Analysis of the transcriptional regulation of the acarbose biosynthesis gene cluster in *Actinoplanes* sp. SE50/110

Dissertation

to obtain the Academic Degree

Doctor of Natural Sciences (Dr. rer. nat.)

Faculty of Biology

Bielefeld University, Germany

submitted by

Julian Droste

born in Bielefeld, Germany

Bielefeld, October 2020

Supervisor:

Prof. Dr. Jörn Kalinowski

Research Group: Microbial Genomics and Biotechnology

Center for Biotechnology (CeBiTec), Bielefeld University

Publications and manuscripts used for this dissertation

Droste, J., Ortseifen, V., Schaffert, L., Persicke, M., Schneiker-Bekel, S., Pühler, A., Kalinowski, J., 2020. The expression dynamics of the *acb* gene cluster in *Actinoplanes* sp. SE50/110 is dependent on the growth phase. BMC genomics. DOI: 10.21203/rs.3.rs-41287/v1 [status: accepted for publication]

Droste, J.*, Wolf, T.*, Gren, T., Ortseifen, V., Schneiker-Bekel, S., Zemke, T., Pühler, A., Kalinowski, J., 2017. The MalR type regulator AcrC is a transcriptional repressor of acarbose biosynthetic genes in *Actinoplanes* sp. SE50/110. BMC genomics 18 (1), 562. DOI: 10.1186/s12864-017-3941-x [status: published] * Authors contributed equally to this study

Droste, J., Kulisch, M., Wolf, T., Schaffert, L., Schneiker-Bekel, S., Pühler, A., Kalinowski, J., 2020. A maltose-regulated large genomic region is activated by MalT in *Actinoplanes* sp. SE50/110. Applied Microbiology and Biotechnology. DOI: 10.1007/s00253-020-10923-2 [status: published]

Additional thematically related publications

Schaffert, L., März, C., Burkhardt, L., **Droste, J.**, Brandt, D., Busche, T., Rosen, W., Schneiker-Bekel, S., Persicke, M., Pühler, A., Kalinowski, J., 2019. Evaluation of vector systems and promoters for overexpression of the acarbose biosynthesis gene acbC in *Actinoplanes* sp. SE50/110. Microbial cell factories, 18(1), 114. DOI: 10.1186/s12934-019-1162-5

Schaffert, L., Schneiker-Bekel, S., Dymek, S., **Droste, J.**, Persicke, M., Busche, T., Brandt, D., Pühler, A., Kalinowski, J., 2019. Essentiality of the maltase AmlE in maltose utilization and its transcriptional regulation by the repressor AmlR in the acarbose-producing bacterium *Actinoplanes* sp. SE50/110. Frontiers in Microbiology, 10, 2448. DOI: 10.3389/fmicb.2019.02448

Schaffert, L., Schneiker-Bekel, S., Gierhake, J., **Droste, J.**, Persicke, M., Rosen, W., Pühler, A., Kalinowski, J., 2020. Absence of the highly expressed small carbohydrate binding protein Cgt improves the acarbose formation in *Actinoplanes* sp. SE50/110. Applied Microbiology and Biotechnology 104(12), 5395-5408. DOI: 10.1007/s00253-020-10584-1 [status: published]

Droste, J., Rückert, C., Kalinowski, J., Hamed, M. B., Anné, J., Simoens, K., Bernaerts, K., Economou, A., Busche, T., 2020. Extensive reannotation of the genome of the model Streptomycete *Streptomyces lividans* TK24 based on transcriptome and proteome information. Frontiers in Microbiology. [status: submitted for publication]

Contents

List of Abbreviations

A	adenine
acb	acarbose biosynthesis
BLAST	basic local alignment search tool
C	cytosine
Cas	CRISPR associated proteins
cDNA	complementary DNA
CDS	coding DNA sequence
CRISPR	clustered regulatory interspaced short palindromic repeats
G	guanine
G+C	guanine + cytosine
gRNA	guide RNA
mRNA	messenger RNA
ncRNA	non-coding RNA
NRPS	nonribosomal peptide synthetase
ORF	open reading frame
PKS	polyketide synthase
RBS	ribosomal binding site
RNA-seq	cDNA sequencing
rRNA	ribosomal RNA
RT-qPCR	reverse transcription quantitative PCR
SNPs	single-nucleotide polymorphism
sp.	species, singular
spp.	species, plural
sRNA	small RNA
T	thymine
TLS	translation start site
tRNA	transfer RNA
TSS	transcription start site
U	uracil
UTR	untranslated region

List of Figures

1. Summary

Actinoplanes sp. SE50/110 is an industrial relevant producer of the diabetes drug acarbose (acarviosyl-1,4-maltose). Acarbose has been used in the treatment of diabetes since 1990 due to its inhibitory effect on α-glucosidases in the human intestine. In the last decades acarbose biosynthesis has been intensively studied. However, a key aspect to better understand acarbose biosynthesis has been missing so far: The enlightenment of the transcriptional regulation of the acarbose biosynthesis.

In this project, important steps for elucidation of the acarbose biosynthesis regulation were made. Since acarbose is produced in a growth-dependent manner, a genome-wide study analyzing the expression profile of all genes during growth was performed. This way, important information on transcription, protein abundancies and co-regulation of several genes and operons were collected. The transcriptome data set was used to generate a high-quality TSS database revealing the operon structure, 5´-UTRs and promoter motifs in the *Actinoplanes* sp. SE50/110 genome. It could be shown that the transcription of all *acb* genes shows a similar course as the specific acarbose formation rate. Thereby, the growth dependency of acarbose formation could be confirmed. However, several Acb proteins were found to be more stable compared to others based on the observation that protein abundancies of some gene products decrease stronger during growth. This could indicate bottlenecks in the acarbose biosynthesis pathway.

Nevertheless, no hints on transcriptional regulation of the acarbose biosynthesis gene cluster are available, since regulator genes are missing inside or close to the *acb* gene cluster. Therefore, in a second part of the project the comparison of different acarbose biosynthesis gene clusters were chosen to identify candidates for regulatory genes of the *acb* gene cluster. Strikingly, a regulator gene could be identified in the acarbose biosynthesis gene clusters of *Streptomyces glaucescens* GLA.O and *Streptomyces coelicoflavus* ZG0656, to which a homolog (AcrC) was found in the genome of *Actinoplanes* sp. SE50/110. The deletion of this regulator gene (*acrC*) revealed an effect on the two acarbose biosynthesis genes *acbD* and *acbE*. Both genes show an increased transcription in the regulator mutant. This regulatory effect could be confirmed by band shift assays. Thereby, a conserved binding site of AcrC upstream of the TSS of *acbD* and *acbE* was found.

Based on previous findings on dependencies between acarbose biosynthesis and carbon source, in the last part of this project the effect of two different sugars on transcription of the *acb* genes was analyzed. Acarbose production in *Actinoplanes* sp. SE50/110 is highly activated on maltose as a carbon source, whereby the addition of glucose as single carbon source resulted in the absence of acarbose formation. Therefore, a differential transcriptome analysis was performed growing the wild type strain on either maltose or glucose. This experiment revealed a large genomic region, which is highly transcribed when maltose was supplemented to the medium. However, no connection to acarbose or maltose metabolism could be identified by *in silico* functional analyses of the corresponding gene products. A transcriptional regulator (ACSP50_3915) similar to MalT regulators in other bacteria was confirmed as the responsible transcription factor of these genes. It could be shown that MalT is a transcriptional activator dependent on maltose or other maltose-derived metabolite as an effector. A MalT binding site was identified in the -35 promoter region of these genes, which is similar to MalT binding sites identified in other bacteria, like *E. coli*. Therefore, this region was named maltose-regulated large genomic region (MRLGR). It can be assumed that this genomic region harbors genes important for specific habitats of *Actinoplanes* sp. SE50/110.

In conclusion, in this work important steps for understanding transcriptional regulation in response to growth and the available carbon source in *Actinoplanes* sp. SE50/110 with specific focus on acarbose biosynthesis were made. Potential bottlenecks could be identified through the analysis of the *acb* expression profile during cultivation. Furthermore, the first transcription factor of *acb* genes was identified and characterized in this project. This knowledge will help to better understand regulatory effects during growth and the connection of different pathways of *Actinoplanes* sp. SE50/110. Analyzing the expression profile in different growth phases is very important for the identification of genes or gene clusters as potential candidates for genome reduction or the reduction of metabolic burden during strain development. These findings will help to optimize *Actinoplanes* sp. SE50/110 for the production of acarbose or other high-value products in the future.

2. Introduction

2.1 The acarbose producing strain *Actinoplanes* sp. SE50/110

The genus *Actinoplanes* is part of the family *Actinomycetes*, which in turn belongs to the phylum *Actinobacteria* (Frommer et al., 1979; Ludwig et al., 2012). Many members of the genus *Actinoplanes* could be isolated from soil samples in various areas of the world. A classification in species and subgenera first took place in 1963 by Couch. *Actinoplanes* spp. are aerobic living Gram positive bacteria, which are characterized by a slow filamentous hyphae growth. The slimy appearing mycelium has similarities to the fungi, *Streptomyces* and *Micromonospora*. In contrast to the members of *Micromonospora* there is no aerial mycelium formed (Bland et al., 1981; Couch, 1950; Parenti and Coronelli, 1979). The color of the *Actinoplanes* mycelium was described as orange (Szaniszlo, 1967), which is probably due to carotenoid biosynthesis. However, also other colors like yellow, brown, red, blue, violet, green, or even black were found (Parenti and Coronelli, 1979; Vobis, 2006).

An optimal growth is observed at a temperature of 28 to 30 °C. The natural habitats are mainly in the field of marine areas with fresh water springs, such as ponds or brooks (Lee, 2002). On solid medium *Actinoplanes* forms small compact colonies with defined contours and a diameter of 3.5 to 4 mm (Parenti and Coronelli, 1979). The mycelium consists of thin, highly branched hyphae with a diameter of 0.5 to 1.0 μm. The spherical spores are flagellated and are produced in round or globular sporangia, which are formed by the decay of internal hyphae directly at or shortly after the branch (Buchanan and Gibbons, 1986; Lee, 2002; Parenti and Coronelli, 1979; Uribe, 2001).

The cell wall of *Actinoplanes* spp. contains the rare cell wall components 2,6-diaminopimelic acid (DAP) and hydroxy-diaminopimelic acid (HDAP). The ratio of the two substances varies between the different strains (Parenti and Coronelli, 1979). It was found to be resistant to lysozyme (Vobis, 1989).

Bacteria of the genus *Actinoplanes* undergo a more complex lifecycle compared to other bacteria. Cell alter between vegetative growth in mycelia and the formation of sporangia to reach new habitats. This lifecycle is dependent on environmental conditions, like aquatic or terrestrial habitats (Vobis et al., 2015). A sufficient amount of moisture lead to the release of flagellated motile spores from sporangia. These sporangia are equipped with chemotactic properties (Palleroni, 1976). This way,

Actinoplanes spp. are able to reach a broad range of habitats. Different strains could be isolated from rivers (Willoughby, 1971), shores (Jensen et al., 1991), but also deserts and sand dunes (Garrity et al., 1996). However, their favorable habitats are located in tropical and subtropical regions, since they prefer frequent drying and moisture cycles (Vobis, 2006).

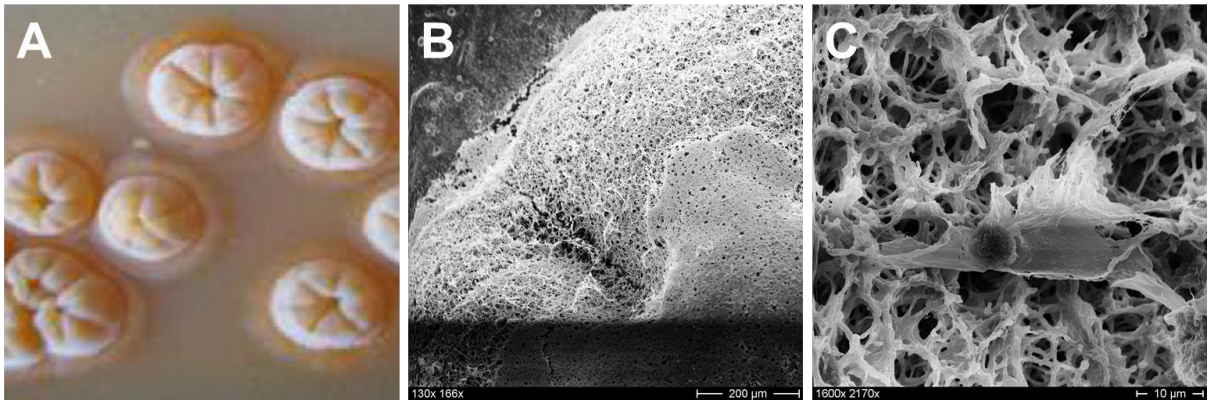

Figure 1: Morphology of *Actinoplanes* sp. SE50/110 colonies grown on SFM medium (A) and electron microscopy pictures of these colonies (B and C). (**A**) Orange, "flower" shaped, round single and joined colonies of *Actinoplanes* sp. SE50/110. (**B**) Colony surface abundantly covered with sporangia-like structures. (**C**) Colony surface with typical structures, presumably round sporangia and substrate mycelia. Pictures were provided by K. Niehaus, H. Bednarz, S. Wendler and V. Ortseifen.

Actinoplanes spp. produce a broad range of pharmaceutically relevant secondary metabolites like antibiotics (lipiarmycin, teichomycin or taitomycin) (Cooper et al., 1992; Parenti and Coronelli, 1979). Moreover, members of the genus became industrial relevant for their ability to produce α-glucosidase inhibitors as in the 70s the strain SE50 (ATCC 31042) was isolated (Schmidt et al., 1977). The strain SE50/110 (ATCC 31044) is a spontaneous mutant thereof and shows a particularly good acarbose production of up to 1.0 g L^{-1} (Frommer et al., 1979).

Culture supernatants of these strains showed an inhibitory effect on microbial α-glucosidases, like α-amylases, sucrases and maltases. The inhibitory effect also shows a clear dependence on the respective cultivation conditions. Mainly the selected carbon source is essential for productivity and the specificity of the inhibitor (Frommer et al., 1979; Schmidt et al., 1977). The inhibitory effect was attributed to the pseudotetrasaccharide acarbose in 1981 by Truscheit et al.

For this reason, *Actinoplanes* sp. SE50 and its natural variant *Actinoplanes* sp. SE50/110 (ATCC 31044) serve as model organism for both the production of α-glucosidase inhibitors as well as the analysis of the acarbose biosynthesis (Hemker et al., 2001; Stratmann, 1997; Stratmann et al., 1999).

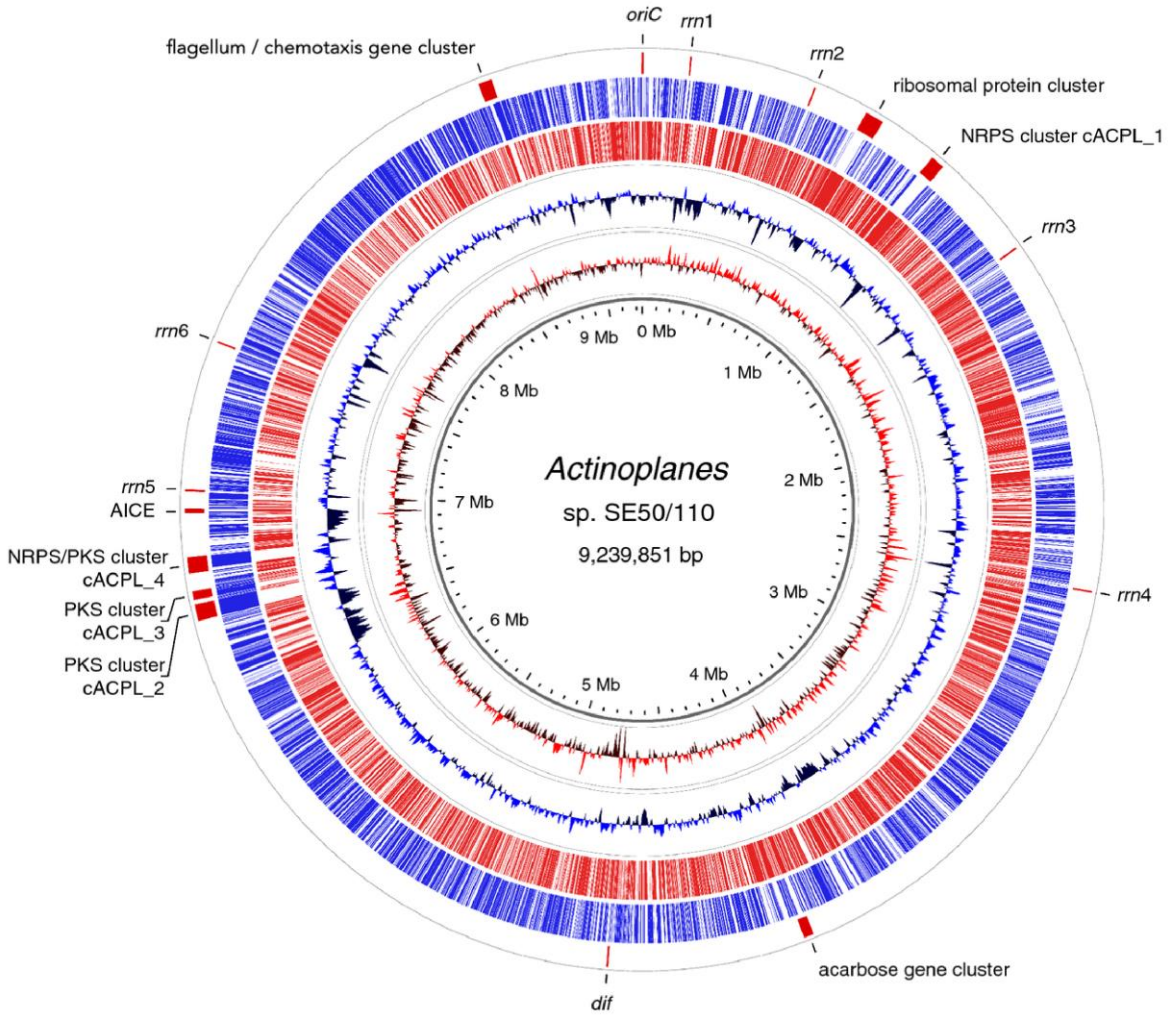

Figure 2: Plot of the complete genome of *Actinoplanes* sp. SE50/110. The genome consists of 9,239,851 base pairs and 8,270 predicted coding sequences. The circles represent from the inside: 1, scale in million base pairs; 2, GC skew; 3, GC content (blue above and black below genome average); 4, genes in backward direction; 5, genes in forward direction; 6, gene clusters and other sites of special interest. Abbreviations were used as follows: oriC, origin of replication; dif, chromosomal terminus region; rrn, ribosomal operon; NRPS, nonribosomal peptide synthetase; PKS, polyketide synthase; AICE, actinomycete integrative and conjugative element (Schwientek et al., 2012).

