation, representing the selected time points (36 h and 54 h for *S. coelicolor* M512/cpzLK09; 48 h and 72 h for *S. coelicolor* M512/lpmLK01), were used to isolate the whole-cell proteome for the DACA. Wearing gloves during the isolation process minimized possible protease contamination. Cell pellets were thawed at room temperature, resuspended in pre-cooled TGED-buffer (2,5 mL buffer/g cell wet weight) and DTT (1 mM), PMSF (1 mM), benzamidin (1 mM) and pepstatin A (1 µM) were added to further block protease activity. When the cell pellets were thawed completely, cells were disrupted by 3-4 times passage through the french press (at approximately 1000 bar). Subsequently, the lysate was centrifuged at 18000 rpm at 4 °C for 45 min. The resulting supernatant was transferred to a fresh centrifuge tube and centrifuged for 15 min again applying the same conditions. The supernatant was then transferred to a falcon tube to quantify the volume. For exact determination of the protein concentration a Bradford assay was conducted (see 3.2.3.4). The protein solution was stored on ice until being used in the DNA-affinity capturing assay.

3.2.3.4 Bradford-Assay

In order to quantify the protein amounts that were available for the DACAs, a calibration line was achieved by using definite amounts of BSA. One mL of Bradford

reagent was mixed together with increasing volumes of a BSA stock solution (1 mg/mL) in TGED buffer resulting in a calibration line ranging from 0 mg/mL, 0,125 mg/mL, 0,25 mg/mL, 0,5 mg/mL, 0,75 mg/mL and 1 mg/mL. After mixing and incubation for approximately 5 minutes, 200 µL of each concentration sample were transferred to a 96-well plate and absorption was measured at 600 nm in a plate reader. The cell lysate used for the DACAs was diluted 1:10 and 1:100 and measured under the same conditions as the BSA standards. Protein concentrations could be calculated using the calibration line.

3.2.3.5 DNA-affinity capturing assay

Before the actual DACA could be conducted, the biotinylated promoter DNA sequences needed to be bound to streptavidin-covered Dynabeads for immobilization.

For every promoter region which was tested in the assay, 5 mg Dynabeads M-280 Streptavidin (Thermo Fisher) were pipetted into a 2 mL Eppendorf tube. The beads were magnetically separated from the storage solution and instead washed with an equal volume (500 µL) of B+W buffer by pipetting along the side of the Eppendorf tube. This washing step was repeated two additional times. After removal of the last wash fraction, an additional volume of B+W buffer was added together with the required volume of the biotinylated promoter DNA regions (120 µg of each fragment). This mixture was gently shaken at room temperature for one hour in order to facilitate binding between the biotinylated DNA and the streptavidin-covered beads. The beads were then magnetically collected at one side of the tube to remove the supernatant. The beads carrying the DNA fragments were gently washed three times with the initial volume of B+W buffer before they were resuspended in the same volume of TGED buffer. In this form the beads were either ready to use for the DACA or could be stored at 4°C.

After coupling the DNA fragments to the Dynabeads and isolation of the whole-cell protein extract from the cell pellets, the actual DNA affinity capturing assay could be conducted.

At first, salmon sperm DNA (100 µg/mL) was added to the protein extract to immediately sort out proteins binding randomly to DNA. This mixture was then divided into falcons for every promoter fragment to be tested. In the DACAs conducted with promoter sequences from the caprazamycin cluster, 146 mg of protein were set in for every promoter (both time points, respectively). DACAs with the promoter sequences from the liposidomycin cluster were conduceted with 127 mg protein (48 h time point)

and 131 mg protein (72 h time point) for each promoter, respectively. Falcons containing salmon sperm DNA and the protein solutions were incubated on ice for 15 minutes until the Dynabeads (with bound DNA fragments) were added. The DNA-protein extract mixture was incubated at room temperature with gentle agitation for 45 minutes. Afterwards, by magnetic separation, the supernatant containing unbound proteins was stepwise removed from the Dynabeads. Specifically bound proteins should now be bound to the DNA, which in turn was coupled to the Dynabeads. The beads were subsequently washed with 500 µL of TGED buffer, 500 µL TGED buffer containing 400 µg salmon sperm DNA and finally another 500 µL of TGED buffer. In order to separate the proteins from the DNA, two times 350 µL of an elution buffer containing 2 M NaCl was used to wash the beads (salt-out strategy). The eluted proteins needed to be precipitated to be applicable to mass-spectrometry measurement.
In order to regenerate the Dynabeads, two washing steps with TGED containing 2 M NaCl and two further wash repetitions with TGED buffer were conducted. The last wash fraction was checked by Bradford assay for the complete absence of proteins.

3.2.3.6 Protein precipitation

Protein precipitation was achieved by addition of 0,25 volumes of trichloroacetic acid (100 % *w/v* in ddH_2O) to the eluates of each DACA. After incubation on ice for 30 min, the precipitates were spun down by centrifugation at 4°C and 20.240 xg for 15 minutes. The supernatant was removed using the pipette and the (visible) protein pellet was washed with 500 µL ice-cold acetone. The centrifugation step was repeated and the acetone supernatant was removed by pipetting. The resulting protein pellet was air-dried for a couple of minutes and stored at -80 °C until analysis at the proteome center (University of Tübingen).

3.2.4 Methods of protein biochemistry

3.2.4.1 Protein overexpression and purification

For recombinant expression of His-tagged Sco4385 (pSW21) and LpmG (pSW22) electrocompetent cells of the *E.coli* strain BL21 (DE3) were transformed with the corresponding plasmid carrying the gene of interest under the control of the T7 promoter. Expression of the T7 RNA- polymerase in the *E. coli* (DE3) genome is facilitated via induction of the *lacUV5* promoter by IPTG. Single colonies were used to inoculate a 10 mL over-night culture (grown under standard conditions) containing

kanamycin. This culture was used to inoculate the main culture of 100 mL TB medium containing kanamycin to an OD_{600} of 0,075. Induction of the T7 promoter was achieved by addition of 0,5 mM IPTG at an OD_{600} of 0,6-0,7 (samples for SDS-PAGE were taken: centrifuged cell pellet from 1 mL culture, stored at -20 °C). After induction, the temperature was adjusted to 20 °C and incubation was continued over night (approximately 20 h). When the culture was harvested the next morning, OD_{600} values were measured and samples for SDS-PAGE analysis were taken. The harvested cell pellets of the total culture (centrifugation 15 min, 4 °C, 4000 rpm) were washed in 20 mL TEGED buffer, centrifuged again and then stored at -20 °C until protein purification.

For purification of Sco4385 and LpmG the cell pellets were thawed on ice and resuspended in 2,5 mL lysis buffer/g cell wet weight. Disruption of the cell suspension was achieved by four-times passaging through a french press at a pressure of approximately 1000 psi. The resulting lysate was centrifuged for 1 h at 4 °C, 15000 xg and the clear supernatant was applied to a Ni-NTA hand column (CV 500 µL). The column was washed with 5 CV of running buffer and the bound protein was finally eluted using increasing concentrations of imidazole in the elution buffer (1x 1 mL 100 mM imidazole, 1x 1 mL 150 mM imidazole, 1x 1 mL 200 mM imidazole, 2x 1 mL 250 mM imidazole). In parallel, a quick bradford assay (mixing 11 µL ddH_2O, 4 µL bradford reagent and 5 µL sample) was carried out to assess when no protein was present in the eluates anymore. Afterwards the column was washed with 10 CV of running buffer and stored in this buffer at 4 °C. Eluates containing the purified protein were stored at -80 °C (buffer contains 10 % glycerol). The soluble and insoluble fraction of the lysate, flow through, wash and the eluates were applied to SDS-PAGE (see 3.2.4.2) to assess the protein quality and success of the purification process. For determination of the protein quantity, the eluates were measured at the NanoDrop (Sco4385: extinction coefficient 19,48, MW 24,69 kDa; LpmG: extinction coefficient 19,73, MW 39,14 kDa) and in a Bradford assay. Due to the fact that both methods gave highly similar results, the protein quantity was only assessed at the NanoDrop for the following experiments.

While the yield of purified Sco4385 protein was very high (approximately 10,4 mg protein from a 100 mL culture) it was not possible to overexpress and isolate LpmG protein although different culture conditions were tested.

3.2.4.2 Polyacrylamide gel electrophoresis (SDS-PAGE)

Samples from the protein purification procedure as well as cell samples taken before and after induction of the culture were monitored by SDS-PAGE according to the method of Lämmli (1970). Table 17 shows the pipetting scheme for two 12 % polyacrylamide gels.

Table 17: Composition of stacking and resolving gels for SDS-PAGE.

Component	Resolving gel (12 %)	Stacking gel
Acrylamide (30 %)	5 mL	666 µL
1,5 M Tris-HCl, pH 8,8	2,5 mL	-
1 M Tris-HCl, pH 6,9	-	625 µL
MilliQ Water	1,9 mL	3,41 mL
SDS (20 %)	500 µL	250 µL
APS (10 %)	100 µL	50 µL
TEMED	10 µL	5 µL

To load cell pellet samples onto the gel, the pellet was resuspended in a volume of ddH_2O corresponding to the OD_{600} at the time point when it was harvested. 20 µL of this suspension or other protein samples were mixed with 5 µL of sample buffer (4:1 ratio) and boiled at 98 °C for 7 min following centrifugation at 13000 rpm for 5 min. 15 µL of the resulting supernatant were loaded onto the gel. 6 µL of PAGE Ruler Plus prestained protein Ladder were applied as a marker. SDS-PAGE was run at 100 V for the first hour or until the samples reached the bottom border of the stacking gel. The voltage was then increased to 120 V for another 1-2 h. Once the lowest marker band reached the bottom edge of the gel it was stopped, taken out of the chamber and stained in staining solution for approximately 3 h under slight agitation. Subsequent destaining of the gel using destaining solution was carried out under slight agitation until the protein bands were well visible.

3.2.4.3 Size exclusion chromatography (SEC)

After purification of Sco4385 via a Ni-NTA hand column, the protein was further purified by size exclusion chromatography (SEC) to circumvent the possibility of contaminating protein in the BLI measurements. SEC was conducted with the kind help of Tim Orthwein at the Department of Microbiology (Institute of Biology, University of Tübingen). Prior to SEC, the buffer of the protein was exchanged in order to remove the glycerol that could cause high pressure during SEC. The buffer

exchange was facilitated using Amicons (Sigma-Aldrich) and the protein was finally present in the original volume of 3 mL in buffer (50 mM Tris-Hcl pH 7,5, 150 mM NaCl). The same buffer composition, except containing 250 mM NaCl, also served as the running buffer during SEC. For chromatographic separation, the flow rate was set to 0,75 mL/min and eluates of 1 mL each were collected. The run was carried out at 4 °C. Sco4385 eluted as a single peak at an elution volume of 50-60 mL. Eight fractions (8 mL) were pooled and concentrated to a volume of 1 mL using Amicons. Glycerol was added in a final concentration of 10 % to store protein aliquots at -80 °C until biolayer-interferometry was conducted

3.2.4.4 Biolayer interferometry (BLI)

Biolayer-interferometry is an emerging method to analyze different types of molecular interactions. It is widely used to detect protein-protein interactions in real-time and can also be used to calculate binding kinetics and dissociation constants. A coated sensor surface is loaded with the bait molecule which is dipped into the analyte solution. Binding of the analyte to the bait results in a wave length shift of a light beam that is constantly emitted on the sensor surface and the reflected light beam is detected by the device again. The wave length shift is thereby proportional to the slice thickness und molecular weight bound to the sensor surface and can be used to characterize the molecular binding. Some of the major advantages are that only tiny amounts of both protein (analyte) and DNA (bait) are needed, it is a high throughput technique able to scan 16 DNA fragments within only 2 h, and it is significally more sensitive compared to other methods such as EMSA.

In this study, BLI was applied to measure protein-DNA interaction between the purified TetR regulator Sco4385 and the promoter regions P*lpmH* and P*cpz10* at which it was found to bind in the DACAs.

All measurements were performed on a K2 Octet system (FortéBiosciences Inc., Dallas, USA) in cooperation with Tim Orthwein at the Department of Microbiology (Institute of Biology, University of Tübingen). In order to determine the binding site of Sco4385 within the promoter regions more precisely, the whole promoter fragments (P*lpmH*: 369 bp; P*cpz10*: 365 bp) were devided into 14 segments, respectively, each 50 bp in length and overlapping by 25 bp to either up- or downstream segments. All tested DNA oligonucleotides are listed in Table 10. Double stranded DNA (dsDNA) segments were coupled to the streptavidin-covered sensor surface via a single-stranded (ss) biotinylated linker (ReDCaT linker), thus, each DNA segment needed to

carry a single-stranded overhang complementary to the linker sequence. In total, three types of DNA oligonucleotides were used:

a) ReDCaT (Re-usable DNA capture technique) linker: biotinylated 20 nt sequence (Stevenson *et al*., 2013)
b) Forward oligonucleotide: 50 nt coding strand by default
c) Reverse oligonucleotide: 50 nt non-coding strand by default with 20 nt overhang complementary to the ReDCaT linker attached to the 3`-end (70 nt in total)

For annealing the forward and reverse oligonucleotide, both were mixed in a ratio of 1,2:1 (forward:reverse) to reduce the amount of free reverse oligonucleotide in the mixture, which could compete with the dsDNA for binding to the ReDCaT linker. The mixture was incubated in 1,5 mL Eppendorf tubes at 95 °C for 10 min on a thermoshaker (no agitation) with a lid. After this time the thermoshaker was switched off and the Eppendorf tube was allowed to slowly cool down on the device for at least 2,5 h. When the mixture was cooled to RT, the volume was adjusted (with binding buffer for BLI, see 3.1.11.10) to achieve a final concentration of 1 µM for the dsDNA oligonucleotide.
Based on the assays conducted by Dr. Paulina Bekiesch using surface plasmon resonance (SPR) (Bekiesch *et al*., 2016b), it was first tried to apply the ReDCaT method (Stevenson *et al*., 2013) to the BLI. This method aims to minimize the number of required sensors by binding the ReDCaT linker to the sensor once and then sequentially binding different dsDNA oligonucleotides to the linker. This procedure requires complete regeneration of the sensor/linker by denaturating the bound dsDNA and thereby disassociation from the linker. Several conditions and regernation solutions were tested but unfortunately, it was not possible to reliably separate the dsDNA from the linker without damaging the sensor for the following measurement. Denaturation of the dsDNA could be achieved by applying high alkaline pH (>pH 11) but already the second tested DNA oligonucleotide gave a weaker signal than the first one when bound to the linker, indicating some damage to the sensor/linker. Finally, every DNA oligonucleotide was measured using a new sensor which allowed to monitor the sequential binding of the linker, dsDNA oligonucleotide and, if applicable, eventually Sco4385.
The measurement itself was carried out using SA Steptavidin “Dip-and-read” sensors (FortéBiosciences Inc., Dallas, USA) in a black 96-well plate at a constant

temperature of 30 °C and with agitation of the plate at 1000 rpm. The concentration of all DNA samples was 1 µM, both for the linker and the dsDNA oligonucleotides. For each measurement 5,68 µg (1 µM) of purified Sco4385 was applied and every well contained a volume of 230 µL. Prior to the measurement, the sensors were incubated in binding buffer for 5 min at RT. Table 18 displays all steps of the measurement, that was repeated for every dsDNA oligonucleotide. All measurements were conducted in duplicates.

Table 18: Measurement protocol for DNA-protein interaction using BLI.

Step No.	Step	Duration [sec]
1	Equilibration of the sensor in binding buffer (baseline)	60
2	Binding of the linker	180
3	Equilibration of the sensor/linker in binding buffer (baseline)	60
4	Binding of the dsDNA to the linker	180
5	Equilibration of the sensor/linker/dsDNA in binding buffer (baseline)	60
6	(Test for) binding of Sco4385	180
7	Dissociation in binding buffer	180

3.2.5 Chemical and analytical methods

3.2.5.1 Extraction of liponucleosides from *Streptomyces*

For the production of liponucleosides from *Streptomyces*, pre-cultures were inoculated and grown according to the standard protocol. One mL of the pre-cultures was transferred to 300 mL baffled flasks containing a stainless-steel spiral and 70 mL of P-medium. The production culture was supplemented with trace elements solution and, in case of the strains harboring the replicative pUWL-vector or derivative thereof, apramycin for selection. Main cultures were grown under standard conditions (see 3.1.10.2) for 7 days. After this time, samples for cell dry weight determination in technical duplicates were collected (see 3.2.1.8) and 50 mL of the culture were harvested by centrifugation. 40 mL of the supernatant were transferred to a new falcon tube to be extracted. The supernatant was acidified with HCl to a pH value of 4. Afterwards, an equal volume of *n*-butanol was added and mixed thoroughly with the supernatant. Phase separation was speeded up by centrifugation at 4 °C and 3000 xg for 15 min. The organic phase was transferred to rotary evaporation flasks and the solvent was removed using rotary evaporators. Finally, the residual substance was

dissolved in one mL of LC-MS grade methanol, stored at -80 °C for 30 min (or over night at -20 °C) and centrifuged (4 °C, 13300 rpm, 15 min). Finally, the resulting supernatant was applied to LC-MS measurement.
For quantification of caprazamycin-aglyca, purified caprazamycin-aglycon A was used as a reference and a calibration line with concentrations of 0,5 mg/L, 10 mg/L, 40 mg/L, 120 mg/L and 240 mg/L was measured in technical triplicates using the same LC-MS method as for the actual samples.

3.2.5.2 HPLC-MS measurements and evaluation

Samples were kindly measured by Andreas Kulik (Institute of Biology, University of Tübingen) on an Agilent 1200 system (Agilent Technologies, USA) using a Luna Omega Polar C18 column (5 µm, 150 x 2,1 mm) as stationary phase. A linear gradient was applied within 20 min from 100 % aqueous phase containing 0,1 % formic acid to 100 % acetonitrile containing 0,06 % formic acid which was hold for another 3 min. Routinely, an injection volume of 5 µL was applied. The flow rate was set to 400 µL/min and the UV detection wavelength was at 262 nm. Mass spectrometry was carried out using a LC/MSD Ultra Trap System XCT 6330 (Agilent Technologies, USA) with positive electrospray ionization (performed with a capillary voltage of 3.500 V, heated capillary temperature 350 °C, acquired mass range from 100 to 2200 *m/z*).
Acquired LC-MS data were evaluated using the DataAnalysis software (Bruker Daltonik GmbH, Germany). The following masses for unsulfated and sulfated caprazamycin-aglyca and liposidomycins were detected: 958,5 *m/z* $[M+H]^+$ for unsulfated caprazamycin aglyca A/B and liposidomycin L/M at Rt 14,9 min; 930,5 *m/z* $[M+H]^+$ for unsulfated caprazamycin aglyca E/F and liposidomycin B/C at Rt 13,8 min; 944,5 *m/z* $[M+H]^+$ for unsulfated caprazamycin aglyca C/D/G and liposidomycin H at Rt 14,3 min; 1038,5 *m/z* $[M+H]^+$ for sulfated aglyca caprazamycin A/B and liposidomycin L/M at Rt 16,4 min; 1010,5 *m/z* $[M+H]^+$ for sulfated caprazamycin aglyca E/F and liposidomycin B/C at Rt 15,1 min and 1024,5 *m/z* $[M+H]^+$ for sulfated caprazamycin aglyca C/D/G and liposidomycin H at Rt 15,7 min. Peaks of the extracted ion chromatograms (EICs) were integrated and the area under the curve was calculated. Finally, all areas were summed up for each strain and the value was routinely normalized against the cell dry weight of the culture determined at the end of the cultivation time (see 3.2.1.8).

4 Results

4.1 RNA-sequencing of *S. coelicolor* M512, *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01

In order to study the transcriptional operon structure of the caprazamycin and liposidomycin BGC upon heterologous expression in *S. coelicolor* M512, the respective strains *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01 were subjected to RNA-sequencing (RNAseq). Additionally, to enable a global view on the regulatory changes that come along with heterologous expression of the BGCs, strain *S. coelicolor* M512 was included for comparison. For RNA-sequencing (RNAseq), total RNA was isolated, digested by DNase and purified (see 3.2.2.9). Quality control, rRNA depletion, library preparation and sequencing was accomplished in cooperation with the NGS competence center Tübingen (NCCT). Evaluation of the raw sequencing data and assignment to the annotated genes of the caprazamycin and liposidomycin BGC was kindly carried out by Theresa Harbig from the working group of Prof. Kay Nieselt (Interfaculty institute for bioinformatics and medical informatics, University of Tübingen). Prepared RNAseq results were visualized using the Integrated Genome Browser (IGB) software.

4.1.1 Preparation and quality assessment of isolated RNA

Two time points, 2 and 4 days of main culture growth, were selected to isolate RNA, representing a sample at the beginning and at the peak of antibiotic production, respectively. Three independent biological replicates of *S. coelicolor* M512, *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01 each were inoculated and cultures were grown as described in 3.2.5.1, with incubation terminated after 4 days of growth in the main culture. Furthermore, a backup culture of each strain was set up which was used to test further purification steps in order to improve subsequent rRNA depletion rates. Based on previous RNAseq attempts, the rRNA depletion step was identified as crucial and the most critical step to achieve high-quality sequencing results.

At the selected time points, cell pellets of the cultures were harvested and used to isolate total RNA (see 3.2.2.9). After DNase treatment and final purification of the RNA, its integrity and amount was assessed by agarose-gelelectrophoresis, which is shown in Figure S 1. All RNA samples clearly exhibit bands at approximately 1500 bp and 1000 bp length, corresponding to 23 S and 16 S rRNA, indicating high levels of integrity and no degradation. In order to test a beneficial impact on the subsequent

rRNA depletion, isolated RNA of the backup cultures was used to conduct an additional clean-up by ethanol precipitation, as suggested by our cooperation partners from the NCCT. However, rRNA depletion did not perform better on the samples after precipitation, therefore, this step was dismissed for the other samples. The remaining 18 samples (3 strains in 3 biological replicates at 2 time points) were handed to our cooperation partners and quality control was carried out. RNA concentrations measured by the NCCT and remaining rRNA contents after rRNA depletion for each sample are listed in Table 19.

Table 19: Quality assessment of RNA samples used for RNA-sequencing. RNA concentrations were measured on a Qubit device at the NCCT. Each sample contained a volume of 50 µL. cpz: cosmid cpzLK09; lpm: cosmid lpmLK01; 2d and 4d: RNA isolated after 2 and 4 days of cultivation, respectively. Numbers at sample names indicate biological replicates.

Sample name	RNA concentration [ng/µL]	Remaining rRNA content [%]
S. coelicolor M512_2d_1	1270	23
S. coelicolor M512_2d_2	1440	18
S. coelicolor M512_2d_3	3040	19
S. coelicolor M512/cpz_2d_1	1040	8
S. coelicolor M512/cpz_2d_2	1930	8
S. coelicolor M512/cpz_2d_3	2940	28
S. coelicolor M512/lpm_2d_1	2180	14
S. coelicolor M512/lpm_2d_2	1040	11
S. coelicolor M512/lpm_2d_3	1700	13
S. coelicolor M512_4d_1	1850	50
S. coelicolor M512_4d_2	2360	34
S. coelicolor M512_4d_3	2960	31
S. coelicolor M512/cpz_4d_1	2940	66
S. coelicolor M512/cpz_4d_2	3320	58
S. coelicolor M512/cpz_4d_3	2140	19
S. coelicolor M512/lpm_4d_1	2360	35
S. coelicolor M512/lpm_4d_2	1220	22
S. coelicolor M512/lpm_4d_3	2960	39

4.1.2 Analysis of operon structures in liponucleoside antibiotic gene clusters

4.1.2.1 The caprazamycin BGC

In order to select putative promoter regions of the heterologously expressed gene clusters for the DACAs not only based on the localization and length of the intergenic regions, but also on transcriptional data, the operon structure and thereby promoter regions were analyzed. Additionally, it allowed to evaluate the impact of the individual promoter regions on transcription of the entire gene cluster.

At first, the transcriptional variance between the three biological replicates of *S. coelicolor* M512/cpzLK09 was assessed. Figure 6 compares transcription of the caprazamycin BGC for all 3 replicates after 2 days of growth. Strikingly, the transcriptional pattern is extremely similar for all replicates and varies only slightly in the absolute amounts of transcripts. This is also the case for the transcription rates measured after 4 days of growth, but the variance in absolute read counts between the biological replicates is higher (Figure S 2). In all measurements, the read count (reflecting the quality of the data) correlates with the percentage of remaining rRNA in the respective sample (Table 19). For a more detailed overview of the caprazamycin BGC operon structure, Figure 7 and Figure 8 show transcription rates for both time points, 2 and 4 days, of the biological replicate exhibiting the lowest rRNA content, respectively.

These data were used to identify promoter regions of the caprazamycin BGC which were then applied in subsequent DNA-affinity-capturing-assay (DACA) experiments.

The intergenic region (302 bp) between genes *cpz5* and *cpz6* is required as a starting point for transcription as these two neighbouring genes are encoded in divergent orientation. This transcriptional start site is also well visible in the RNAseq data. While *cpz6* encodes for a type III polyketide-synthase, *cpz5* encodes for a 3-hydoxy-3-methyl-glutaryl-CoA-synthase, involved in biosynthesis of the 3-methylglutaryl moiety in caprazamycins and liposidomycins. The next promoter region of the gene cluster was assumed to be right upstream of the cluster-situated regulatory gene *cpz9* (encoding for an AraC-type regulator) with an intergenic area of 103 bp. Surprisingly, the RNAseq data do not indicate that *cpz9* is regulated by its own promoter, but rather shows that *cpz9* is transcribed as an operon together with genes *cpz6*-*cpz8*. Furthermore, the data reveal that, as expected, the promoter upstream of *cpz10* (intergenic region of 149 bp length) facilitates transcription of the biosynthetic genes *cpz10*-*cpz23* (possibly up to- *cpz32*) as one long polycistronic mRNA. Transcription of *cpz10* and *cpz11* reaches the highest read counts within the whole BGC, with values up to 115000, which then gradually decrease with progressing length of the

transcript. Therefore, regulators binding to this promoter could directly alter transcription of the biosynthetic operon and thus production of caprazamycin-aglyca. The last presumed promoter region of the gene cluster is located upstream of *cpz24* (intergenic region of 280 bp length), which represents a gene with unknown function. Indeed, RNAseq indicates a minor additive transcriptional effect in *cpz24* (read counts increasing up to 30000) by facilitating re-binding of the RNA-polymerase at this position (Figure 7 and Figure 8).

Although the region upstream of *cpz9* was not confirmed as a (strong) promoter, this intergenic region was included in the subsequent DACA studies as well, because Cpz9 constitutes the only cluster-situated regulator of the gene cluster and the intergenic area is large enough to be bound be regulatory proteins. The promoters/intergenic regions of the caprazamycin BGC selected for the DACAs were assigned P*cpz6*, P*cpz9*, P*cpz10* and P*cpz24*.

Lastly, when comparing the transcriptional rates of the caprazamycin BGC from one time point (2 days) to the other (4 days), the transcription strength shows different trends between the biological replicates. While for replicate 1 and 2, expression of the gene cluster is higher after two days of growth compared to the four day sample, it increases from day 2 to 4 in replicate 3.

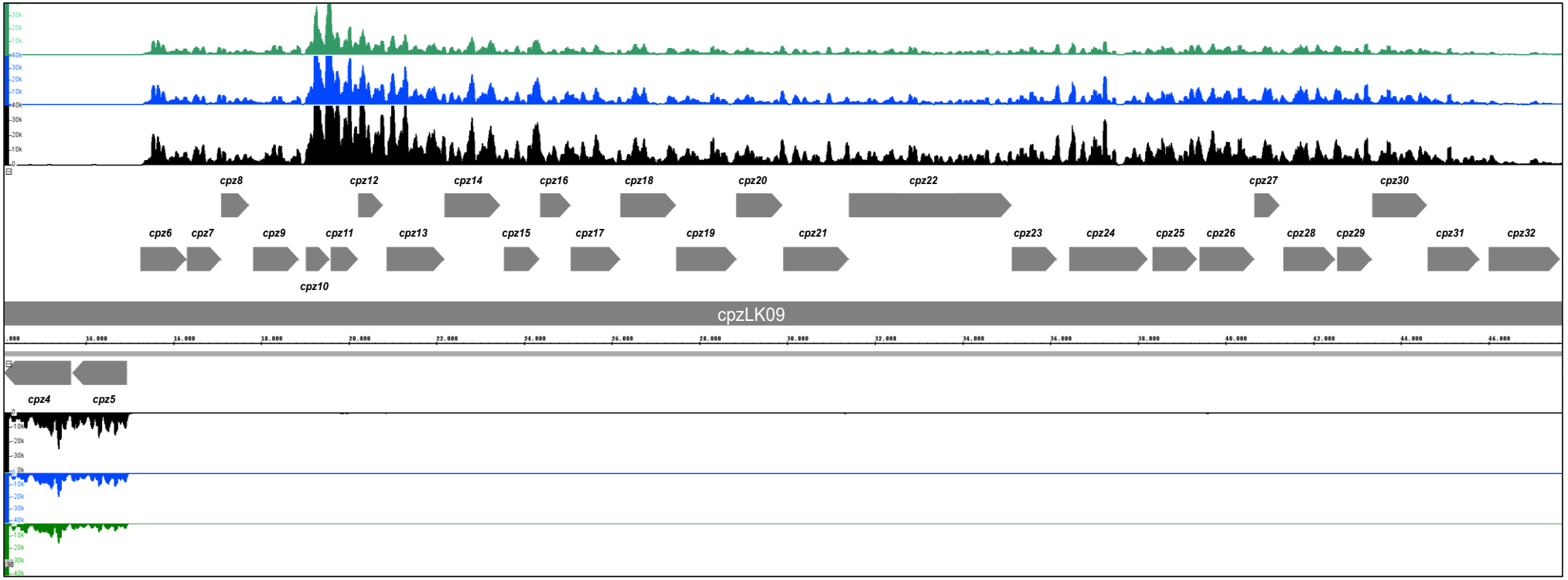


Figure 6: Transcription of the caprazamycin gene cluster in *S. coelicolor* M512/cpzLK09 after two days of cultivation. The transcriptional pattern of three independent biological replicates is shown. Replicate one is illustrated in black and has a remaining rRNA content of 8 %, replicate two in blue has a rRNA content of 8 % and replicate three in green an rRNA content of 28 %. The genes of the cluster, depicted as arrows, are divided into sense and anti-sense strand and respective coverage plots are mapped against them. To allow a comparative assessment of the data, a threshold value of 40000 read counts is applied to all replicates.