The genome of *Actinoplanes* sp. SE50/110 has a high content of guanine and cytosine (G+C of 71.32 %). This is characteristic for a member of the *Actinobacteria*. The

genome was the first time completely sequenced and annotated in 2012 by Schwientek et al. with a size of 9.239.851 bp including 8.270 protein-coding sequences. An illustration of the genome with some important gene clusters is shown in Figure 2. The genome sequence and annotation was refined recently by Wolf et al., 2017b.

Optimized strains of *Actinoplanes* sp. were used for the industrial production of acarbose since 1990 by Bayer AG. Acarbose is marketed worldwide under the name Glucobay® and is used in the treatment of diabetes (Schwientek, 2012; Wehmeier and Piepersberg, 2004; Wendler et al., 2013).

2.2 Biosynthesis and biological function of acarbose

In the 1970s, while searching for new inhibitors of microbial and mammalian glucosidases, a new group of oligosaccharide-based inhibitors could be identified (Schmidt et al., 1977). It was shown, that all substances of this class are produced by Actinomycetes (Wehmeier and Piepersberg, 2004). In 1979 the inhibitory effect could be attributed to the pseudotetrasaccharide acarbose (O {4,6-dideoxy-4 [1S-(1,4,6/5)-4,5,6-trihydroxy-3-hydroxymethyl-2-cyclohexen-1-yl]-amino-α-D-glucopyranosyl}-(1→4)-O-α-D-glucopyranosyl-(1→4)-D-glucopyranose) (Truscheit et al., 1981). The structure of acarbose is shown in Figure 3.

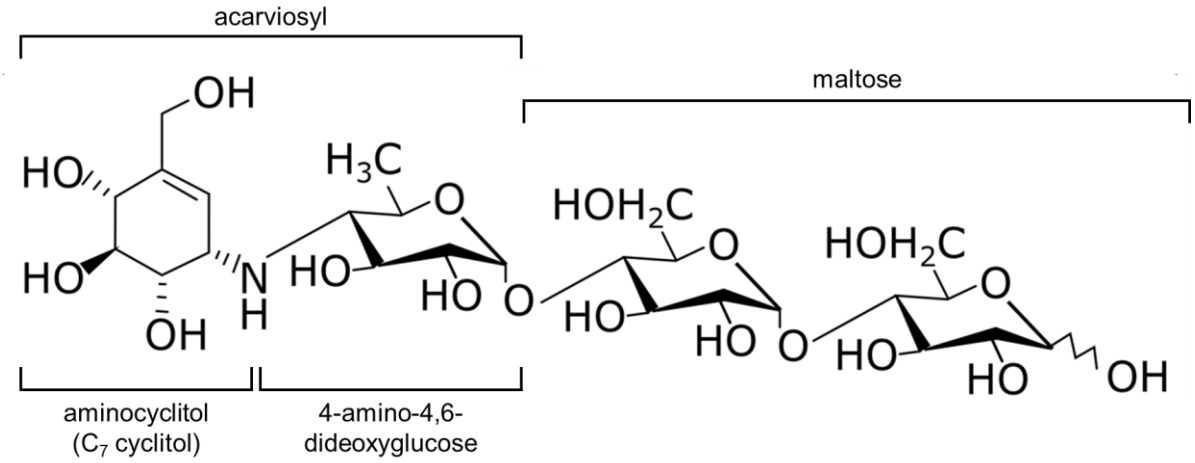

Figure 3: Structure of the pseudotetrasaccharide acarbose, which is naturally produced by *Actinoplanes* sp. SE50/110. The backbone consists of a pseudodisaccharide with a C_7-cyclitol unit (valienamine) and a N-glycoside linked 4-amino-4,6-deoxyglukose. In case of acarbose the core structure is linked via α-1,4-bond to maltose (two glucose units) (Bowers et al., 2002).

The inhibitory effect of acarbose is based on the acarviosyl subunit (acarviose), which consists of an unsaturated aminocyclitol (valienamine) and 4,6-dideoxyglucose, that are N-glycosidically linked. This bond cannot be hydrolyzed. As a result, acarviose can serve as an inhibitor of α-glucosidases (Heiker et al., 1981; Nahoum et al., 2000; Wehmeier, 2004). The affinity of the pseudotetrasaccharide acarbose is up to 15,000-fold higher for sucrases than that of the natural substrate sucrose. Thus, a strong enzyme-inhibitor-complex is formed (Wehmeier, 2004).

In addition to acarbose more pseudooligosaacharides were produced by *Actinoplanes* and closely related species, which differ in the number of α-1,4-linked glucose units at

the conserved acarviose backbone (Wehmeier, 2004). Depending on the number and position of substituents, various enzymes can be inhibited. The strength of the inhibition varies depending on the respective residue (Frommer et al., 1979; Schmidt et al., 1977; Truscheit et al., 1981). The number of glucose residues varies with the used carbon source. While carbon sources with low molecular weight (e.g. glucose) result primarily in the production of inhibitors of maltases and disaccharidases. Efficient inhibitors of α-amylases can be formed by adding starch or similar high molecular weight substances (Frommer et al., 1979; Wehmeier, 2004).

2.2.1 The clinical relevance of acarbose

Acarbose is marketed since 1990 by Bayer AG under the name Glucobay® and is used for the treatment of diabetes in more than 40 countries (Bischoff et al., 1994; Wehmeier, 2004). Diabetes is a chronic disease in which the patient is unable to form new or to use the existing insulin. Insulin promotes the uptake of glucose from the blood to muscles and fat tissue. Due to lack of insulin, the blood sugar levels of diabetic patients is increased after ingestion, leading to more serious sequela, such as renal failure, obesity and heart attack (Schatz, 2006).

The number of diabetic patients is increasing steadily. In 2020, there are approximately 463 million sufferers and by 2035 this number will presumably increase up to 529 million (International Diabetes Federation, 2014). This could be due to today´s diet, which often contains a high proportion of easily degradable carbohydrates (Caspary and Graf, 1979).

Acarbose functions as an α-glucosidase inhibitor (Truscheit et al., 1981). In this way, the intestinal hydrolyzing enzymes (α-glucosidases) are inhibited (Puls et al., 1977) and the cleavage of carbohydrates in the intestines is limited (Schmidt et al., 1977). Acarbose inhibits glucoamylases as well as sucrases and maltases in the human intestine. The inhibition is characterized by an increased affinity (10,000 to 100,000-fold) compared to the regular substrates (Caspary and Graf, 1979). Thus, a rapid breakdown of carbohydrates to monosaccharides (e.g. glucose) is prevented (Wehmeier, 2004).

2.2.2 Biosynthesis and metabolism of acarbose

Despite the well-established production process of acarbose, the biosynthesis of acarbose in *Actinoplanes* sp. SE50/110 is not fully clarified (Wehmeier, 2004; Wendler et al., 2013). Already in 1987, it was postulated that the biosynthesis of acarbose begins with the formation of the cyclitol unit, which is formed by the ring closure of heptulose derived from the pentose phosphate pathway (Degwert et al., 1987). This was described as the transition from the primary to secondary metabolism (Bowers et al., 2002; Degwert et al., 1987). The C7 cyclitol synthase AcbC was identified as the responsible enzyme for the key reaction of sedo-heptulose-7-phosphate to 2-epi-5-epi-valiolon (Stratmann et al., 1999). Starting from sedo-heptulose-7-phosphate all steps of acarbose biosynthesis in *Actinoplanes* sp. SE50/110 are illustrated in Figure 4, which is based on models of Zhang et al. 2002, 2003; Wehmeier and Piepersberg, 2004; Zhang et al., 2020 and Zhao et al., 2020.

The model in Figure 4 shows, that after the formation of 2-epi-5-epi-valiolon a phosphorylation of the primary hydroxyl group takes place by the ATP-dependent kinase AcbM (Zhang et al., 2002). This protects the cell during intracellular synthesis of acarbose against the inhibitory effect of acarbose (Drepper and Pape, 1996). After that, epimerization is carried out by the cyclitol-7-phosphate-2-epimarase AcbO resulting in the intermediate 5-epi-valiolon-7-phosphate (Zhang et al., 2003). The further steps are catalyzed by the enzymes AcbL, AcbN, AcbU and AcbR, thereby finally formed NDP-1-epi-valienol-7-phosphate (Wehmeier and Piepersberg, 2004; Wendler et al., 2013; Zhang et al., 2002).

The synthesis of the deoxy sugar by the dTDP hexose pathway (Piepersberg and Distler, 2001) is parallel to the above-described cyclitol biosynthesis. For this purpose, D-glucose-1-phoshate is converted to dTDP-4-amino-4,6-dideoxy-D-glucose by AcbA, AcbB and AcbV (Piepersberg et al., 2002; Stratmann et al., 1999; Wehmeier, 2003; Wehmeier and Piepersberg, 2004).

Finally, the enzyme AcbS catalyzes the reaction to form the pseudodisaccharide acarviose from NDP-1-epi-valienol-7-phosphate and dTDP-4-amino-4,6-dideoxy-D-glucose (Zhang et al., 2002). The formation of acarbose-7-phosphate is probably catalyzed by the enzymes AcbI and AcbJ which transfer the maltose residue by an α-1,4-N-glycosidic bond to acarviose (Rockser and Wehmeier, 2009; Wendler et al., 2013). The export of acarbose-7-phosphate and simultaneous dephosphorylation is

mediated by the ABC transporter AcbWXY (Piepersberg et al., 2002; Wehmeier and Piepersberg, 2004).

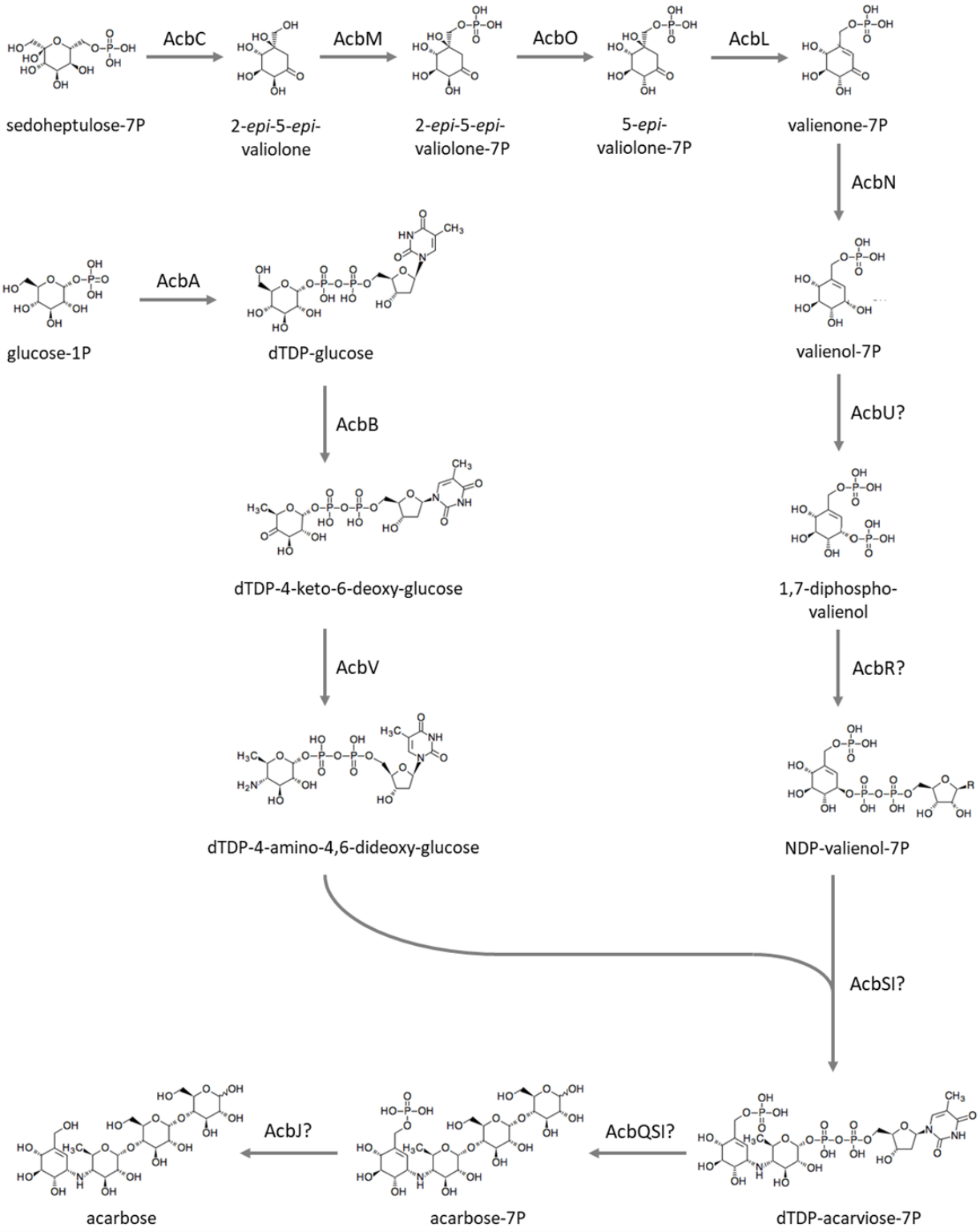

Figure 4: Model of acarbose biosynthesis pathway in *Actinoplanes* sp. SE50/110 according to Zhang et al., 2002; Zhang et al., 2020; Zhao et al., 2020. Enzymatic steps marked with an "?" are postulated but not experimentally proven.

Acarbose is supposed to have two different functions in *Actinoplanes* sp. SE50/110. On the one hand acarbose is responsible for the cells sugar supply (Wehmeier and Piepersberg, 2009) and on the other hand the inhibition of α-glucosidases so that nutrient competitors of *Actinoplanes* sp. SE50/110 can be hindered in growing (Merettig, 2009). Because it is possible to attach further glucose units on acarbose, *Actinoplanes* can be supplied with glucose by reimport of loaded acarbose. For this, carbohydrates are cleaved by acarbose-resistant α-amylases (AcbE and AcbZ) of *Actinoplanes* (Wehmeier, 2003). The resulting monosaccharides are transferred to acarbose by the acarviosyl transferase AcbD (Hemker et al., 2001; Ortseifen, 2016). The import of loaded acarbose is probably realized by the transport system MalEFG (Licht et al., 2011; Wendler et al., 2013). Due to the inhibitory effect to α-amylases acarbose must be re-phosphorylated to protect the cytoplasmic enzymes. This reaction is catalyzed by the acarbose-7-kinase AcbK. The release of the monosaccharides in the cell is performed by the amylomaltase AcbQ (Drepper and Pape, 1996; Rockser and Wehmeier, 2009; Wehmeier and Piepersberg, 2004). This is also known as “carbophor“ function of acarbose (Merettig, 2009).

Extracellular acarbose biosynthesis was investigated in previous studies (Ortseifen, 2016; Wendler et al., 2013). The proposed model is illustrated in Figure 5.

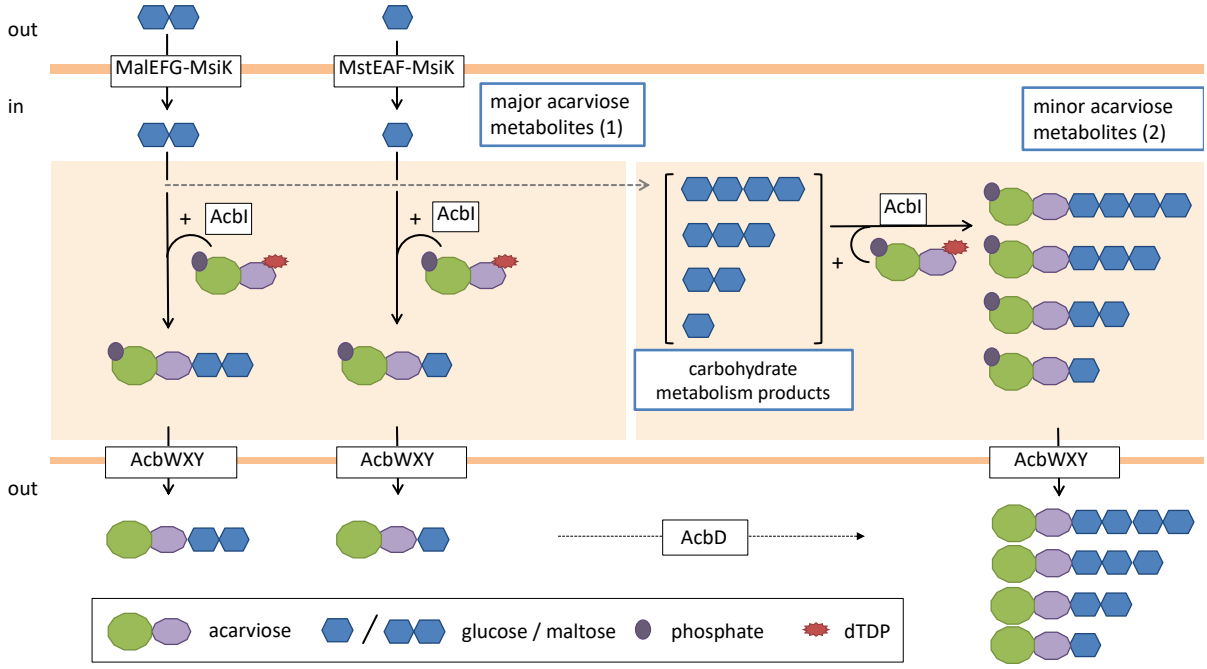

Figure 5: Proposed model of the intra- and extracellular biosynthesis of various acarviose-based metabolites dependent on the available carbon source (left: maltose, right: glucose) in *Actinoplanes* sp. SE50/110 (modified from Wendler et al., 2014).

2.2.3 The acarbose biosynthesis gene cluster (*acb* gene cluster)

Already prior whole genome sequencing of *Actinoplanes* sp. SE50/110 was performed (Schwientek et al., 2012) the *acb* gene cluster (GenBank: Y18523.4) was identified as acarbose biosynthesis gene cluster. In 1999 Stratmann et al. could determine the first genes of this cluster and assign the respective function in connection with the acarbose biosynthesis. The organization of genes in a cluster, which are involved in the biosynthesis of such a metabolite, is typical of organisms of the order actinomycetales (Martin, 1992) and can be observed for example in the production of antibiotics in *Streptomyces* spp.

With the sequencing of the entire genome by Schwientek et al., 2012, all genes of the *acb* gene cluster could be identified, having a total length of about 32 kb. However, of 25 annotated genes in this region only 22 belong to the *acb* gene cluster. In 2011, to the genes *acbFGH* the function of a galactose transporter could be assigned, whereby they are not functionally involved in acarbose biosynthesis (Licht et al., 2011). Previously it was assumed that the gene products of *acbFGH* have the function of an ABC transporter for the import of acarbose (Brunkhorst et al., 2005; Wehmeier, 2003; Wehmeier and Piepersberg, 2004). Alternatively, MalEFG has been suggested as a possible transport system of acarbose (Wendler et al., 2013).

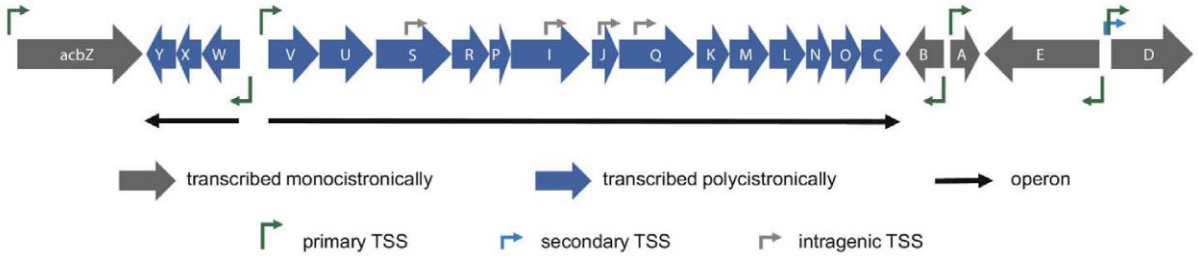

Figure 6: The acarbose biosynthesis gene cluster (*acb* gene cluster) of *Actinoplanes* sp. SE50/110 with the containing transcription start sites (TSS) and the operon structure based on the data of RNA sequencing. The genes are also categorized according to their transcriptional organization: transcribed monocistronically (grey) and transcribed polycistronically (blue) (Wolf et al., 2017b).

The 22 genes of the *acb* gene cluster encoding various enzymes for biosynthesis and a transport system (Figure 6). This includes enzymes for the synthesis of the deoxy sugar and the cyclitol unit of acarbose, enzymes for the extracellular starch degradation and enzymes for the intracellular modification of acarbose. Furthermore,

there are glycosyl transferases and an ATP-dependent transporter (Wehmeier, 2003). Figure 4 shows the *acb* gene cluster with the corresponding genes and the gene products with their function and localization.

The transcription of 22 genes occurs in 7 different transcription units: *acbZ*, *acbWXY*, *acbVUSRPIJQKMLNOC*, *acbB*, *acbA*, *acbE* and *acbD* (Wolf et al., 2017b). The genes *acbA* and *acbB* just like *acbD* and *acbE* are located in opposite direction and share a common intergenic region. This upstream region potentially harbors regulatory elements for transcriptional control (Wehmeier and Piepersberg, 2004). The operons *acbWXY* and *acbVUSRPIJQKMLNOC* are in the same orientation and are each transcribed as an operon (Wolf et al., 2017b). In this case also the transcription of sub operons, like *acbKMLNOC*, is possible (Wehmeier, 2003; Zhang et al., 2002). With the help of RNA sequencing the transcriptional start sites (TSS) of the *acb* gene cluster could be identified (Figure 6, Wolf et al., 2017b).

2.2.4 Regulation of acarbose biosynthesis

Today, the regulation of acarbose biosynthesis and the regulation of the *acb* gene clusters transcription is poorly understood. Already in the second half of the 1970s it was found that the acarbose biosynthesis depends on the provided carbon source. Cells grown on maltose produce more acarbose than cells, which are grown on glucose (Frommer et al., 1975; Frommer et al., 1979; Rauenbusch and Schmidt, 1978; Schmidt et al., 1977). The acarbose formation begins during the exponential growth phase (Thomas, 2001). It shows that the product formation is linked in this case closely with the carbohydrate metabolism and thus the growth of the cells (Drepper and Pape, 1996; Schwientek, 2012; Thomas, 2001), which indicates, that acarbose seems not to be a secondary metabolite in the strict sense.

Later, it was shown that glucose acts as a repressor for the *acb* genes (Brunkhorst and Schneider, 2005; Stratmann, 1997; Virolle and Gagnat, 1994). However, an induction of the cluster could be shown under conditions with maltose or maltodextrin as a carbon source (Virolle and Gagnat, 1994; Wehmeier and Piepersberg, 2004). This was first attributed to the role of maltose as a precursor for acarbose biosynthesis. But it could be shown that maltose or maltodextrin acts as an inducer for the extracellular enzymes of acarbose biosynthesis, especially for the acarviosyl transferase AcbD and

the acarbose-resistant α-amylase AcbE (Merettig, 2009; Wehmeier and Piepersberg, 2004).

The positive regulation of these genes by maltotriose could be shown by heterologous expression of the genes *acbD* and *acbE* together with the associated upstream elements in *Streptomyces lividans* 66 (Virolle and Gagnat, 1994; Wehmeier and Piepersberg, 2004). The induction of maltose-based carbohydrates was also postulated by Stratmann 1997. In 2013 Schwientek et al. verified these findings by a comparative cultivation in glucose- and maltose-based media followed by transcriptome analysis with RNA sequencing (Schwientek et al., 2013). Schwientek et al. showed that the transcription of the *acb* genes is increased when growing on maltose.

Already a few years earlier dyadic symmetry element (DSE) was identified as a potential operator for transcriptional control of some *acb* genes (Stratmann, 1997). This DSE were found upstream of both *acbA* and *acbB* as well as between *acbD* and *acbE*. Besides, this DSE is attributed to a carbohydrate binding activity, which is associated with the repression of these *acb* genes by glucose (Stratmann, 1997; Virolle and Gagnat, 1994; Wehmeier and Piepersberg, 2004). Furthermore, an interaction of this DSE with a Reg-1-like protein – a pleiotropic regulator – was shown (Frederick and Tatchell, 1996; Nguyen et al., 1997; Wehmeier and Piepersberg, 2004). For the Reg-1 protein of *Streptomyces lividans* 66, having an N-terminal HTH-motif, a homology with the LacI/GalR-type repressor family was found. In addition, the sequence of this Reg-1 protein shows 95 % amino acid identity compared to the transcriptional regulator MalR, the repressor of the *malEFG* operon in *Corynebacterium glutamicum* and *Streptomyces coelicolor* (Krause et al., 2012; Nguyen et al., 1997). In this context, a potential hexanucleotide recognition sequence (5´-C/ATTGCT/A-3´) of the LacI/GalR-type transcriptional regulators, especially MalR, was found upstream of these genes in *Streptomyces lividans* (Schlösser et al., 2001). A similar motif, a so-called maltose box, for transcriptional activators such as MalT is known from *E. coli* (Stratmann, 1997).

2.3 Transcriptional regulation in bacteria

Microorganisms are confronted with constantly changing environmental conditions and stress, which effects cell growth and metabolism. Therefore, one of the most important mechanisms is the fine-tuned regulation of gene expression. During evolution cells have developed different mechanisms to control gene expression. Bacteria have the ability to respond very fast to environmental stimuli. The control of bacterial gene expression can be achieved on many levels: The transcription initiation or elongation of a gene, mRNA stability and availability, the translation, protein turnover and, if necessary, post-translational modifications (Lloyd et al., 2001). This way, cells can control whether, when and how much protein is produced (Gottesman, 1984). In the course of evolution, various mechanisms have been developed which allow the cell to influence the protein biosynthesis in order to react to changing environmental conditions and maintain the metabolic balance (Brinkrolf, 2004). One of the most important mechanisms is the regulation of transcription in response to extracellular and intracellular signals (Matic et al., 2004).

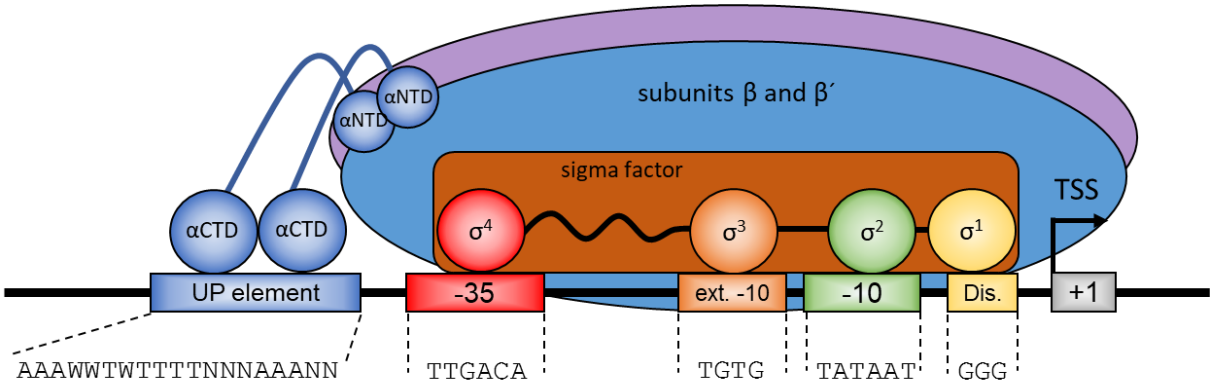

Figure 7: Interaction of the RNA polymerase holoenzyme with different promoter elements in bacteria forming a closed complex. The schematic illustration shows the binding of the different sigma factor subunits to the -35 region (consensus sequence TTGACA), the extended -10 region (TGn consensus motif) and the -10 region (TATAAT consensus sequence). Furthermore, the interaction of the alpha subunit with the upstream promoter element (UP element) is shown. Modified from Browning and Busby, 2004.

Bacterial transcription is initiated through the binding of the RNA polymerase to the promoter sequence located upstream of the transcription start site (TSS). This process is influenced by several factors like DNA sequence and topology, proteins, and small molecules (Seshasayee et al., 2011). The eubacterial RNA polymerase is a

multicomponent enzymatic complex, which composed of at least five subunits: $\alpha_2\beta\beta'\sigma$ (Rojo, 1999). The two large subunits (β and β´) represent the structural and catalytic center of the RNA polymerase and together with the two α subunits, which interact with upstream promoter elements, form the core enzyme. Recently a small ω subunit was found to belong to the core structure of the RNA polymerase (Murakami and Darst, 2003). Many other factors interact with the RNA polymerase and therefore influence its affinity to the promoter region. The most important factor is the RNA polymerase sigma factor (σ), which forms the RNA polymerase holoenzyme (Figure 7). But also other proteins or ligands can affect the formation, activity or promoter preference of the RNA polymerase (Browning and Busby, 2016).

The formation of the transcription initiation complex is one of the most important steps in control of transcription. The complex undergoes a lot of changes until the RNA polymerase starts the transcription. First, the RNA polymerase core enzyme, the sigma factor and the respective promoter region form a closed complex (Figure 7). Afterwards, DNA strand was melted leading to an open complex (deHaseth et al., 1998) and the transcription starts (initiation complex). However, elongation of the transcript only takes place after sigma factor dissociation from the initiation complex. Terminator structures in the produced transcript lead to termination of the transcription (Browning and Busby, 2016; Lloyd et al., 2001).

Bacterial promoters consist of several different sequence motifs. The most prominent are the -10 and the -35 region upstream of the TSS (Browning and Busby, 2016). Usually, the housekeeping sigma factor (σ^{70} in *E. coli*) recognizes the -35 region, the extended -10 region (upstream of the consensus -10 hexamer), the -10 region (consensus hexamer) and the discriminator region (located downstream of the -10 hexamer). The different sigma factor subunits bind to these different elements (Paget, 2015), shown in Figure 7. The relative distance and contribution of the different elements differ with each promoter. Since lack of one element is not necessarily associated with a lower promoter strength, it can be assumed that one element can be compensated by another. The predominantly reason for sigma factor binding is the recruitment of the RNA polymerase (Murakami and Darst, 2003). Therefore, not all elements are necessary for transcription initiation.

In bacteria, in addition to sigma factors, most regulation processes are executed with the help of DNA-binding transcription factors (Nguyen and Saier, 1995). These proteins

can act as transcriptional activators, repressors or dual regulators acting as both repressors and activators. Transcription factors specifically recognize and bind to transcription factor binding sites located in the upstream region of target genes (Pabo and Sauer, 1992). The basic process of repression and activation is shown in Figure 8. The interaction of the respective regulator protein with the specific DNA sequence is mediated by a DNA binding domain. Prokaryotic DNA binding domains consist of helix-turn-helix (HTH), winged helix (WH) and β-ribbon structures (Huffman and Brennan, 2002). The regulator binding affinity is influenced through several further factors like effector molecules (e.g. sugars) or other proteins, which can lead to activation or deactivation of the respective regulator by inducing conformational changes (Figure 8). In *E. coli*, over 50% of the transcription factors harbor a regulatory domain (Madan Babu and Teichmann, 2003). Transcriptional regulators often bind as homodimers or dimers of dimers (tetramers) to palindromic or pseudo-palindromic nucleotide sequences (Huffman and Brennan, 2002).

Regulator proteins can be grouped into families. Most regulators are categorized according to their DNA binding domain based on sequencing results (Luscombe et al., 2000). Besides, regulators can be classified by their respective regulon. Some regulators control a large number of genes (global regulators), whereas other regulate only single genes or an operon (local regulators) (Martínez-Antonio and Collado-Vides, 2003). Regulation of gene expression is also influenced by expression level of the regulators themselves (Browning and Busby, 2004).

All genes or operons, which are regulated by the same transcription factor belong to the same regulon. All regulons form the transcriptional regulatory network of the cell (Rodionov, 2007). For *Escherichia coli* K-12 a minimal set of 314 regulatory DNA-binding proteins was estimated comprising of 35% activators, 43% repressors and 22% dual regulators (Perez-Rueda and Collado-Vides, 2000).

In addition to that, regulation of gene expression can occur by various regulatory RNA structures, like riboswitches and attenuator structures in the leader region of target genes, or by post-translational events like proteolysis (Browning and Busby, 2004).

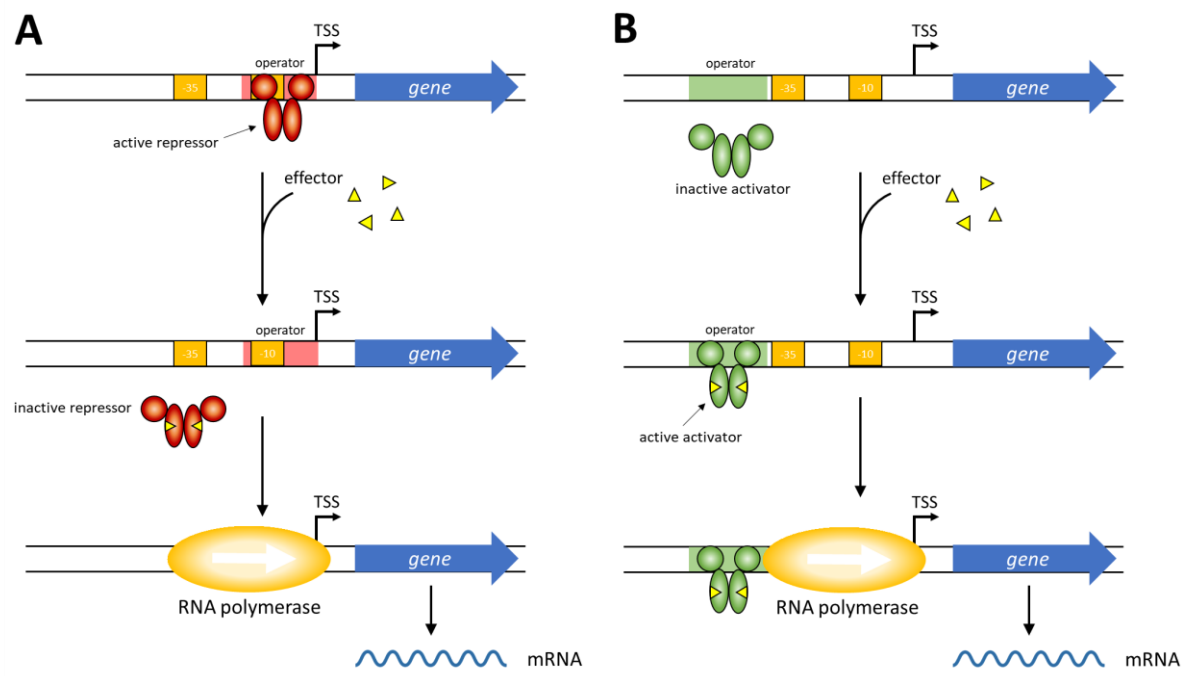

Figure 8: Positive and negative transcriptional gene regulation in bacteria. The involved molecules and binding sites are color coded: repressor = red; activator = green; effector = yellow; RNA polymerase = orange; gene = blue. (**A**) The active repressor is binding the operator site and blocks the transcription of the gene by the RNA polymerase. The repressor is deactivated by binding the respective effector molecule. (**B**) The transcriptional activator is activated by binding the effector molecule. After activation it can bind the operator site and thereby enables the transcription of the target gene by the RNA polymerase.

2.3.1 Transcriptional repressors

Transcriptional repressors are the most common form of DNA-binding transcription factors (Ishihama, 1997). Most of them are characterized by the location of their DNA-binding domain. Most of the transcriptional regulators acting as repressors show an N-terminal HTH motif, whereas activators more often have a C-terminal HTH domain (Perez-Rueda and Collado-Vides, 2000).

In general, it is described that transcriptional repressors inhibit transcription of a specific gene by binding to the promoter region and therefore impedes subsequent binding of the RNA polymerase. Several studies have shown that this steric hinderance of the RNA polymerase is one, but not the only method in bacterial transcriptional repression (Rojo, 1999). The most prominent mechanisms of transcriptional repression are shown in Figure 9.