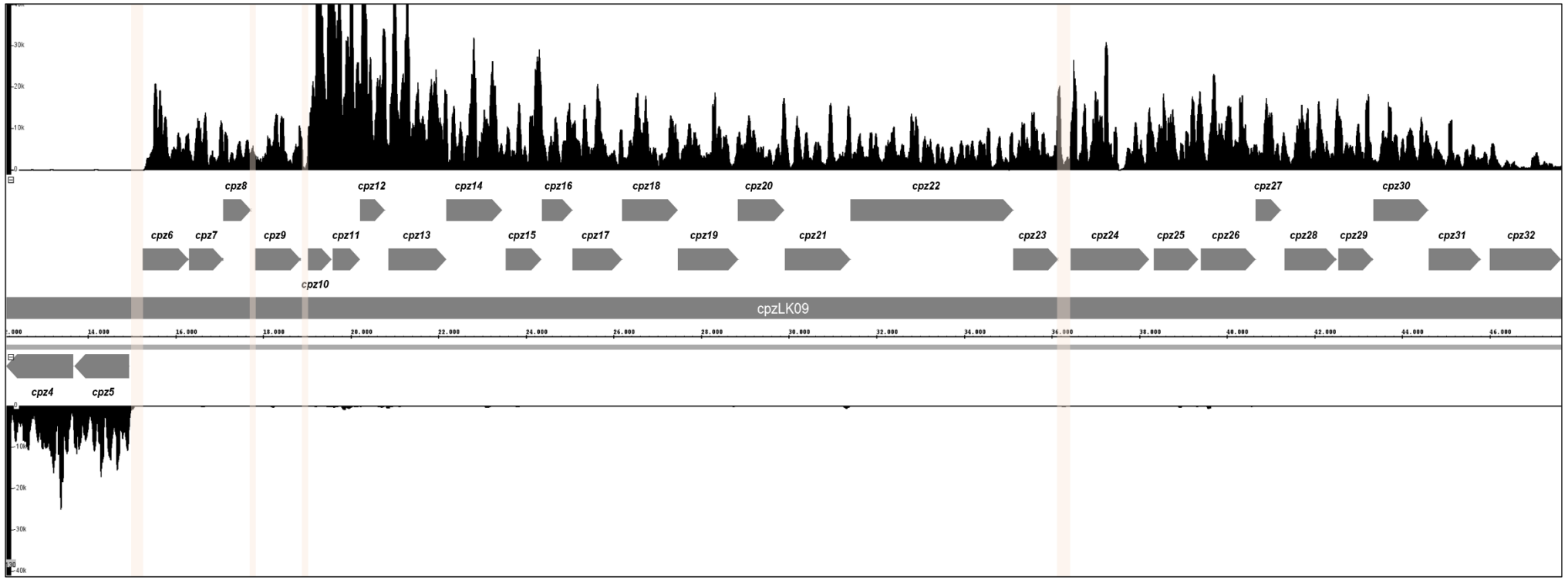


Figure 7: Transcription of the caprazamycin gene cluster in *S. coelicolor* M512/cpzLK09 (replicate one) after two days of cultivation. The transcriptional pattern of replicate one, which exhibits the lowest amount of remaining rRNA (8 %) is shown. The genes of the cluster, depicted as arrows, are divided into sense and anti-sense strand and respective coverage plots are mapped against them. A threshold value of 40000 read counts is applied. Intergenic regions selected for DNA affinity capturing assays are highlighted with light red bars (P*cpz6*, P*cpz9*, P*cpz10*, P*cpz24*).

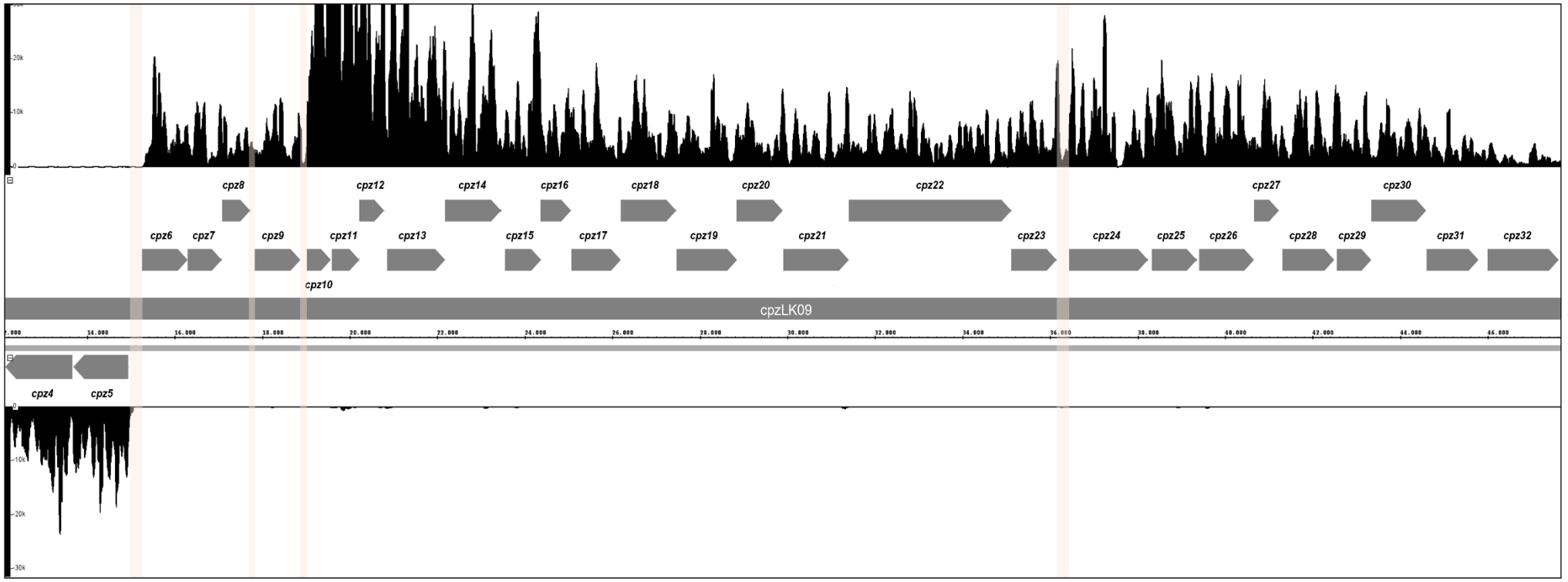


Figure 8: Transcription of the caprazamycin gene cluster in *S. coelicolor* M512/cpzLK09 (replicate three) after four days of cultivation. The transcriptional pattern of replicate three, which exhibits the lowest amount of remaining rRNA (19 %) is shown. The genes of the cluster, depicted as arrows, are divided into sense and anti-sense strand and respective coverage plots are mapped against them. A threshold value of 30000 read counts is applied. Intergenic regions selected for DNA affinity capturing assays are highlighted with light red bars (P*cpz6*, P*cpz9*, P*cpz10*, P*cpz24*).

4.1.2.2 The liposidomycin BGC

Due to the highly similar organization of the caprazamycin and liposidomycin gene cluster and their intergenic regions being located at the same positions, promoters of the liposidmycin BGC are expected at the same positions as described for the caprazamycin gene cluster (see 2.3).

Figure 9 and Figure S 3 show, that the transcriptional pattern of the liposidomycin gene cluster in *S. coelicolor* M512/lpmLK01 is extremely similar between the biological replicates both after 2 and 4 days of cultivation. In all three replicates, the transcriptional rates of the cluster are considerably higher after 4 days than after 2 days of growth, indicating that biosynthesis of liposidomycins starts later than of caprazamycins. For reasons of better clarity, RNAseq data of the replicate with the lowest rRNA content are depicted for both time points in Figure 10 and Figure 11, respectively.

The intergenic region (431 bp) between genes *lpmC* and *lpmD* constitutes a bidirectional promoter which is also reflected by the RNAseq data. LpmD, like its corresponding counterpart Cpz6, is a type III polyketide synthase whereas gene *lpmC* is of unknown function and shows no homology to any gene of the caprazamycin gene cluster. Noteworthy, in relation to the other biosynthetic genes, the first genes of the liposidomycin BGC, especially *lpmB*, is strongly expressed (read counts of up to 31000). The gene *lpmB* encodes for a sulfotransferase and is therefore responsible for the high proportion of sulphated liposidomycins produced by *S. coelicolor* M512/lpmLK01. In comparison, the homologous gene in the caprazamycin BGC, *cpz4*, is only expressed with read counts up to approximately 25000. Although the total caprazamycin cluster expression is much higher than expression of the liposidomycin BGC in regard to total read counts. Thus, the different distribution of sulphated caprazamycins and liposidomycins is mirrored on the transcriptional level.

Based on the analysis of intergenic regions in the liposidomycin BGC, the next promoter would be expected upstream of the regulatory gene *lpmG* (intergenic region of 259 bp in length) which is homologous to *cpz9*. Similar to *cpz9*, the RNAseq data reveal that *lpmG* is rather transcribed as a continuous mRNA with *lpmD*-*lpmF* than from its own promoter. Moreover, analogously to the caprazamycin BGC, the biosynthetic genes *lpmH*-*lpmU* (and even *lpmH*-*lpmY*) are transcribed as one polycistronic mRNA. The intergenic region upstream of *lpmH* is 128 bp in length. Right at the start of transcription, genes *lpmH* and *lpmI*, encoding for a hydroxylase and a N-methyltransferase, respectively, show the highest expression levels (read counts up to 38000) within the whole gene cluster. Lastly, a small increase in read counts is

recorded in gene *lpmV*, therefore the intergenic area upstream of *lpmV* (286 bp in length) could to some extent facilitate re-binding of the RNA polymerase. LpmV is the homolog to Cpz24 and is of unknown function.

Based on these RNAseq results the intergenic regions/promoters of the liposidomycin BGC were selected for the DACA studies and assigned as P*lpmD*, P*lpmG*, P*lpmH* and P*lpmV*.

On nucleotide level, the corresponding promoters from the caprazamycin and liposidomycin gene cluster exhibit 59 %, 52 %, 58 % and 72 % sequence identity for P*cpz6*/P*lpmD*, P*cpz9*/P*lpmG*, P*cpz10*/P*lpmH* and P*cpz24*/P*lpmV*, respectively. By applying their DNA sequences in DACAs, proteins from the host *S. coelicolor* M512 being able to bind to either one or both heterologously expressed gene clusters, could be identified.

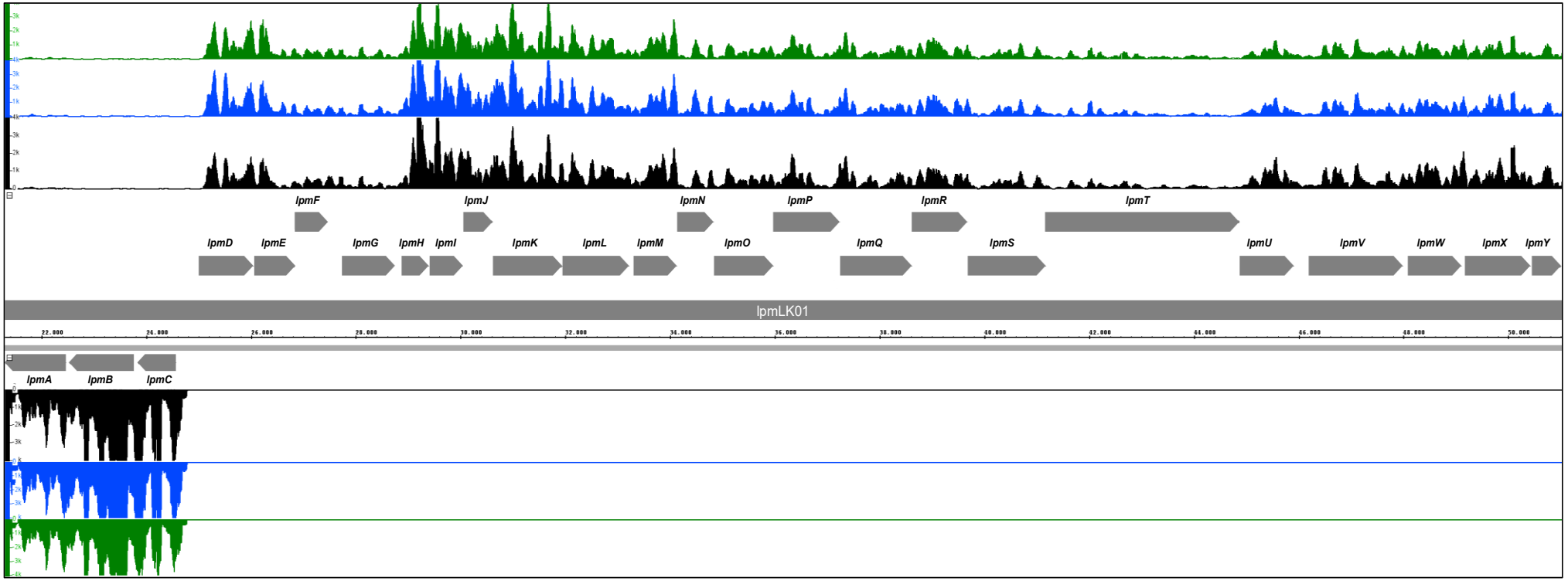


Figure 9: Transcription of the liposidomycin gene cluster in *S. coelicolor* M512/lpmLK01 after two days of cultivation. The transcriptional pattern of three independent biological replicates is shown. Replicate one is illustrated in black and has a remaining rRNA content of 14 %, replicate two in blue has a rRNA content of 11 % and replicate three in green an rRNA content of 13 %. The genes of the cluster, depicted as arrows, are divided into sense and anti-sense strand and respective coverage plots are mapped against them. To allow a comparative assessment of the data, a threshold value of 4000 read counts is applied to all replicates.

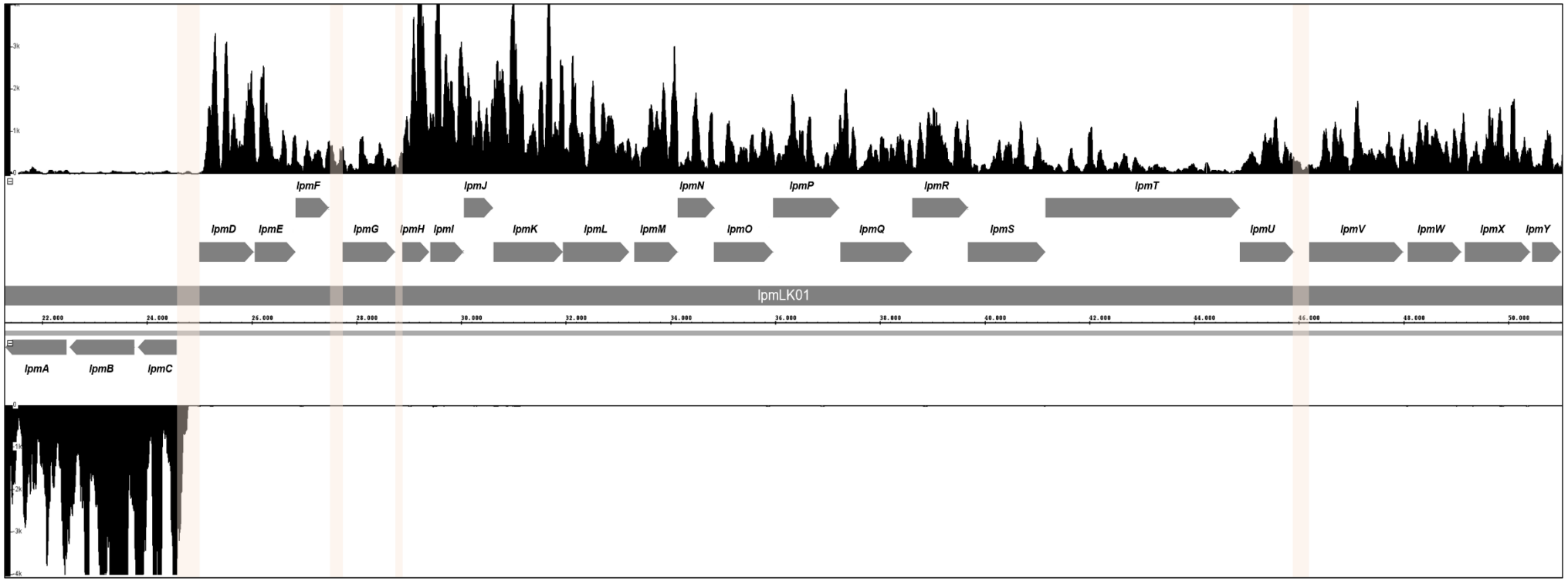


Figure 10: Transcription of the liposidomycin gene cluster in *S. coelicolor* M512/lpmLK01 (replicate two) after two days of cultivation. The transcriptional pattern of replicate two, which exhibits the lowest amount of remaining rRNA (11 %) is shown. The genes of the cluster, depicted as arrows, are divided into sense and anti-sense strand and respective coverage plots are mapped against them. A threshold value of 4000 read counts is applied. Intergenic regions selected for DNA affinity capturing assays are highlighted with light red bars (P*lpmD*, P*lpmG*, P*lpmH*, P*lpmV*).

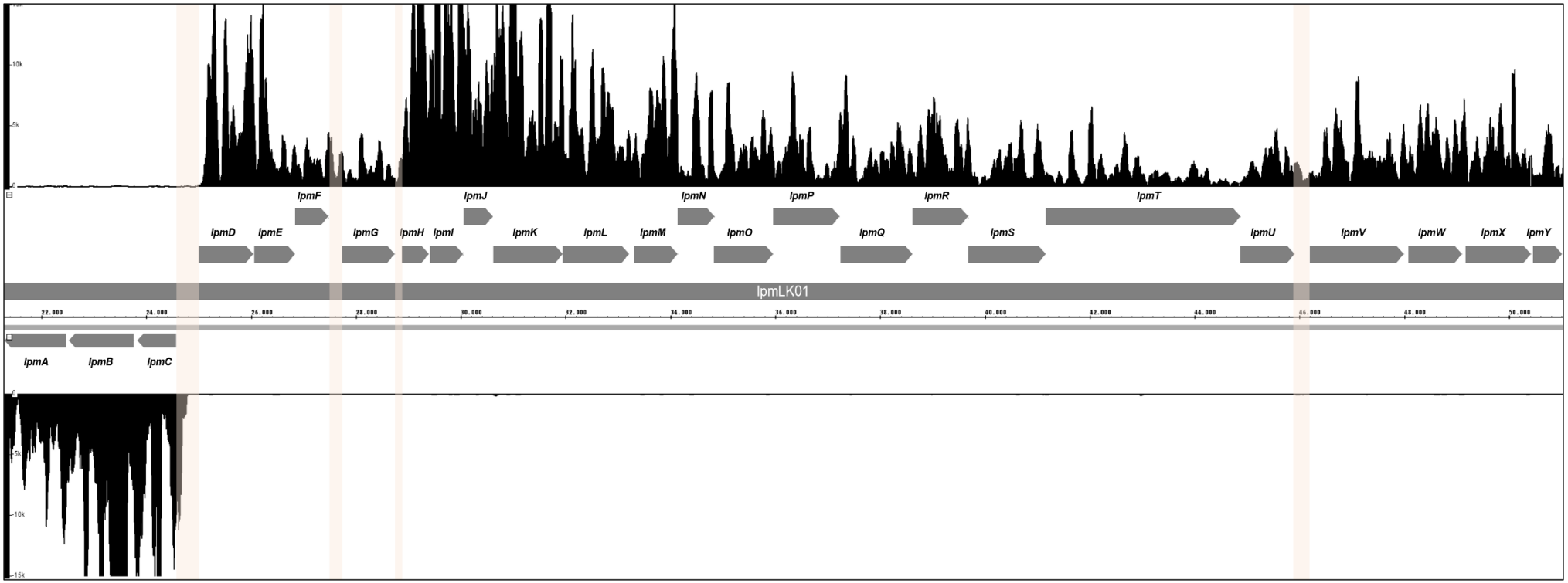


Figure 11: Transcription of the liposidomycin gene cluster in *S. coelicolor* M512/lpmLK01 (replicate two) after four days of cultivation. The transcriptional pattern of replicate two, which exhibits the lowest amount of remaining rRNA (22 %) is shown. The genes of the cluster, depicted as arrows, are divided into sense and anti-sense strand and respective coverage plots are mapped against them. A threshold value of 15000 read counts is applied. Intergenic regions selected for DNA affinity capturing assays are highlighted with light red bars (P*lpmD*, P*lpmG*, P*lpmH*, P*lpmV*).

4.2 Fermentation of *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01

In order to obtain sufficient protein amounts for the DNA affinity capturing assays (DACAs), the heterologous strains *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01 were cultivated in a 20-L fermenter, respectively. The total cultivation time was 90 h and two time points were determined: one before and one right at maximum production of the liponucleosides, at which the protein samples should be applied in the DACA. By selection of two time points, proteins either activating or repressing the heterologously expressed BGCs should be detected. Samples were collected at nine time points (18 h, 24 h, 36 h, 42 h, 48 h, 54 h, 60 h, 72 h, 90 h after inoculation) and the sampling volume was adjusted to cell density (smaller volume at later time points).

Most of the sample was processed and stored for future protein extraction, but the residual sample volume was used to determine the cell dry weight and liponucleoside production levels (see 3.2.3.2). Furthermore, RNA was isolated from the samples and RT-qPCR was applied to detect transcription rates of the first biosynthetic genes of the caprazamycin and liposidomycin operon (see 3.2.2.9 and 3.2.2.10). Figure 12 shows cell dry weight, liponucleoside production and transcription of the biosynthetic genes, *cpz10* and *lpmH*, during growth in the fermenters for both heterologous strains. Regarding *S. coelicolor* M512/cpzLK09, the strain enters exponential growth after 18 hours and reaches a maximum cell dry weight of 8-9 mg/mL 54 hours after inoculation.This value remains constant until the end of the cultivation. Simultaneously, caprazamycin-aglycon levels increase gradually at each sampling time point, first being detectable 36 h after inoculation (mean value of both technical duplicates 6,6 mg/L) and reaching a maximum concentration of 65 mg/L (Figure 12 A). This value is in accordance to previously reported caprazamycin-aglycon titers (59 mg/L) produced by *S. coelicolor* M512/cpzLK09 (Flinspach *et al*., 2010).

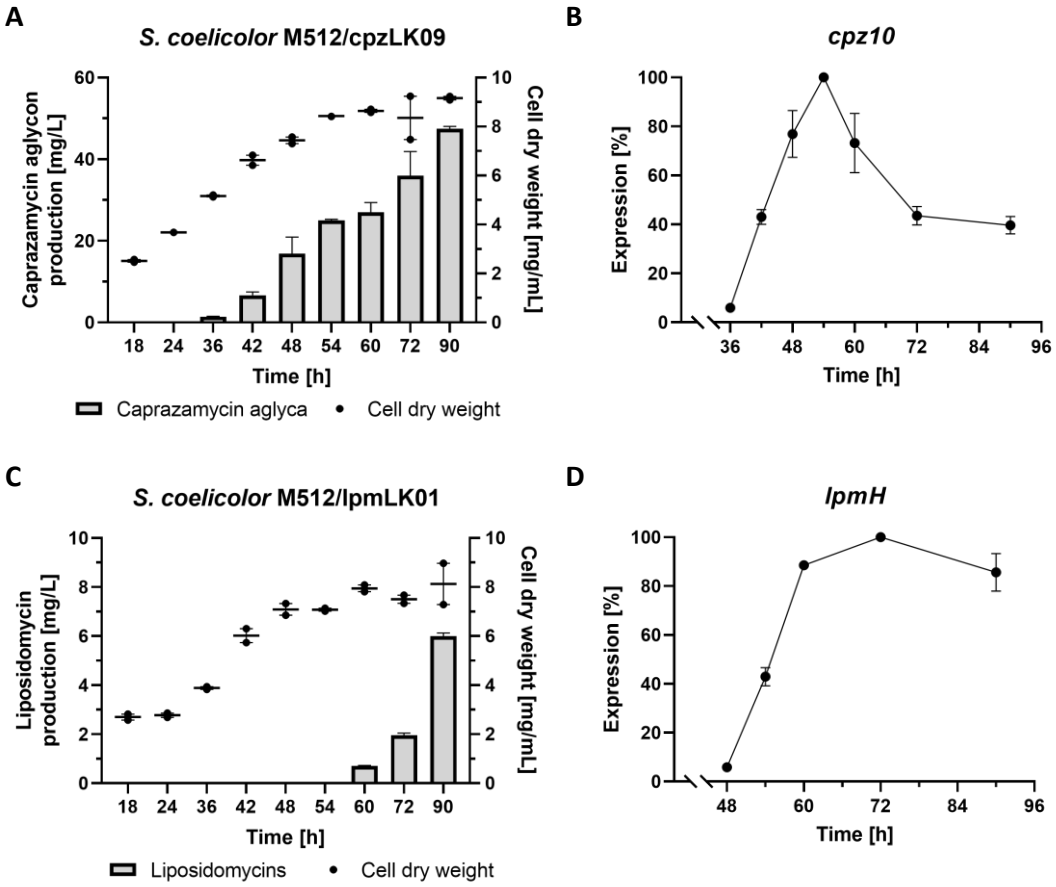


Figure 12: 20-L fermentation of *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01. Cell dry weight and liponucleoside production of *S. coelicolor* M512/cpzLK09 (A) and *S. coelicolor* M512/lpmLK01 (C) are depicted. Transcription rates were determined by RT-qPCR for the biosynthetic genes *cpz10* (B) and *lpmH* (D) during fermentation. Highest expression was set to 100 %, respectively. All error bars indicate the range of technical duplicates.

The growth curve of *S. coelicolor* M512/lpmLK01 is similar to the one of *S. coelicolor* M512/cpzLK09 but it exhibits a lag-phase until 36 h after inoculation. The following exponential growth phase lasts until 60 h after inoculation and a maximum cell dry weight of 8 mg/mL is reached at the end of fermentation (time point 90 h). As expected, liposidomycin titers in the growth medium of *S. coelicolor* M512/lpmLK01 are considerably lower compared to the amounts of caprazamycin-aglyca. During growth in Erlenmeyer flasks under standard conditions, it was observed previously in this study, that liposidomycin production rates were in average 6-8 times lower than those of caprazamycin-aglycons. This is also reflected on the transcriptional level of both BGCs as revealed by RNAseq (see 4.1.2). In the fermenter, production of liposidomycin was only detectable in considerable amounts 60 h after inoculation (0,7 mg/L) and increased gradually to a maximum titer of 6 mg/L at the end of

fermentation, 90 hours after inoculation (Figure 12 C). Therefore, total production rates of caprazamycin-aglyca were 10-fold higher than those of liposidomycins during fermentation, which is an even stronger difference than during cultivation in flasks. Furthermore, it became apparent that, under the conditions of fermentation, biosynthesis of caprazamycin starts 24 h earlier than liposidomycin production which matches the transcription rates of the gene clusters as determined by RNAseq (see 4.1). This finding had to be considered for the selection of time points for the DACAs. Overall, time points for the DACA were not only determined based on growth and production values but to a large extent on transcription data of the biosynthetic genes *cpz10* and *lpmH* during fermentation. As it can be seen in Figure 12 B and D, both genes exhibit clear expression maxima at 54 h and 72 h after inoculation, respectively. These values were set to 100 % to allow a relative correlation of the lower values. The expression pattern of *lpmH* shows high values of over 80 % at time points 60 h, 72 h and 90 h which spans the time frame at which production is detectable. In the case of caprazamycin, production is measured when relative transcription rates of *cpz10* are still low in regard to the maximum value, indicating strong absolute expression of the biosynthetic genes in the caprazamycin BGC. Correlated to the maximum expression values, transcription of *cpz10* and *lpmH* was only 6 % at time point 36 h and 5,9 % at time point 48 h, respectively. Thus, these samples represent low expression of the heterologous BGCs. Taken together, protein samples from time points 36 h (low gene expression) and 54 h (highest gene expression) were applied in the DACA with *S. coelicolor* M512/cpzLK09 and samples from time points 48 h (low gene expression) and 72 h (highest gene expression) in the DACA with *S. coelicolor* M512/lpmLK01.