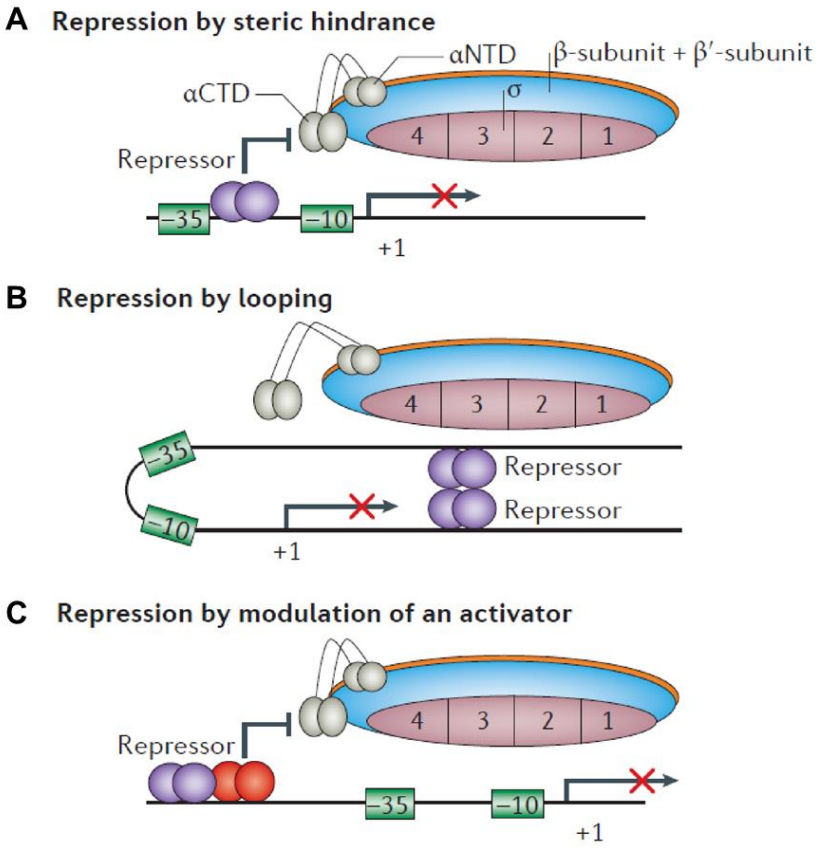

Figure 9: Different types of transcriptional repression by transcription factors. (**A**) Repression by steric hinderance. (**B**) Repression by looping. The operator sites are located upstream and downstream of the promoter region. A loop is built by protein-protein interaction of two or more repressor molecules and the promoter is not accessible for the RNA polymerase. (**C**) Repression by modulation of an activator occurs, when repressor molecules modulate activators, which are afterwards not able to recruit RNA polymerase. Modified from Browning and Busby, 2016.

Repression by steric hinderance occurs through transcription factor binding at the operator site, which often overlaps with the -35 or -10 promoter elements (Figure 9A). Thus, RNA polymerase binding is blocked (Browning and Busby, 2004). Bacterial promoters could contain multiple operator sites, which are located at different positions inside region of transcription initiation. Multiple operator copies can increase the repression strength (Browning and Busby, 2016).

A further common mechanism is repression by DNA loops. This occurs through operator sites located upstream and downstream of the promoter region. Each operator is bound by a repressor and a loop is formed through protein-protein interaction (Figure 9B). By this, the promoter region is not accessible for the RNA polymerase and transcription initiation is blocked (Lloyd et al., 2001). Especially

members of the LacI/GalR regulator family are described to form DNA loops in the promoter region by tetramerization (Rutkauskas et al., 2009). A prominent example is the lac operon in *E. coli* , which was found to be repressed by tetrameric structure of LacI repressors binding to mainly two of the three available operator sites located upstream (operator 3) and downstream (operator 2) of the *lac* promoter (Lewis et al., 1996; Oehler et al., 1990; Reznikoff et al., 1974).

A more complex mechanism is the transcriptional repression by activator modulation (Figure 9C). Repressors can prevent activator binding to their respective operator sites. Thereby, transcriptional activation is blocked. A prominent example is the *E. coli* CytR repressor, which interacts directly with the catabolite activator protein (CAP) and prevents CAP-dependent activation (Gerlach et al., 1990). Therefore, CytR is often called anti-activator rather than a repressor (Lloyd et al., 2001).

Furthermore, some repressors can directly interact with the RNA polymerase, which lead to RNA polymerase binding, but often promoter clearance is prevented (Browning and Busby, 2016). The p4 regulatory protein from the *Bacillus subtilis* phage phi 29 repress the strong promoter A2c by preventing the initiation of elongation step after RNA polymerase binding. The p4 protein binds to both the α-subunit of the RNA polymerase and the DNA upstream of the promoter and by that prevents promoter clearance (Monsalve et al., 1996).

2.3.2 Transcriptional activators

Activating regulators can increase the transcription levels from a basal level up to strong transcription. The three most common activator mechanisms are shown in Figure 10: Class I activation, class II activation and the activation by promoter conformation change (Lee et al., 2012). Transcriptional activation is based on an increase of the RNA polymerase affinity to the promoter region.

For class I transcriptional activation, an activator binds upstream of the UP element interacting with the C-terminus of the RNA polymerase α-subunit (α-CTD) by protein-protein interaction. By this, the recruitment of the RNA polymerase is enhanced (Figure 10A). This event is often observed for promoters which harbor suboptimal promoter motifs (Browning and Busby, 2004). Since the α-subunit is flexible, positioning of the

activator binding site is less important in class I activation. A prominent example is the cAMP receptor protein (CRP or CAP) of the *lac* operon in *E. coli* (Zhou et al., 2014).

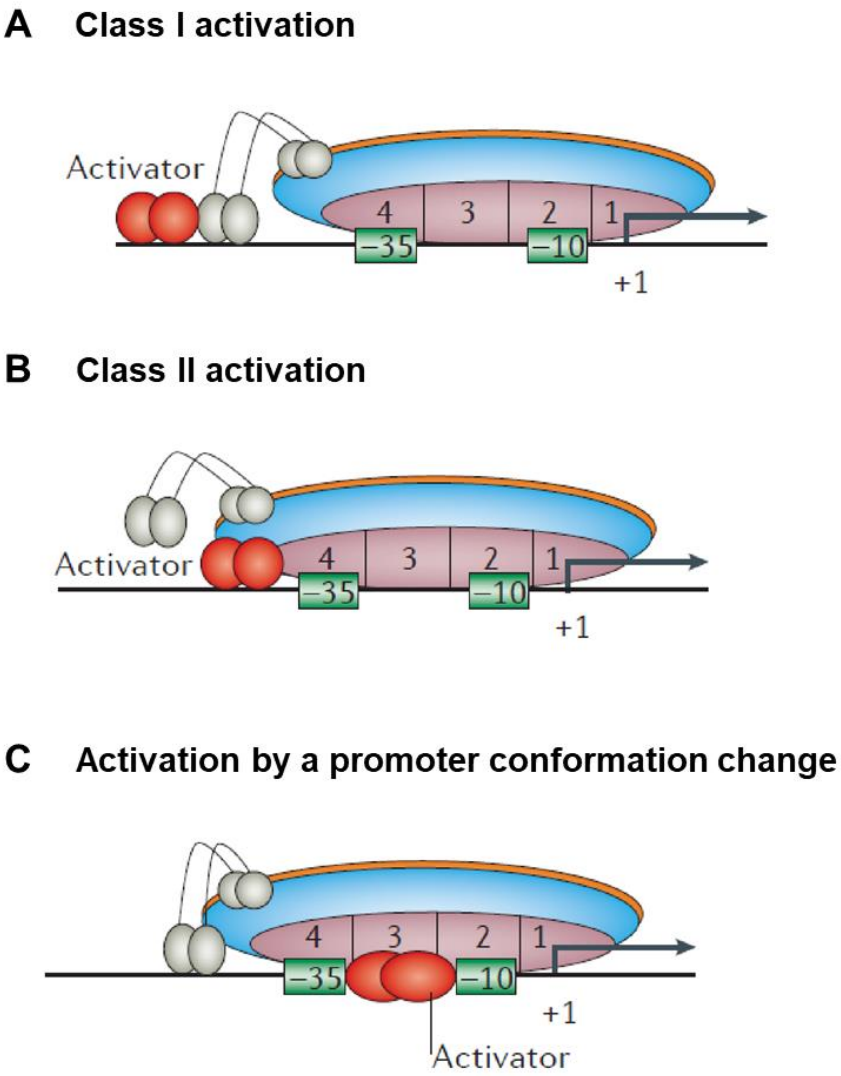

Figure 10: Different types of transcriptional activation by transcription factors. (**A**) Class I activation is performed by the activator binding to an operator site located upstream of the UP element. The activator interacts with the C-terminal domain of the RNA polymerase α subunit (α-CTD) and supports recruitment of the polymerase. (**B**) During class II activation an activator binds to an operator site, which is located close to the -35 promoter element. In this case, interaction with the σ_4 subunit of the sigma factor can be observed, which increases transcription initiation efficiency. (**C**) Promoter conformation can be altered by transcriptional activators, which lead to activation of transcription. Modified from Browning and Busby, 2016.

In contrast to that, class II activation is less flexible, since the operator site is often overlapping with the -35 region resulting in a more strict distance between operator site and TSS (Browning and Busby, 2004). During class II activation the activator

interacts with the subunit 4 of the bacterial σ factor (σ_4), which leads to an increased transcription initiation (Lee et al., 2012). Since class II operators are located close to the -35 region, it is possible for these activators to work together with class I activators (Browning and Busby, 2016). In this way, two different signals can be combined at one promoter by a synergetic mechanism (Browning and Busby, 2004).

Finally, transcription can be activated through conformational changes in the promoter region caused by activator proteins (Ghosh et al., 2010). In contrast to the direct activation mechanisms described above (class I and II), this mechanism is based on promoter modification to make the promoter “more attractive” for the RNA polymerase. Promoters, which are activated in this way, often show suboptimal spacer between -35 and -10 region, which can be changed by activator proteins to increase RNA polymerase affinity (Philips et al., 2015). This kind of activation can be found in some σ^{54}-dependent promoters (Lloyd et al., 2001). Another example are regulators of the MerR family, which can alter the spacer between -10 and -35 region through DNA bending (Brown et al., 2003).

2.3.3 Transcriptional regulation by promoter modifications

Transcriptional regulation not only takes place by binding of transcription factors or alternative sigma factors. There are several, often underestimated, mechanisms, which can affect bacterial transcription. These mechanisms include promoter modifications, like base modifications or spacer length. By this, operator sites for transcription factors or essential promoter elements influencing the binding of the RNA polymerase, can alter transcription.

A prominent modification is DNA methylation (Casadesús and Low, 2006). In *E. coli*, DNA methylation lead to the loss of OxyR repressor binding to the operator site upstream of *ag43* gene (van der Woude and Henderson, 2008). This regulation process is shown in Figure 11a.

Another example of promoter modification is the inversion of DNA segments or the whole promoter region of a gene. Due to this, gene expression can be switched *on* or *off* (Henderson et al., 1999). Those promoters are often identified upstream of genes coding for extracellular gene products or antibiotic resistance (Jiang et al., 2019). A prominent example in *E. coli* is the transcription of the *fim* operon (Gally et al., 1993) shown in Figure 11b.

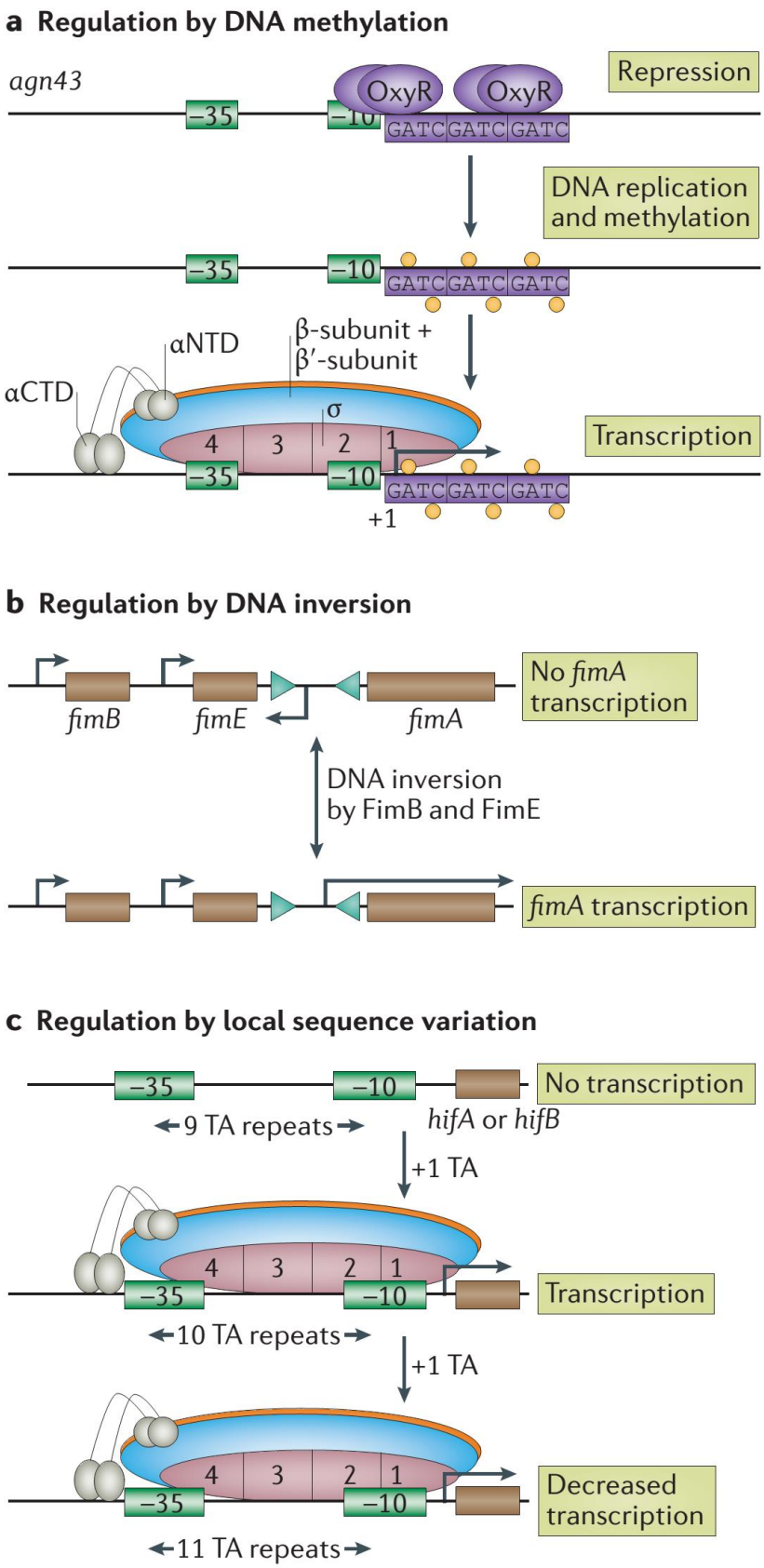

Figure 11: Transcriptional regulation by promoter DNA modification. (a) Regulation by DNA methylation. (b) Regulation by DNA inversion. (c) Regulation by local sequence variation (Browning and Busby, 2016).

In contrast to change the promoter sequence or orientation, a more advanced strategy is the variation of the spacer length between the -10 and the -35 region. Since the spacer length is important for sigma factor binding, transcription can be controlled by variation of this length. An example is given in Figure 11c. In *Haemophilus influenzae*, the promoters of *hifA* and *hifB* were found to have a variable number of TA repeats. This results in different transcription levels of the *hifA* and *hifB* genes. The number of TA repeats changes between different generations or populations (Power et al., 2009). This more complex regulatory mechanism indicates, that transcription factors may not essential for regulation of bacterial transcription under specific environmental conditions (Browning and Busby, 2016).

2.3.4 Regulation by premature termination of transcription

Another common regulatory mechanism in bacteria is the premature termination of transcription. This process is also called attenuation. Attenuation needs a 5´-untranslated region, which harbors an RNA element, which can perform a conformational change by sensing specific signals. This way, a premature terminator or anti-terminator structure can be formed and the genes located downstream are transcribed or not (Lyubetskaya et al., 2003; Merino and Yanofsky, 2005).

The most prominent examples of attenuator structures are leader peptides and riboswitches (Naville and Gautheret, 2009). These two processes differ in their working mechanism. Riboswitches contain RNA elements, which directly interact with small molecules (metabolites) forming terminator or anti-terminator structures (Nudler and Mironov, 2004; Winkler and Breaker, 2005). In contrast to that, leader peptides represent a connection between transcription and translation, which is often found upstream of genes involved in amino acid biosynthesis (Elf et al., 2001) but also more complex natural products, like lantibiotics, microcins and thiopeptides (Oman and van der Donk, 2010). A short peptide sequence, which is enriched of codons of the respective amino acid. In case of an excess of the respective amino acid, translation occurs fast and the corresponding attenuator structure in the mRNA leads to termination of transcriptional. In contrast, in case of deficient amino acid supply the translation rate is low and the ribosome stagnates at the regulatory codons, which in turn leads to the formation of an anti-terminator structure and prevents termination (Henkin and Yanofsky, 2002). Thereby, the downstream genes are increased transcribed, when the amino acid level is low and vice versa (Elf et al., 2001).

3. Aims of this work

The overall goal of this work is the analysis of the transcriptional regulation of the acarbose biosynthesis gene cluster in *Actinoplanes* sp. SE50/110. By analyzing the transcriptional landscape of the *Actinoplanes* sp. SE50/110 genome and in particular the *acb* gene cluster, targets for potential regulators of the acarbose biosynthesis should be identified. Since acarbose is produced in a growth-dependent manner, the expression profile of all genes with focus on those involved in biosynthesis of acarbose should be analyzed during growth. The aim was to use different omics methods to gain knowledge about transcription and protein abundancies and their respective changes during growth. Thereby, bottlenecks of the acarbose formation and regulatory effects during growth should be detected.

Based on the generated omics data potential regulator targets should be characterized through deletion of the respective genes. The corresponding regulon of these transcription factors should be studied and subsequent the effect on acarbose biosynthesis should be investigated. This is intended to provide a better understanding of acarbose biosynthesis and improve acarbose formation by targeted strain development in the future.

4. Results and Discussion

4.1 The transcription of the acarbose biosynthesis genes is dependent on the growth of *Actinoplanes* sp. SE50/110

This chapter summarize and discuss the following publication:

Droste, J.; Ortseifen, V.; Schaffert, L.; Persicke, M.; Schneiker-Bekel, S.; Pühler, A.; Kalinowski, J. (2020): The expression of the acarbose biosynthesis gene cluster in *Actinoplanes* sp. SE50/110 is dependent on the growth phase. BMC genomics [status: accepted for publication]

The expression profile analysis is a suitable method to determine bottlenecks and regulatory effects in cells metabolism. This technique was used for the analysis of several model organisms like *Escherichia coli* (Lempp et al., 2019) or *Streptomyces coelicolor* A3(2) (Nieselt et al., 2010). In this study, transcriptomic and proteomic data from seven time points of controlled bioreactor cultivations in maltose minimal medium were used to analyze the expression dynamics during growth of *Actinoplanes* sp. SE50/110. Subsequent cluster analysis revealed co-regulated genes, which show a similar transcription course over the cultivation time. By this, a typical switch from primary to secondary metabolism during transition phase could be observed.

Furthermore, a significantly decreasing transcript abundance of all acarbose biosynthetic genes, with the strongest decrease for the monocistronically transcribed genes *acbA, acbB, acbD* and *acbE* was found. These data confirm a similar trend for *acb* gene transcription and acarbose formation rate. Interestingly, the proteome dynamics does not follow the respective transcription for all *acb* genes. This suggests different protein stabilities or post-transcriptional regulation of the Acb proteins, which in turn could indicate bottlenecks in the acarbose biosynthesis. Finally, several genes co-expressed with the *acb* gene cluster were identified.

4.1.1 Acarbose production of *Actinoplanes* sp. SE50/110 steadily decreases during the growth phase and almost ceases in stationary phase

Controlled bioreactor cultivations are well-suited for the analysis of changes in the transcriptome or proteome pattern of bacterial cells. Furthermore, product formation can be monitored in correlation to biomass formation. By this, a previous reported connection of acarbose formation and the course of biomass was confirmed (Wendler et al., 2014). Acarbose is produced in the lag phase (24 h), during growth phase (48 h to 96 h) until transition phase (120 h to 144 h). However, the specific product formation rate increases until the early growth phase and decreases during remaining cultivation time (Figure 12). In the stationary phase (144 h to 168 h), no further acarbose production was observed. This course was reported in previous *Actinoplanes* sp. SE50/110 studies (Wendler et al., 2014; Wolf et al., 2017a).

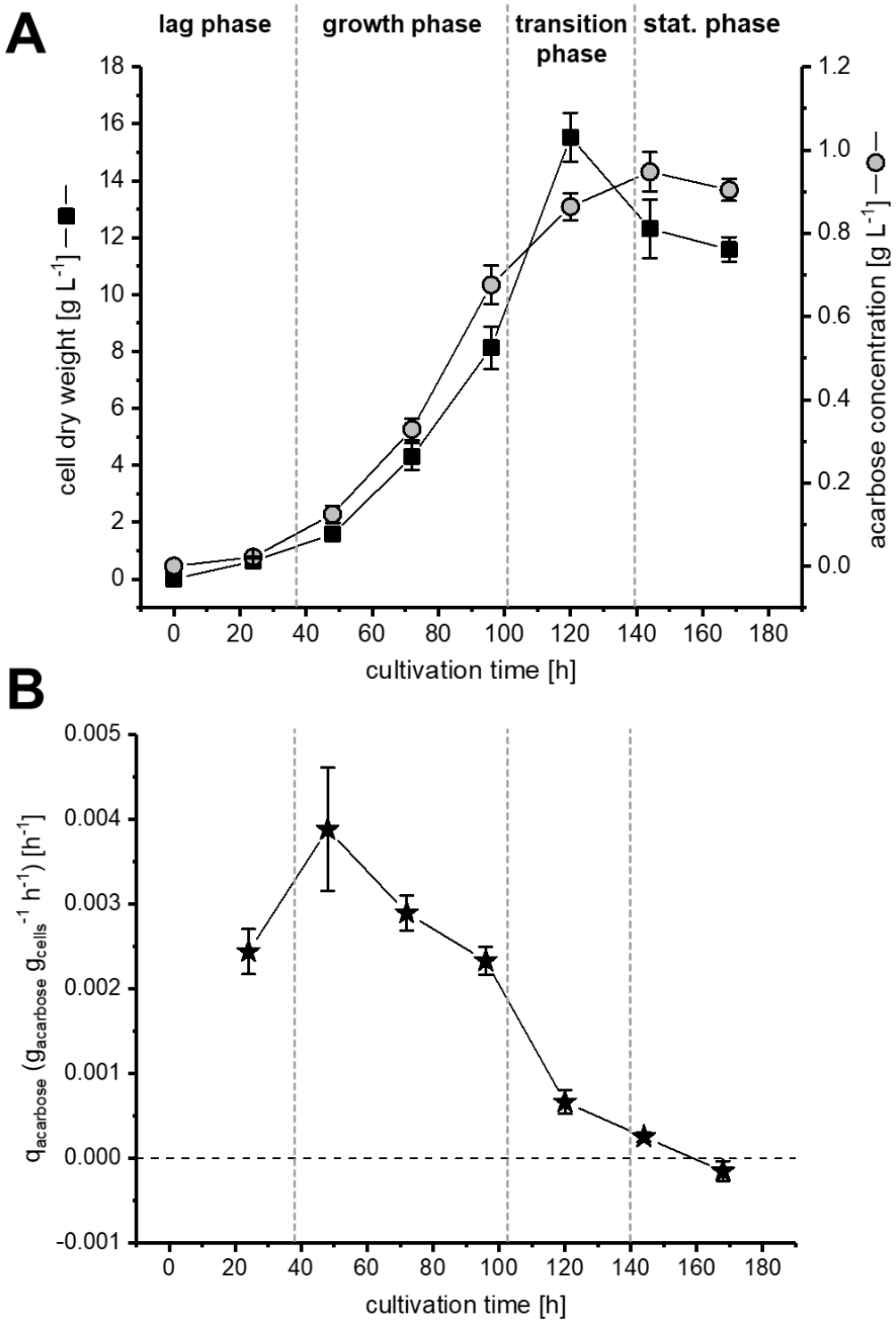

Figure 12: Characterization of growth and acarbose production of *Actinoplanes* sp. SE50/110 in controlled fermenter conditions. Different growth phases (lag, growth, transition and stationary phase) were indicated by vertical dashed lines. (**A**) Cell dry weight (black boxes) and acarbose concentration (grey circles) over the cultivation course. Plotted are the means and standard deviations of three biological replicates, each of which were measured in three technical replicates. (**B**) Specific product formation rates ($q_{Acarbose}$) defined as produced acarbose normalized on the mean cell dry weight and cultivation time difference.

Transcriptome and proteome analysis of each time point revealed transcription and proteome dynamics in comparison to the average transcript or protein amount over all time points (Figure 13). Most genes are transcribed in the growth phase and therefore show their mean transcription level in the mid growth phase. The observed trend of differences in transcription is in good accordance to the expectation as a minimal number of differentially transcribed genes is expected in the growth phase (Jeong et al., 2016). In contrast to that, the greatest difference regarding transcription was observed for the late stationary phase (168 h). This could be a hint for a typical switch from primary to secondary metabolism (Jeong et al., 2016; Nieselt et al., 2010). The proteome data (Figure 13B) confirm the transcriptome data overview (Figure 13A).

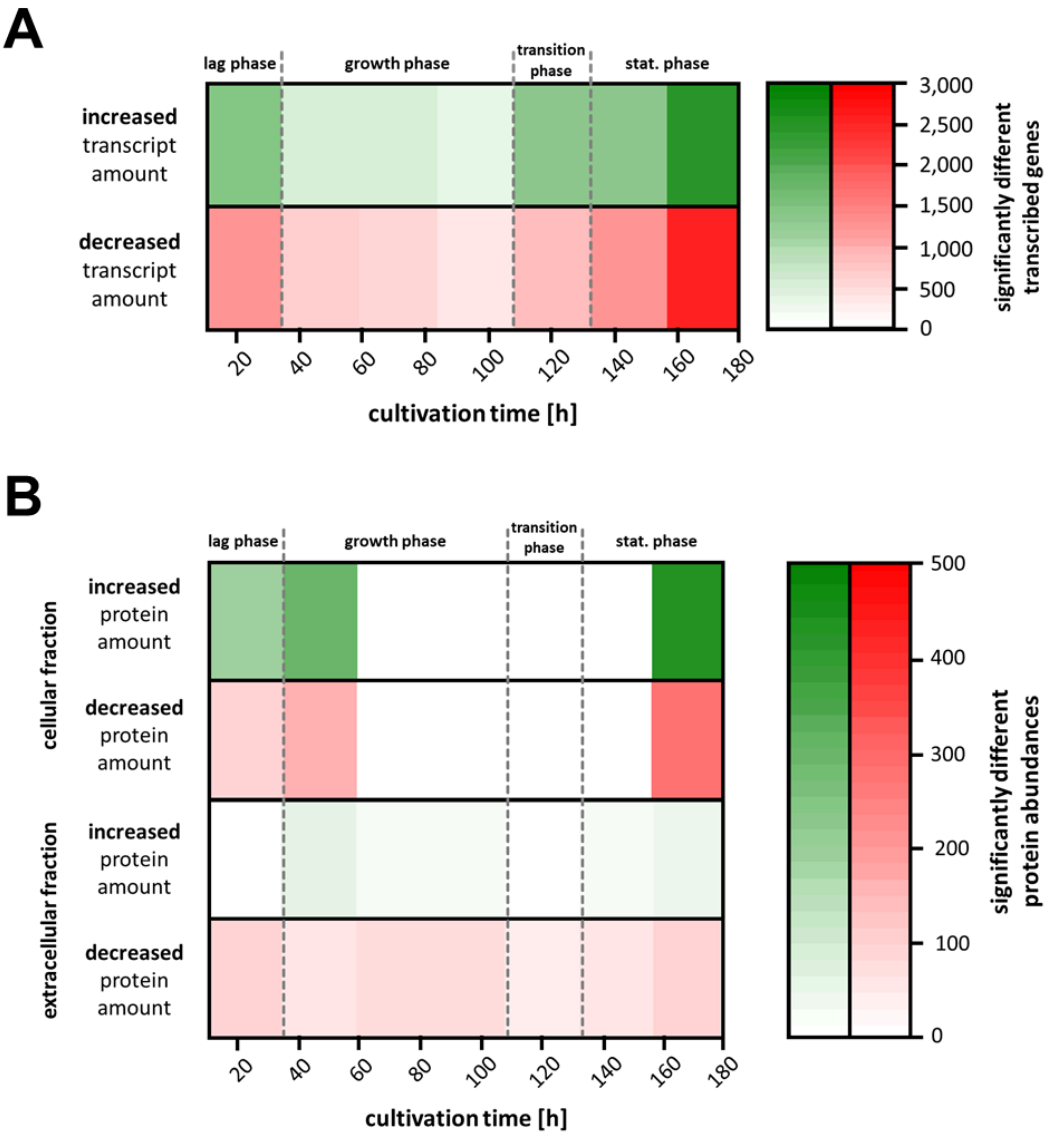

Figure 13: Overview about transcriptome (A) and proteome (B) dynamics in *Actinoplanes* sp. SE50/110 during cultivation time. Number of transcripts/proteins with significantly (p_{adj}-value < 0.05) increased

(green) and decreased (red) transcript/protein abundances during cultivation at the given time points. Protein dynamics was analyzed in cellular and extracellular fraction. Growth phases are indicated with dashed lines.

Bacterial gene expression is regulated on several stages including transcriptional, post-transcriptional, translational and post-translational level. By comparing transcriptome and proteome data, these regulatory mechanisms become apparent. We performed correlation of transcriptome and proteome data by calculation of Pearson coefficient of each transcript/protein data pair. If transcription and translation are not in correlation with each other, this could be a hint for protein degradation or other above-mentioned mechanisms (Koussounadis et al., 2015).

Pearson coefficient was found to range from 0.10 to 0.63, which seem to be a broad range compared to other bacteria with a coefficient range of 0.4 to 0.5 (Bathke et al., 2019; Koussounadis et al., 2015; Picard et al., 2012) or in yeast like *Saccharomyces cerevisia*e with 0.66 to 0.76 (Greenbaum et al., 2003). This could be due to technical or methodological constrains, but also to translational and posttranslational regulation processes (Berghoff et al., 2013; Maier et al., 2011), which is why correlation of transcript and protein abundance is often poor (Koussounadis et al., 2015; Vogel and Marcotte, 2012).

Nevertheless, several interesting effects are visible in the correlation data. Strongest correlation of transcript and protein data was found in the early growth phase (47.8 h and 72.3 h) with a Pearson coefficient of 0.63 and 0.48 respectively, whereas lowest correlation was observed in the transition phase (120 h) with a coefficient of 0.1. Strikingly, when comparing the transcriptomic data with proteome data of the following sampling point (transcriptomic data from time point X compared to proteome data from time point X+1). This results in Pearson coefficients of 0.61, 0.61, 0.50, 0.24, 0.21 and 0.20 respectively. This increased correlation could be due to an offset caused by protein folding and processing or higher protein stability compared to its respective transcript (Christiano et al., 2014; Wang et al., 2002).

4.1.2 Identification of co-transcribed genes by hierarchical cluster analysis of transcription dynamics data

The identification of co-transcribed genes by hierarchical cluster analysis was performed within the software Omics Fusion (Brink et al., 2016) and resulted in an optimal cluster amount of 36 (Figure 14; Supplementary Figures 9 and 10). The clusters contain 45 to 645 genes.

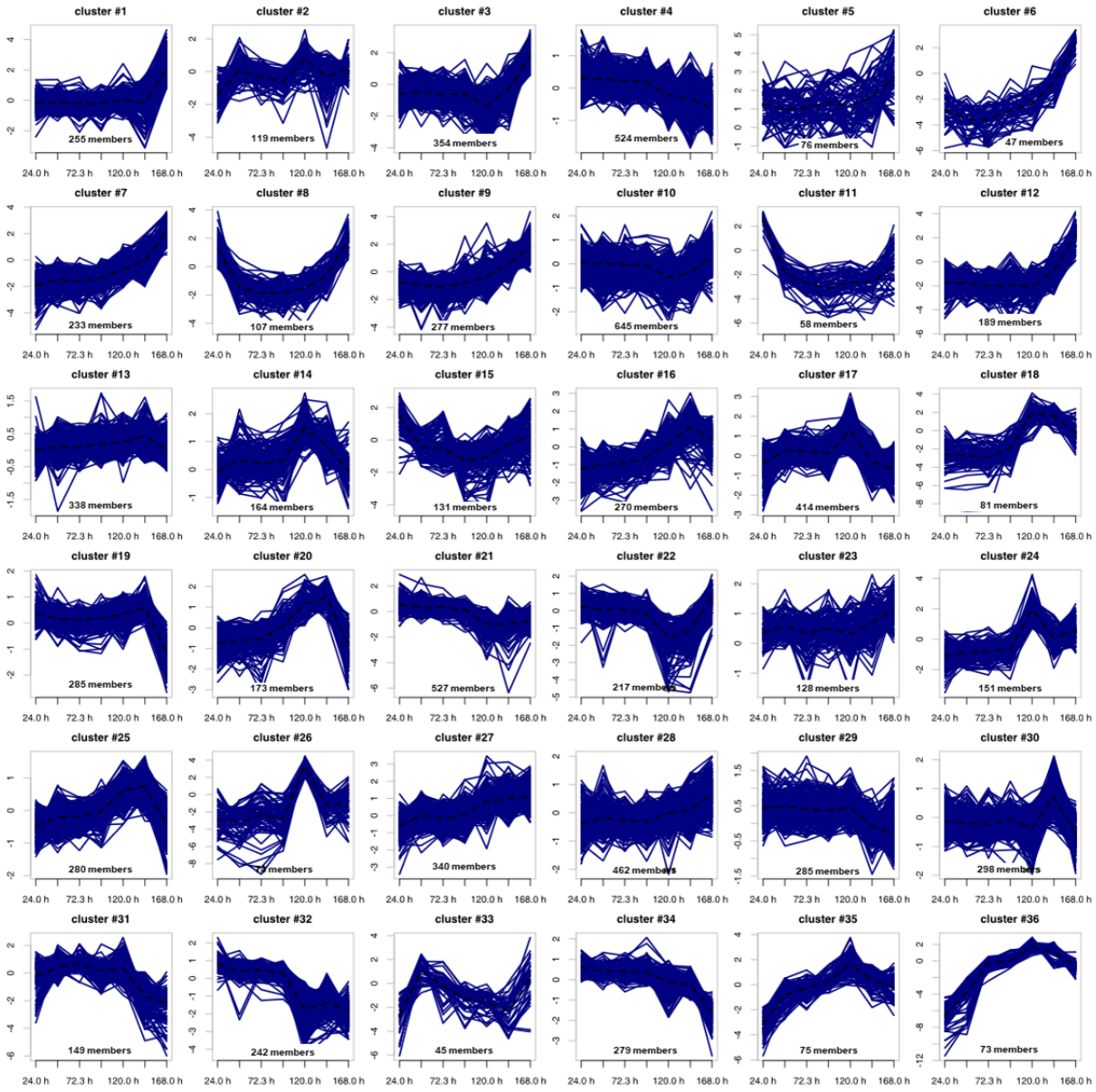

Figure 14: Hierarchical cluster analysis cluster analysis of transcription dynamics with calculation of optimal cluster size and cluster grouping using the tool *Omics Fusion* (Brink et al., 2016). The log2(fold-changes) values for each time point and gene were used as input and are shown in the cluster graphs. The respective mean value of each cluster is visualized by a dashed line. For better visualization the trend of each cluster is scaled differently.

The different cluster were analyzed regarding enrichment of protein functions using the COG (clusters of orthologous groups) database (Galperin et al., 2015) and the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Kanehisa and Goto, 2000). Each cluster was scanned for overrepresented protein functions and metabolic pathway affiliation. Furthermore, changes in the transcription dynamics could be a hint for metabolic switches or different stages of the *Actinoplanes* sp. SE50/110 life cycle (Palleroni, 1976).

The cluster analysis revealed the co-regulation of about 65% of the genes involved in chemotaxis, motility and flagellum associated proteins (COG class N). These genes were found in clusters 8, 11 and 15, which show an increased transcription in the lag phase. This effect can be explained by the life cycle of *Actinoplanes* sp. SE50/110 (Palleroni, 1976; Uchida et al., 2011), since the cultivation was inoculated with spores. Genes involved in flagellar development have been identified and described to be highly expressed in spores in *Actinoplanes* spp. (Jang et al., 2016; Mouri et al., 2017). Due to the fact, that these genes are only expressed for a short time after inoculation, spore formation and cell motility seem to play a minor role in the further course of cultivation under the tested conditions.

In clusters 4, 21, 32 and 34 genes with a decreasing transcript abundance were found. Many of these cluster members are involved in translation, ribosomal structure and biogenesis (COG class J). The ribosomal proteins of 30S and 50S ribosomal subunits are located in clusters 32 and 34. This transcription dynamics was previously reported in other Gram positive bacteria, like *S. coelicolor* A3(2) (Nieselt et al., 2010; Strauch et al., 1991) or *Corynebacterium glutamicum* (Brockmann-Gretza and Kalinowski, 2006; Ruwe et al., 2019).

In total, 20 different secondary metabolite gene cluster were predicted in the *Actinoplanes* sp. SE50/110 genome (Wolf et al., 2017b). Most of the involved genes were found to be highly transcribed in the transition (clusters 17, 20 and 30; Figure 14) and stationary phase (cluster 1, 3, 7, 10, 12 and 16; Figure 14). This is a common observation for secondary metabolite gene cluster expression in actinomycetes (Jeong et al., 2016; Nieselt et al., 2010). This shows the typical switch from primary to secondary metabolism. Interestingly, the acarbose biosynthesis (*acb*) gene cluster is highly transcribed during early growth phase, but less in the stationary phase. This indicates, that acarbose is not a secondary metabolite by definition.

4.1.3 The genes of the acarbose biosynthetic gene cluster are transcriptionally and post-transcriptionally regulated during growth

The transcription dynamics of the acarbose biosynthesis (*acb*) genes (Figure 15B) seem to be similar to the course of acarbose formation rate (Figure 12B). Transcript abundance increases during transition phase and decreases afterwards.

All *acb* genes are distributed over two clusters (Figure 14). The transcription profile of the monocistronically transcribed genes (*acbZ*, *acbB*, *acbA*, *acbE* and *acbD*) are grouped in cluster 32, whereas all other genes organized in the two operons *acbVUSRPIJQKMLNOC* and *acbWXY* can be found in cluster 31.