4.3 DNA-affinity capturing assay

4.3.1 Preparation of biotinylated promoter sequences as bait molecules

As described in section 4.1.2, corresponding promoter regions P*cpz6*/P*lpmD*, P*cpz9*/P*lpmG*, P*cpz10*/P*lpmH* and P*cpz24*/P*lpmV* were selected as bait DNA sequences in the pull-down assays. Furthermore, the promoter region of the vegetative sigma-factor gene *hrdB* (assigned as P*hrdB*) served as a negative control to evaluate the specificity of DNA-binding of the trapped proteins. P*hrdB* is a promoter region commonly used as a negative control in DNA binding studies (Bekiesch *et al*., 2016b; Li *et al*., 2023), however, the gene is known to be regulated to some extend as well (Hesketh *et al*., 2007). 120 µg of biotinylated DNA were amplified for each

promoter fragment in two-step PCRs (see 3.2.3.1) and were applied in each DACA. All four promoter regions from the caprazamycin cluster and the P*hrdB* sequence were exposed to protein extract from *S. coelicolor* M512/cpzLK09 (146 mg total protein per promoter) and promoters from the liposidomycin cluster to the protein extract from *S. coelicolor* M512/lpmLK01 (129 mg total protein per promoter). The general workflow of the assay is explained in detail in 3.2.3.5 and is depicted in Figure 5 (page 21).

4.3.2 Label-free and semi-quantitative identification of DNA-bound proteins by mass-spectrometry

Eluted proteins trapped during the DACA were kindly processed and identified via mass-spectrometry in cooperation with Dr. Mirita Franz-Wachtel at the Proteome Center (University of Tübingen, working group of Prof. Boris Macek).
A total number of 2214 different proteins were identified on all tested DNA fragments. 1906 and 1540 proteins were found in the single DACAs with *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01, respectively. 674 proteins were exclusively identified at the caprazamycin BGC (and the negative control P*hrdB* bound by proteins from *S. coelicolor* M512/cpzLK09) while it was 308 proteins at the promoters of the liposidomycin cluster, leaving 1232 proteins binding at least once in both clusters. The vast majority of proteins bound unspecifically to more than one promoter of each cluster, accounting for 83 % (1575 proteins) and 85 % (1305 proteins) in the caprazamycin and liposidomycin BGC, respectively (Figure 13, page 88). Protein abundances were measured as intensities which ranged between 1,67E+07 and 1,2E+12. These values do not allow any statement about physiological abundance of the protein in the cell or affinity to the bound DNA segment. A very even distribution of proteins identified at the individual promoter regions was observed within both data sets (Table 20). The aquired data were systematically scanned for matching certain criteria.

Table 20: Amount of proteins detected on promoter regions in the DACAs. The numbers indicate proteins that bound the respective promoter region either at one or at both time points.

***S. coelicolor* M512/cpzLK09**		***S. coelicolor* M512/lpmLK01**	
Promoter region:	Amount of proteins:	Promoter region:	Amount of proteins:
P*cpz6*	1431	P*lpmD*	1028
P*cpz9*	1437	P*lpmG*	1032
P*cpz10*	1436	P*lpmH*	1106
P*cpz24*	1553	P*lpmV*	981
P*hrdB*	1437	P*hrdB*	1319

4.3.2.1 Binding of known global regulators to the caprazamycin and liposidomycin BGC

Firstly, well-known global regulators such as AdpA, NdgR, SlbR, Rok7B7 and AtrA were identified among the binding proteins. Most of these proteins were detected in high amounts reaching intensities of up to 3,42E+11 (in case of the AraC regulator AdpA). These global regulators were found to bind unspecifically to every DNA sequence. The only exception being AtrA, which bound to all tested promoters of the liposidomycin BGC, but only to P*cpz9* within the caprazamycin BGC (and the negative control P*hrdB*). While the presence of these proteins validated binding of known regulators to the promoter sequences, they were excluded for further studies due to their highly unspecific binding properties.

4.3.2.2 Cluster-situated regulators Cpz9 and LpmG

Both aquired data-sets, from the DACA of *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01, were screened for the presence of the cluster-situated regulators (CSRs) Cpz9 and LpmG, respectively. Cpz9 was found to bind to all tested promoters of the caprazamycin cluster and also to the negative control. However, P*hrdB* was only bound at the later time point (54 h) and Cpz9 intensity in this sample was lowest. All cluster-internal promoters were bound in moderate intensities between 3,88E+08 and 2,73E+09 at both time points (Table 21, page 91). The ubiquitous binding of Cpz9 to all tested DNA sequences makes it difficult to speculate on its regulatory mechanism for caprazamycin biosynthesis or its binding site within the cluster.

Strikingly, the homologous regulator from the liposidomycin cluster, LpmG, was not identified in any of the samples. Based on the high amino-acid similarity between Cpz9 and LpmG (69 % identity, 77 % similarity by BLASTp sequence alignment), a targeted search for LpmG in the collected mass-spectrometry data was conducted, but yielded no hit (kindly carried out by Dr. Mirita Franz-Wachtel, Proteome Center, University of Tübingen).

4.3.2.3 Proteins binding specifically within the caprazamycin or liposidomycin BGC

For the subsequent analyses, binding was strictly defined as specific if only one of all tested promoters was bound. If not otherwise stated, the term “specific” is only used based on this definition. Of the 2214 total proteins, this criterion was only met by a

few proteins and reduced the absolute numbers of protein candidates drastically (Figure 13).

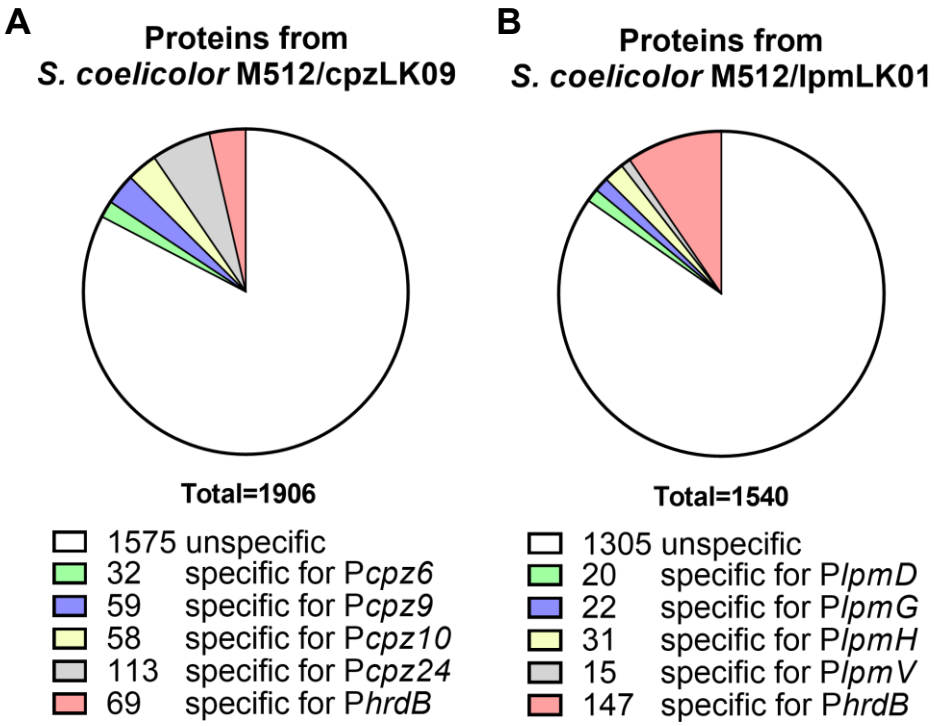


Figure 13: Distribution and numbers of proteins binding unspecifically and specifically in the DACAs. Specific binding was defined as binding of proteins solely to one of the tested promoters in the DACA with *S. coelicolor* M512/cpzLK09 (A) or *S. coelicolor* M512/lpmLK01 (B).

Within the caprazamycin BGC, most proteins were found to bind specifically to P*cpz24* (113 i.e. 5,9 %), whereas fewest proteins were trapped by P*cpz6* (32 i.e. 1,7 %). The residual promoter regions upstream of *cpz9* and *cpz10* were specifically bound by 59 (3,1 %) and 58 (3,0 %) proteins, respectively. Matching the generally lower number of proteins identified at the liposidomycin cluster, the numbers of proteins binding specifically to one promoter within this cluster were also smaller. Here, most proteins bound specifically to P*lpmH* (31 i.e. 2,0 %) and the lowest number was found to bind to P*lpmV* (15 i.e. 1,0 %). At P*lpmD* and P*lpmG* 20 (1,3 %) and 22 (1,4 %) proteins were trapped, respectively (Figure 13).

Of the specifically binding proteins, the majority of them are annotated as (putative) regulatory proteins, but enzymes (such as kinases) and hypothetical proteins are also among them. The reduced numbers of proteins binding specifically within the caprazamycin or liposidomycin gene cluster allowed to screen all candidates annotated as regulators or hypothetical proteins (maximum 50 proteins per promoter,

if applicable) in more detail and subject them to BLAST analysis and, for some of them, to literature research.

In the course of this screening, protein Sco7543 was discovered as an interesting protein candidate for further characterization. Sco7543 bound specifically to P*lpmH* at both tested time points and was found at an average intensity of 3,2E+08 (Table 21, page 91). BLAST analysis revealed Sco7543 to be a ROK- (repressor, open reading frame, kinase) family type regulator which is a regulator family known for its involvement in both primary and secondary metabolism. Moreover, a proposed binding sequence 5´ TGTGNNNNNNNNCACA 3´ was already published for this protein (Iqbal *et al*., 2012) and screening of P*lpmH* revealed to contain an almost perfectly matching sequence (5´ TGTGNNNNNNNNCACC 3´). The presence of the putative binding site in P*lpmH* and specific binding of the protein to this promoter increased the possibility of this binding to cause a transcriptional impact.

4.3.2.4 Proteins binding to corresponding promoters in the caprazamycin and liposidomycin BGC

The number of proteins binding specifically to the corresponding promoters of both clusters, is even more restricted. Of the total number of 2214 proteins only three meet this criterion: A MarR regulator (Sco2987) bound to the corresponding promoter pair P*cpz6*/P*lpmD* and two putative TetR-family regulators (Sco4385 and Sco5956) to P*cpz10*/P*lpmH* (Table 21).

A literature search revealed that Sco2987, designated as OhrR, was already described as a regulator of the *ohrA* gene involved in organic hydroperoxide resistance in *S. coelicolor*. The genes *ohrA* and *ohrR* share a divergent promoter. In dependence of its oxidation state Sco2987 either acts as a repressor of *ohrA* and its own gene or, in an oxidized form, acts as an activator of its own transcription (Oh *et al*., 2007). This is a commonly known regulatory pattern of multiple-antibiotic-resistance-regulators (MarR) which often control genes encoded divergently from their own gene by direct DNA binding and thereby regulate interaction of the RNA polymerase with the respective promoter (Deochand and Grove, 2017). In the present DACA, Sco2987 bound to P*cpz6* only at time point 54 h and to P*lpmD* at both time points (48 h and 72 h).

The TetR-family regulators (TFRs) Sco4385 and Sco5956 are both completely uncharacterized proteins at the beginning of this study. Sco4385 was found to bind P*cpz10* and P*lpmH* at both time points, respectively (at approximately 10-fold higher intensities at the promoter of the liposidomycin gene cluster, than at the caprazamycin

promoter region). In contrast, Sco5956 was only detected to bind at P*cpz10* and P*lpmH* at the earlier time points, respectively, with intensity values being twice as high at the promoter of the liposidomycin cluster than of the caprazamycin BGC (Table 21).

4.3.3 Selection of putative regulatory proteins for further characterization

On one hand, protein binding patterns determined in the DACAs laid the basis for the selection of putative regulatory proteins, that should be further characterized by overexpression and deletion of their corresponding genes. Based on this approach the putative ROK regulator Sco7543 was selected due to its specific binding to P*lpmH* and the presence of a matching proposed binding site and regulators Sco2987, Sco4385 and Sco5956 due to their specific interaction with the corresponding promoter regions of both clusters.

On the other hand, the DACA data set could also be exploited to search for proteins that were already known to play a regulatory role in caprazamycin and/or liposidomycin biosynthesis. Based on the work of Dr. Daniel Bär in his dissertation, the leucine/isovalerate utilization (*liu*-) pathway was recently discovered as an important precursor supply pathway of the 3-methylglutaryl (3-MG) moiety in caprazamycins (Bär *et al*., 2022). Literature research aiming to identify regulators of possible precursor pathways revealed the *liu*-pathway to be strongly up-regulated by the cAMP-receptor protein (CRP), Sco3571 (Gao *et al*., 2012). Scanning the DACA data-set for Sco3571 showed that it indeed bound to both clusters, however, in an unspecific manner (Table 21). Nevertheless, due to the known involvement in regulation of the precursor supply, Sco3571 was included in the study and thus five proteins in total were selected for subsequent characterization: Sco7543, Sco2987, Sco4385, Sco5956 and Sco3571.

Table 21: Binding pattern and intensities of selected proteins detected by DNA-affinity capturing assay (DACA). Listed are proteins binding to the corresponding promoters of both clusters, ROK family regulator Sco7543 binding specifically to P*lpmH*, CRP regulator Sco3571 involved in precursor supply, the CSR of the caprazamycin BGC, Cpz9, and examples of global regulators of secondary metabolism in *Streptomyces*.

Protein name/Sco No.	Regulator family	P*cpz6*		P*cpz9*		P*cpz10*		P*cpz24*		P*hrdB* (Cpz DACA)		P*lpmD*		P*lpmG*		P*lpmH*		P*lpmV*		P*hrdB* (Lpm DACA)	
		36 h	**54 h**	**36 h**	**54 h**	**36 h**	**54 h**	**36 h**	**54 h**	**36 h**	**54 h**	**48 h**	**72 h**	**48 h**	**72 h**	**48 h**	**72 h**	**48 h**	**72 h**	**48 h**	**72 h**
Sco2987	MarR		1,35 E+08									3,41 E+08	7,52 E+08								
Sco4385	TetR					4,87 E+08	1,48 E+08									5,77 E+09	4,66 E+09				
Sco5956	TetR					1,02 E+08										2,39 E+08					
Sco7543	ROK/CRP															4,88 E+08	1,53 E+08				
Sco3571	CRP	3,86 E+09	7,61 E+09	4,74 E+09	7,46 E+09	4,55 E+09	7,07 E+09	5,54 E+09	9,23 E+09	4,42 E+09	8,22 E+09	4,05 E+09	1,52 E+09	4,45 E+09	1,36 E+09	3,86 E+09	1,45 E+09	4,49 E+09	1,30 E+09	6,35 E+09	2,58 E+09
Cpz9	AraC	9.68 E+08	2.73 E+09	3.88 E+08	4.00 E+08	5.39 E+08	1.24 E+09	7.00 E+08	6.01 E+08		3.07 E+08	Not applicable									
AdpA (Sco2792)	AraC	2.50 E+11	2.76 E+10	1.09 E+11	1.18 E+10	1.39 E+11	2.16 E+10	1.03 E+11	8.49 E+09	3.42 E+11	3.53 E+10	2.08 E+11	8.39 E+10	1.31 E+11	6.47 E+10	1.00 E+11	8.58 E+10	2.25 E+10	1.46 E+10	1.83 E+11	4.89 E+10
AtrA (Sco4118)	TetR			3.11 E+08	2.68 E+08					5.78 E+09	3.66 E+09	2.64 E+09	1.77 E+09	9.78 E+08	9.97 E+08	3.91 E+09	2.42 E+09	7.61 E+07		2.54 E+09	1.55 E+09
NdgR (Sco5552)	IclR	1.45 E+11	4.35 E+10	7.71 E+10	1.69 E+10	1.15 E+11	4.07 E+10	7.68 E+10	1.26 E+10	1.25 E+11	2.33 E+10	3.68 E+10	1.40 E+10	4.21 E+10	1.58 E+10	3.75 E+10	1.44 E+10	4.75 E+10	2.90 E+10	3.46 E+09	1.11 E+09
Rok7B7 (Sco6008)	ROK	1.71 E+09	6.42 E+08	5.89 E+08	1.08 E+08	2.31 E+08	2.20 E+08	2.00 E+08	9.80 E+07			7.87 E+08	1.06 E+09	5.70 E+08	5.36 E+08	6.85 E+08	7.23 E+08			1.86 E+08	
SlbR (Sco0608)	γ-butyro-lactone-binding regulator	1.55 E+10		1.75 E+10		1.56 E+10		7.61 E+08		3.09 E+08		2.40 E+10		1.62 E+10		1.60 E+10		2.16 E+10		1.09 E+09	

4.4 Overexpression of candidate regulatory genes

4.4.1 Overexpression of *sco7543* in *S. coelicolor* M512/lpmLK01

In the DACAs, the ROK-family regulator Sco7543 only bound to P*lpmH*, which contains an almost perfectly matching recognition site (TGTG$(N)_8$CACC) as proposed for this protein (TGTG$(N)_8$CACA) by Iqbal *et al*., (2012). A genome-wide search for alternative binding sites in *S. coelicolor* M512 revealed 17 putative binding sites for Sco7543. In order to closer characterize its interaction with the liposidomycin BGC, gene *sco7543* was amplified and cloned into the replicative vector pUWL-apra-oriT (pUWL) under the transcriptional control of the strong, constitutive promoter *ermE** (P*ermE**), resulting in plasmid pSW10. The empty vector pUWL and pSW10 were introduced into the heterologous strain *S. coelicolor* M512/lpmLK01 via conjugation yielding *S. coelicolor* M512/lpmLK01/pUWL and *S. coelicolor* M512/lpmLK01/pSW10. Three independent clones, respectively, were cultivated to determine liposidomycin production. Extracts of the culture supernatant were measured by LC-MS and quantified by integration of the extracted ion chromatogram (EIC) peaks. The resulting values were normalized to the cell dry weight of the respective production culture. Figure 14 shows production levels of *S. coelicolor* M512/lpmLK01/pSW10 referenced to the control strain carrying the empty vector pUWL. Overexpression of *sco7543* resulted in a slight decrease in liposidomycin production to 77 % (in average) compared to the control strain. However, this impact is insignificant and therefore no meaningful change in production could be stated for overexpression of *sco7543*.

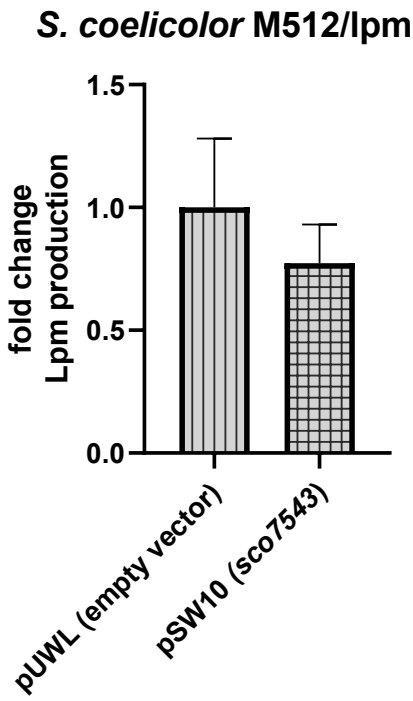


Figure 14: Mean liposidomycin (Lpm) production of *S. coelicolor* M512/lpmLK01 carrying control vector pUWL-apra-oriT or pSW10. Production of the pSW10 (encoding for ROK- regulator Sco7543) containing strain is referenced to the empty vector control strain. Error bars indicate production range of three independent biological replicates.

4.4.2 Overexpression of *sco2987*, *sco3571*, *sco4385* and *sco5956* in *S. coelicolor* M512/cpzLK09

Furthermore, selected genes *sco2987*, *sco3571*, *sco4385* and *sco5956* were individually cloned into the pUWL-apra-oriT plasmid, resulting in plasmids pSW4, pSW5, pSW6 and pSW7, respectively. Due to generally higher production levels, the impact of these overexpression plasmids on the caprazamycin BGC should be investigated first. Therefore, overexpression strains *S. coelicolor* M512/cpzLK09/pSW4, *S. coelicolor* M512/cpzLK09/pSW5, *S. coelicolor* M512/cpzLK09/pSW6 and *S. coelicolor* M512/cpzLK09/pSW7 were generated by biparental conjugation. Caprazamycin-aglyca production levels were determined, normalized to cell dry weight and compared to the empty vector control strain *S. coelicolor* M512/cpzLK09/pUWL. Figure 15 sums up the production levels of all overexpression mutants, related to the control strain.

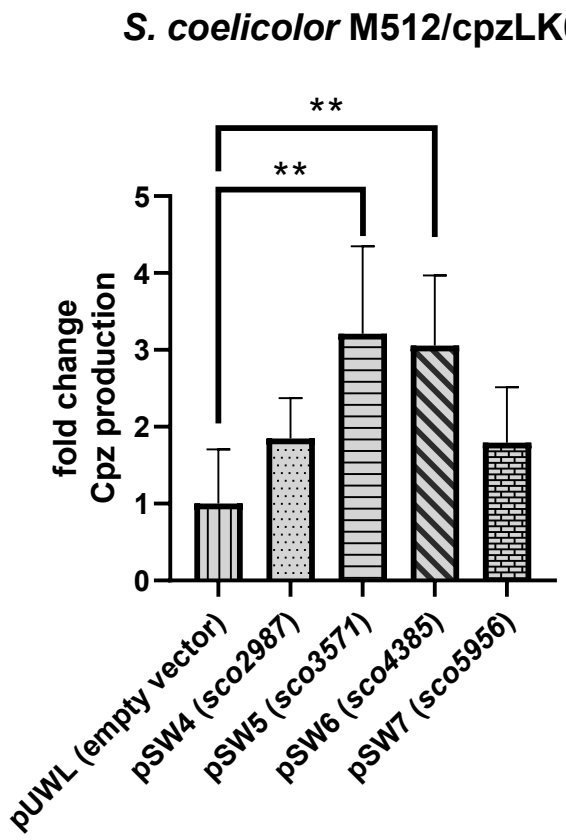


Figure 15: Mean caprazamycin-aglycon (Cpz) production of *S. coelicolor* M512/cpzLK09 strains containing vectors pSW4, pSW5, pSW6, pSW7 or the empty vector pUWL-apra-oriT. pSW4 (encoding for MarR regulator Sco2987), pSW5 (encoding for CRP regulator Sco3571), pSW6 (encoding for TetR regulator Sco4385) or pSW7 (encoding for TetR regulator Sco5956) were introduced into *S. coelicolor* M512/cpzLK09 and production was referenced to the control strain carrying the empty vector pUWL-apra-oriT. Unpaired, two-tailed t-tests were conducted to assess statistically significant differences. Two asterisks signify p-value <0,01. Error bars represent production range of three independent biological replicates.

Independent overexpression of all four putative regulatory genes in *S. coelicolor* M512/cpzLK09 resulted in an increase in production of caprazamycin-aglyca. Production was, when referred to the mean production levels of the control strain, in average increased to 185 %, 321 %, 306 % and 179 % for *S. coelicolor* M512/cpzLK09 harboring pSW4, pSW5, pSW6 and pSW7, respectively. Statistical significance of the differential production rates between the overexpression strains and the control strain was assessed applying two-tailored, unpaired t-tests. While production rates of *S. coelicolor* M512/cpzLK09/pSW4 and *S. coelicolor* M512/cpzLK09/pSW7 were within the boundaries of biological variance of *S. coelicolor* M512/cpzLK09/pUWL, the production rates of the heterologous caprazamycin producer carrying pSW5 or pSW6 were significantly enhanced (p-value 0,0061 and 0,0052, respectively). These findings revealed the CRP regulator

Sco3571 and the uncharacterized TetR regulator Sco4385 as new regulators of heterologous caprazamycin production in *S. coelicolor* M512.

4.4.3 Overexpression of *sco3571* and *sco4385* in *S. coelicolor* M512/lpmLK01

Both regulatory genes, *sco3571* and *sco4385*, that showed a significant impact on caprazamycin-aglycon production upon overexpression in *S. coelicolor* M512/cpzLK09, were subsequently overexpressed in strain *S. coelicolor* M512/lpmLK01 in an analogous manner. Vectors pSW5 (carrying *sco3571*) and pSW6 (carrying *sco4385*) were introduced by conjugation and produced liposidomycin levels of three independent exconjugants were determined. These data are visualized in Figure 16. The presence of pSW5 in *S. coelicolor* M512/lpmLK01 showed the same outcome as in *S. coelicolor* M512/cpzLK09, elevating liposidomycin production levels significantly to 279 % compared to the control strain. Interestingly, the presence of pSW6 did not alter liposidomycin rates (insignificant increase by 23 %) showing a differential role of Sco4385 within the heterologous caprazamycin and the liposidomycin producer strains.

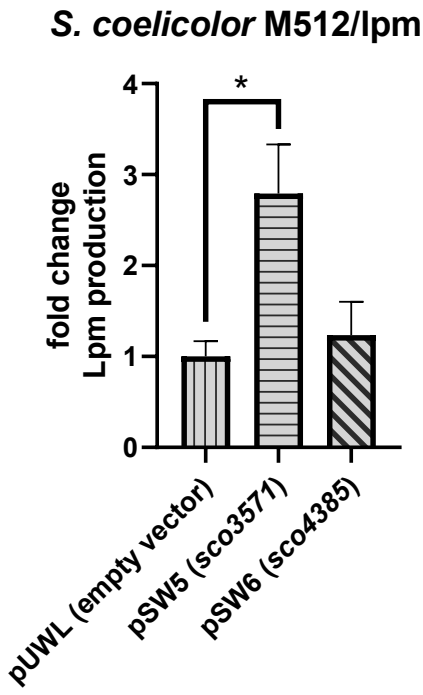


Figure 16: Mean liposidomycin (Lpm) production of *S. coelicolor* M512/lpmLK01 overexpression strains. pSW5 (encoding for CRP regulator Sco3571) and pSW6 (encoding for TetR regulator Sco4385) were introduced into *S. coelicolor* M512/lpmLK01 and production was referenced to the control strain carrying the empty vector pUWL-apra-oriT. Unpaired, two-tailed t-tests were conducted to assess statistically significant differences. One asterisk signifies p-value <0,05. Error bars represent production range of three independent biological replicates.

4.4.4 Overexpression of *sco3571* in *liu*-pathway deficient *S. coelicolor* M1154/cpzLK09

The fact that *sco3571* (CRP regulator) overexpression led to a 3-fold overproduction of both caprazamycin-aglyca and liposidomycins (Figure 15 and Figure 16) raised the question if the enhanced 3-methyl-glutaryl (3-MG) precursor supply from the *liu*-pathway is the underlying reason for that finding. In order to address this question, the CRP regulatory gene was individually overexpressed in a Δ*liu* (deletion of the complete *liu*-pathway; Δ*sco2776-sco2779*), Δ*cpz5* and respective double-deletion strain of *S. coelicolor* M1154/cpzLK09. *S. coelicolor* M1154 was originally found to produce the highest caprazamycin-aglycon titers (152 mg/L in average in 24-square deepwell plates) as a heterologous host carrying cosmid cpzLK09 (Flinspach *et al*., 2010). The mentioned deletion strains had been constructed and were kindly provided by Dr. Daniel Bär. Unfortunately, conjugation of these strains with the replicative overexpression vector pUWL or pSW5 (vector containing *sco3571*) resulted in extremely low numbers of exconjugants and upcoming colonies showed slightly impaired growth. Thus, the integrative vector pMVM3100 equipped with the *sco3571* gene under the control of P*ermE**, designated as pSW24, was used to enable overexpression in these strains. Integration of the vector into the φBT1 *attB* integration site was facilitated by standard biparental conjugation (see 3.2.1.7) and was verified by colony-PCR of the final mutant spore stocks. Subsequently, three independent exconjugants of each strain were grown in production cultures following the standard protocol (see 3.2.5.1) and caprazamycin-aglycon levels were determined (Figure 17). In case of the control strain *S. coelicolor* M1154/cpzLK09 carrying no genetic modifications, a mean caprazamycin-aglyca production of 35 mg/L was determined. Strikingly, overexpression of *sco3571* from the integrated plasmid (pSW24) in *S. coelicolor* M1154/cpzLK09 resulted in a mean production of 38,5 mg/L. Thus, no significantly enhanced caprazamycin-aglycon levels, as it was the case with the replicative overexpression vector in *S. coelicolor* M512/cpzLK09, were observed (see 4.4.2). This shows, that high amounts of Sco3571 are needed to cause the strong overproduction of liponucleosides in the heterologous host.

For this reason, no influence of *sco3571* overexpression could be expected in the mutant strains and thus no conclusion could be drawn.