The similar course of transcription during growth and similar fold changes by comparing the growth and stationary phase indicate a co-regulation of these genes. Although the genes are transcribed monocistronically, co-regulation was reported for the pair *acbD* and *acbE* (Wolf et al., 2017a). The genes *acbE* and *acbD* as well as *acbB* and *acbA* are located in opposite to each other in the genome sharing an intergenic region (Figure 15). Therefore, it can be assumed, that the intergenic regions between the pairs *acbA* & *acbB* and *acbD* & *acbE* harbor binding sites for transcriptional regulators. This could be the reason for the observed co-regulation of these genes (Stratmann, 1997; Wehmeier and Piepersberg, 2004). For *acbE* and *acbD* it was shown, that the MalR type transcriptional regulator AcrC (ACSP50_6387) is the repressor of these two genes in *Actinoplanes* sp. SE50/110 (Wolf et al., 2017a).

The two operons inside the *acb* gene cluster are also located in a head-to-head arrangement. The data indicate, a co-regulation of these operon as well. So far, no regulators of these two operons are described.

As mentioned above, the course of the transcript abundance for all *acb* genes (Figure 15B) are more or less in accordance with the course of the specific product formation rate (Figure 12B). However, the protein abundances are not in correlation with the specific product formation rate for all Acb proteins (Figure 15B). Especially the alpha-amylase AcbZ, the dTDP-4-keto-6-deoxy-glucose dehydratase AcbB and the acarviose transferase AcbD are not following their respective transcript signals on protein level. Their protein levels seem to be constant during growth. This could be a hint for a post-transcriptional regulation, high protein stability due to the secretion of AcbZ and AcbD. This could be an indication for the essentiality of AcbD and AcbZ for acarbose biosynthesis. For AcbD, this was shown previously with an *acbD* deletion

mutant, which show no acarbose formation (Gren, 2017). However, according to the current models of acarbose biosynthesis, it is described that acarbose is formed intracellularly (Ortseifen, 2016; Zhang et al., 2002) and the proteins AcbZ, AcbE and AcbD are not involved in these reactions (Wendler et al., 2014).

AcbB is involved in the biosynthesis of dTDP-4-amino-4,6-dideoxy-D-glucose. It is the only intracellular Acb enzyme whose protein dynamics highly differs from its transcription profile. Its transcription strongly decreases, whereas the protein level stays stable during the cultivation. Interestingly, *acbA* shows the same transcription pattern, but on protein level, a strong decrease can be observed for AcbA. This shows, that acbA and acbB are regulated in a similar way on transcription level, but their protein stability differs. Since AcbA shows a strong decrease on proteome level, it would be an interesting target for overexpression to improve acarbose formation, since low AcbA amounts might be a bottleneck in the sub-pathway composed of AcbA, AcbB and AcbV.

Different protein dynamics were also found between the genes of the large operon *acbVUSRPIJQKMLNOC.* Interestingly, the proteins AcbC, AcbO and AcbM are catalyzing the first steps in acarbose biosynthesis. These proteins were found to be less stable during cultivation, since their respective protein abundancies decrease more strongly compared to the other proteins encoded by this operon. This could be a further indication for a regulation on a post-transcriptional or protein level. This is a strategy to save energy by preventing the first steps of the biosynthesis.

Finally, several genes showing the same transcription dynamics or both transcription and protein dynamics, were identified. In total, 9 transcriptional regulators (ACSP50_1631, ACSP50_2235, ACSP50_4697, ACSP50_5005, ACSP50_6401, ACSP50_6463, ACSP50_8007, ACSP50_8120 and ACSP50_8287), a two-component regulator system (ACSP50_3744, ACSP50_3745) and 2 sigma factor genes (ACSP50_0644, ACSP50_6006) were determined to show the same transcription dynamics as the monocistronic *acb* genes located in cluster 32.

By combined clustering of transcriptomic and proteomic data, the transcriptional regulator gene *ACSP50_0424* was found to be an interesting target for further analyses regarding a potential effect on acarbose biosynthesis.

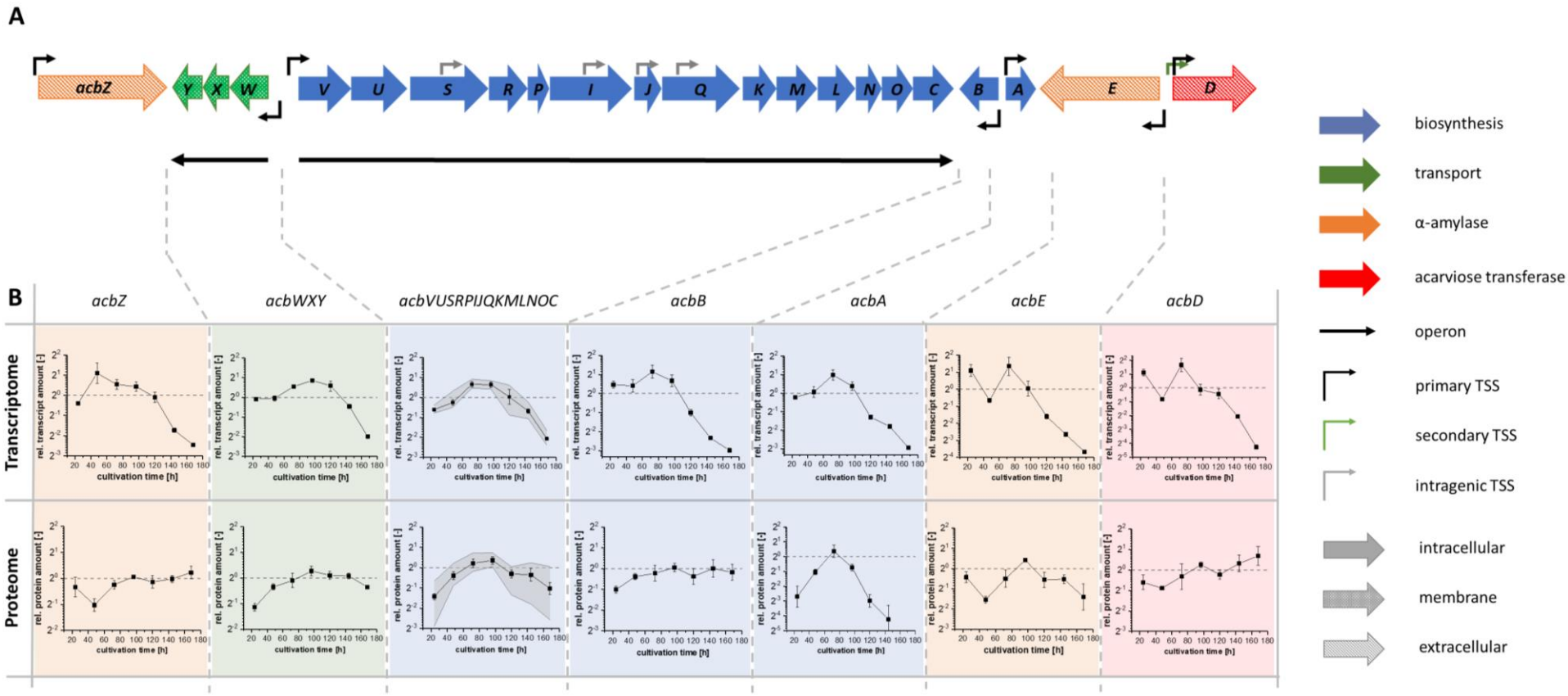

Figure 15: Overview about the expression dynamics of the *acb* gene cluster in *Actinoplanes* sp. SE50/110. **(A)** The *acb* gene cluster with its transcriptional landscape including operon structure and TSS (modified from Wolf et al., 2017b). The function of all genes and operons are color-coded. The sub-cellular localization (according to Wendler et al., 2015) of the corresponding gene products are encoded by filled, dotted and striped arrow content. **(B)** Dynamics of the relative transcript abundances and the relative protein amounts of the *acb* genes and Acb proteins. A relative abundance of one corresponds to the average amount of RNA or protein over all time points. Mean values and standard deviation of three biological replicates are shown for each time point. For the operon *acbWXY* only protein abundancies of AcbW are shown. For the operon *acbVUSRPIJQKMLNOC* the maximum and minimum values are shown by grey area.

4.2 The MalR type regulator AcrC is a transcriptional repressor of acarbose biosynthetic genes in *Actinoplanes* sp. SE50/110

This chapter summarize and discuss the following publication:

Wolf, T*.; **Droste, J.***; Gren, T.; Ortseifen, V.; Schneiker-Bekel, S.; Zemke, T.; Pühler, A.; Kalinowski, J. (2017): The MalR type regulator AcrC is a transcriptional repressor of acarbose biosynthetic genes in *Actinoplanes* sp. SE50/110. *BMC genomics*, 18(1), pp. 562. [status: published] *shared first authorship

In contrast to acarbose biosynthetic gene clusters in *Streptomyces* spp., the corresponding gene cluster of *Actinoplanes* sp. SE50/110 lacks genes for transcriptional regulators. The transcription factor acarbose regulator C (AcrC) was identified through an *in silico* approach by comparing the LacI family regulators of the acarbose biosynthetic gene clusters from *S. glaucescens* GLA.O (Ortseifen et al., 2017) and *S. coelicoflavus* ZG0656 (Guo et al. 2012) with the regulator repertoire of the *Actinoplanes* sp. SE50/110 genome. The LacI/GalR family regulator gene *acrC* is located head-to-head with the maltose/maltodextrin ABC transporter *malEFG* operon. Transcriptional regulators of this family are often involved in carbohydrate and nucleotide metabolism (Swint-Kruse and Matthews, 2009; Weickert and Adhya, 1992).

The gene of *acrC* was functional deleted by an replacement with an apramycin resistance gene using PCR targeting (“Redirect”) technique (Gust et al., 2003). The constructed regulator mutant was characterized by cultivation experiments, transcriptomic analyses using genome-wide microarrays and RT-qPCR as well as electrophoretic mobility shift assays (EMSA). It could be shown, that the transcription of the acarbose biosynthesis genes *acbD* and *acbE* was elevated in the Δ*acrC* strain (Figure 16A). The transcriptomic profile could be reconstituted through a complementation of the deleted *acrC* gene (Wolf et al., 2017a).

Furthermore, a regulatory sequence motif for the binding of AcrC was identified upstream of the genes *acbD* and *acbE* (Figure 16B). Interestingly, AcrC does not regulate the *malEFG* operon.

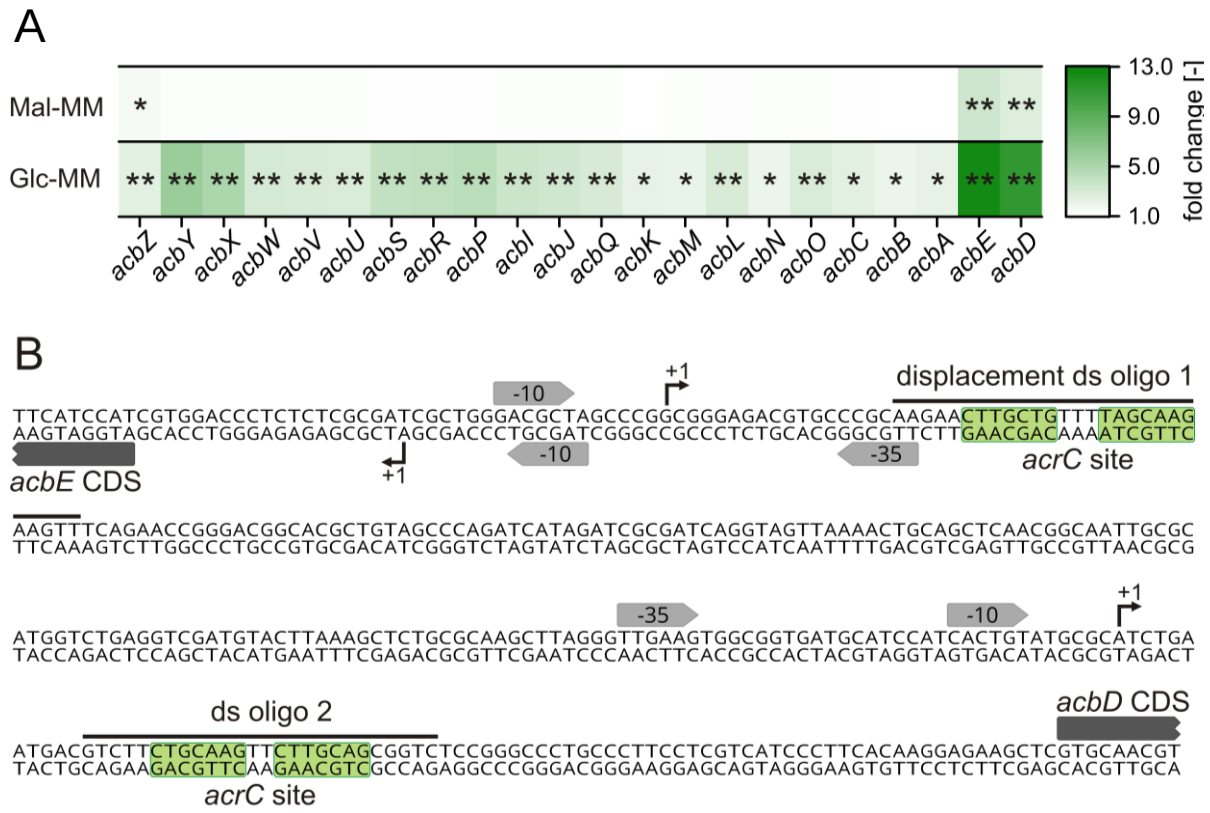

Figure 16: Differential transcriptional analysis of the acb genes in the deletion mutant Δ*acrC* compared to the wild type and intergenic region of *acbE* and *acbD*. (A) Heatmap of the fold change of transcript abundance for the genes of the *acb* gene cluster. Data gained from whole genome microarrays of the strain *Actinoplanes* sp. SE50/110 Δ*acrC* compared to the *Actinoplanes* sp. SE50/110 wild type grown in maltose minimal medium (Mal-MM) and glucose minimal medium (Glc-MM). Significance of $p < 0.05$ is marked with a single asterisk, significance of $p < 0.01$ with two asterisks (*t*-test, two-sample, Holm). (B) Intergenic region of *acbE* and *acbD* with the promoter structures (data obtained from Wolf et al., 2017b) and the AcrC binding sites determined by electrophoretic mobility shift assays.

It was shown that the use of the Redirect technology for the deletion of the *acrC* gene leads to polar effects on the neighboring gene *dapE2* (*ACSP50_6388*) through the highly transcribed antibiotic resistance gene. This effect could not be reversed by complementation of *acrC*. Since it was shown, that AcrC does not bind the upstream region of dapE2, the increased transcription of this gene in the regulator mutant Δ*acrC* is caused by polar effects (Wolf et al., 2017a). Therefore, further deletion mutants were constructed using CRISPR/Cas9 technique, which has been shown to have no off-target effects in *Actinoplanes* sp. SE50/110 (Wolf et al., 2016).

Interestingly, the transcript abundance of the maltose/maltodextrin ABC transporter operon *malEFG* was not significantly influenced by the deletion of *acrC*. This was

unexpected, as the acrC gene is located in a head-to-head arrangement to this operon, which is often accompanied with a direct regulation. In addition, acrC was predicted as a MalR-type regulator and members of this regulator family act as a repressor of the *malEFG* operon in other Actinobacteria (Schlösser et al., 2001; van Wezel et al., 1997b).

Further analyses revealed, that the MalR-type transcriptional repressor in *S. lividans* was shown to not only bind upstream of *malEFG*, but also to operator sites upstream of α-amylase genes (Nguyen et al., 1997; Nguyen, 1999). The repression of α-amylase genes through LacI/GalR type regulators depending on glucose was also reported for other Gram-positive bacteria (Afzal et al., 2015; Henkin et al., 1991; Virolle and Bibb, 1988). For *S. coelicolor* the deletion of *malR* leads to a glucose-insensitive transcription of *malE* (van Wezel et al., 1997b; van Wezel et al., 1997a).

In this work it could be shown, that the deletion of the MalR-type regulator gene *acrC* in *Actinoplanes* sp. SE50/110 leads to an elevated transcription of *acbD* and *acbE*. Additionally, the AcrC protein binds to a palindromic sequence in the intergenic region of these genes (Figure 16B). AcbD is an acarviosyl transferase, which was described to catalyze the transfer of glucose or short malto-oligosaccharides onto acarbose (Ortseifen, 2016). Furthermore, the catalytic site of AcbD was described as similar to other enzymes with an α-amylase activity (Leemhuis et al., 2004). The *acbE* gene encodes an acarbose-resistant α-amylase, which is responsible for extracellular starch and maltodextrin degradation (Wehmeier and Piepersberg, 2009). Both enzymes are involved in extracellular starch and malto-oligosaccharide metabolism. These findings fit the beforementioned results gained from literature research, that MalR-type regulators are involved in the regulation of α-amylase genes and similar enzymes.

The AcrC binding site identified upstream of the genes *acbD* and *acbE* consists of a palindromic 7 bp repeat (5'-CTTGC(A/T)G-3') (Figure 16B). This binding sequence resembles the core binding motif described for MalR in *S. lividans* (5'-CTTGCAG-3'), which is located upstream of *malE* but downstream of the corresponding promoter site (Schlösser et al., 2001). Additional sites are reported upstream of other amylase genes as direct or inverted repeat (spacer of 3-15 bp) in *S. lividans* (Nguyen, 1999).

In *Actinoplanes* sp. SE50/110 the AcrC binding site was found upstream of the TSS of *acbE*, but downstream of the *acbD* TSS. The proximity of these two binding sequences (182 bp) could be a hint for a tetrameric repressor structure, where two homodimers

bind on each site and build a tetramer resulting in a loop in the DNA in this area. Tetrameric structures are typical for LacI/GalR-type regulators (Lewis et al., 1996). This way, the promoter sites of both genes is blocked for the RNA polymerase. This DNA-looping with two operator sites is known for other LacI-type regulators, like the repressor of the lac operon in *E. coli* (Oehler et al., 1990; Rutkauskas et al., 2009; Wong et al., 2008). A similar process could be the case for the intergenic region of *acbD* and *acbE* (Wolf et al., 2017a). A further copy of the palindromic AcrC binding sequence was found in the intergenic region of *malE* and *acrC* itself. It is located upstream of the *malE* TSS and downstream of two of the three identified *acrC* TSS (Wolf et al., 2016). Band shift assays show the binding of AcrC to this sequence. However, since there is no increased transcription of *malE* observed in the Δ*acrC* deletion mutant, it can be assumed that only autoregulation of *acrC* occurs.