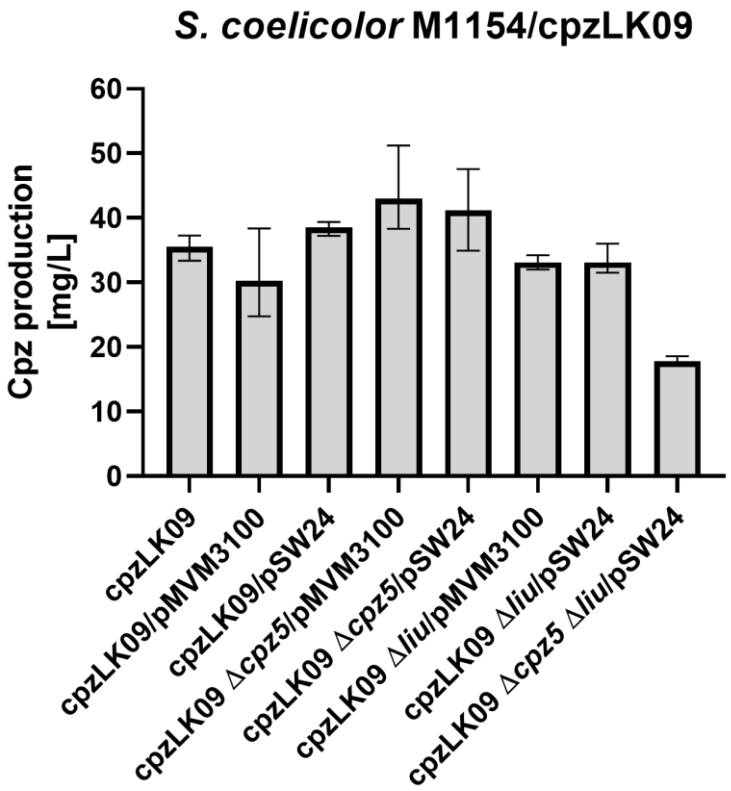


Figure 17: Mean caprazamycin-aglycon (Cpz) production of *sco3571* overexpression mutants of *S. coelicolor* M1154/cpzLK09 deficient in 3-MG supply. pSW24 and the empty vector control pMVM3100 were introduced into strain *S. coelicolor* M1154 carrying the heterologously expressed Cpz cluster. Derivates thereof carry the deletion of *cpz5*, of the complete *liu*- pathway or the double-knockout. Error bars represent the production range of three independent mutants.

4.5 Deletion of *sco3571* and *sco4385* in *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01 by CRISPR-Cas9

In order to characterize the role of Sco3571 and Sco4385 in heterologous liponucleoside biosynthesis in more detail, the genes were independently knocked-out in *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01. Scar- and markerless deletions were achieved by applying a plasmid-based CRISPR-Cas9 method. The pCRISPR-TT vector was equipped with the sgRNA sequence and homology domains specific for the target gene (see 3.2.2.11). Successfully cloned vectors were introduced into the target strains by biparental conjugation and Cas9 induction was facilitated by addition of theophylline to the growth medium (see 3.2.2.13). After several repetitions, in case of the *sco3571* deletion, only two independent exconjugants could be verified with a clean genotype in *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01, respectively. Phenotypically, these

mutant strains were characterized by accelerated growth and sporulation on agar plates. Following plasmid curing, the strains were grown as production cultures and liponucleoside levels were normalized against the cell dry weight of the respective culture (see 3.2.1.2 and 3.2.5.1). Figure 18 summarizes the resulting changes in production upon deletion of *sco3571* and *sco4385* in reference to the intact heterologous strains (control). In *S. coelicolor* M512/cpzLK09, deletion of *sco3571* resulted in a reduction of caprazamycin-aglycon amounts to 15 % and 53 % compared to the control strain. Applying a two-tailored, unpaired t-test this change was found to be statistically significant, however, due to the small sample size of *S. coelicolor* M512/cpzLK09 Δ*sco3571* (n=2) this finding is not indicated in Figure 18. Deletion of *sco4385* in the same strain did not alter liposidomycin production at all (average 91 % of the control strain production).

Regarding the *S. coelicolor* M512/lpmLK01 deletion strains, variance between the biological replicates was very high. Mutants carrying the *sco3571* deletion produced 68 % and 639 % of the control strain level. Deletion of *sco4385* led to slightly increased production levels for two replicates (122 % and 119 % of the control strain) while the third mutant produced 3-fold higher liposidomycin levels compared to the control strain. Repetition of the cultivation resulted in even higher variance between the biological replicates and therefore allowed no conclusive interpretation. Here, is has to be noted that the mutant strains were not subjected to whole-genome sequencing after CRISPR-Cas9 editing, leaving the possibility of off-target effects in the genome, that could influence liponucleoside biosynthesis as well.

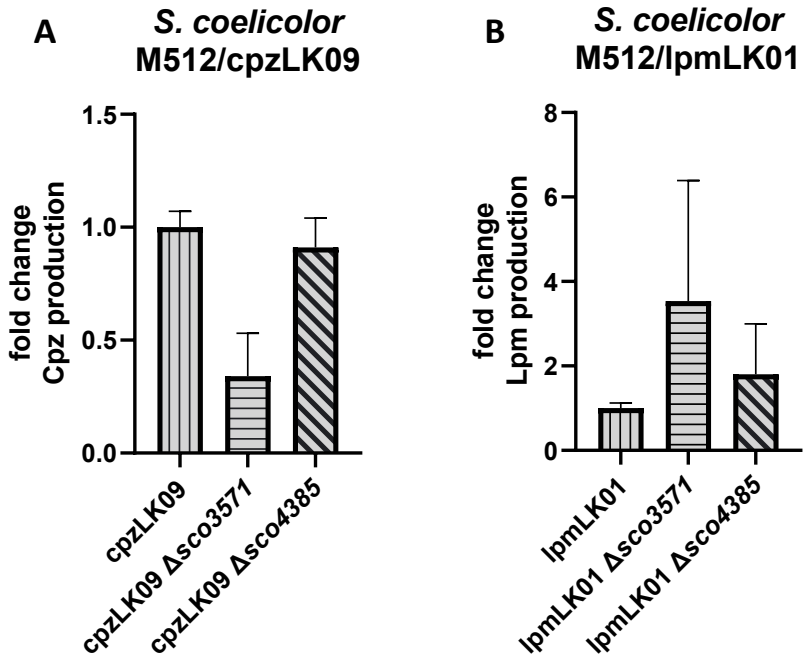


Figure 18: Mean liponucleoside production of *sco3571* and *sco4385* deletion strains. Both regulatory genes were deleted individually in *S. coelicolor* M512/cpzLK09 (A) and *S. coelicolor* M512/lpmLK01 (B) by CRISPR-Cas9 editing. Generally, error bars represent production range of three independent mutants, but for *sco3571* mutants production of only two mutants for both strains is shown, respectively. Cpz: caprazamycin-aglyca, Lpm: liposidomycin.

4.6 Interaction of Sco4385 with P*cpz10* and P*lpmH*

While deletion of the TetR regulatory gene *sco4385* did not show any impact on either caprazamycin-aglyca nor liposidomycin production, the differential role of this protein upon overexpression remained an open question. Therefore, interaction of Sco4385 with the P*cpz10* and P*lpmH* regions should be validated *in-vitro* and closer investigated by determination of the exact binding location within both promoters. An emerging method to detect and characterize protein-DNA interactions is biolayer-interferometry (BLI) which, in the present study, requires biotinylated DNA and the protein of interest in a highly purified form (Sultana and Lee, 2015).

4.6.1 Overexpression and purification of Sco4385

In order to isolate Sco4385 protein in high amounts, the gene was cloned into the overexpression vector pHis8, facilitating the production of N-terminally 8x his-tagged Sco4385 (MW incl. his-tag 24,69 kDa; extinction coefficient 19,48). This vector,

designated pSW21, was used to transform *E. coli* BL21 (DE3) cells and to inoculate a production culture in 100 mL TB medium (see 3.2.4.1). After harvesting the culture, proteins were extracted by cracking the cells via french-press. The supernatant resulting from centrifugation was applied to affinity chromatography via a Ni-NTA resin column and Sco4385 was separated taking advantage of the attached his-tag. The protein was finally eluted using increasing concentrations of imidazole in the buffer (see 3.2.4.1). Samples from both the overexpression and purification procedure were applied to SDS-PAGEs (see 3.2.4.2), which are presented in Figure 19. The sample taken from the production culture after induction clearly shows a band with increased intensity at the expected height around 25 kDa (8x his-tagged Sco4385 MW 24,69 kDa) compared to the sample taken prior to induction. SDS-PAGE also reveals that Sco4385 is well soluble and only a small proportion is present in the insoluble fraction. Furthermore, loss of protein in the flow-through and wash fraction is minimal while elution from the column with high imidazole concentrations yields large amounts of Sco4385 without any visible contaminations.

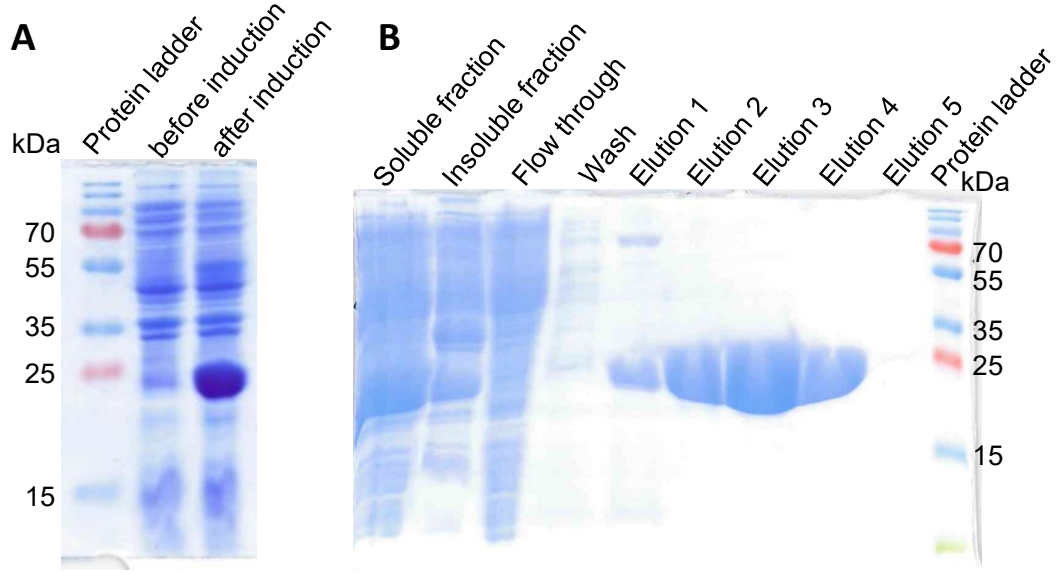


Figure 19: SDS-PAGE of Sco4385 overexpression (A) and purification (B). Samples before and after induction were normalized to OD values of the culture. The calculated molecular weight of 8x his-tagged Sco4385 protein is 24,69 kDa. Imidazole concentration was stepwise increased in the elution buffer: Eluate 1: 100 mM, Eluate 2: 150 mM, Eluate 3: 200 mM, Eluate 4 and 5: 250 mM. kDa: kilo Dalton.

Protein concentrations of all eluates were quantified by Bradford assay and NanoDrop spectrometry yielding highly comparable results. Therefore, the total protein amount was measured by NanoDrop. 10,4 mg Sco4385 protein were isolated from a 100 mL culture. Subsequently, 7,8 mg of protein was applied to size-exclusion

chromatography (SEC; see 3.2.4.3) to reach a higher degree of protein purity. SEC was carried out with the kind help of Tim Orthwein (Department of Microbiology, Institute of Biology, University of Tübingen). During SEC, Sco4385 was detected as a single peak as visible in the chromatogram depicted in Figure 20.

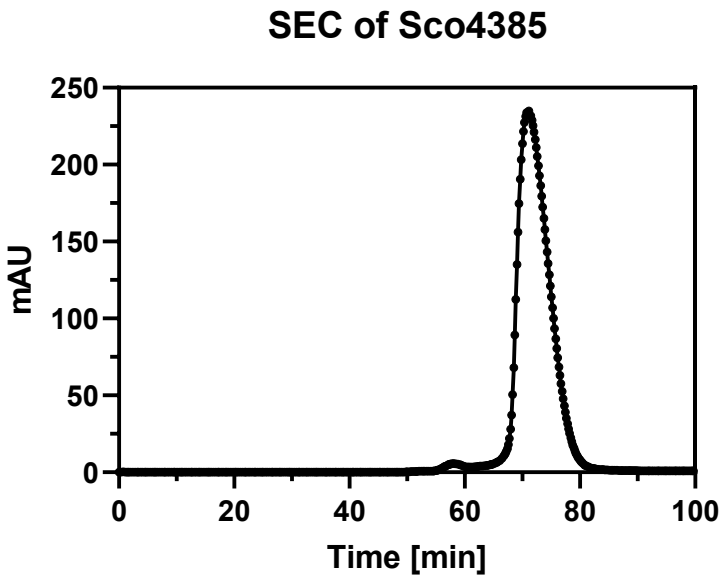


Figure 20: Size-exclusion chromatogram of Sco4385.

All fractions from the SEC containing Sco4385 were pooled and concentrated to 1 mL volume using Amicon® Centrifugal Filter Units. NanoDrop quantification of the solution revealed a total protein amount of 4,2 mg which were available for the upcoming protein-DNA interaction measurements.

4.6.2 Identification of binding locations within P*cpz10* and P*lpmH*

Numerous methods are available to study protein-DNA interactions, such as electrophoretic mobility shift assay (EMSA), surface plasmon resonance (SPR), DNase-footprinting and biolayer interferometry (BLI) which was used in this study. Some of the several advantages of BLI are its high sensitivity and throughput. In addition, it requires only very small amounts of both protein and DNA. The underlying principle of BLI is that an empty sensor chip, covered, for instance, with streptavidin, reflects an emitted light beam back to the device at the same wave length (λ). As soon as a molecule, such as biotinylated DNA, is bound to the sensor surface, the wave length of the emitted light beam is shifted which is detected by the device. Furthermore, an additional shift occurs, if the DNA attached to the sensor is bound by

the target protein. Due to the wave-length shift (Δλ) being proportional to the slice thickness and quantity of proteins bound to the surface, BLI also allows to determine kinetic parameters and binding affinities of the protein of interest.

In order to figure out the binding location of Sco4385 within the promoters of *cpz10* and *lpmH,* the regions were divided into 14 parts, respectively. Each of these segments was 50 bp in length, covering the whole promoter region as it was applied in the DACA. These 50 bp sequences were purchased as single-stranded DNA (ssDNA) oligonucleotides (assigned as forward oligo), hence annealing of the complementary DNA strain (assigned as reverse oligo) was required prior to BLI measurement to obtain double-stranded DNA (dsDNA). A 50 bp segment from the middle of the P*hrdB* region was included as a negative control. The sequences of all oligonucleotides are listed in Table 10. By heating up and slowly cooling down (see 3.2.4.4), the forward and reverse oligo were fused together. The reverse oligo carrys an additional single-stranded overhang complementary to a biotinylated linker sequence, the ReDCaT (re-usable DNA capture technique) linker. This linker was especially designed to form no secondary structures (Stevenson *et al*., 2013) and was successfully used in previous studies to sequentially bind different DNA fragments for protein-DNA interaction studies by SPR (Bekiesch *et al*., 2016b; Tran *et al*., 2018). The advantage of the ReDCaT method is that only one streptavidin-covered sensor is required to measure protein binding to several different DNA fragments. Therefore, the ReDCaT method was tested for its application on the BLI device. However, attempts to regenerate the sensor chip with the linker bound to it, failed, due to the senor/linker being damaged by the harsh alkaline conditions required to reliably separate the DNA segment from the linker. Thus, the ReDCaT method was found to be incompatible with the experimental setup in BLI. Consequently, each dsDNA segment was attached to a new sensor via the biotinylated ReDCaT linker. An exemplary BLI measurement curve with all assigned steps is depicted in Figure 21. The underlying measurement protocol is described in Table 18 (see 3.2.4.4).

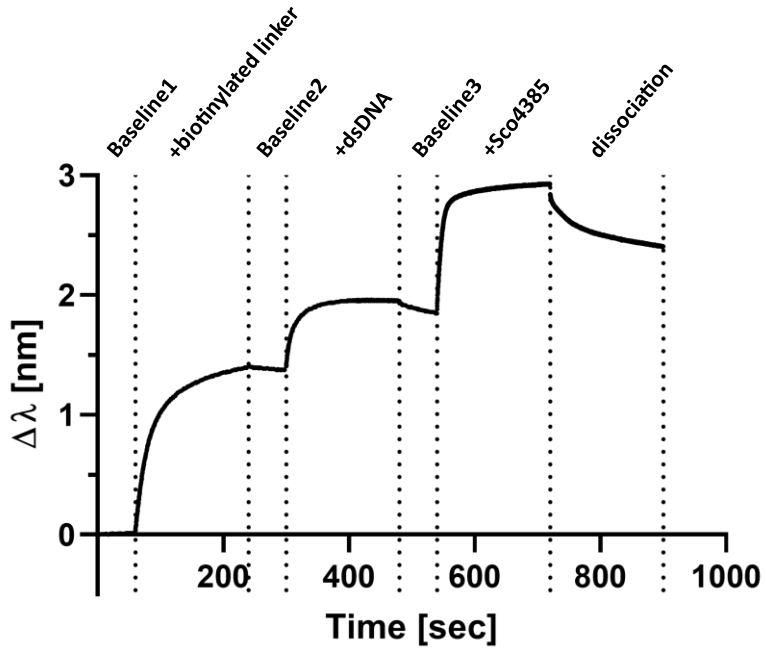


Figure 21: Binding steps during biolayer-interferometry.

All acquired binding curves of Sco4385 at the segments constituting P*cpz10* are depicted in Figure S 4 and Figure S 5 and those constituting P*lpmH* in Figure S 6 and Figure S 7. Indeed, some segments within both promoters were bound by Sco4385, validating on the one hand the designed BLI protocol for measurement of Sco4385-DNA interactions and on the other hand the results from the DACAs (Figure S 8).
Within P*cpz10*, binding of Sco4385 at segments 1 (1,05 nm shift) and 4 (0,74 nm shift) resulted in a stronger wave-length shift compared to the negative control DNA from P*hrdB* (0,39 nm shift). These segments were designated as P*cpz10*_1 and P*cpz10*_4, respectively. Furthermore, in P*lpmH,* interaction of Sco4385 with segments 1 (0,9 nm shift), 7 (0,99 nm shift) and 8 (1,08 nm shift) was detected (negative control 0,43 nm shift). Analogously to the P*cpz10* promoter, these segments were designated as P*lpmH*_1, P*lpmH*_7 and P*lpmH*_8. The respective binding curves as well as the position of the bound segments within the promoter regions are visualized in Figure 22. All other segments of P*cpz10* and P*lpmH* with Sco4385 yielded almost identical measurement curves as segment P*hrdB*.

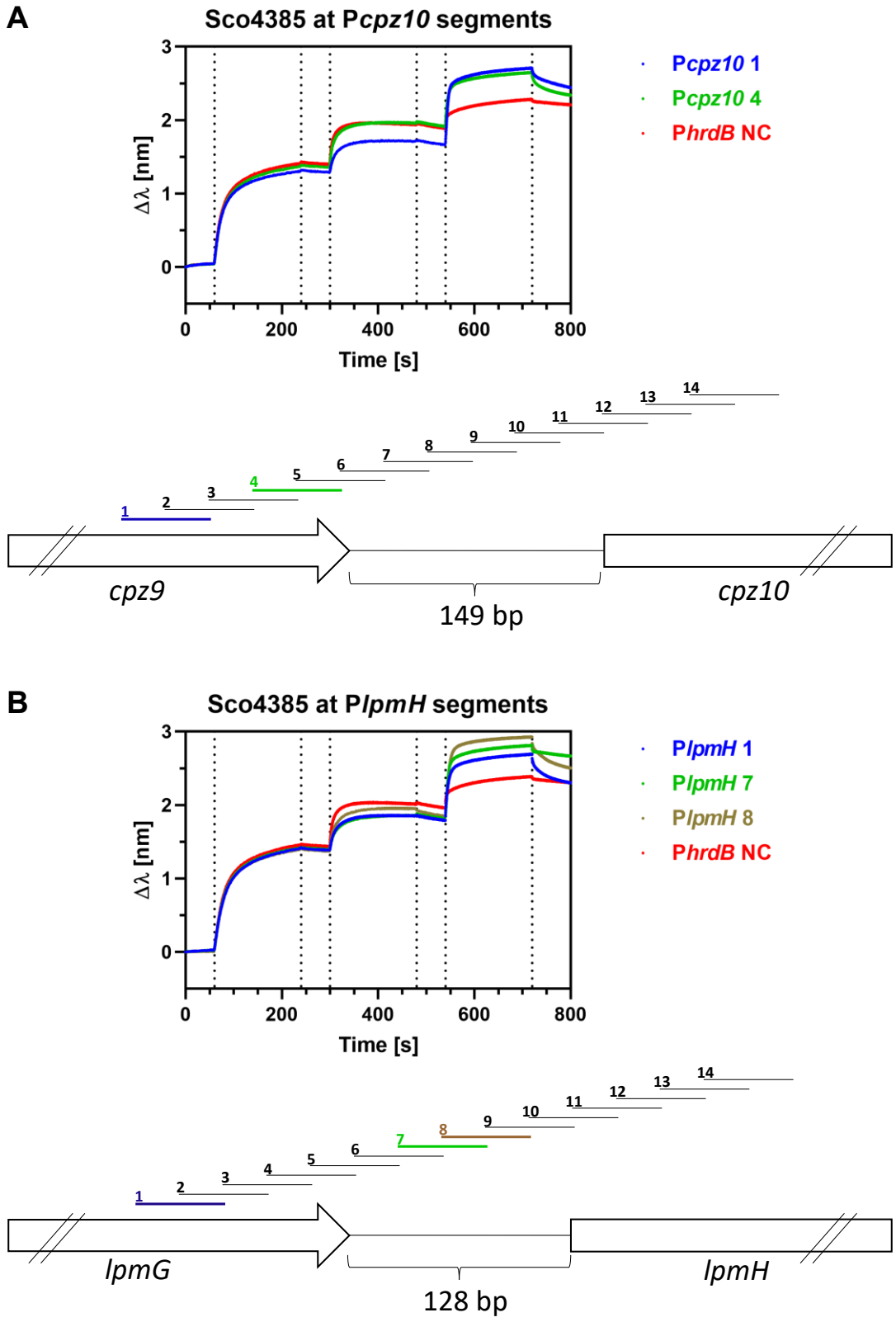


Figure 22: BLI binding curves of Sco4385 interacting with segments of P*cpz10* and P*lpmH* and their positions within the whole promoter fragment. Each promoter was divided into 14 segments of 50 bp length, respectively. Within the promoter of *cpz10*, Sco4385 binds to segments 1 and 4, both located within the coding region of *cpz9* (A), while in P*lpmH,* Sco4385 binds to segments 1, 7 and 8, being located within the *lpmG* gene (segment 1) and in the middle of the intergenic region (segment 7 and 8; B).

Interestingly, considering the location of the segments bound by Sco4385 as depicted in Figure 22, the TetR-family regulator bound P*cpz10* only within the coding region of *cpz9* while in P*lpmH*, besides binding within the *lpmG* gene as well, it additionally bound two overlapping segments in the middle of the intergenic region, namely P*lpmH*_7 and P*lpmH*_8. This striking difference in the interaction of Sco4385 with both promoters could be responsible for the differential regulatory role of this protein upon overexpression (Figure 15 and Figure 16). Moreover, these five DNA sequences could be exploited to search for a common motif that could serve as the recognition and binding sequence for Sco4385.

4.6.3 Determination and validation of a consensus sequence for Sco4385 binding

The identification of five different 50 bp DNA segments (with P*lpmH*_7 and P*lpmH*_8 overlapping by 25 bp) bound by Sco4385 provided the opportunity to screen the sequences for common motives serving as recognition sites. For this purpose the online tool GLAM2 (gapped local alignment of motives; (Frith *et al*., 2008) which belongs to the MEME Suite (https://meme-suite.org/meme/index.html; Bailey *et al*., 2015) was used. MEME Suite is a website providing numerous motif-based sequence analysis tools for different applications. Indeed, GLAM2 revealed a 23 bp consensus sequence (Figure 23) being present in all segments except for P*cpz10*_4 (which also showed the smallest wave length shift). In P*lpmH*_7 and P*lpmH*_8 the sequence was completely covered in their shared overlapping region. The respective 23 bp of every segment were designated as P*cpz10*_1_consensus, P*lpmH*_1_consensus and P*lpmH*_7/8_consensus. Matching the overall high DNA identity between *cpz9* and *lpmG* (77,8 %), P*cpz10*_1_consensus and P*lpmH*_1_consensus (which are both positioned within the coding regions of *cpz9* and *lpmG*, respectively) share 83 % DNA identity to each other, while they both share only 43 % DNA identity with P*lpmH*_7/8_consensus.

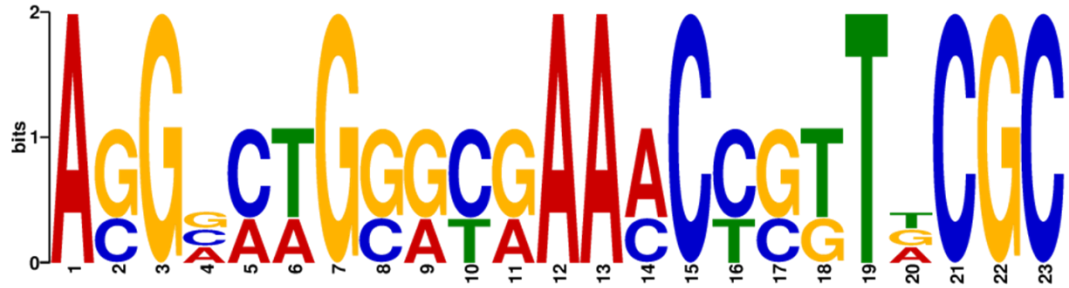


Figure 23: 23 bp consensus sequence calculated by GLAM2 for Sco4385 binding. The sequence has a GC content of 65 %.

In order to experimentally validate the 23 bp sequences as recognition and binding sites for Sco4385, the individual sequences of P*cpz10*_1_consensus, P*lpmH*_1_consensus and P*lpmH*_7/8_consensus were purchased as oligonucleotides (Table 10). dsDNA was achieved as described above (see 3.2.4.4) and tested for Sco4385 binding by BLI, analogously to the previous measurements. The achieved binding curves are shown in Figure 24. As a positive control, P*lpmH*_8 was included, at which binding of Sco4385 had resulted in the strongest wave length shift so far. Binding of Sco4385 to this positive control, indeed, was highly reproducible, resulting in a 1,07 nm shift (compared to 1,08 nm shift in the first measurement).

At P*cpz10*_1_consensus and P*lpmH*_1_consensus, binding of Sco4385 resulted in wave-length shifts of 1,13 nm and 0,93 nm, respectively, confirming that this 23 bp region is sufficient for Sco4385 recognition and binding. In the case of P*cpz10*_1_consensus binding of the predicted recognition sequence resulted even in a slightly higher wave-length shift than with the whole 50 bp segment (1,05 nm). Strikingly, the P*lpmH*_7/8_consensus sequence was not bound by Sco4385 (Figure 24) showing that this 23 bp sequence was not sufficient to allow proper Sco4385 binding. Based on the experimentally validated P*cpz10*_1_consensus and P*lpmH*_1_consensus sequences as Sco4385 binding sites (narrowed down to 23 bp), an even more defined consensus sequence was deduced which is depicted in Figure S 9. Hereafter, this sequence is referred to as the “re-defined” consensus sequence.

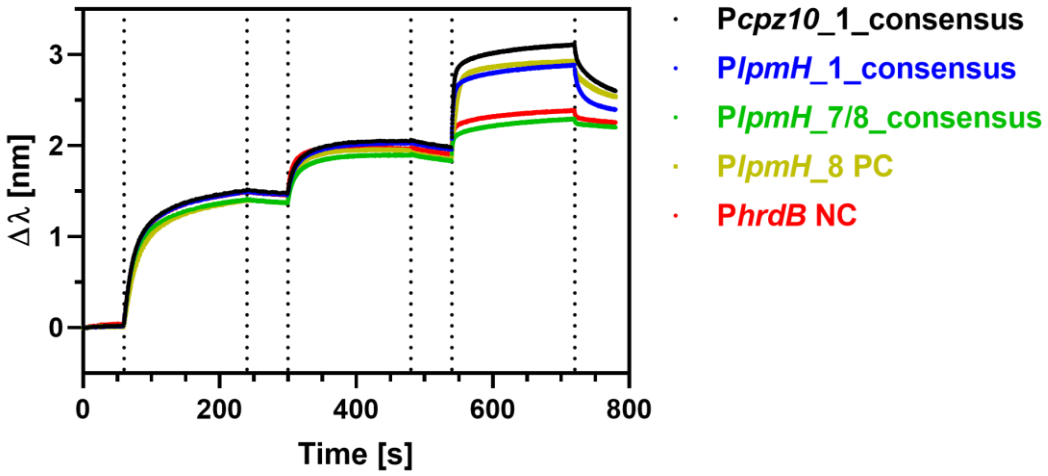


Figure 24: BLI measurement of Sco4385 binding to the predicted recognition sites.

4.6.4 Scanning of the *S. coelicolor* M512 genome with the consensus sequence

Having a (validated) consensus binding sequence at hand, it could be used in a whole-genome search of *S. coelicolor* M512 to possibly identify further target genes of Sco4385. Besides searching for common motives in various DNA sequences, the MEME Suite website also facilitates scanning a whole genome with supplied DNA motives using tool GLAM2Scan (Frith *et al*., 2008). Application of this tool with the re-defined sequence (5´ ASGSCTGGGCRAAACCGTTRCGC 3´; based solely on the two experimentally validated sequences), exceeded the maximum of 200 hits that could be displayed by the program. However, GLAM2Scan allows an open number of mismatches and the best achieved hit, with the re-defined consensus sequence by default, contained already four mismatches. Therefore, the analysis was repeated using the tool “dna-pattern” which is implemented in the regulatory sequence analysis tool RSAT (Santana-Garcia *et al*., 2022). This program enables to specify the number of allowed mismatches within the motif for the genome-wide search. Allowing the maximum of up to two mismatches in the re-defined consensus sequence, the search yielded, in accordance with the GLAM2Scan result, no hit within the whole *S. coelicolor* M512 genome. However, using the originally determined more unspecific consensus sequence 5` ASGVMWGSRYRAAMCYSKTDCGC 3´ (Figure 23), six hits were found when allowing two mismatches. Strikingly, these six hits were all located within coding regions of genes instead of intergenic regions. The genes which contain the putative Sco4385 recognition site encode for two putative glycoside

hydrolases (Sco6944 and Sco1777), an putative aminohydrolase (Sco1861), a LuxR-family regulator (Sco0132) and two hypothetical proteins (Sco6386 and Sco6331) as predicted by BLAST analysis. None of the proteins has been experimentally characterized so far.

4.7 Attempts of LpmG overexpression and purification

As its regulatory mode of action and potential binding site are unknown, the cluster-situated regulator of the liposidomycin cluster was subjected to overexpression and purification attempts as well. *lpmG* was amplified and cloned into vector pHis8, for expression of N-terminally his-tagged LpmG protein, and into vector pET28a(+), facilitating the attachment of a C-terminal his-tag, yielding plasmids pSW22 and pSW23, respectively (LpmG incl. tag MW 39,14 kDa). *E. coli* BL21 (DE3) harboring pSW22 was cultivated according to the same protocol that was successfully applied for the overexpression of Sco4385 (see 3.2.4.1). However, SDS-PAGE revealed that no protein band intensity was increased by induction of the culture with IPTG. For the overexpression of LpmG from pET28a(+) in *E. coli* BL21 (DE3), the temperature was adjusted to 30 °C after induction and cells were harvested 1,5 h later to circumvent degradation of the target protein. Again, the respective SDS-PAGE indicated no LpmG overexpression to be observed after induction. Nevertheless, harvested cell pellets of the overexpression cultures were subjected to protein purification as described in 3.2.4.1, but in none of the samples and eluates LpmG could be detected by SDS-PAGE.

4.8 Impact of *cpz9* and *lpmG* overexpression on heterologous liponucleoside production

Besides the identification of new regulators, the regulatory impact of the cluster situated regulators (CSRs) Cpz9 and LpmG should be investigated in the heterologous host *S. coelicolor* M512 carrying either the caprazamycin or liposidomycin gene cluster. *cpz9* and *lpmG* were amplified from genomic DNA of *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01 and cloned into vector pUWL-apra-oriT, resulting in pSW8 and pSW9, respectively. In order to find out whether the AraC-type regulators could also influence production levels of the respective other BGC, both plasmids were conjugated into both heterologous strains individually. Strikingly, no exconjugants of strain *S. coelicolor* M512/cpzLK09, carrying pSW9 (for *lpmG* overexpression) could be generated after several attempts.

Cultivation of three individual clones of the successful conjugations as well as the control strains carrying the empty pUWL vector was carried out following the standard protocol (see 3.2.5.1). Liponucleoside amounts were determined by LC-MS and normalized to the cell dry weight of each culture. Figure 25 shows caprazamycin-aglycon and liposidomycin levels in reference to the empty vector control. While in the heterologous caprazamycin producer strain, in which production levels are already high, introduction of additional *cpz9* copies did barely impact caprazamycin-aglycon amounts (average 1,35-fold increase), overexpression of *lpmG* in *S. coelicolor* M512/lpmLK01 resulted in average 3,5-fold enhanced liposidomycin amounts. The homologous regulator Cpz9 increased liposidomycin levels as well, but to an insignificant extent (1,94-fold), which is also due to the high variance in production levels between the individual biological replicates.

In average, absolute amouts of caprazamycin-aglyca and liposidomycins produced by the control strains *S. coelicolor* M512/cpzLK09/pUWL and *S. coelicolor* M512/lpmLK01/pUWL corresponded to 35,12 mg/L and 11,56 mg/L, respectively. Overexpression of *lpmG* in *S. coelicolor* M512/lpmLK01 consequently increased liposidomycin production to levels similar to caprazamycin-aglyca production by *S. coelicolor* M512/cpzLK09/pUWL. Therefore, overexpression of the cluster-situated regulatory gene *lpmG* resulted in the highest liposidomycin titers measured in this study.

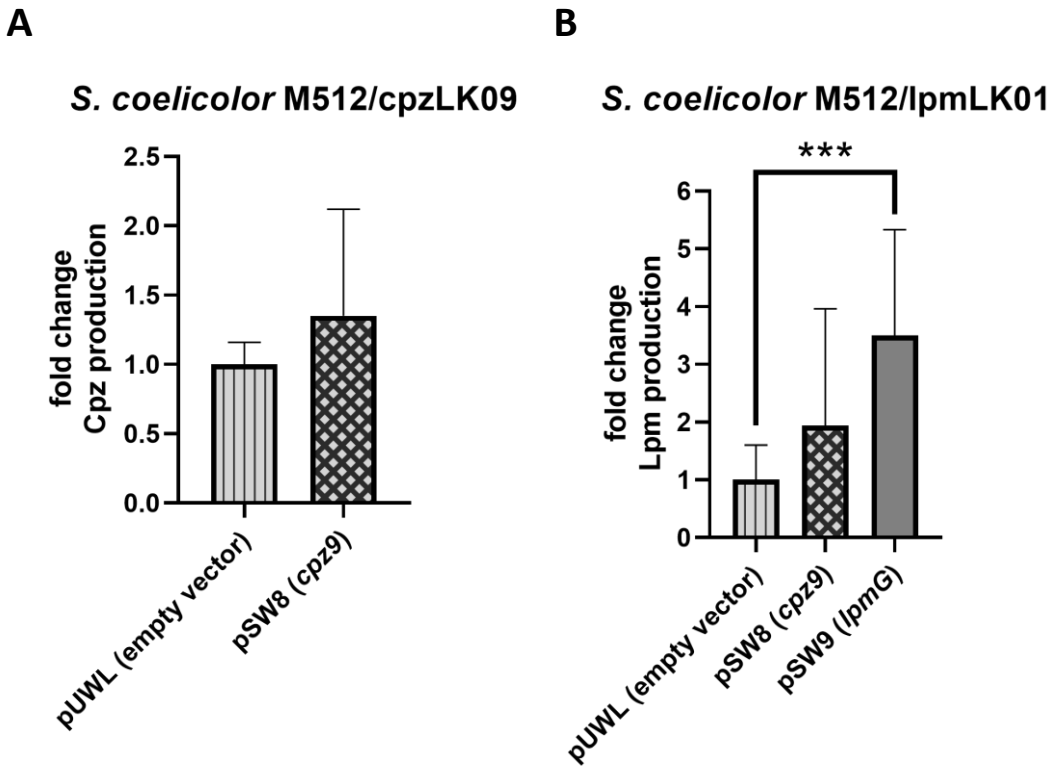


Figure 25: Mean liponucleoside production of *S. coelicolor* M512/cpzLK09 (A) and *S. coelicolor* M512/lpmLK01 (B) overexpressing the cluster-situated regulators Cpz9 and LpmG. Quantified liponucleoside amounts were normalized by cell dry weight and related to the empty vector control strains carrying pUWL-apra-oriT. Error bars indicate the production range of three individual exconjugants. Two-sided, unpaired t-tests were conducted to assess statistical significance. Three asterisks signify p-value <0,001. Cpz: caprazamycin-aglyca, Lpm: liposidomycin.

5 Discussion

5.1 RNA-sequencing

RNA-sequencing (RNAseq) was applied to define the operon structure, and therefore promoter regions, of the heterologously expressed caprazamycin and liposidomycin BGCs. Overall transcription of the caprazamycin BGC was found to be much stronger than the liposidomycin BGC. This was expected due to the higher liponucleoside production in *S. coelicolor* M512/cpzLK09 compared to *S. coelicolor* M512/lpmLK01 (see 4.1.2). Comparing the read counts of the first biosynthetic genes of both operons, *cpz10* expression was up to 15-fold higher than *lpmH* transcription in the samples collected after 2 days. In the samples collected after 4 days, when overall transcription of the caprazymcin BGC had decreased and liposidomycin BGC expression had increased, *cpz10* was still expressed up to 3,3-fold stronger than *lpmH*. Accordingly, average expression (from the three biological replicates) of the cluster-situated regulatory gene *cpz9* was 8700 read counts after 2 days and 6700 read counts after 4 days, while it was only in average <1000 and 3000 read counts for *lpmG* at the two time points, respectively. This could indicate, that *lpmG* expression in the heterologous host is directly repressed by host regulatory proteins or that *cpz9* transciption is activated in the same strain. However, as long as expression of the gene clusters and the cluster-situated regulators in the native strains is unknown, a regulatory effect from the heterologous host remains hypothetical.

Another interesting outcome of the RNAseq data is, that genes *cpz22* and *lpmT*, encoding for ABC transporters, are expressed at low levels compared to the other genes of the BGCs. These transporters are presumably involved in self-resistance, excreting the antibiotics, before reaching toxic intracellular titers. Due to their low expression, this resistance mechanism could be a restricting factor in the accumulation of caprazamycins and liposidomycins. Therefore, a promising approach in order to improve production levels in the heterologous hosts, would be the increase of *cpz22* and *lpmT* expression e.g. by cloning the highly active promoter region of P*cpz10*/P*lpmH* upstream of the respective ABC transporter gene.

Furthermore, the location of promoter sites within the BGCs largely met the predictions. Promoter pairs P*cpz6*/P*lpmD* and P*cpz10*/Pl*pmH* could be verified as strong promoters, while the intergenic regions P*cpz9*/P*lpmG* and P*cpz24*/P*lpmV* showed only a minor role as transcription start sites and therefore may only have a slight additional effect on the overall transcription rates.

In addition to the gene cluster expression, differential genome-wide expression of the heterologous strains, *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01,

in comparison to the wild-type host strain, *S. coelicolor* M512, was preliminarily assessed as well. Due to the data complexity, an advanced bioinformatic evaluation is required, which was not subject of this study. However, some interesting conclusions could be drawn by comparing the absolute numbers of differentially expressed genes from the three strains at two selected time points (Table 22).

Table 22: Differential genome-wide expression of genes in *S. coelicolor* M512, *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01. Absolute numbers of significantly differentially expressed genes are given. Cosmid-encoded genes, only present in the heterologous strains, were excluded from the analysis. Cpz: *S. coelicolor* M512/cpzLK09; Lpm: *S. coelicolor* M512/lpmLK01; WT: "wild-type" strain *S. coelicolor* M512; 2 d: time point at two days; 4 d: time point at four days.

Compared conditions	Up-regulated genes	Down-regulated genes	Total differentially expressed genes
Cpz 2 d vs. WT 2 d	103	345	448
Lpm 2 d vs. WT 2 d	5	2	7
Cpz 4 d vs. WT 4 d	317	508	825
Lpm 4 d vs. WT 4 d	3	4	7
Cpz 4 d vs. Cpz 2 d	141	52	193
Lpm 4 d vs. Lpm 2 d	471	533	1004
WT 4 d vs. WT 2 d	198	104	302

As indicated in Table 22, incorporation of the caprazamycin or liposidomycin BGC results in very different responses of the host on transcription level. At both time points, the heterologous caprazamycin producer showed significantly more differentially expressed genes (448 and 825 at 2 and 4 days, respectively) than the liposidomycin producer (7 at both time points), when compared to the wild-type strain *S. coelicolor* M512. This finding may be the result of high antibiotic production in *S. coelicolor* M512/cpzLK09, which affects the host strain in several ways: firstly, the strong production of the new compound leads to a high cellular stress level, which might result in the activation of genes, required to adapt to stress responses or in detoxification. As produced liposidomycin titers are significantly lower than those of caprazamycin-aglycons, the dose-dependent stress response is expected to be less severe in the heterologous liposidomycin producer. Secondly, the biosynthetic machinery for production of the heterologous compound may compete with host-owned biosynthetic pathways for building blocks from the (primary) metabolism. Therefore, the host might react by up-regulating primary metabolism pathways or

down-regulate other secondary metabolism pathways, due to limiting precursor supply. As the scope of this competition is again dependent on the total production titers, a stronger impact may be expected in *S. coelicolor* M512/cpzLK09 than in *S. coelicolor* M512/lpmLK01.

Finally, these findings challenge the hypothesis, that introduction of highly similar BGCs entails similar (transcriptional) responses in the same heterologous host. It rather indicates that the scope of regulatory and metabolic adaptations might be more dependent on other factors, such as biosynthesis rates of the new compound and thus the associated physiological and metabolic stress responses within the host.

The differentially expressed genes listed in Table 22 were screened for Sco3571 and Sco4385, which were identified as regulators of heterologous liponucleoside biosynthesis in the present study. Only *sco3571* was found among the differentially expressed genes in the data set comparing *S. coelicolor* M512/lpmLK01 global expression at the two time points. Transcription of *sco3571* was found to be 2,58-fold higher in this strain after 4 days than after 2 days, which meets the expections. Furthermore, the gene encoding for the molecular target of liponucleoside antibiotics, *mraY* (*sco2087*), is not listed among the differentially expressed genes.

5.2 DNA-affinity capturing assay

According to Bentley *et al*. (2002), there are 965 predicted regulatory proteins encoded within the genome of *S. coelicolor* A(3), which accounts for 12,3 % of the total proteome. In the present study, four promoter regions of the caprazamycin and liposidomycin BGCs, respectively, and the promoter region of the negative control gene *hrdB* were used to trap putative regulatory proteins from the heterologous strains *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01. A total number of 2214 different proteins were identified from both strains. Considering the DACAs of both strains individually, the promoter regions (and negative control) of the caprazamycin gene cluster were bound by 1906 proteins and the liposidomycin promoter regions (and the negative control) by 1540 proteins. Both exceed the number of predicted regulatory proteins encoded in the host strain, which indicates a substantial percentage of bound proteins in the data set, that do not fulfill a regulatory function or are at least not assigned as a regulator. However, these numbers correlate quite well with the amounts of proteins identified in previously conducted DACAs with *Streptomyces* strains. In her PhD thesis, Dr. Tomke Bakker also applied five promoters (including a negative control) of the tacrolimus BGC to trap regulators from the native producer *S. tsukubaensis* and identified 1044 different proteins (Bakker,

2021). Similarly to the present study, Dr. Paulina Bekiesch used the novobiocin BGC heterologously expressed in *S. coelicolor* M512 to identify 2475 proteins from the host to bind to four selected promoter regions, including P*hrdB* as negative control (Bekiesch, 2015). Thus, in the latter study, a much larger quantity of proteins was detected, while applying less promoter regions as bait.

5.2.1 Cluster-situated regulators Cpz9 and LpmG

Unexpectedly, the cluster-situated regulator (CSR) of the caprazamycin BGC, Cpz9, was found to bind rather unspecifically to all tested promoter regions, including the negative control P*hrdB*. In contrast, in the aforementioned study by Dr. Paulina Bekiesch, the CSR of the novobiocin cluster, NovG, was found to bind specifically to one promoter region of its associated gene cluster (Bekiesch, 2015). However, in the study conducted by Dr. Tomke Bakker, one of the CSRs of the tacrolimus BGC, FkbN, bound to only one of the cluster-derived promoters, although binding sites for this regulator were predicted in all of the applied promoter regions. Similar to Cpz9, FkbN also bound to the negative control (Bakker, 2021).
Moreover, the finding that LpmG could not be identified in any of the samples was also unexpected, as *lpmG* transcripts could be detected in the fermenter samples by RT-qPCR. Nevertheless, the RT-qPCR results do not allow absolute but only relative quantification of the transcripts, which is why the resulting protein could be present on a very low level. This hypothesis is supported by the RNAseq results, showing low transcription of *lpmG*, at least in relation to the homologous gene *cpz9* and the residual genes of the BGC (see 4.1.2 and 5.1). Therefore, the amount of protein, that probably bound to the DNA fragments in the DACA, might have been too low to be detected by semi-quantitative mass-spectrometry. An additional possibility is, that LpmG protein is instable or highly susceptible to protease activity and thus challenging to be detected by mass-spectrometry. *In silico* analysis using ProtParam (Expasy) predicted LpmG to be instable, which was probably also the reason for the unsuccessful attempts of recombinant overexpression and purification of LpmG (see 4.7).

5.3 Regulators of liponucleoside antibiotics biosynthesis

Four putative regulatory proteins were overexpressed in *S. coelicolor* M512/cpzLK09, using a pUWL201-based replicative overexpression plasmid. Determination of the produced antibiotic levels revealed two of the mutant strains as significant

overproducers of caprazamycin-aglyca. Thus, CRP-family regulator Sco3571 and TetR-family regulator Sco4385 were newly identified as regulators of caprazamycin-alycon biosynthesis in the heterologous host. While overexpression of *sco3571* likewise enhanced liposidomycin levels in *S. coelicolor* M512/lpmLK01, the overexpression of *sco4385* did not affect production levels in this strain.

5.3.1 CRP-family regulator Sco3571

The regulator family of cyclic-AMP-receptor proteins (CRPs) is a major subgroup of transcriptional regulators and is widespread among the bacterial kingdom (Körner *et al*., 2003). CRP was first described as a central regulator in carbon catabolite repression (CCR) in *E. coli*, where it represses expression of gene *cya,* encoding for the adenylate cyclase (Mori and Aiba, 1985) and activates transcription of the glucose transporter gene *ptsG*, allowing the preferred uptake of glucose as the carbon source (Kimata *et al*., 1997). Binding of cAMP at the N-terminus of the protein is required for CRP dimerization and activation of the C-terminal DNA-binding domain, consisting of a helix-turn-helix (HTH) motif (Kim *et al*., 1992).
In contrast to *E. coli*, in *Streptomyces*, CRP is not involved in CCR (Derouaux *et al*., 2004a), but is a global regulator of both primary and secondary metabolism and furthermore mediates crucial developmental processes, such as sporulation and colony formation (Derouaux *et al*., 2004b; Piette *et al*., 2005). One member of the CRP-family is Sco3571, which was previously shown to influence the expression of a number of secondary metabolite BGCs, such as actinorhodin, the calcium-dependent antibiotic and yellow-pigmented polyketide. Moreover, overexpression of *sco3571* in a set of *Streptomyces* species resulted in the awakening of silent gene clusters and thus the production of novel metabolites (Gao *et al*., 2012). Therefore, in order to enhance metabolite production levels, the CRP regulator became a promising target for genetic manipulation strategies. Recently, the Sco3571 homolog in *S. tsukubaensis* was investigated in detail and its overexpression led to a 2-fold increase of tacrolimus production (Schulz *et al*., 2023).
In the present study, Sco3571 was revealed as a regulator of heterologous liponucleoside antibiotic biosynthesis in *S. coelicolor* M512, as its gene overexpression resulted in an average 3-fold overproduction (and deletion in a reduction) of caprazamycin-aglyca and liposidomycins. Furthermore, morphological characteristics in the *sco3571* deletion mutants were observed, especially accelerated sporulation, that started several days earlier than in the non-modified wild-type strain. The fact, that Sco3571 is a known global regulator, strongly impacting

physiology and fundamental processes during the bacterial life cycle, could be the reason for the challenges encountered in generating the deletion mutants (see 3.2.2.13).

The increase in antibiotic production in the *sco3571* overexpression mutants is probably based on an enhanced transcription rate of precursor pathways, as Sco3571 is predicted to mediate between primary and secondary metabolism. Additionally, one recently discovered precursor pathway of caprazamycin biosynthesis, the leucine-isovalerate-utilization (*liu*-) pathway (encompassing genes *sco2776*-*sco2779*), is directly controlled by Sco3571. Overexpression of *sco3571* from an inducible plasmid in *S. coelicolor* M145 increases expression of *liuB* (*sco2776*), the first gene of the *liu*-operon, by approximately 18-fold (Gao *et al*., 2012). The *liu*-pathway was recently shown to be one out of two routes for the supply of the 3-methyl-glutaryl (3-MG) precursor molecule, 3-methyl-glutaconyl-CoA, incorporated in caprazamycins and liposidomycins. The other supply route is mediated via the cluster-encoded hydroxy-methylglutaryl synthase Cpz5 (and its homolog LpmA in the liposidomycin BGC; Bär *et al*., 2022). However, supply of the 3-MG moiety has yet not been described as a possible bottleneck of liponucleoside production, which suggests, that Sco3571 might additionally regulate one or several other, more limiting precursor pathways for caprazamycin and liposidomycin biosynthesis.

Although the exact reason for enhanced liponucleoside biosynthesis upon *sco3571* overexpression remains unknown, it is another example for a regulator from the host, that binds to the heterologously expressed gene cluster and is additionally involved in the regulation of a precursor pathway. Such a regulatory “cross-talk” between the hosts primary and secondary metabolism was also investigated by Bekiesch *et al*. (2016b) with the novobiocin BGC heterologously expressed in *S. coelicolor* M512. Therein, a ROK family regulator, RokB, was shown in DACAs to bind specifically to one of the promoter regions of the novobiocin BGC and its exact binding site was determined by surface-plasmon-resonance (SPR). Bekiesch *et al.* (2016a) applied the newly identified binding sequence in a search for alternative binding sites in the *S. coelicolor* genome and suggested a putative role in the regulation of amino acid metabolism. Notably, L-tyrosine is the common precursor of the aminocoumarin moiety of novobiocin. Both *rokB* deletion and overexpression led to a strong reduction of novobiocin production, indicating a pleiotropic regulatory effect. However, it was shown, that feeding of casamino acids to the growth medium partly restored novobiocin production in the *rokB* deletion mutant, while an increase in production was not observed in the intact heterologous novobiocin producer. This finding indicated a regulation of heterologous novobiocin biosynthesis through mediation of

the amino acid precursor supply by the RokB regulator (Bekiesch *et al*., 2016b). So far, RokB and Sco3571 are the only two examples, that potentially link a heterologously expressed BGC to a precursor supply pathway by regulatory cross-talk.

Using chromatin immunoprecipitation-microassays, Gao *et al*. (2012) determined 393 Sco3571-associated sequences within the genome of *S. coelicolor* M145. Among the predicted target genes of Sco3571, 9,9 % are annotated as transcriptional regulators and 17,6 % are putatively involved in metabolism, of which 5,1 % account for genes of secondary metabolism. Strikingly, 8 out of 22 predicted BGCs of *S. coelicolor* M145 were found to contain CRP associated binding sites, emphasizing its strong interlinking in secondary metabolism. Alignment of three predicted binding sites, that were additionally verified by DNase I footprinting, allowed to create two consensus sequences for Sco3571, a highly conserved (GTGWSMTSYGNCMSA) and a less conserved one (SYSMNSWNNKNSAS; Gao *et al.*, 2012).

Using the conserved binding sequence predicted by Gao *et al*. (2012), a whole genome search in *S. coelicolor* A(3) yielded 3532 hits (MEME FIMO result with a p-value <0,0001). Amongst them, the putative binding site was found in the promoter region of *sco3571* itself, indicating an auto-regulatory mechanism. On the contrary, the binding site is not present in the promoter region of *liuB*, although a regulatory effect on this gene was shown (Gao *et al*., 2012). This could indicate, that Sco3571 regulates *liuB* (*sco2776*) via an indirect mechanism, for instance, via another transcriptional regulator, forming a regulatory cascade. However, it is more likely, that Sco3571, as a DNA-binding transcription factor, is flexible towards alternative binding sites. This hypothesis is supported by its unspecific binding to all tested DNA sequences in the DACAs of the present study. However, the putative binding site was identified only in P*cpz6* and P*cpz9* of the caprazamycin BGC and in the corresponding promoter regions of the liposidomycin gene cluster P*lpmD* and P*lpmG*, but not in the other promoter regions. Altogether, these findings suggest, that Sco3571 harbors more flexible DNA binding properties, than covered by the proposed consensus sequence.

As Sco3571 is a wide-spread and broadly conserved protein among *Streptomyces* species (Gao *et al*., 2012) the question arised, whether a homologous protein is encoded within the genome of the native caprazamycin and liposidomycin producer strains. Native strains *S*. sp. MK730-62F2 and *S*. sp. SN-1061M served as the source for the isolation of the BGCs for heterologous expression and their genomes were previously sequenced (unpublished). A BLAST search among all proteins encoded in their genomes revealed homologs to Sco3571 (224 aa) in both strains with extremely

high similarity. *S*. sp. MK730-62F2 contains a protein homolog with 97 % identity and similarity (157 aa; 70 % coverage; E-value $4E^{-110}$) and a protein from *S*. sp. SN-1061M exhibits 98 % identity and similarity (224 aa; 100 % coverage; E-value $2E^{-161}$) to Sco3571. Therefore, a regulation of liponucleoside biosynthesis by the homologous proteins in the native strains is highly probable. Interestingly, the amino acid sequences of the two homologous proteins from the native hosts are identical (100 % identity), indicating either a high evolutionary proximity between *S*. sp. MK730-62F2 and *S*. sp. SN-1061M or a high selection pressure on this global regulatory protein.

5.3.2 TetR-family regulator Sco4385

One of the most abundant and influential regulator group among bacteria is the TetR-family of regulators. The name-giving TetR regulator was first reported in *E. coli* to convey resistance against tetracycline (Izaki *et al*., 1966). Only almost three decades later, the molecular mode of action of TetR was revealed in detail. In the absence of tetracycline, TetR binds to the promoter of *tetA*, which is orientated divergently from its own gene, thereby inhibiting transcription of *tetA*. As soon as tetracycline is present, it binds to TetR as a ligand, which results in a conformational change, leaving TetR unable to bind to DNA. Thus, *tetA*, encoding for the tetracycline efflux pump, is expressed and tetracycline resistance is conferred (Kisker *et al*., 1995; Cuthbertson and Nodwell, 2013). Until today, hundreds of bacterial TetR-family regulators have been studied. Most of them act as transcriptional repressors, however, a substantial percentage has been characterized as activators. The first bacterial TetR-family regulator (TFR) described to enhance, rather than repress gene transcription, was DhaS from *Lactococcus lactis* (Christen *et al*., 2006). More recent examples for activating TFRs in *Streptomyces* are DepR1 from *S. roseosporus*, accelerating and significantly increasing production of the industrially relevant antibiotic daptomycin (Yuan *et al*., 2016), or MilR2 from *S. hygroscopius*, enhancing biosynthesis of the antihelminthic and insecticidal metabolite 5-oxomilbemycin (Wei *et al*., 2018). Furthermore, TFRs may perform dual functions, such as GouR from *S. graminearus*, involved in gougerotin biosynthesis (Wei *et al*., 2014), or GdmRIII, which simultaneously activates geldanamycin biosynthesis and represses elaiophylin production in *S. autolyticus* (Jiang *et al*., 2017). In *S. coelicolor*, TFRs are the most abundant type of regulators, as they account for 15,8 % of all predicted regulatory genes (Cuthbertson and Nodwell, 2013). One of the best studied regulators of antibiotic biosynthesis in *Streptomyces* is AtrA, a TFR that positively impacts, among others, actinorhodin, lincomycin and daptomycin biosynthesis (Uguru *et al*., 2005; Xu

et al., 2023; Wu *et al*., 2023). AtrA was also detected at promoter regions of the heterologously expressed caprazamycin and liposidomycin BGCs in the DACAs presented here, however, in an unspecific manner and was thus not considered for further characterization.

Another predicted TFR, Sco4385, was detected specifically at corresponding promoter regions of the caprazamycin and liposidomycin BGC (P*cpz10*/P*lpmH*) in the present study. Overexpression studies of the yet uncharacterized protein revealed a significant positive effect on caprazamycin, but not on liposidomycin biosynthesis. Therefore, Sco4385 constitutes another example for a TFR acting as an activator in the regulation of a secondary metabolite BGC. In her PhD thesis, Dr. Tomke Bakker likewise identified and characterized two TFRs (IFCAGPOI_05107 and IFCAGPOI_06489), that showed specific DNA-binding properties in DACAs and significantly enhanced tacrolimus production in gene overexpression experiments. Overexpression of the TetR-type regulator IFCAGPOI_05107 even resulted in a 3,5-fold increase of tacrolimus production (on a 100 mL scale), which represents the maximum tacrolimus production level obtained in her study (Bakker, 2021). As Sco4385 overexpression led to a 3-fold enhancement of caprazamycin-aglyca production in *S. coelicolor* M512/cpzLK09, the overall importance of activating TFRs in secondary metabolite regulation is emphasized by these results.