Although an effector dependent regulation of MalR-type regulator is described in the literature, no effector molecule could be identified during band shift assays. However, the effect of *acrC* deletion was stronger in glucose compared to maltose minimal medium (Figure 16A), an effector molecule related to maltose or a metabolic product of it would be possible. In this case, maltodextrins are described to be possible candidates for effector molecules (Schlösser et al., 2001).

In the early growth phase of the regulator mutant Δ*acrC* an increased acarbose formation compared to the wild type was observed, which results in a higher maximal specific product formation rate regarding acarbose. This effect can be attributed to increased transcription of *acbD* and *acbE*. However, no elevated final acarbose concentration was detected for the mutant strain. The reason is that acarbose is formed intracellular and is afterwards exported and be loaded with oligosaccharides by AcbD. Thus, AcbD and AcbE are not involved in direct biosynthesis of acarbose.

In this study, important findings for the elucidation of the transcriptional regulation of the acarbose biosynthesis genes were reached. The first repressor of the *acb* gene cluster in *Actinoplanes* sp. SE50/110 was described and characterized in relation to the regulon and binding motifs. Furthermore, important techniques, like gene deletion using PCR-targeting technology or whole-genome microarrays were established for *Actinoplanes* sp. SE50/110, which will be of high value for further analyses and the characterization of other transcriptional regulators and the regulatory network of *Actinoplanes* sp. SE50/110.

4.3 A maltose-regulated large genomic region is activated by the transcriptional regulator MalT in *Actinoplanes* sp. SE50/110

This chapter summarize and discuss the following publication:

Droste, J; Kulisch, M.; Wolf, T.; Schaffert, L..; Schneiker-Bekel, S.; Pühler, A.; Kalinowski, J. (2020): A maltose-regulated large genomic region is activated by the transcriptional regulator MalT in *Actinoplanes* sp. SE50/110. Applied Microbiology and Biotechnology [status: published]

During analysis of the expression dynamics in *Actinoplanes* sp. SE50/110 (Droste et al., 2020) a large genomic region of 51 genes ranging from *ACSP50_3900* to *ACSP50_3950* was identified to be co-regulated (Figure 17). 41 of these genes show a similar transcription with a continuously increasing transcript amount during growth. This indicates that these genes might belong to the same regulon. Since acarbose production seem to be coupled to bacterial growth, it was assumed that these genes might be involved in acarbose metabolism or in close relation with it.

Further experiments showed a stronger transcription of this large genomic region on maltose compared to glucose as a carbon source (Figure S1 and Table S4 of this publication). The similar transcription pattern, the close proximity of these genes, as well as the maltose-dependent expression indicates a maltose-dependent co-regulation of this genomic region during growth. Therefore, this genomic region (Figure 17) was named maltose-regulated large genomic region (MRLGR).

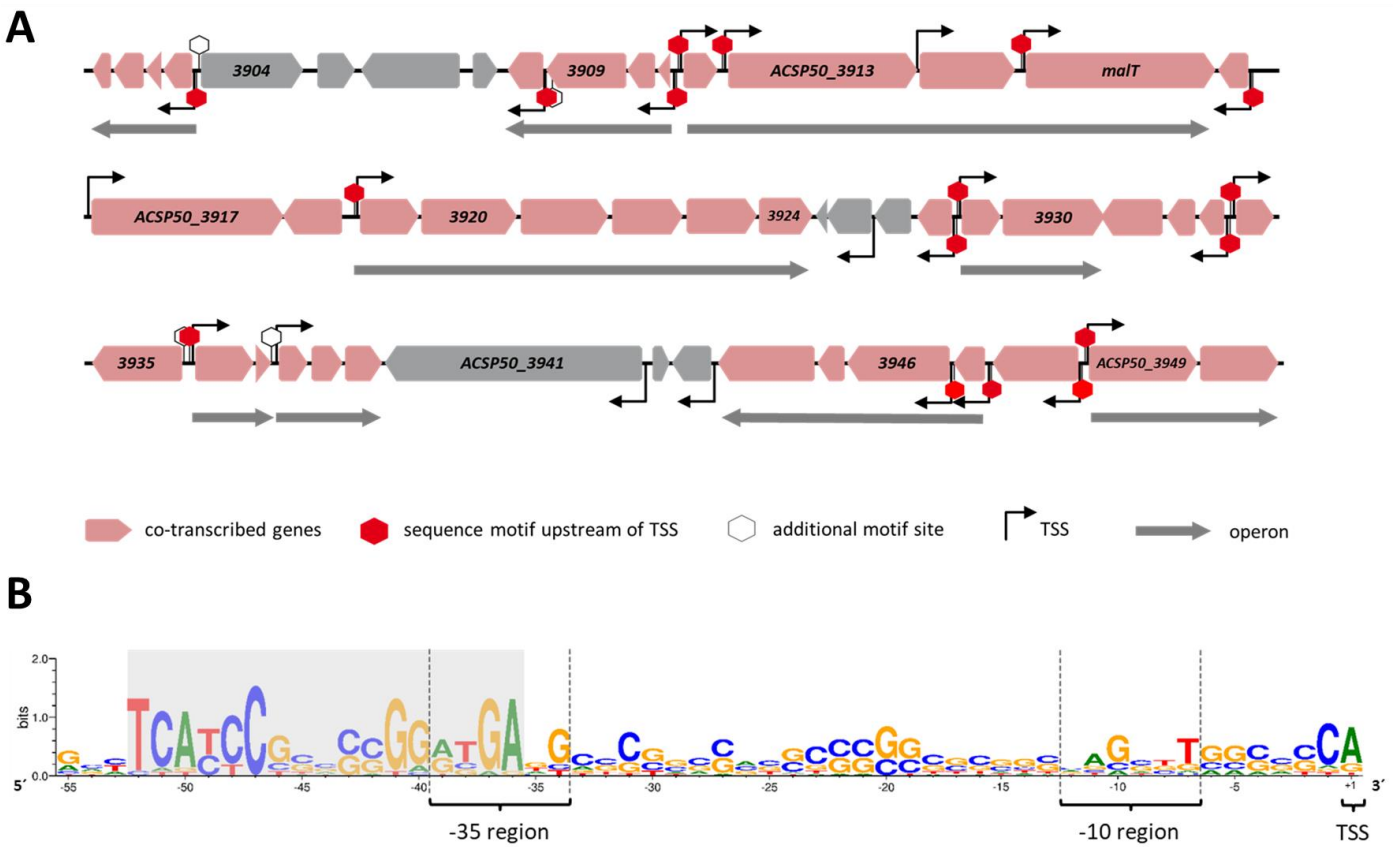

Figure 17: Maltose-regulated large genomic region (MRLGR) ranging from *ACSP50_3900* to *ACSP50_3950* in *Actinoplanes* sp. SE50/110 and the identified palindromic sequence motif. (**A**) Genes, which were found to be co-regulated during growth are marked in light red (data obtained from (Droste et al., 2020)). Transcription start sites (TSS) and operon structure are indicated by black and grey arrows respectively. Additionally, locations of the sequence motif (shown in B) are marked with hexagons. If the sequence motif was found upstream of a TSS, the position is marked with a red hexagon, whereas additional locations of the motif are visualized in white, black edged hexagons. (**B**) Consensus sequence motif upstream of 17 co-regulated MRLGR genes with an assigned TSS according to Droste et al., 2020. The promoter motif (-10 and -35 region) are marked with dashed lines and the conserved palindromic sequence motif overlapping the -35 region is highlighted in grey.

Such a strict co-regulation of several genes in bacteria is likely to be caused by a transcriptional regulator. Therefore, TSS upstream sequences of the MRLGR genes were scanned for potential regulator binding sites. Since several genes are organized in operons (Droste et al., 2020), a TSS could not be identified for all 51 genes.

We identified a conserved sequence motif in the upstream region of 17 TSS inside the MRLGR region each overlapping the -35 promoter region of the respective gene (e-value of 5.4×10^{-29}). The distance to the TSS was found to be 34.4 ± 1.0 bases. The corresponding sequence (5' TCATCC-5nt-GGATGA 3') shows high similarities to reported MalT binding sites in *Escherichia coli* and *Klebsiella pneumoniae*. For these bacteria, MalT homologs are described as a global activator of genes of the maltose metabolism. Interestingly, no genes coding for maltose metabolism were identified in the MRLGR, although these genes seem to be transcribed by maltose availability.

Strikingly, two MalT homologs were identified in the MRLGR (ACSP50_3915 and ACSP50_3917). Since only ACSP50_3915 was found to be up-regulated on maltose and shows high similarity to the MalT regulator in *E. coli* (42% similarity, Figure S2 of this publication), it was assumed, that ACSP50_3915 might be the regulator of the MRLGR.

4.3.1 The transcriptional regulator MalT (ACSP50_3915) is the activator of the MRLGR region in *Actinoplanes* sp. SE50/110

The *malT* gene was deleted using CRISPR/Cas9 technique (Wolf et al., 2016), as well as overexpressed using the strong heterologous promoter P_{gapDH} from *Eggerthella lenta* (Schaffert et al., 2019a). Differential transcriptome analyses of these strains using microarrays and RT-qPCR was performed to analyze the effect of *malT* expression level on the genes inside the MRLGR region (Figure 18). Interestingly, the transcription of the MRLGR region was strongly influenced by the deletion of the *malT* gene (ACSP50_Δ*malT*) on maltose minimal medium but not on glucose (Supplementary material of this publication). This can be explained by the fact, that the genes are not transcribed on glucose at all, which is why no effect on the respective genes is visible on glucose. However, functional analysis of the MRLGR region revealed no connection to the maltose metabolism. Since most of the MRLGR genes are poorly annotated, the function of this region is farther unclear.

In total, 247 genes were found to be differentially transcribed in ACSP50_Δ*malT* compared to the wild type strain on maltose as a carbon source, whereas 141 genes were found to be influenced on glucose. Strikingly, among these genes which seem to be effected through *malT* deletion, no genes involved in maltose metabolism or transport were found (Schaffert et al., 2019b).

All genes of the MRLGR, which were identified to be less transcribed in ACSP50_Δ*malT*, were also previously described as co-regulated, except of ACSP50_3907. Additionally, 23 of the 31 genes, which were both described as co-regulated as well as significantly down-regulated in ACSP50_Δ*malT*, were found to be significantly up-regulated on maltose compared to glucose (Supplementary material of this publication). In conclusion, 42 of 51 genes inside the MRLGR were found to be significantly less transcribed in the deletion mutant ACSP50_Δ*malT* (Figure 18). In contrast, overexpression of *malT* (ACSP50_OE*malT*) leads to up-regulation of 39 of the 51 MRLGR genes compared to the empty vector control (Figure 18). All of these genes were previously found in the deletion mutant to be down-regulated. All genes, which are proposed to be located in one operon, show the same transcriptional behavior dependent on malT expression level.

In conclusion, 37 genes of the MRLGR were identified to be influenced by the expression level of MalT. These findings indicate that MalT is the maltose-dependent transcriptional activator of these genes.

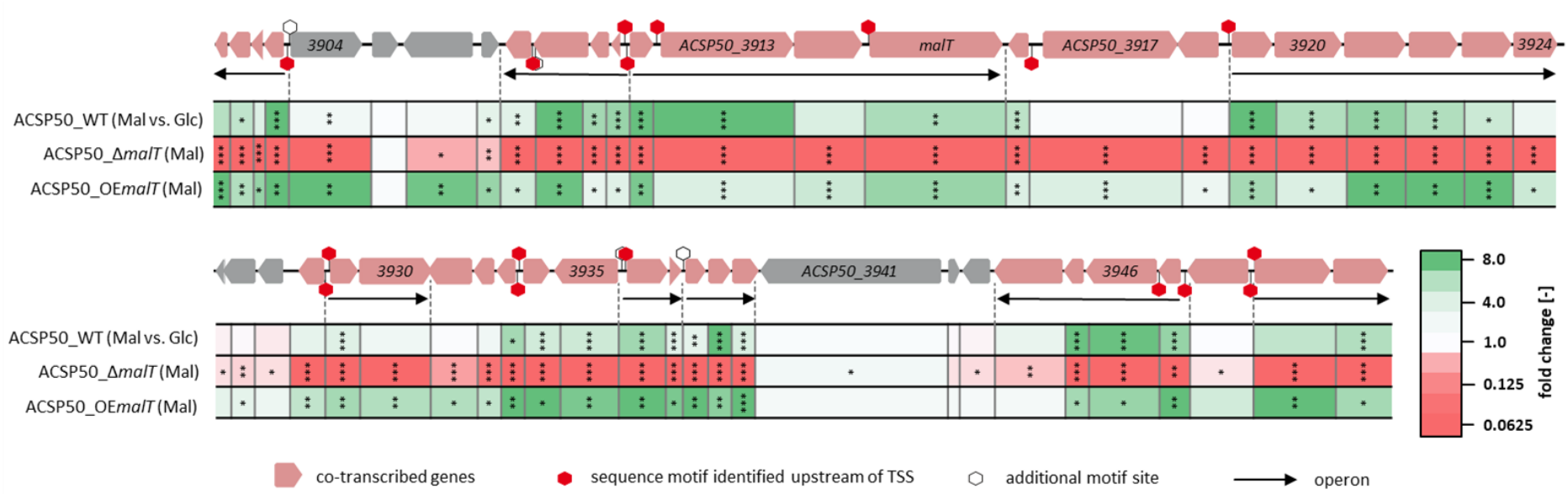

Figure 18: Differential transcriptional analysis of the MRLGR of ACSP50_WT under different expression levels of the transcriptional activator MalT (ACSP50_3915) and on different carbon sources. The values for ACSP50_WT on maltose (Mal) compared to glucose (Glc) (green color indicates increased transcription on maltose) as well as the deletion mutant ACSP50_Δ*malT* and the *malT* overexpression strain ACSP50_OE*malT* on maltose both compared to ACSP50_WT are shown. Heatmap of the fold changes of transcript abundance for the genes in the genomic region surrounding *malT* was derived from whole genome microarray (Mal vs. Glc) and reverse transcription quantitative PCR (RT-qPCR) data (ACSP50_Δ*malT* (Mal) and ACSP50_OE*malT* (Mal)). Green color indicates increased transcription (for "Mal vs. Glc": green = increased on maltose). Significance value of $p < 0.05$ is marked with a single asterisk, $p < 0.01$ with two asterisks and $p < 0.001$ with three asterisks (t-test, two-sample, Holm).

4.3.2 Functional analysis of the proteins encoded by the MRLGR region in *Actinoplanes* sp. SE50/110

The genes of the MRLGR were analyzed regarding their annotation and supposed function. No relation to maltose metabolism could be identified, although the genes seem to be regulated dependent on maltose. However, most of the genes are poorly annotated by the automatic prokka pipeline (Seemann, 2014). Therefore, the tools KEGG mapper (Kanehisa et al., 2016a; Kanehisa and Sato, 2020) and GhostKOALA (Kanehisa et al., 2016b) were used to further analyze the potential function of the genes of the MRLGR region in *Actinoplanes* sp. SE50/110 (full list Is available in supplementary material of this publication). Interestingly, no common pathway or metabolism could be identified containing a significant number of proteins encoded by the MRLGR region, although co-regulation of these genes was observed. However, the strongest accordance was found for 10 proteins, which contain domains with significant similarity to enzymes of the amino acid metabolism (Table 1). It can be assumed, that parts of the MRLGR products are involved in the amino acid metabolism, such as arginine biosynthesis. However, for most of these proteins at least one homologous gene/protein was identified in the *Actinoplanes* sp. SE50/110 genome (Table 1). This could be an indication that these MRLGR genes might be not essential in *Actinoplanes* sp. SE50/110, since a further enzyme is available to take over its function. A maltose-dependent regulation of these amino acid biosynthesis genes is difficult to explain. A possible explanation could be the natural habitat of *Actinoplanes* sp. SE50/110. It is a soil bacterium, isolated from a coffee plantation in Kenia (Frommer et al., 1975). Therefore, a special nutrient supply could be a reason for a sugar-dependent regulation of genes involved in amino acid uptake, peptide degradation and amino acid biosynthesis. A close connection of sugar and amino acid metabolism has been reported for prokaryotes (Gänzle et al., 2007), as well as eukaryotes (Binder, 2010; Rennie and Tipton, 2000). In plants, several regulatory effects of sugar on specific parts of the amino acid metabolisms have been shown (Pratelli and Pilot, 2014). A further explanation could be, that gene products of the MRLGR region are involved in biosynthesis of a metabolite, which is not essential under laboratory conditions. Therefore, it could be regulated dependent on availability of maltose as an indicator of good environmental conditions. However, an analysis of the MRLGR for secondary metabolite genes using the web tool antiSMASH 5.0 (Blin et al., 2019) revealed no significant hits (data not shown).

Table 1: Annotated function of 10 genes inside the MRLGR region. The putative metabolic pathway and homologous genes in the genome of ACSP50_WT were listed.

Locus tag	annotated function (Wolf et al., 2017b)	Metabolic pathway	Homologous genes[1] in ACSP50_WT
ACSP50_3919	class II glutamine amidotransferase	Amino acid metabolism	ACSP50_6409
ACSP50_3920	amino acid permease	Amino acid metabolism	ACSP50_2706; ACSP50_3876
ACSP50_3921	arginine deiminase	Arginine Biosynthesis	ACSP50_8316
ACSP50_3922	ornithine carbamoyltransferase	Arginine Biosynthesis	ACSP50_4060
ACSP50_3923	carbamate kinase	Arginine Biosynthesis	ACSP50_6398
ACSP50_3924	cyclic nucleotide-binding protein (phosphodiesterase)	put. Serine/Threonine Biosynthesis	
ACSP50_3944	beta-Ala-His dipeptidase	Amino acid metabolism	ACSP50_1214
ACSP50_3946	amino acid permease	Amino acid metabolism	
ACSP50_3948	threonine/serine exporter family protein	Serine/Threonine Biosynthesis	
ACSP50_3950	aminopeptidase P family protein	Amino acid metabolism	ACSP50_1832

[1]Revealed by BLASTP analysis, e-value < $7e^{-14}$

Interestingly, the identified genomic region seems to be conserved in different other close relative actinobacteria, like *Pseudosporangium ferrugineum, Couchioplanes caeruleus and Krasilnikovia* cinnamomea (Figure 19). Indeed, the arrangement and order of the homologous genes in these soil bacteria differs from *Actinoplanes* sp. SE50/110. However, for all strains, at least 23 genes homologous to genes from ACSP50_3900 to ACSP50_3950 were found to be located in close

proximity to each other. Nevertheless, not all genes were found in the same order and direction. Some genes, which seem to be organized in operons in *Actinoplanes* sp. SE50/110 are rearranged in the other strains. Even between the two *Actinoplanes* spp. clear differences were identified regarding this genomic region. According to the expectations, in all analyzed genomes harboring parts of MRLGR homologs, a MalT-like regulator could be found. Furthermore, the palindromic sequence motif, the potential MalT binding site, identified in the -35 region of the MRLGR genes (Figure 17B) could also be identified upstream of several TLS in the analyzed genomic regions of *A. missouriensis, P. ferrugineum, C. caeruleus* and *K. cinnamomea* (Figure 19). This confirms the close relation of these genomic regions and the role of MalT as an important regulator of this genomic region.