TFRs show a highly conserved tertiary structure, consisting of nine α-helices. Regarding their functional assignment, helices 1-3 are involved in formation of the DNA-binding domain, while helices 4-9 constitute a ligand-binding domain. In almost all cases, TFRs are predicted to rely on small molecules as ligands, altering the proteins' ability to bind DNA. However, only for few TFRs the associated ligand is known, due to its challenging experimental assessment and the large number of possible ligand candidates. DNA binding by helix 2 and 3, forming a HTH-motif, is specified by helix 3, recognizing binding sequences in the major groove of the DNA strand (Cuthbertson and Nodwell, 2013). Figure 26 shows, that Sco4385 indeed exhibits the characteristic structural features of a TFR, based on a protein model generated using Colabfold (Mirdita *et al*., 2022).

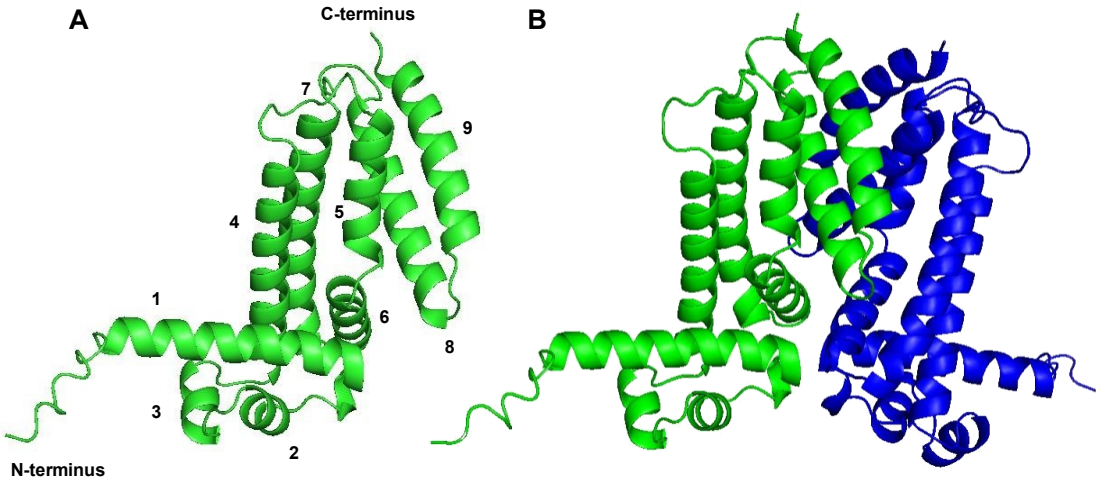


Figure 26: Predicted structure of Sco4385 by Colabfold. The Sco4385 monomer (A) shows the typical structure of a TFR consisting of nine α-helices (indicated by numbers). Helices 1-3 constitute the DNA binding domain, helices 5-7 the ligand-binding domain and helices 8 and 9 of each monomer are involved in the formation of the dimer interface (B).

The genomic landscape of *sco4385* suggests, that it is likely part of a gene cluster, encompassing genes *sco4374* to *sco4386*. These genes are all orientated in the same direction and exhibit either overlapping open reading frames or intergenic regions <135 bp.

While many genes of this putative cluster are of unknown function (*sco4374*, *sco4376*, *sco4378*, *sco4379* and *sco4386*), the predicted functions of the other genes indicate a role in fatty acid metabolism: *sco4380* and *sco4381* are annotated as acetyl-CoA carboxylase (ACCase) subunits α and β, respectively, whereas *sco4382* encodes for an acyl-CoA dehydrogenase and *sco4384* for an enoyl-CoA hydratase. These enzymes play important roles in fatty acid metabolism, such as β-oxidation of fatty acids, which are in turn building blocks for caprazamycin and liposidomycin biosynthesis. Therefore, Sco4385 could potentially be involved in the regulation of the supply of β-hydroxy fatty acids.

Interestingly, several genes of this putative cluster show (functional) homology to genes of the aforementioned *liu*-pathway involved in 3-MG biosynthesis (see 5.3.1). *sco4380* and *sco4381* encode for one of only four ACCases (subunit α and β) present in the *S. coelicolor* genome, with *liuD* and *liuB* (*sco2777* and *sco2776*) being one of the other three α- and β-subunits, respectively (Gago *et al*., 2011; Bär *et al*., 2022). Furthermore, gene *sco4384* is one out of four candidate genes in the *S. coelicolor* genome, that were identified as potential homologes to *liuC*, which is not part of the *liu*-gene cluster in *Streptomyces*, but is present in *Pseudomonas* (Bär *et al*., 2022). The gene *liuC* encodes for a 3-methylglutaconyl-CoA hydratase, converting 3-

methylglutaconyl-CoA to 3-hydroxy-3-methylglutaryl-CoA. To rule out the possibility that *sco4384* takes over *liuC* activity in 3-MG biosynthesis, Bär *et al*. (2022) generated deletion mutants with the four putative *liuC* homologs deleted individually by transposon insertions. It was shown that deletion of *sco4384* alone does not replace LiuC activity in the biosynthesis of caprazamycin-aglyca. Nevertheless, it could be interesting to delete genes *sco4380-sco4384* in a *liu*-pathway deficient *S. coelicolor* M1154/cpzLK09 strain and investigate the impact on caprazamycin biosynthesis on a quantitative level. If Sco4385 regulates its neighboring genes, it could be partially involved in the regulation of the 3-MG precursor supply as well. However, it is known, that genes *sco4374-sco4386* alone can not be capable of synthesizing the 3-MG moiety, as simultaneous deletion of the *liu*-pathway and *cpz5* in *S. coelicolor* M1154/cpzLK09 (see 2.3) resulted in the accumulation of hydroxyacyl-caprazols, lacking the 3-MG moiety (Bär *et al*., 2022).

According to Ahn *et al*. (2012), the genome context of a TFR gene can be used to predict its regulatory target gene, which could be applied to *sco4385*. In *S. coelicolor*, 69 % of all TFRs are encoded in the opposite direction of their neighboring gene and a regulatory control of this gene by the TFR is highly probable, if the intergenic region between both genes is <200 bp. Unfortunately, *sco4385* belongs to the 14 % of TFRs in the *S. coelicolor* genome, that are orientated in the same direction as their neighboring genes (up- and downstream) and their intergenic regions are <35 bp (intergenic region *sco4384-sco4385*: 8 bp; *sco4385-sco4386*; 27 bp). Although this arrangement strongly indicates a co-transcription and therefore common regulation of *sco4383-sco4386*, it does not allow to predict a regulatory target gene of Sco4385. This question should therefore be addressed by determination of the binding site in protein-DNA interaction experiments.

Finally, as for Sco3571, a search for homologs of Sco4385 within the native caprazamycin strain *S*. sp. MK730-62F2 and the native liposidomycin strain *S*. sp. SN-1061M was performed. Again, both strains carry highly conserved homologs to Sco4385 from *S. coelicolor*, however, the similarity is slightly lower than assessed for Sco3571. The protein homolog from *S*. sp. MK730-62F2 shows 87 % identity and 90 % similarity (203 aa; 99 % coverage; E-value $5E^{-122}$) and the one encoded by *S*. sp. SN-1061M exhibits 86 % identity and 90 % similarity (200 aa; 99 % coverage; E-value $5E^{-118}$) to Sco4385 (202 aa). BLAST analysis of the amino acid sequences of the two homologs from the native strains revealed 89 % identity and 93 % similarity (98 % coverage; E-value $2E^{-125}$) to each other. Due to the high degrees of similarity, it may be assumed, that the protein homologs fulfill the same roles in *S*. sp. MK730-62F2, *S*. sp. SN-1061M and *S. coelicolor*. Therefore, the regulatory influences of Sco4385

on liponucleoside antibiotics biosynthesis in the heterologous host, may probably be transferable to the native strains as well.

5.4 Sco4385-DNA binding studies

The differential effect of *sco4385* overexpression on liponucleoside biosynthesis in the heterologous caprazamycin and liposidomycin producer strains and its genetic proximity to genes potentially involved in liponucleoside precursor supply, laid the focus on this protein for further studies. By the implementation of DNA-protein assays, several goals were addressed. At first, the binding of Sco4385 to P*cpz10* and P*lpmH*, as observed in the DACAs, could be verified. Thereby the binding location within the promoter regions could be determined, which potentially provides explanations for the differential impact on biosynthesis levels in *S. coelicolor* M512/cpzLK09 and *S. coelicolor* M512/lpmLK01 upon overexpression of *sco4385*. Unravelling a consensus binding sequence could additionally fuel speculations on additional functions carried out by Sco4385 in the *Streptomyces* metabolism.

Using biolayer-interferometry measurements, the TFR was found to dominantly bind two segments within P*cpz10* and three segments within P*lpmH*, with two of them overlapping in the latter promoter region. Therefore, the specific binding of Sco4385 found by the DACAs was confirmed. Strikingly, both binding regions in P*cpz10* (P*cpz10*_1 and P*cpz10*_4) are located within the coding region of *cpz9*. In P*lpmH*, Sco4385 bound within the corresponding coding region of *lpmG* as well (P*lpmH*_1), however, the main binding site was found in the intergenic region between *lpmG* and *lpmH*, where strong binding occurred at overlapping segments P*lpmH*_7 and P*lpmH*_8 (see 4.6.2).

This finding allows to hypothesize, that the different main binding locations in P*cpz10* and P*lpmH* could be responsible for the varying impacts of Sco4385 in both BGCs. In P*cpz10*, Sco4385 presumably activates transcription of downstream genes from its binding location in *cpz9*, while the intergenic region remains available for binding of Cpz9 or other regulatory proteins. Therefore, *sco4385* overexpression and increased DNA binding do not interfere with the “natural” regulation of the downstream genes in the heterologous producer. It rather adds an activating effect, resulting in enhanced caprazamycin-aglycon production. In *S. coelicolor* M512/lpmLK01, in which *sco4385* overexpression does not increase production levels, the protein likely acts as a transcriptional activator as well, but at the same time potentially blocks the intergenic region of P*lpmH*, based on the positioning of the binding site. The putative activation and blocking, that possibly hinders LpmG or other regulators from binding, could result

in an unaffected liposidomycin production. However, this hypothesis is highly speculative and requires the aquisition of further experimental data to be verified e.g. by RT-qPCR and determination of the binding sites for Cpz9 and LpmG.

Using the bound DNA segments as templates for the search of a common consensus sequence, a 23-bp region (5` ASGVMWGSRYRAAMCYSKTDCGC 3´), contained in four out of five segments, was identified (see 4.6.3). Interestingly, there are six matches of this sequence detected within the genome of *S. coelicolor* M512 (when allowing two mismatches), which are all located in coding regions of genes.

Thus, it is most likely, that the binding sequences occur at these positions by chance and do not represent the natural recognition sites of Sco4385 (completely).

One of the six genes containing the consensus sequence is *sco0132*, which encodes for a large ATP-binding regulator of the LuxR family (LAL-family). Representatives of this underinvestigated regulator group were shown to impact antibiotic production and influence several essential processes in *Streptomyces* (Guerra *et al*., 2012). Therefore, interference of Sco4385 with *sco0132* expression could potentially lead to a larger-scale regulatory cascade, that may eventually also influence heterologous liponucleoside biosynthesis. As the LAL gene *sco0132* is under the control of at least one other global regulator, TopA (Szafran *et al*., 2019), its regulatory network may be highly complex and the impact of Sco4385 binding within this gene may hardly be predicted. The gene products of the residual five genes, that contain the predicted Sco4385 consensus recognition sequence, show no obvious relation to liponucleoside biosynthesis (see 4.6.4) and are not discussed in literature so far.

Furthermore, the consensus sequence was not found within the putative transcriptional cluster encompassing the *sco4385* gene itself and exhibiting wide-ranging parallels to several genes of the *liu*-pathway (see 5.3.2). However, to fully exclude the possibility of a regulatory impact of Sco4385 on this cluster, further research would be required e.g. the implementation of RT-qPCR measurements.

5.5 The cluster-situated regulators Cpz9 and LpmG

Cpz9 and LpmG are both putative AraC-type regulators. Generally, members of this important regulator family are transcriptional activators, but representatives showing an inhibiting effect on transcription are found in literature as well (Barbosa and Levy, 2000). AraC family regulators normally bind DNA as functional dimers with the HTH motif located at the C-terminus of the protein and are further characterized by a N-terminal ligand binding site within the dimerization domain (Schleif, 2010). The best studied member of the AraC regulator family in *Streptomyces* is AdpA (A-factor

dependent protein). It is a global regulator, that, for example, positively controls chloramphenicol biosynthesis in *S. venezuelae* (Płachetka *et al*., 2021), or indirectly increases streptomycin biosynthesis in *S. griseus* by activation of *strR* expression, the cluster-situated regulatory gene of the streptomycin BGC (Ohnishi *et al*., 2005). Furthermore, AdpA is reported to be involved in morphological development in *Streptomyces* (Romero-Rodríguez *et al*., 2016).

The closest related gene clusters to the caprazamycin and liposidomycin BGCs with already characterized cluster-situated regulators, are the muraymycin and sansanmycin BGCs. Xu *et al*. (2013) described Mur34 as a strong negative regulator of muraymycin biosynthesis in both the native producer *S*. sp. NRRL30471 and in the heterologous host *S. lividans* TK24. Upon *mur34* deletion, in both strains, production of muraymycin was significantly increased (e.g. 10-fold enhanced muraymycin titers in *S*. sp. NRRL30471). By electrophoretic-mobility-shift-assays (EMSAs), Mur34 was detected on the promoter region of *mur33*, where binding led to an approximately 10-fold decreased transcription rate of the downstream gene. Mur33 exhibits homology to cluster-situated regulators of the related uridyl peptide antibiotic BGCs, such as PacA (from the pacidamycin gene cluster), NpsM (from the napsamycin gene cluster) and SsaA (from the sansanmycin gene cluster; see 2.1.2). As all these homologous regulators positively influence the expression of their respective gene cluster and reduced *mur33* expression leads to a decrease in muraymycin production, Mur33 most likely represents a cluster-situated activator of the muraymycin gene cluster (Xu *et al*., 2013). Therefore, the cascade-like regulation of the muraymycin BGC is obviously different from the direct regulation via Cpz9 and LpmG, proposed for caprazamycin and liposidomycin gene cluster regulation.

The aforementioned cluster-situated regulator of the sansanmycin BGC, SsaA, was the first regulator of uridyl peptide antibiotics biosynthesis, that was studied in detail. A gene deletion of *ssaA* was reported to lead to a complete abolishment of sansanmycin production by *S*. sp. SS, due to an absent expression of the biosynthetic genes. On the other hand, overexpression of the regulatory gene resulted in a 100 % increase of sansanmycins A and H formation (Li *et al*., 2013). An interesting approach was applied by Jiang *et al*. (2015), who used overexpression of *ssaA* from *S*. sp. SS in *S. roseosporus* NRRL15998, to successfully activate a cryptic mureidomycin gene cluster (see 2.1.2). Strikingly, awakening of this BGC by SsaA indeed led to the production of mureidomycins, while overexpression of the endogenous *ssaA* homolog in *S. roseosporus* did not stimulate mureidomycin biosynthesis. Based on these findings, the potential for a “cross-regulation” of the caprazamycin and liposidomycin BGC by their cluster-situated regulators was assessed in the present study.

Genes *cpz9* and *lpmG* were overexpressed in both associated heterologous strains. Overexpression of *lpmG* in *S. coelicolor* M512/cpzLK09 was not possible, which could indicate that caprazamycin-aglyca titers were elevated to toxic levels. RNAseq results showed, that the ABC transporter gene *cpz22* is expressed only at low levels in relation to the residual genes of the caprazamycin gene cluster (see 5.1). Therefore, the putative resistance mechanism via Cpz22 could be a limiting factor for the accumulation of caprazamycin-aglyca. However, if LpmG enhances caprazamycin biosynthesis to a greater extend than the endogenous regulator Cpz9, it would imply the ability of LpmG to bind to promoter regions of the caprazamycin BGC and that it may be a more potent activator. The notion of LpmG as a highly potent transcriptional activator, which is only expressed at very low levels (see 5.1), would be a possible reason why it was not detected in the DACA. Furthermore, it showed a strong effect upon overexpression in *S. coelicolor* M512/lpmLK01 (3,5-fold increase in liposidomycin production), while *cpz9* overexpression resulted only in a moderately (1,94-fold) increased production in this strain. Thus, the endogenous regulator LpmG elevated liposidomycin titers stronger than its homolog Cpz9. A similar effect as observed for *ssA* overexpression in *S. roseosporus* can therefore not be transferred to the liponucleoside BGCs.

Since strain *S. coelicolor* M512/lpmLK01/pSW9, carrying the *lpmG* overexpression plasmid, produced similar amounts of liposidomycins compared to caprazamycin-aglycon production rates by *S. coelicolor* M512/cpzLK09 (see 4.8), the low expression of *lpmG* can be considered as the main cause for low liposidomycin production levels in the heterologous host.

6 Bibliography

Ahn, S. K., Cuthbertson, L., and Nodwell, J. R. (2012). Genome context as a predictive tool for identifying regulatory targets of the TetR family transcriptional regulators. *PLoS One* **7**:e50562.

Almahayni, K., Spiekermann, M., Fiore, A., Yu, G., Pedram, K., *et al*. (2022). Small molecule inhibitors of mammalian glycosylation. *Matrix Biol Plus* **16**:100108.

Arai, M., Haneishi, T., Enokita, R., and Kayamori, H. (1974). Aspiculamycin, a new cytosine nucleoside antibiotic I. Producing organism, fermentation and isolation. *J Antibiot (Tokyo)* **27**:329–333.

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res* **43**:W39–W49.

Bakker, T. (2021). Regulation studies of tacrolimus biosynthesis and generation of derivatives. *University of Tübingen*.

Bär, D., Konetschny, B., Kulik, A., Xu, H., Paccagnella, D., *et al*. (2022). Origin of the 3-methylglutaryl moiety in caprazamycin biosynthesis. *Microb Cell Fact* **21**:232.

Barbosa, T. M., and Levy, S. B. (2000). Differential expression of over 60 chromosomal genes in *Escherichia coli* by constitutive expression of MarA. *J Bacteriol* **182**:3467–3474.

Barnard-Britson, S., Chi, X., Nonaka, K., Spork, A. P., Tibrewal, N., *et al*. (2012). Amalgamation of nucleosides and amino acids in antibiotic biosynthesis: Discovery of an L-threonine:uridine-5′-aldehyde transaldolase. *J Am Chem Soc* **134**:18514–18517.

Bekiesch, P. (2015). Regulation of novobiocin biosynthesis by the heterologous host strain *Streptomyces coelicolor* M512. *University of Tübingen*.

Bekiesch, P., Franz-Wachtel, M., Kulik, A., Brocker, M., Forchhammer, K., *et al*. (2016a). DNA affinity capturing identifies new regulators of the heterologously expressed novobiocin gene cluster in *Streptomyces coelicolor* M512. *Appl Microbiol Biotechnol* **100**:4495–4509.

Bekiesch, P., Forchhammer, K., and Apel, A. K. (2016b). Characterization of DNA binding sites of RokB, a ROK-Family regulator from *Streptomyces coelicolor* reveals the RokB regulon. *PLoS One* **11**:e0153249.

Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A.-M., Challis, G. L., Thomson, N. R., *et al*. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**:141–147.

Boyle, D. S., and Donachie, W. D. (1998). *mraY* is an essential gene for cell growth in *Escherichia coli*. *J Bacteriol* **180**:6429–6432.

Brandish, P. E., Kimura, K.-I., Inukai, M., Southgate, R., Lonsdale, J. T., *et al*. (1996). Modes of action of tunicamycin, liposidomycin B, and mureidomycin A: Inhibition of phospho-N-acetylmuramyl-pentapeptide translocase from *Escherichia coli*. *Antimicrob Agents Chemother* **40**:1640–1644.

Browning, D. F., and Busby, S. J. W. (2004). The regulation of bacterial transcription initiation. *Nat Rev Microbiol* **2**:57–65.

Bruntner, C., Lauer, B., Schwarz, W., Möhrle, V., and Bormann, C. (1999). Molecular characterization of co-transcribed genes from *Streptomyces tendae* Tü 901 involved in the biosynthesis of the peptidyl moiety of the peptidyl nucleoside antibiotic nikkomycin. *Mol Gen Genet* **262**:102–114.

Bugg, T. D. H., and Kerr, R. V. (2019). Mechanism of action of nucleoside antibacterial natural product antibiotics. *Journal of Antibiotics* **72**:865–876.

Chatterjee, S., Nadkarni, S. R., Vljayakumar, E. K. S., Patel, M. V, Ganguli, B. N., *et al*. (1994). Napsamycins, new *Pseudomonas* active antibiotics of the mureidomycin family from *Streptomyces sp.* HIL Y-82,11372. *Journal of Antibiotics* **47**:595–599.

Chen, W., Qu, D., Zhai, L., Tao, M., Wang, Y., *et al*. (2010). Characterization of the tunicamycin gene cluster unveiling unique steps involved in its biosynthesis. *Protein Cell* **1**:1093–1105.

Cheng, L., Chen, W., Zhai, L., Xu, D., Huang, T., *et al*. (2011). Identification of the gene cluster involved in muraymycin biosynthesis from *Streptomyces* sp. NRRL 30471. *Mol Biosyst* **7**:920–927.

Chi, X., Pahari, P., Nonaka, K., and Van Lanen, S. G. (2011). Biosynthetic origin and mechanism of formation of the aminoribosyl moiety of peptidyl nucleoside antibiotics. *J Am Chem Soc* **133**:14452–14459.

Chi, X., Baba, S., Tibrewal, N., Funabashi, M., Nonaka, K., *et al*. (2013). The muraminomicin biosynthetic gene cluster and enzymatic formation of the 2-deoxyaminoribosyl appendage. *Med. Chem. Commun* **4**:239–243.

Christen, S., Srinivas, A., Bähler, P., Zeller, A., Pridmore, D., *et al*. (2006). Regulation of the Dha operon of *Lactococcus lactis*: A deviation from the rule followed by the TetR family of transcription regulators. *Journal of Biological Chemistry* **281**:23129–23137.

Chung, B. C., Zhao, J., Gillespie, R. A., Kwon, D.-Y., Guan, Z., *et al*. (2013). Crystal structure of MraY, an essential membrane enzyme for bacterial cell wall synthesis. *Science (1979)* **341**:1012–1016.

Chung, B. C., Mashalidis, E. H., Tanino, T., Kim, M., Matsuda, A., *et al*. (2016). Structural insights into inhibition of lipid I production in bacterial cell wall synthesis. *Nature* **533**:557–560.

Cui, Z., Wang, X.-C., Liu, X., Lemke, A., Koppermann, S., *et al*. (2018). Self-resistance during muraymycin biosynthesis: a complementary nucleotidyltransferase and phosphotransferase with identical modification sites and distinct temporal order. *Antimicrob Agents Chemother* **62**:e00193--18.

Cui, Z., Overbay, J., Wang, X., Liu, X., Zhang, Y., *et al*. (2020). Pyridoxal-5′-phosphate-dependent alkyl transfer in nucleoside antibiotic biosynthesis. *Nat Chem Biol* **16**:904–911.

Cuthbertson, L., and Nodwell, J. R. (2013). The TetR family of regulators. *Microbiology and Molecular Biology Reviews* **77**:440–475.

Dähn, U., Hagenmaier, H., Höhne, H., König, W. A., Wolf, G., *et al*. (1976). Stoffwechselprodukte von Mikroorganismen. Nikkomycin, ein neuer Hemmstoff der Chitinsynthese bei Pilzen. *Arch Microbiol* **107**:143–160.

Deochand, D. K., and Grove, A. (2017). MarR family transcription factors: dynamic variations on a common scaffold. *Crit Rev Biochem Mol Biol* **52**:595–613.

Derouaux, A., Halici, S., Nothaft, H., Neutelings, T., Moutzourelis, G., *et al*. (2004a). Deletion of a cyclic AMP receptor protein homologue diminishes germination and affects morphological development of *Streptomyces coelicolor*. *J Bacteriol* **186**:1893–1897.

Derouaux, A., Dehareng, D., Lecocq, E., Halici, S., Nothaft, H., *et al*. (2004b). Crp of *Streptomyces coelicolor* is the third transcription factor of the large CRP-FNR superfamily able to bind cAMP. *Biochem Biophys Res Commun* **325**:983–990.

Doroghazi, J. R., Ju, K. S., Brown, D. W., Labeda, D. P., Deng, Z., *et al*. (2011). Genome sequences of three tunicamycin-producing *Streptomyces strains*, *S. chartreusis* NRRL 12338, *S. chartreusis* NRRL 3882, and *S. lysosuperificus* ATCC 31396. *J Bacteriol* **193**:7021–7022.

Engels, S., Schweitzer, J. E., Ludwig, C., Bott, M., and Schaffer, S. (2004). *clpC* and *clpP1P2* gene expression in *Corynebacterium glutamicum* is controlled by a regulatory network involving the transcriptional regulators ClgR and HspR as well as the ECF sigma factor σH. *Mol Microbiol* **52**:285–302.

Erb, A., Luzhetskyy, A., Hardter, U., and Bechthold, A. (2009). Cloning and sequencing of the biosynthetic gene cluster for saquayamycin Z and galtamycin B and the elucidation of the assembly of their saccharide chains. *ChemBioChem* **10**:1392–1401.

Filipe, S. R., Severina, E., and Tomasz, A. (2001). Functional analysis of *Streptococcus pneumoniae* MurM reveals the region responsible for its specificity in the synthesis of branched cell wall peptides. *Journal of Biological Chemistry* **276**:39618–39628.

Flinspach, K., Westrich, L., Kaysser, L., Siebenberg, S., Gomez-Escribano, J. P., *et al.* (2010). Heterologous expression of the biosynthetic gene clusters of coumermycin A1, clorobiocin and caprazamycins in genetically modified *Streptomyces coelicolor* strains. *Biopolymers* **93**:823–832.

Floriano, B., and Bibb, M. (1996). *afsR* is a pleiotropic but conditionally required regulatory gene for antibiotic production in *Streptomyces coelicolor* A3(2). *Mol Microbiol* **21**:385–396.

Fox, J. J., Kuwada, Y., and Watanabe, K. A. (1968). Nucleosides LVI. On the structure of the nucleoside antibiotic gougerotin. *Tetrahedron Lett* **57**:6029–6032.

Frith, M. C., Saunders, N. F. W., Kobe, B., and Bailey, T. L. (2008). Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol* **4**:e1000071.

Fronko, R., Lee, M., Lee, V. J., and Leger, R. (2001). Pacidamycins produced by *Streptomyces coeruleorubidus*.

Fujita, Y., Kizuka, M., Funabashi, M., Ogawa, Y., Ishikawa, T., *et al.* (2011). A-90289 A and B, new inhibitors of bacterial translocase I, produced by *Streptomyces* sp. SANK 60405. *Journal of Antibiotics* **64**:495–501.

Fujita, Y., Kagoshima, Y., Masuda, T., Kizuka, M., Ogawa, Y., *et al.* (2019). Muraminomicins, new lipo-nucleoside antibiotics from *Streptosporangium* sp. SANK 60501-structure elucidations of muraminomicins and supply of the core component for derivatization. *Journal of Antibiotics* **72**:943–955.

Funabashi, M., Nonaka, K., Yada, C., Hosobuchi, M., Masuda, N., *et al.* (2009). Identification of the biosynthetic gene cluster of A-500359s in *Streptomyces griseus* SANK60196. *Journal of Antibiotics* **62**:325–332.

Funabashi, M., Baba, S., Nonaka, K., Hosobuchi, M., Fujita, Y., *et al.* (2010). The biosynthesis of liposidomycin-like A-90289 antibiotics featuring a new type of sulfotransferase. *ChemBioChem* **11**:184–190.

Gago, G., Diacovich, L., Arabolaza, A., Tsai, S. C., and Gramajo, H. (2011). Fatty acid biosynthesis in actinomycetes. *FEMS Microbiol Rev* **35**:475–497.

Gao, C., Hindra, Mulder, D., Yin, C., and Elliot, M. A. (2012). Crp is a global regulator of antibiotic production in *Streptomyces*. *mBio* **3**:e00407–e00412.

Global tuberculosis report (2023). *World health organization*.

Guerra, S. M., Rodríguez-García, A., Santos-Aberturas, J., Vicente, C. M., Payero, T. D., *et al*. (2012). LAL regulators Sco0877 and Sco7173 as pleiotropic modulators of phosphate starvation response and actinorhodin biosynthesis in *Streptomyces coelicolor*. *PLoS One* **7**:e31475.

Guxi, X., Houzhi, H., Jiaren, C., Weixin, C., and Linsen, W. (1995). A new agricultural antibiotic-Ningnanmycin. *Acta Microbiol Sin* **35**:368–374.

Hakulinen, J. K., Hering, J., Brändén, G., Chen, H., Snijder, A., *et al*. (2017). MraY-antibiotic complex reveals details of tunicamycin mode of action. *Nat Chem Biol* **13**:265–267.

Hesketh, A., Chen, W. J., Ryding, J., Chang, S., and Bibb, M. (2007). The global role of ppGpp synthesis in morphological differentiation and antibiotic production in *Streptomyces coelicolor* A3(2). *Genome Biol* **8**:R161.1-R161.18.

Hirano, S., Ichikawa, S., and Matsuda, A. (2008). Synthesis of caprazamycin analogues and their structure-activity relationship for antibacterial activity. *Journal of Organic Chemistry* **73**:569–577.

Huszár, S., Singh, V., Polcicová, A., Baráth, P., Barrio, M. B., *et al*. (2017). N-acetylglucosamine-1-phosphate transferase, WecA, as a validated drug target in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **61**:1–46.

Igarashi, M., Nakagawa, N., Doi, N., Hattori, S., Naganawa, H., *et al*. (2003). Caprazamycin B, a novel anti-tuberculosis antibiotic from *Streptomyces sp. J Antibiot* **56**:580–583.

Igarashi, M., Takahashi, Y., Shitara, T., Nakamura, H., Naganawa, H., *et al*. (2005). Caprazamycins, novel lipo-nucleoside antibiotics, from *Streptomyces sp. J. Antibiot.* **58**:327–337.

Ikeda, M., Wachi, M., Jung, H. K., Ishino, F., and Matsuhashi, M. (1991). The *Escherichia coli mraY* gene encoding UDP-*N*-acetylmuramoyl-pentapeptide:undecaprenyl-phosphate phospho-*N*-acetylmuramoyl-pentapeptide transferase. *J Bacteriol* **173**:1021–1026.

Ikeuchi, T., Kitame, F., Kikuchi, M., and Ishida, N. (1972). An antimycoplasma antibiotic asteromycin: its identity with gougerotin. *J Antibiot (Tokyo)* **25**:548–550.

Inukai, M., Isono, F., and Takatsuki, A. (1993). Selective inhibition of the bacterial translocase reaction in peptidoglycan synthesis by mureidomycins. *Antimicrob Agents Chemother* **37**:980–983.

Iqbal, M., Mast, Y., Amin, R., Hodgson, D. A., Wohlleben, W., *et al*. (2012). Extracting regulator activity profiles by integration of *de novo* motifs and

expression data: Characterizing key regulators of nutrient depletion responses in *Streptomyces coelicolor*. *Nucleic Acids Res* **40**:5227–5239.

Ishizaki, Y., Hayashi, C., Inoue, K., Igarashi, M., Takahashi, Y., *et al*. (2013). Inhibition of the first step in synthesis of the mycobacterial cell wall core, catalyzed by the GlcNAc-1-phosphate transferase WecA, by the novel caprazamycin derivative CPZEN-45. *Journal of Biological Chemistry* **288**:30309–30319.

Isono, K., Uramoto, M., Kusakabe, H., Kimura, K., Isaki, K., *et al*. (1985). Liposidomycins: novel nucleoside antibiotics which inhibit bacterial peptidoglycan synthesis. *J Antibiot (Tokyo)* **38**:1617–1621.

Isono, F., Katayama, T., Inukai, M., and Haneishi, T. (1989). Mureidomycins A-D, novel peptidylnucleoside antibiotics with spheroplast forming activity. III. Biological properties. *J Antibiot (Tokyo)* **42**:674–679.

Iwasaki, H. (1962). Studies on the structure of gougerotin. *Yakugaku Zasshi* **82**:1358–1361.

Izaki, K., Kiuchi, K., and Arima, K. (1966). Specificity and mechanism of tetracycline resistance in a multiple drug resistant strain of *Escherichia coli*. *J Bacteriol* **91**:628–633.

Jez, J. M., Ferrer, J.-L., Bowman, M. E., Dixon, R. A., and Noel, J. P. (2000). Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. *Biochemistry* **39**:890–902.

Jiang, L., Wang, L., Zhang, J., Liu, H., Hong, B., *et al*. (2015). Identification of novel mureidomycin analogues via rational activation of a cryptic gene cluster in *Streptomyces roseosporus* NRRL 15998. *Sci Rep* **5**.

Jiang, M. X., Yin, M., Wu, S. H., Han, X. L., Ji, K. Y., *et al*. (2017). GdmRIII, a TetR family transcriptional regulator, controls geldanamycin and elaiophylin biosynthesis in *Streptomyces autolyticus* CGMCC0516. *Sci Rep* **7**:4803.

Kagoshima, Y., Tokumitsu, A., Masuda, T., Namba, E., Inoue, H., *et al*. (2019). Muraminomicins, novel ester derivatives: *in vitro* and *in vivo* antistaphylococcal activity. *Journal of Antibiotics* **72**:956–969.

Karwowski, J. P., Jackson, M., Theriault, R. J., Chen, R. H., Barlow, G. J., *et al*. (1989). Pacidamycins, a novel series of antibiotics with anti-*Pseudomonas aeruginosa* activity. I. Taxonomy of the producing organism and fermentation. *J Antibiot (Tokyo)* **42**:506–511.

Kaysser, L., Lutsch, L., Siebenberg, S., Wemakor, E., Kammerer, B., *et al*. (2009). Identification and manipulation of the caprazamycin gene cluster lead to new

simplified liponucleoside antibiotics and give insights into the biosynthetic pathway. *Journal of Biological Chemistry* **284**:14987–14996.

Kaysser, L., Siebenberg, S., Kammerer, B., and Gust, B. (2010a). Analysis of the liposidomycin gene cluster leads to the identification of new caprazamycin derivatives. *ChemBioChem* **11**:191–196.

Kaysser, L., Wemakor, E., Siebenberg, S., Salas, J. A., Sohng, J. K., *et al*. (2010b). Formation and attachment of the deoxysugar moiety and assembly of the gene cluster for caprazamycin biosynthesis. *Appl Environ Microbiol* **76**:4008–4018.

Kaysser, L., Tang, X., Wemakor, E., Sedding, K., Hennig, S., *et al*. (2011). Identification of a napsamycin biosynthesis gene cluster by genome mining. *ChemBioChem* **12**:477–487.

Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., and Hopwood, D. A. (2000). Practical *Streptomyces* genetics. John Innes Foundation Norwich.

Kim, J., Adhya, S., Gargest, S., and Pastan, I. (1992). Allosteric changes in the cAMP receptor protein of *Escherichia coli*: Hinge reorientation. *Proc. Natl. Acad. Sci. USA* **89**:9700–9704.

Kim, B., Lee, H. J., Jo, S. H., Kim, M. G., Lee, Y., *et al*. (2022). Study of SarA by DNA affinity capture assay (DACA) employing three promoters of key virulence and resistance genes in methicillin-resistant *Staphylococcus aureus*. *Antibiotics* **11**.

Kimata, K., Takahashi, H., Inada, T., Postma, P., and Aiba, H. (1997). cAMP receptor protein-cAMP plays a crucial role in glucose-lactose diauxie by activating the major glucose transporter gene in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **94**:12914–12919.

Kimura, K. I., Miyata, N., Kawanishi, G., Kamio, Y., Izaki, K., *et al*. (1989). Liposidomycin C inhibits phospho-*N*-acetylmuramyl-pentapeptide transferase in peptidoglycan synthesis of *Escherichia coli* y-10. *Agric Biol Chem* **53**:1811–1815.

Kimura, K., Ikeda, Y., Kagami, S., Yoshihama, M., Suzuki, K., *et al*. (1998). Selective inhibition of the bacterial peptidoglycan biosynthesis by the new types of liposidomycins. *Journal of Antibiotics* **51**:1099–1104.

Kirillov, S. V, Porse, B. T., and Garrett, R. A. (1999). Peptidyl transferase antibiotics perturb the relative positioning of the 3´-terminal adenosine of P/P´-site-bound tRNA and 23S rRNA in the ribosome. *RNA* **5**:1003–1013.

Kisker, C., Hinrichs, W., Tovar, K., Hillen, W., and Saenger, W. (1995). The complex formed between Tet repressor and tetracycline-Mg^{2+} reveals mechanism of antibiotic resistance. *J. Mol. Biol* **247**:260–280.

Körner, H., Sofia, H. J., and Zumft, W. G. (2003). Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: Exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol Rev* **27**:559–592.

Kumar, G., and Amrutha, C. (2023). Natural products and their analogues acting against *Mycobacterium tuberculosis*: A recent update. *Drug Dev Res* **84**:779–804.

Lämmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680–685.

Larwood, D. J. (2020). Nikkomycin Z—ready to meet the promise? *Journal of Fungi* **6**:1–14.

Lee, D. J., Minchin, S. D., and Busby, S. J. W. (2012). Activating transcription in bacteria. *Annu Rev Microbiol* **66**:125–152.

Li, Q., Wang, L., Xie, Y., Wang, S., Chen, R., *et al*. (2013). SsaA, a member of a novel class of transcriptional regulators, controls sansanmycin production in *Streptomyces* sp. strain SS through a feedback mechanism. *J Bacteriol* **195**:2232–2243.

Li, X., Hao, G., Wang, Q., Chen, Z., Ding, Y., *et al*. (2017). Ningnanmycin inhibits tobacco mosaic virus virulence by binding directly to its coat protein discs. *Oncotarget* **8**:82446–82458.

Li, C., Urem, M., Du, C., Zhang, L., and van Wezel, G. P. (2023). Systems-wide analysis of the ROK-family regulatory gene *rokL6* and its role in the control of glucosamine toxicity in *Streptomyces coelicolor*. *Appl Environ Microbiol* **89**:1–23.

Liu, Y. Q., Zhang, J., Feng, G., Li, L. H., Yang, L., *et al*. (2012). Ribavirin, a nucleoside with potential insecticidal activity. *Pest Manag Sci* **68**:1400–1404.

Liu, X., Wang, D., Zhang, Y., Zhuang, X., and Bai, L. (2023). Identification of multiple regulatory genes involved in TGase production in *Streptomyces mobaraensis* DSM 40587. *Engineering Microbiology* **3**:100098.

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**:402–408.

MacNeil, D. J., Gewain, K. M., Ruby, C. L., Dezeny, G., Gibbons, P. H., *et al*. (1992). Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**:61–68.

Madigan, M. T., Martinko, J. M., Stahl, D. A., and Clark, D. P. (2013). *Brock Mikrobiologie*. 13th ed. Hallbergmoos: Pearson.

Martin, R. G., and Rosner, J. L. (2001). The AraC transcriptional activators. *Curr Opin Microbiol* **4**:132–137.

Mashalidis, E. H., Kaeser, B., Terasawa, Y., Katsuyama, A., Kwon, D. Y., *et al*. (2019). Chemical logic of MraY inhibition by antibacterial nucleoside natural products. *Nat Commun* **10**:1–12.

McDonald, L. A., Barbieri, L. R., Carter, G. T., Lenoy, E., Lotvin, J., *et al*. (2002). Structures of the muraymycins, novel peptidoglycan biosynthesis inhibitors. *J Am Chem Soc* **124**:10260–10261.

McErlean, M., Liu, X., Cui, Z., Gust, B., and Van Lanen, S. G. (2021). Identification and characterization of enzymes involved in the biosynthesis of pyrimidine nucleoside antibiotics. *Nat Prod Rep* **38**:1362–1407.

Mini, E., Nobili, S., Caciagli, B., Landini, I., and Mazzei, T. (2006). Cellular pharmacology of gemcitabine. *Annals of Oncology* **17**:7–12.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., *et al*. (2022). ColabFold: making protein folding accessible to all. *Nat Methods* **19**:679–682.

Mori, K., and Aiba, H. (1985). Evidence for negative control of *cya* transcription by cAMP and cAMP receptor protein in intact *Escherichia coli* cells. *J Biol Chem* **260**:14838–14843.

Münch, D., Roemer, T., Lee, S. H., Engeser, M., Sahl, H. G., *et al*. (2012). Identification and *in vitro* analysis of the GatD/MurT enzyme-complex catalyzing lipid II amidation in *Staphylococcus aureus*. *PLoS Pathog* **8**:1–11.

Muramatsu, Y. , Muramatsu, A., Ohnuki, T., Ishii, M. M., Kizuka, M. , *et al*. (2003a). Studies on novel bacterial translocase I inhibitors, A-500359s I. Taxonomy, fermentation, isolation, physico-chemical properties and structure elucidation of A-500359 A, C, D and G. *J Antibiot (Tokyo)* **56**:243–252.

Muramatsu, Y., Ishii, M. M., and Inukai, M. (2003b). Studies on novel bacterial translocase I inhibitors, A-500359s II. Biological activities of A-500359 A, C, D and G. *J Antibiot (Tokyo)* **56**:253–258.

Murray, C. J., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., *et al*. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* **399**:629–655.

Niu, G., and Tan, H. (2015). Nucleoside antibiotics: Biosynthesis, regulation, and biotechnology. *Trends Microbiol* **23**:110–119.

Niu, G., Li, L., Wei, J., and Tan, H. (2013). Cloning, heterologous expression, and characterization of the gene cluster required for gougerotin biosynthesis. *Chem Biol* **20**:34–44.

Nix, D. E., Swezey, R. R., Hector, R., and Galgiani, J. N. (2009). Pharmacokinetics of nikkomycin Z after single rising oral doses. *Antimicrob Agents Chemother* **53**:2517–2521.

Oh, S. Y., Shin, J. H., and Roe, J. H. (2007). Dual role of OhrR as a repressor and an activator in response to organic hydroperoxides in *Streptomyces coelicolor*. *J Bacteriol* **189**:6284–6292.

Ohnishi, Y., Yamazaki, H., Kato, J.-Y., Tomono, A., and Horinouchi, S. (2005). AdpA, a central transcriptional regulator in the A-factor regulatory cascade that leads to morphological development and secondary metabolism in *Streptomyces griseus*. *Biosci. Biotechnol. Biochem* **69**:431–439.

Paget, M. S. B., Chamberlin, L., Atrih, A., Foster, S. J., and Buttner, M. J. (1999). Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J Bacteriol* **181**:204–211.

Park, S. S., Ko, B. J., and Kim, B. G. (2005). Mass spectrometric screening of transcriptional regulators using DNA affinity capture assay. *Anal Biochem* **344**:152–154.

Piette, A., Derouaux, A., Gerkens, P., Noens, E. E. E., Mazzucchelli, G., *et al*. (2005). From dormant to germinating spores of *Streptomyces coelicolor* A3(2): New perspectives from the *crp* null mutant. *J Proteome Res* **4**:1699–1708.

Płachetka, M., Krawiec, M., Zakrzewska-Czerwi Nska, J., Wola Nski, M., and Gralnick, J. A. (2021). AdpA positively regulates morphological differentiation and chloramphenicol biosynthesis in *Streptomyces venezuelae*. *Spectrum Microbiology* **9**:1–17.

Qu, S., Kang, Q., Wu, H., Wang, L., and Bai, L. (2015). Positive and negative regulation of GlnR in validamycin A biosynthesis by binding to different loci in promoter region. *Appl Microbiol Biotechnol* **99**:4771–4783.

Rackham, E. J., Grüschow, S., and Goss, R. J. M. (2011). Revealing the first uridyl peptide antibiotic biosynthetic gene cluster and probing pacidamycin biosynthesis. *Bioeng Bugs* **2**:218–221.

Rey, D. A., Pühler, A., and Kalinowski, J. (2003). The putative transcriptional repressor McbR, member of the TetR-family, is involved in the regulation of the metabolic network directing the synthesis of sulfur containing amino acids in *Corynebacterium glutamicum*. *J Biotechnol* **103**:51–65.

Robak, T., Lech-Maranda, E., Korycka, A., and Robak, E. (2006). Purine nucleoside analogs as immunosuppressive and antineoplastic agents: Mechanism of action and clinical activity. *Curr Med Chem* **13**:3165–3189.

Romero-Rodríguez, A., Robledo-Casados, I., and Sánchez, S. (2015). An overview on transcriptional regulators in *Streptomyces*. *Biochim Biophys Acta Gene Regul Mech* **1849**:1017–1039.

Romero-Rodríguez, A., Rocha, D., Ruiz-Villafan, B., Tierrafría, V., Rodríguez-Sanoja, R., *et al*. (2016). Transcriptomic analysis of a classical model of carbon catabolite regulation in *Streptomyces coelicolor*. *BMC Microbiol* **16**:1–16.

Salomon, J. J., Galeron, P., Schulte, N., Morow, P. R., Severynse-Stevens, D., *et al*. (2013). Biopharmaceutical *in vitro* characterization of CPZEN-45, a drug candidate for inhalation therapy of tuberculosis. *Ther Deliv* **4**:915–923.

Sambrook, J., and Russell, D. W. (1989). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press.

Santana-Garcia, W., Castro-Mondragon, J. A., Padilla-Gálvez, M., Nguyen, N. T. T., Elizondo-Salas, A., *et al*. (2022). RSAT 2022: regulatory sequence analysis tools. *Nucleic Acids Res* **50**:W670–W676.

Schleif, R. (2010). AraC protein, regulation of the L-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiol Rev* **34**:779–796.

Schleifer, K. H., and Kandler, O. (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol Rev* **36**:407–477.

Schulz, S., Sletta, H., Fløgstad Degnes, K., Krysenko, S., Williams, A., *et al*. (2023). Optimization of FK-506 production in *Streptomyces tsukubaensis* by modulation of Crp-mediated regulation. *Appl Microbiol Biotechnol* **107**:2871–2886.

Serpi, M., Ferrari, V., and Pertusati, F. (2016). Nucleoside derived antibiotics to fight microbial drug resistance: New utilities for an established class of drugs? *J Med Chem* **59**:10343–10382.

Sheppard, D., and Englesberg, E. (1966). Positive control in the L-arabinose gene-enzyme complex of *Escherichia coli* B/r as exhibited with stable merodiploids. *Cold Spring Harb Symp Quant Biol* **31**:345–347.

Shiraishi, T., and Kuzuyama, T. (2019). Recent advances in the biosynthesis of nucleoside antibiotics. *Journal of Antibiotics* **72**:913–923.

Shubitz, L. F., Roy, M. E., Nix, D. E., and Galgiani, J. N. (2013). Efficacy of Nikkomycin Z for respiratory coccidioidomycosis in naturally infected dogs. *Med Mycol* **51**:747–754.

Sibinelli-Sousa, S., Hespanhol, J. T., and Bayer-Santos, E. (2021). Targeting the achilles' heel of bacteria: Different mechanisms to break down the peptidoglycan cell wall during bacterial warfare. *J Bacteriol* **203**:e00478--20.

Siebenberg, S., Kaysser, L., Wemakor, E., Heide, L., Gust, B., *et al*. (2011). Identification and structural elucidation of new caprazamycins from *Streptomyces* sp. MK730-62F2 by liquid chromatography/electrospray ionization tandem mass spectrometry. *Rapid Communications in Mass Spectrometry* **25**:495–502.

Stenland, C. J., Lis, L. G., Schendel, F. J., Hahn, N. J., Smart, M. A., *et al*. (2013). A practical and scalable manufacturing process for an antifungal agent, nikkomycin Z. *Org Process Res Dev* **17**:265–272.

Stevenson, C. E. M., Assaad, A., Chandra, G., Le, T. B. K., Greive, S. J., *et al*. (2013). Investigation of DNA sequence recognition by a streptomycete MarR family transcriptional regulator through surface plasmon resonance and X-ray crystallography. *Nucleic Acids Res* **41**:7009–7022.

Stewart, I. E., Durham, P. G., Sittenauer, J. M., Barreda, A. P., Stowell, G. W., *et al*. (2022). Optimization and scale up of spray dried CPZEN-45 aerosol powders for inhaled tuberculosis treatment. *Pharm Res* **39**:3359–3370.

Studier, W. F., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. In *Methods in Enzymology*, pp. 60–89.

Sultana, A., and Lee, J. E. (2015). Measuring protein-protein and protein-nucleic acid interactions by biolayer interferometry. *Curr Protoc Protein Sci* **79**:19.25.1-19.25.26.

Szafran, M. J., Gongerowska, M., Małecki, T., Elliot, M., and Jakimowicz, D. (2019). Transcriptional response of *Streptomyces coelicolor* to rapid chromosome relaxation or long-term supercoiling imbalance. *Front Microbiol* **10**:1–18.

Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., *et al*. (2018). Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect Dis* **18**:318–327.

Takahashi, A., Saito, N., Hotta, K., Okami, Y., and Umezawa, H. (1986). Bagougeramines A and B, new nucleoside antibiotics produced by a strain of *Bacillus circulans*. *J Antibiot (Tokyo)* **39**:1033–1040.

Takahashi, Y., Igarashi, M., Miyake, T., Soutome, H., Ishikawa, K., *et al*. (2013). Novel semisynthetic antibiotics from caprazamycins A-G: Caprazene derivatives and their antibacterial activity. *Journal of Antibiotics* **66**:171–178.

Takatsuki, A., Arima, K., and Tamura, G. (1971). Tunicamycin, a new antibiotic. *J Antibiot (Tokyo)* **24**:215–113.

Tang, X., Eitel, K., Kaysser, L., Kulik, A., Grond, S., *et al*. (2013). A two-step sulfation in antibiotic biosynthesis requires a type III polyketide synthase. *Nat Chem Biol* **9**:610–615.

Thomson, J. M., and Lamont, I. L. (2019). Nucleoside analogues as antibacterial agents. *Front Microbiol* **10**:1–11.

Tong, Y., Charusanti, P., Zhang, L., Weber, T., and Lee, S. Y. (2015). CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth Biol* **4**:1020–1029.

Tran, N. T., Stevenson, C. E., Som, N. F., Thanapipatsiri, A., Jalal, A. S. B., *et al*. (2018). Permissive zones for the centromere-binding protein ParB on the *Caulobacter crescentus* chromosome. *Nucleic Acids Res* **46**:1196–1209.

Uguru, G. C., Stephens, K. E., Stead, J. A., Towle, J. E., Baumberg, S., *et al*. (2005). Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in *Streptomyces coelicolor*. *Mol Microbiol* **58**:131–150.

van der Heul, H. U., Bilyk, B. L., McDowall, K. J., Seipke, R. F., and van Wezel, G. P. (2018). Regulation of antibiotic production in Actinobacteria: New perspectives from the post-genomic era. *Nat Prod Rep* **35**:575–604.

Verma, A., Naik, B., Kumar, V., Mishra, S., Choudhary, M., *et al*. (2023). Revolutionizing tuberculosis treatment: Uncovering new drugs and breakthrough inhibitors to combat drug-resistant *Mycobacterium tuberculosis*. *ACS Infect Dis* **9**:2369–2385.

Wang, K., Su, B., Wang, Z., Wu, M., Li, Z., *et al*. (2010). Synthesis and antiviral activities of phenanthroindolizidine alkaloids and their derivatives. *J Agric Food Chem* **58**:2703–2709.

Wang, J., Yue, T., He, C., Zhou, Y., Bai, Y., *et al*. (2023). Biocontrol of tomato bacterial wilt by a combination of *Bacillus subtilis* GSJB-1210 and ningnanmycin. *Sci Hortic* **321**:112296.

Wei, J., Tian, Y., Niu, G., and Tan, H. (2014). GouR, a TetR family transcriptional regulator, coordinates the biosynthesis and export of gougerotin in *Streptomyces graminearus*. *Appl Environ Microbiol* **80**:714–722.

Wei, K., Wu, Y., Li, L., Jiang, W., Hu, J., *et al*. (2018). MilR2, a novel TetR family regulator involved in 5-oxomilbemycin A3/A4 biosynthesis in *Streptomyces hygroscopicus*. *Appl Microbiol Biotechnol* **102**:8841–8853.

Wiegmann, D., Koppermann, S., Wirth, M., Niro, G., Leyerer, K., *et al*. (2016). Muraymycin nucleoside-peptide antibiotics: Uridine-derived natural products as lead structures for the development of novel antibacterial agents. *Beilstein Journal of Organic Chemistry* **12**:769–795.

Wiker, F., Hauck, N., Grond, S., and Gust, B. (2019a). Caprazamycins: Biosynthesis and structure activity relationship studies. *International Journal of Medical Microbiology* **309**:319–324.

Wiker, F., Konnerth, M., Helmle, I., Kulik, A., Kaysser, L., *et al*. (2019b). Identification of novel α-pyrones from *Conexibacter woesei* serving as sulfate shuttles. *ACS Chem Biol* **14**:1972–1980.

Wu, W., Kang, Y., Hou, B., Ye, J., Wang, R., *et al*. (2023). Characterization of a TetR-type positive regulator AtrA for lincomycin production in *Streptomyces lincolnensis*. *Biosci Biotechnol Biochem* **87**:786–795.

Wyszynski, F. J., Hesketh, A. R., Bibb, M. J., and Davis, B. G. (2010). Dissecting tunicamycin biosynthesis by genome mining: Cloning and heterologous expression of a minimal gene cluster. *Chem Sci* **1**:581–589.

Xie, Y., Chen, R., Si, S., Sun, C., and Xu, H. (2007). A new nucleosidyl-peptide antibiotic, sansanmycin. *J Antibiot (Tokyo)* **60**:158–161.

Xie, Y., Xu, H., Sun, C., Yu, Y., and Chen, R. (2010). Two novel nucleosidyl-peptide antibiotics: Sansanmycin F and G produced by *Streptomyces* sp SS. *Journal of Antibiotics* **63**:143–146.

Xu, D., Liu, G., Cheng, L., Lu, X., Chen, W., *et al*. (2013). Identification of Mur34 as the novel negative regulator responsible for the biosynthesis of muraymycin in *Streptomyces* sp. NRRL30471. *PLoS One* **8**:e76068.

Xu, W., Sun, C., Gao, W., Scharf, D. H., Zhu, C., *et al*. (2023). Degradation mechanism of AtrA mediated by ClpXP and its application in daptomycin production in *Streptomyces roseosporus*. *Protein Science* **32**:e4617.

Yamaguchi, H., Sato, S., Yoshida, S., Takada, K., and Itoh, M. (1986). Capuramycin, a new nucleoside antibiotic. *J Antibiot (Tokyo)* **39**:1047–1053.

Yang, Z., Chi, X., Funabashi, M., Baba, S., Nonaka, K., *et al*. (2011). Characterization of LipL as a non-heme, Fe(II)-dependent α-ketoglutarate: UMP dioxygenase that generates uridine-5′-aldehyde during A-90289 biosynthesis. *Journal of Biological Chemistry* **286**:7885–7892.

Yuan, P. H., Zhou, R. C., Chen, X., Luo, S., Wang, F., *et al*. (2016). DepR1, a TetR family transcriptional regulator, positively regulates daptomycin production in an industrial producer, *Streptomyces roseosporus* SW0702. *Appl Environ Microbiol* **82**:1898–1905.

Zadeh, S. M., Chen, M. H., Wang, Z. C., Astani, E. K., Lo, I. W., *et al*. (2022). β-hydroxylation of α-amino-β-hydroxylbutanoyl-glycyluridine catalyzed by a nonheme hydroxylase ensures the maturation of caprazamycin. *Commun Chem* **5**:1–16.

Zhang, W., Ostash, B., and Walsh, C. T. (2010). Identification of the biosynthetic gene cluster for the pacidamycin group of peptidyl nucleoside antibiotics. *PNAS* **107**:16828–16833.

Zhou, J., Cai, Y., Liu, Y., An, H., Deng, K., *et al*. (2022). Breaking down the cell wall: Still an attractive antibacterial strategy. *Front Microbiol* **13**:952633.

7 Appendix

Table S 1: Homology of functionally corresponding gene products of the caprazamycin (Cpz) and liposidomycin (Lpm) BGCs. The proteins encoded in the gene clusters from native producer strains *S*. sp. MK730-62F2 and *S*. sp. SN-1061M were compared on amino acid level using BLASTp alignment analysis.

Cpz BGC	(Putative) function	Lpm BGC	Identity/similarity [%]
Cpz4	Sulfotransferase	LpmB	32/50
Cpz5	3-hydroxy-3-methylglutaryl-CoA synthase	LpmA	81/89
Cpz6	Type III polyketide synthase	LpmD	83/86
Cpz7	PAPS-phosphatase	LpmE	80/84
Cpz8	PAPS-dependent sulfotransferase	LpmF	84/89
Cpz9	AraC-type regulator	LpmG	69/77
Cpz10	β-hydroxylase	LpmH	88/93
Cpz11	*N*-methyltransferase	LpmI	76/84
Cpz12	Kinase	LpmJ	73/82
Cpz13	Aminotransferase	LpmK	81/86
Cpz14	Hydroxymethyltransferase	LpmL	87/91
Cpz15	Dioxygenase	LpmM	83/89
Cpz16	Nucleotidyltransferase	LpmN	82/91
Cpz17	Glycosyltransferase	LpmO	81/87
Cpz18	Aminotransferase	LpmP	82/87
Cpz19	Pyrimidine phosphorylase	LpmQ	83/88
Cpz20	Acyl-CoA-synthase	LpmR	88/91
Cpz21	Carboxyesterase	LpmS	85/90
Cpz22	ABC transporter	LpmT	79/85
Cpz23	Lipase	LpmU	84/93
Cpz24	Hypothetical protein	LpmV	70/80
Cpz25	Acyl-dehydrogease	LpmW	91/95
Cpz26	Methyltransferase	LpmX	76/84
Cpz27	Kinase	LpmY	77/85
Cpz28	*O*-methyltransferase	-	-
Cpz29	*O*-methyltransferase	-	-
Cpz30	*O*-methyltransferase	-	-
Cpz31	Glycosyltransferase	-	-

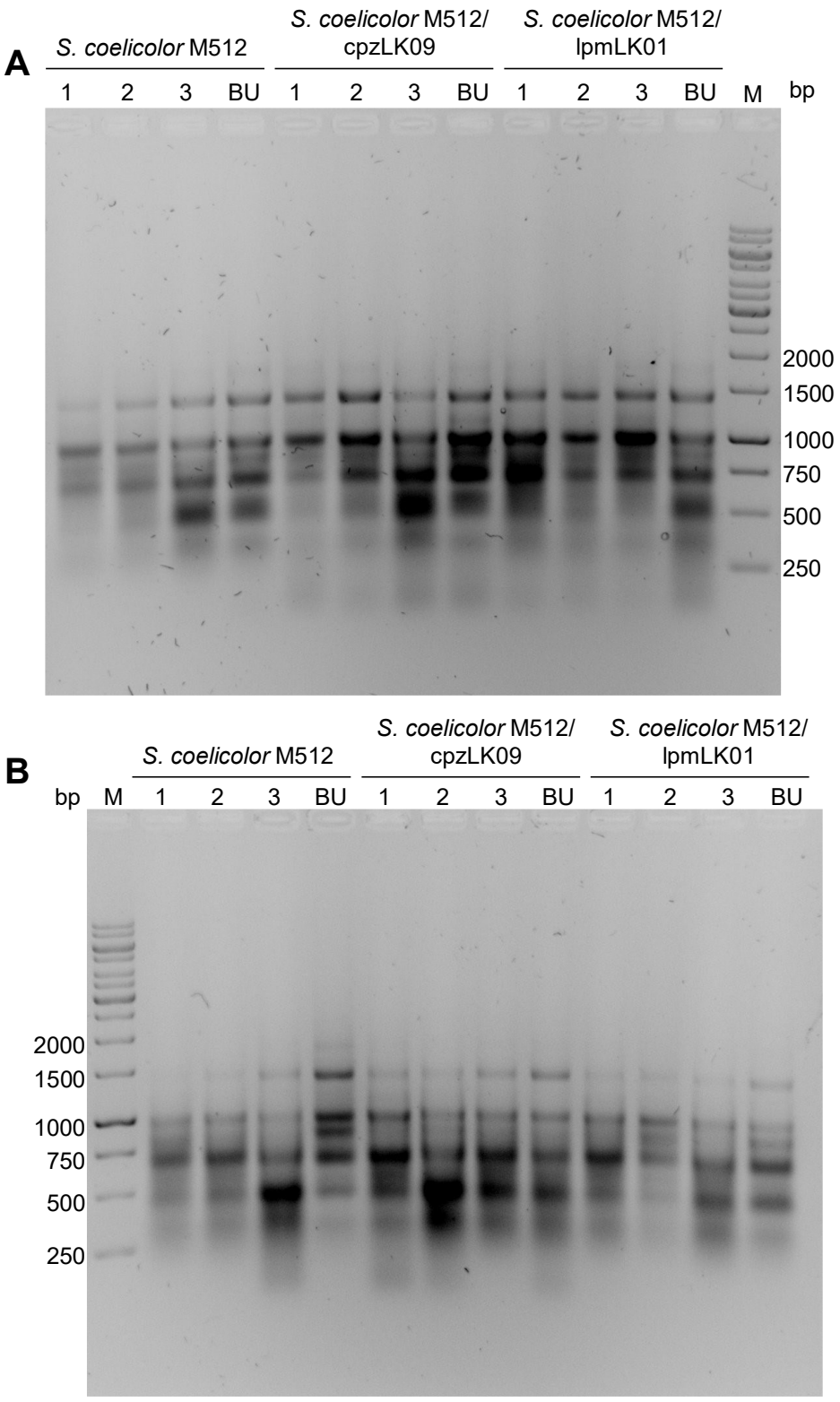


Figure S 1: Agarose gelelectrophoresis of isolated RNA for RNA-sequencing from cultures grown for 2 days (A) and 4 days (B). 4 µL of the isolated and DNase- digested RNA samples was applied to 1 % agarose gels which was run for 1,5 h at 100 V. Bands at the height of approximately 1500 bp and 1000 bp represent 23 S and 16 S rRNA, respectively. BU: backup. These samples were used for preceding optimization of the rRNA depletion process and were not applied in the actual RNA-sequencing.

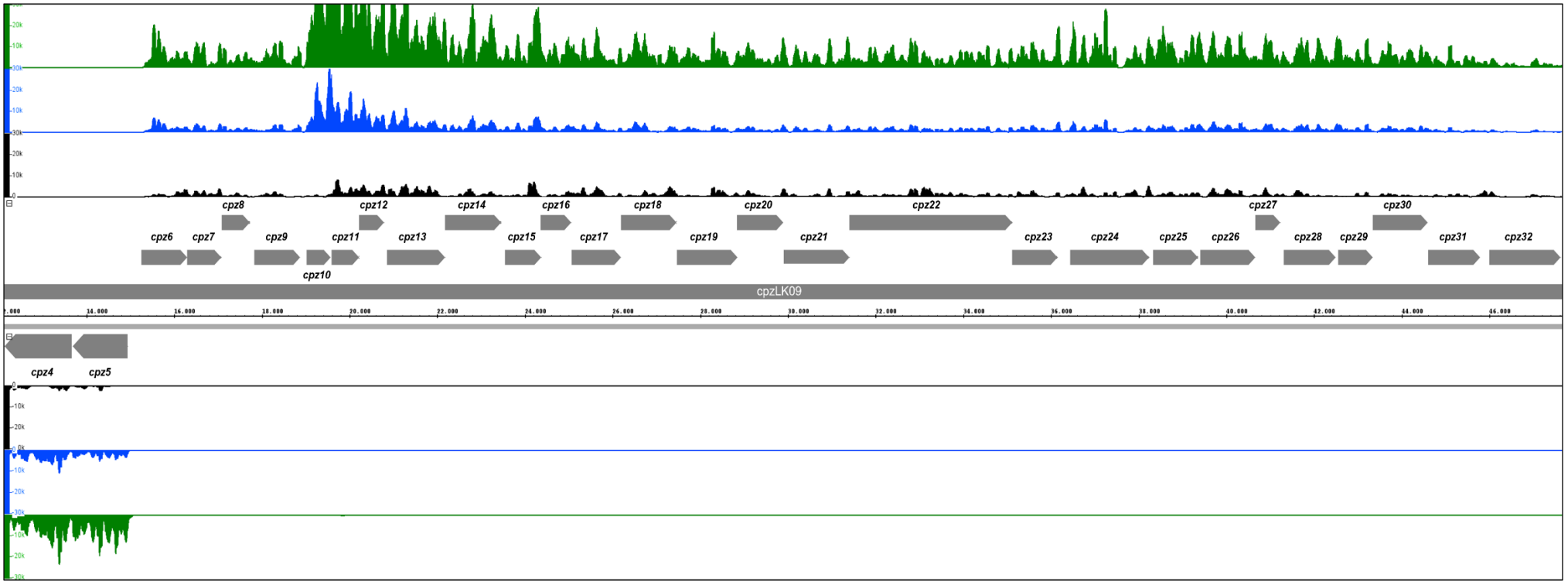


Figure S 2: Transcription of the caprazamycin gene cluster in *S. coelicolor* M512/cpzLK09 after four days of cultivation. The transcriptional pattern of three independent biological replicates is shown. Replicate one is illustrated in black and has a remaining rRNA content of 66 %, replicate two in blue has a rRNA content of 58 % and replicate three in green an rRNA content of 19 %. The genes of the cluster, depicted as arrows, are divided into sense and anti-sense strand and respective coverage plots are mapped against them. To allow a comparative assessment of the data, a threshold value of 30000 transcripts is applied to all replicates.

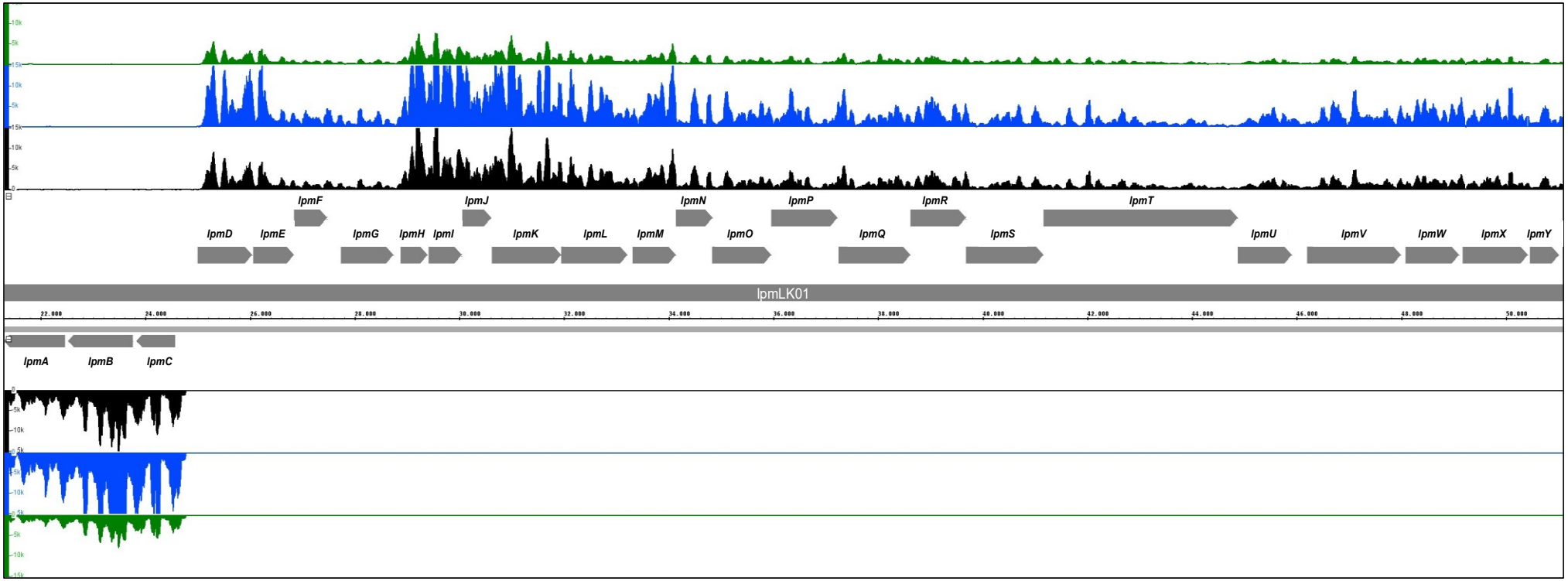


Figure S 3: Transcription of the liposidomycin gene cluster in *S. coelicolor* M512/lpmLK01 after four days of cultivation. The transcriptional pattern of three independent biological replicates is shown. Replicate one is illustrated in black and has a remaining rRNA content of 35 %, replicate two in blue has a rRNA content of 22 % and replicate three in green an rRNA content of 39 %. The genes of the cluster, depicted as arrows, are divided into sense and anti-sense strand and respective coverage plots are mapped against them. To allow a comparative assessment of the data, a threshold value of 15000 transcripts is applied to all replicates.

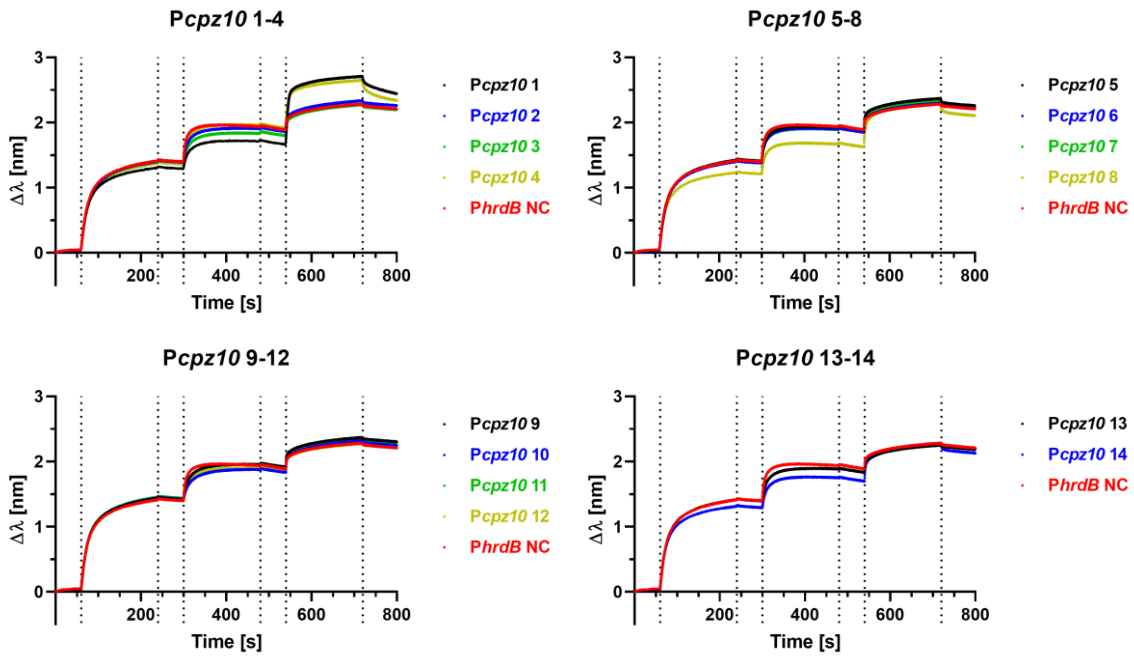


Figure S 4: BLI binding curves of Sco4385 to all segments of P*cpz10*. Dotted lines indicate dipping the sensor into another well. NC: negative control.

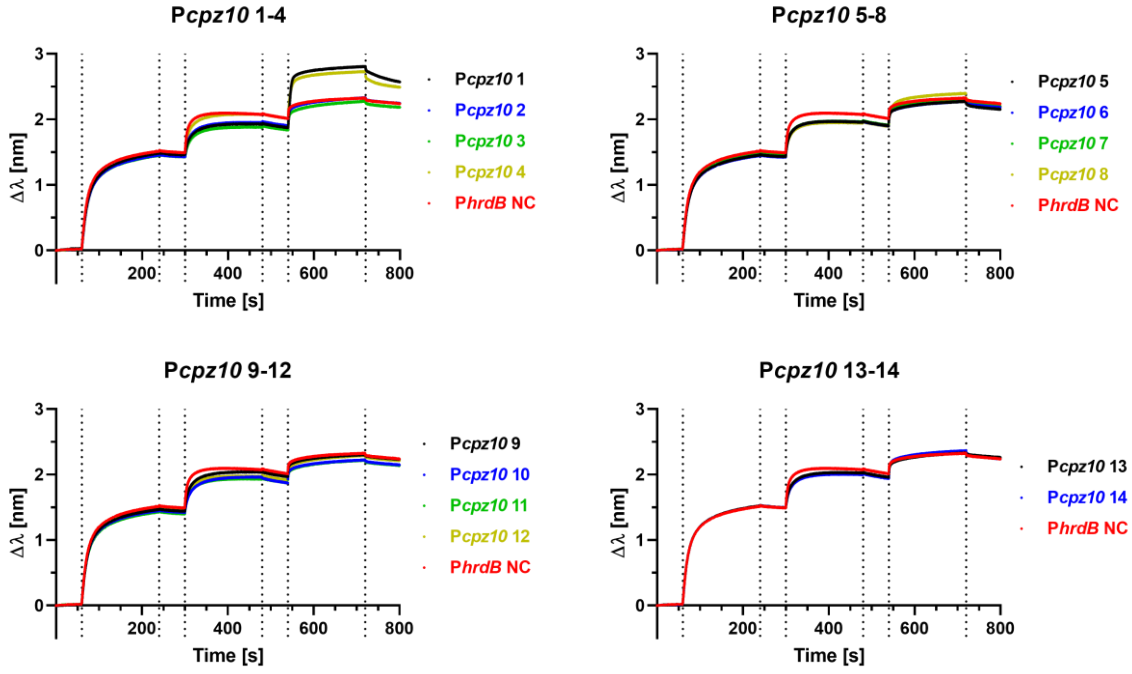


Figure S 5: BLI binding curves of Sco4385 to all segments of P*cpz10* (duplicate). Dotted lines indicate dipping the sensor into another well. NC: negative control.

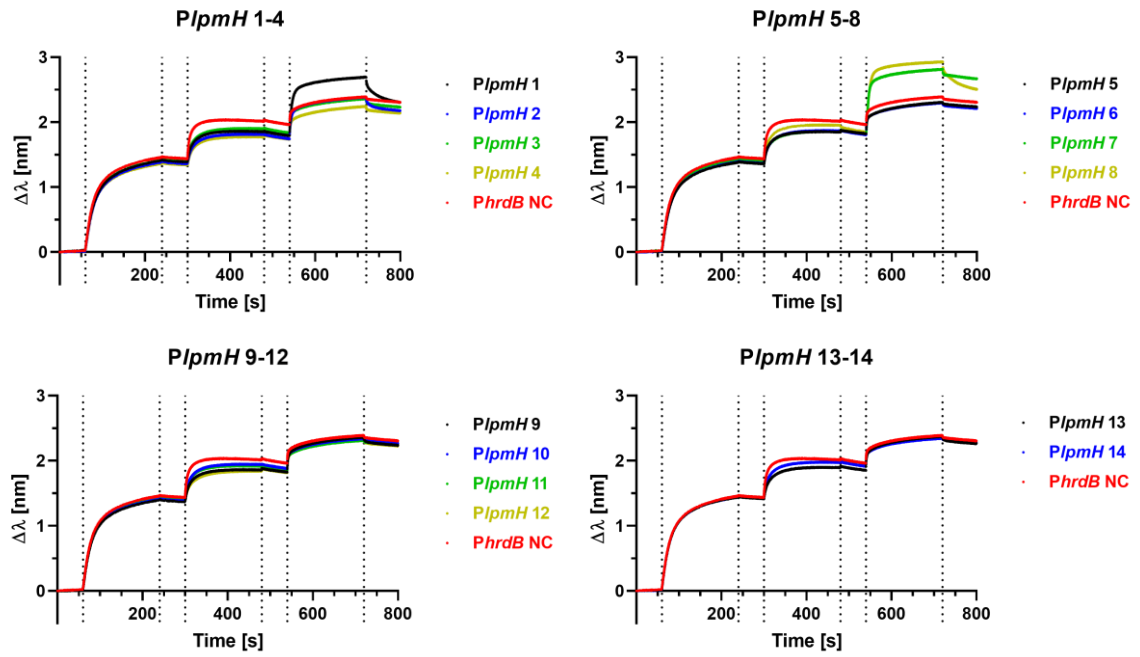


Figure S 6: BLI binding curves of Sco4385 to all segments of P*lpmH*. Dotted lines indicate dipping the sensor into another well. NC: negative control.

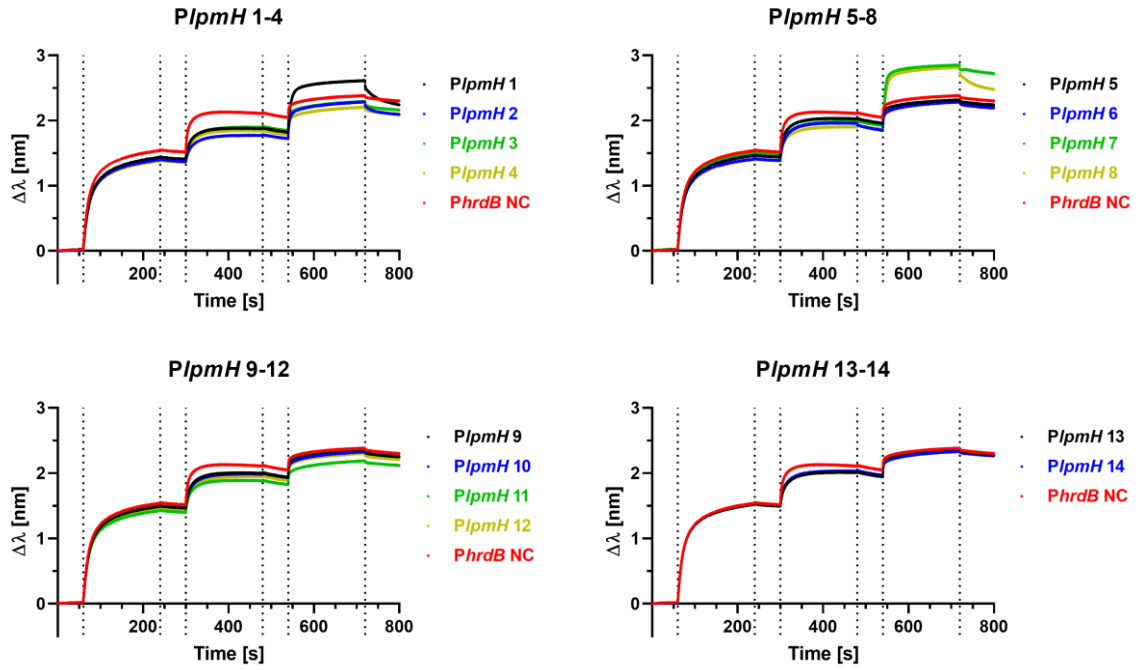


Figure S 7: BLI binding curves of Sco4385 to all segments of P*lpmH* (duplicate). Dotted lines indicate dipping the sensor into another well. NC: negative control.

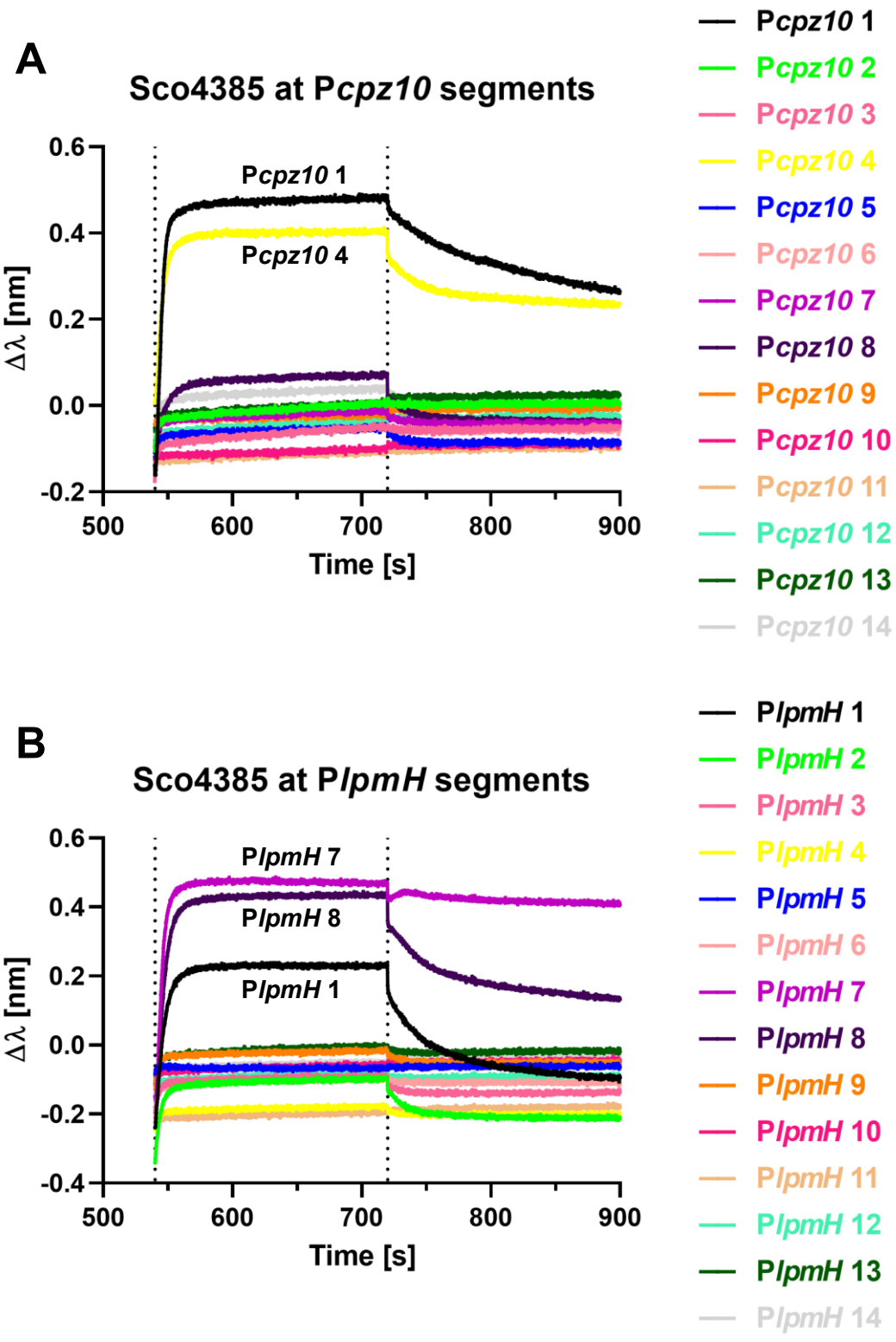


Figure S 8: Wave length shifts (Δλ) and dissociation resulting from Sco4385 binding to segments constituting P*cpz10* (A) and P*lpmH* (B). Values of the negative control P*hrdB* were subtracted from each binding curve, partially resulting in negative binding curves. Bound segments that were used for determination of the consensus sequence are labelled. Dotted lines indicate dipping the sensor into another well.

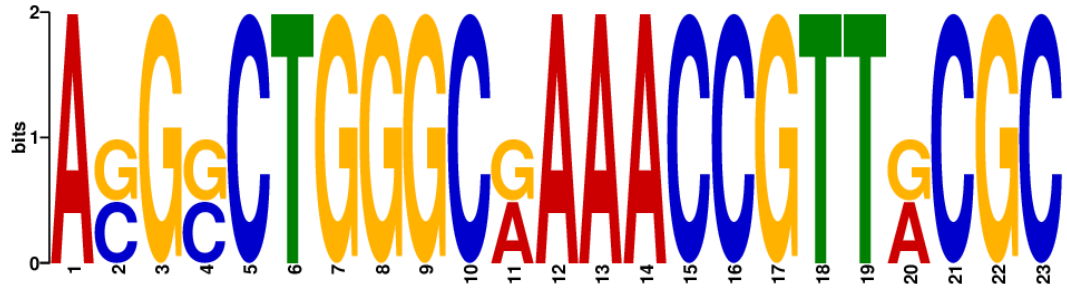


Figure S 9: 23 bp consensus sequence based on P*cpz10*_1_consensus and P*lpmH*_1_consensus, which were both validated as Sco4385 binding sites.

Acknowledgements

Firstly, I would like to thank PD Dr. Bertolt Gust for his excellent supervision over the years, for giving advice, finding supportive words and always having an open door. I would like to thank Prof. Dr. Leonard Kaysser for entrusting me with several wide-ranging projects, for the scientific support and for allowing me to work in great independence.
Further thanks goes to Prof. Dr. Harald Groß for giving helpful scientific advice during group seminar presentations and acting as an examiner.
A huge thank you is dedicated to Andreas Kulik for his kind and competent help with the 20-liter fermentations and conducting a myriad of LC-MS measurements. I would also like to thank Dr. Mirita Franz-Wachtel for mass-spectrometry analysis of the DACA samples and helpful discussions on the results. Prof. Dr. Kay Nieselt and her PhD student Theresa Harbig are acknowledged for supporting me with the analysis and visualization of the RNA-sequencing results.
Furthermore, I thank Prof. Dr. Karl Forchhammer and his PhD student Tim Orthwein for introducing me to biolayer-interferometry and allowing me to use the device independently.
I thank the team of the IGIM Graduate School for offering interesting and helpful seminars and giving me the opportunity to act as their PhD student representative for two years.
A huge thank you goes to Corinna Fischer for her valuable help in all organizational matters.
I would also like to thank my master students Jacob Stock and Felix Emmelmann for their helpful and active cooperation on my projects.
I wholeheartedly thank my colleagues from the pharmaceutical biology department for the wonderful time together, all the tea (and coffee) breaks and the supportive atmosphere. I would like to especially thank my lab-mate Alicia for the fun time together and for introducing me to the basics of working with “Streptos” at the beginning of my studies. Likewise, I thank Tomke for sharing her broad knowledge on DACA studies with me. I warmly thank Hannah, Richard, Patrick, Tarik, Daniel, Irina, Patricia, Roger, Emmanuel, Cathrin, Kerstin, Julia, Johanna, Paula and Isabel for the fantastic time in and outside of the lab. I am sure, many of you will be friends for life.
Lastly, I would like to thank my family from the bottom of my heart for the great support not only during the last four years, but throughout my life. And finally, I am endlessly thankful for you, Simon, for your loving support, for calming me down in stressful moments and simply always being there for me.