In conclusion, it can be assumed that the MRLGR region harbors genes, which are important for the specific habitats of *Actinoplanes* sp. SE50/110. As it can be found partly in other soil bacteria, which occur in similar environments, the proteins encoded in this genomic region could be involved in uptake and degradation of specific nutrients or in production of an optional metabolite.

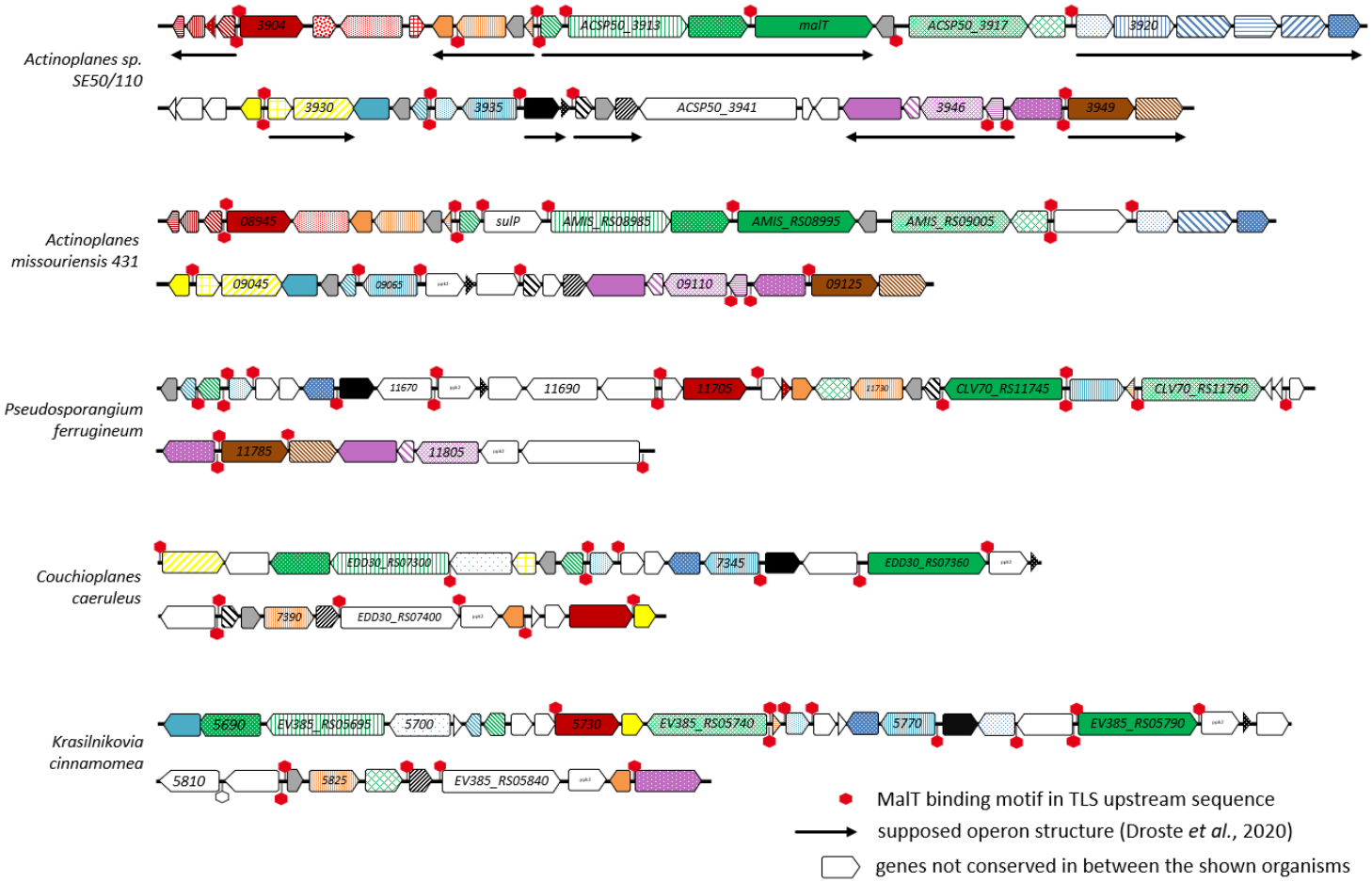

Figure 19: Comparison of the MRLGR region to similar genomic regions of *Actinoplanes missouriensis*, *Pseudosporangium ferrugineum*, *Couchioplanes caeruleus* and *Krasilnikovia cinnamomea* containing homologous gene products identified by BLAST analysis. Genes of homologous proteins are marked in the same color code. The positions of the conserved sequence motif (5'-TCATCC-5 bp-GGATGA-3') in all strains is marked with red (upstream of TLS) and white (additional sites) hexagons.

5. Conclusion and perspectives

In this work, insights into the transcriptional regulation of *Actinoplanes* sp. SE50/110 were gained with specific focus on the acarbose biosynthesis (*acb*) gene cluster. As a starting point, the expression profile of all genes in the *Actinoplanes* sp. SE50/110 genome was analyzed.

In controlled fermenter cultivations it could be shown, that acarbose is predominantly produced during growth, but the specific product formation rate is continuously decreasing until the cells reaching the stationary phase. Therefore, a database of transcription and protein courses during the whole cultivation time was recorded. In order to identify co-regulated genes and operons, cluster analysis was performed and genes showing a similar transcription pattern were identified.

With this in hand regulatory networks in *Actinoplanes* sp. SE50/110 can be elucidated. In addition to that, a genome wide TSS database was generated including promoter motif sites, RBS and 5´-UTR lengths. The TSS data were used to elaborate the operon structure of the *Actinoplanes* sp. SE50/110 genome. This information is indispensable for the detailed analysis of transcriptional regulation.

Moreover, transcriptome and proteome profiles of all CDS within the *acb* gene cluster could confirm the growth-dependent biosynthesis of acarbose. It could be shown that the transcription of all *acb* genes behaves in parallel to the course of the specific product formation rate. Therefore, a co-regulation of all *acb* genes was assumed on a transcriptional level. However, at protein level the picture is different. The course of several Acb proteins is in good accordance with their respective transcription dynamics, but for the proteins AcbB, AcbD and AcbZ this was not observed. These proteins seem to be more stable compared to the other Acb proteins. Moreover, several putative bottlenecks of the acarbose biosynthesis could be identified. It could be shown that abundancies for AcbA and AcbI, AcbJ, AcbM, AcbO and AcbC show a stronger decrease on protein level compared to the other Acb proteins. Furthermore, additional genes could be identified, which seem to be co-regulated to the *acb* gene cluster and were not in focus of acarbose research before. These genes could be further targets for genetic engineering to improve acarbose formation in *Actinoplanes* sp. SE50/110.

The transcriptional regulator AcrC was identified as a repressor of the genes *acbD* and *acbE*. The repressor was identified by comparison of the regulator repertoire of *Actinoplanes* sp. SE50/110 to a regulator gene identified in the acarbose biosynthesis gene clusters of *Streptomyces glaucescens* GLA.O and *Streptomyces coelicoflavus* ZG0656. It could be shown that AcrC binds to a palindromic sequence upstream of the TSS of the adjacently located and oppositely directed genes *acbD* and *acbE*. This could be proven by EMSA. Based on bioinformatic analyses the AcrC regulator has a sugar binding site, which could be a maltose derivative. Therefore, it can be assumed that a sugar molecule acts as an effector. This shows the close connection of acarbose and sugar (maltose) metabolism. The deletion of *acrC* lead to an increased specific acarbose formation rate in the early growth phase. However, the final acarbose concentration seems not to be elevated in the mutant strain. This can be explained by the fact, that AcbD and AcbE are not involved in acarbose biosynthesis itself. This underlines the need of elucidation of the transcriptional regulation of all *acb* genes to further improve acarbose formation.

In addition to growth-dependent production, previous studies show a dependence on the biosynthesis of acarbose to the carbon source provided. Due to the close connection of maltose and acarbose metabolism, differential transcriptome analysis was performed comparing cells grown on maltose and glucose as a carbon source. This way, genes were identified which are highly transcribed on maltose. Interestingly, no genes of the maltose or acarbose metabolism are among the top-scorers of this experiment. However, a large genomic region was identified to be highly transcribed on maltose. These genes were analyzed regarding their potential function in *Actinoplanes* sp. SE50/110. Interestingly, no connection to maltose metabolism or related pathways was found. Nevertheless, this maltose-regulated large genomic region (MRLGR) was further studied and its responsible transcription factor MalT (ACSP50_3915) was characterized. It could be shown that MalT activates the genes of the MRLGR dependent on maltose or another maltose-derived metabolite as an effector. A MalT binding site was found upstream of the regulated genes similar to MalT binding sites in *E. coli*.

Based on the present omics data and bioinformatic analyses, important milestones for understanding transcriptional regulation of the *acb* genes were achieved. A comprehensive time-resolved transcriptomic and proteomic analyses were performed

to gain expression profiles of all genes in the *Actinoplanes* sp. SE50/110 genome. Moreover, the first transcription factor of the acarbose biosynthesis (AcrC) was identified and analyzed in detail. Furthermore, a second regulator (MalT) dependent on maltose but not involved in maltose metabolism was described and elucidated. Those analyses can serve as blue print for future studies on secondary metabolite regulators in *Actinoplanes* sp. SE50/110 or close related bacteria.

Nevertheless, there are several open questions need to be answered for a complete understanding of the transcriptional regulation of all acarbose biosynthesis genes and the related metabolic pathways. Transcriptional regulators of the acarbose transporter operon and the intracellular biosynthesis still need to be identified. Moreover, the data indicate a post-transcriptional regulation of some gene products within the *acb* gene cluster, which should be investigated in more detail in the future. However, the basis for further research in this area was laid through this work. The fundamental knowledge gained in this study will be of high value for further analyses in the field of transcriptional regulation and metabolic engineering of the biotechnologically relevant bacterium *Actinoplanes* sp. SE50/110.

6. References

Afzal, M., Shafeeq, S., Manzoor, I., Kuipers, O.P., 2015. Maltose-Dependent Transcriptional Regulation of the mal Regulon by MalR in *Streptococcus pneumoniae*. PloS one 10 (6), e0127579.

Bathke, J., Konzer, A., Remes, B., McIntosh, M., Klug, G., 2019. Comparative analyses of the variation of the transcriptome and proteome of *Rhodobacter sphaeroides* throughout growth. BMC genomics 20 (1), 358.

Berghoff, B.A., Konzer, A., Mank, N.N., Looso, M., Rische, T., Förstner, K.U., Krüger, M., Klug, G., 2013. Integrative "omics"-approach discovers dynamic and regulatory features of bacterial stress responses. PLoS Genetics 9 (6), e1003576.

Binder, S., 2010. Branched-Chain Amino Acid Metabolism in *Arabidopsis thaliana*. The arabidopsis book 8, e0137.

Bischoff, H., Ahr, H.J., Schmidt, D., Stoltefuß, J., 1994. Acarbose - ein neues Wirkprinzip in der Diabetestherapie. Nachr. Chem. Tech. Lab. 42 (11), 1119–1128.

Bland, C.E., Couch, J.N., Starr, M., Stolp, H., Trüper, H., Balous, A., Schleges, H., 1981. The family *Actinoplanaceae*. The prokaryotes (II), 2004–2010.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., Weber, T., 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47 (W1), W81-W87.

Bowers, S.G., Mahmud, T., Floss, H.G., 2002. Biosynthetic studies on the α-glucosidase inhibitor acarbose: the chemical synthesis of dTDP-4-amino-4,6-dideoxy-α-d-glucose. Carbohydrate Research 337 (4), 297–304.

Brink, B.G., Seidel, A., Kleinbölting, N., Nattkemper, T.W., Albaum, S.P., 2016. Omics Fusion - A Platform for Integrative Analysis of Omics Data. Journal of integrative bioinformatics 13 (4), 296.

Brinkrolf, K., 2004. Charakterisierung transkriptioneller Regulatoren der LacI/GalRFamilie von *Corynebacterium glutamicum* durch komparative Genomik und globale Transkriptionsanalysen mit DNA-Microarrays. Diplomarbeit, Bielefeld.

Brockmann-Gretza, O., Kalinowski, J., 2006. Global gene expression during stringent response in *Corynebacterium glutamicum* in presence and absence of the rel gene encoding (p)ppGpp synthase. BMC genomics 7, 230.

Brown, N.L., Stoyanov, J.V., Kidd, S.P., Hobman, J.L., 2003. The MerR family of transcriptional regulators. FEMS Microbiol Rev 27 (2-3), 145–163.

Browning, D.F., Busby, S.J., 2004. The regulation of bacterial transcription initiation. Nature reviews. Microbiology 2 (1), 57–65.

Browning, D.F., Busby, S.J.W., 2016. Local and global regulation of transcription initiation in bacteria. Nature reviews. Microbiology 14 (10), 638–650.

Brunkhorst, C., Schneider, E., 2005. Characterization of maltose and maltotriose transport in the acarbose-producing bacterium *Actinoplanes* sp. Research in microbiology 156 (8), 851–857.

Brunkhorst, C., Wehmeier, U.F., Piepersberg, W., Schneider, E., 2005. The *acbH* gene of *Actinoplanes* sp. encodes a solute receptor with binding activities for acarbose and longer homologs. Research in microbiology 156 (3), 322–327.

Buchanan, R.E., Gibbons, N.E., 1986. Bergey´s manual of determinative bacteriology. The Williams & Wilkins Co, Baltimore (8. Edition), 657–710.

Casadesús, J., Low, D., 2006. Epigenetic gene regulation in the bacterial world. Microbiology and Molecular Biology Reviews 70 (3), 830–856.

Caspary, W.F., Graf, S., 1979. Inhibition of human intestinal α-glucosidehydrolases by a new complex oligosaccharide. Research in Experimental Medicine (175), 1–6.

Christiano, R., Nagaraj, N., Fröhlich, F., Walther, T.C., 2014. Global proteome turnover analyses of the Yeasts *S. cerevisiae* and *S. pombe*. Cell reports 9 (5), 1959–1965.

Cooper, R., Truumees, I., Gunnarsson, I., Loebenberg, D., Horan, A., Marquez, J., Patel, M., Gullo, V., Puar, M., Das, P., 1992. Sch 42137, a novel antifungal antibiotic from an *Actinoplanes* sp. Fermentation, isolation, structure and biological properties. The Journal of antibiotics 45 (4), 444–453.

Couch, J.N., 1950. *Actinoplanes*, a New Genus of the *Actinomycetales*. Journal of the Mitchell Society, 87–95.

Couch, J.N., 1963. Some New Genera and Species of the *Actinoplanaceae*. Journal of the Mitchell Society, 53–70.

Degwert, U., van Hülst, R., Pape, H., 1987. Studies on the Biosynthesis of the α-Glucosidase Inhibitor Acarbose: Valienamine, a m-C7N Unit not derived from Shikimate Pathway. The Journal of antibiotics 40 (6), 855–861.

deHaseth, P.L., Zupancic, M.L., Record, M.T., 1998. RNA Polymerase-Promoter Interactions: the Comings and Goings of RNA Polymerase. Journal of bacteriology 180 (12), 3019–3025.

Drepper, A., Pape, H., 1996. Acarbose 7-Phosphotransferase from *Actinoplanes* sp.: Purification, Properties and Possible Physiological Function. The Journal of antibiotics 49 (7), 664–668.

Droste, J., Ortseifen, V., Persicke, M., Schneiker-Bekel, Susanne, Pühler, Pühler, A., Kalinowski, J., 2020. The expression of the acarbose biosynthesis gene cluster in *Actinoplanes* sp. SE50/110 is dependent on the growth phase.

Elf, J., Berg, O.G., Ehrenberg, M., 2001. Comparison of repressor and transcriptional attenuator systems for control of amino acid biosynthetic operons. Journal of molecular biology 313 (5), 941–954.

Frederick, D.L., Tatchell, K., 1996. The REG2 gene of *Saccharomyces cerevisiae* encodes a type 1 protein phosphatase-binding protein that functions with Reg1p and the Snf1 protein kinase to regulate growth. Molecular and cellular biology 16 (6), 2922–2931.

Frommer, W., Junge, B., Müller, L., Schmidt, D., Truscheit, E., 1979. Neue Enzyminhibitoren aus Mikroorganismen. Journal of Medicinal Plant Research (3), 195–217.

Frommer, W., Puls, W., Schaefer, D., Schmidt, D., 1975. Glycosidehydrolase enzyme inhibitors.

Gally, D.L., Bogan, J.A., Eisenstein, B.I., Blomfield, I.C., 1993. Environmental regulation of the fim switch controlling type 1 fimbrial phase variation in *Escherichia coli* K-12: effects of temperature and media. Journal of bacteriology 175 (19), 6186–6193.

Galperin, M.Y., Makarova, K.S., Wolf, Y.I., Koonin, E.V., 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic acids research 43 (Database issue), D261-9.

Gänzle, M.G., Vermeulen, N., Vogel, R.F., 2007. Carbohydrate, peptide and lipid metabolism of lactic acid bacteria in sourdough. Food microbiology 24 (2), 128–138.

Garrity, G.M., Heimbuch, B.K., Gagliardi, M., 1996. Isolation of zoosporogenous actinomycetes from desert soils. Journal of industrial microbiology & biotechnology 17 (3-4), 260–267.

Gerlach, P., Valentin-Hansen, P., Bremer, E., 1990. Transcriptional regulation of the cytR repressor gene of *Escherichia coli*: autoregulation and positive control by the cAMP/CAP complex. Mol Microbiol 4 (3), 479–488.

Ghosh, T., Bose, D., Zhang, X., 2010. Mechanisms for activating bacterial RNA polymerase. FEMS microbiology reviews 34 (5), 611–627.

Gottesman, S., 1984. Bacterial regulation: global regulatory networks. Annual review of genetics 18, 415–441.

Greenbaum, D., Colangelo, C., Williams, K., Gerstein, M., 2003. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome biology 4 (9), 117.

Gren, T., 2017. Development and application of genetic engineering methods for *Actinoplanes* sp. SE50/110. PhD thesis.

Gust, B., Challis, G.L., Fowler, K., Kieser, T., Chater, K.F., 2003. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc Natl Acad Sci U S A 100 (4), 1541–1546.

Heiker, F.R., Böshagen, H., Junge, B., Müller, L., Stoltefuß, J., 1981. Studies designed to localize the essential structural unit of glycoside-hydrolase inhibitors of the acarbose type. First international symposium in acarbose, 1981, Amsterdam.

Hemker, M., Stratmann, A., Goeke, K., Schroder, W., Lenz, J., Piepersberg, W., Pape, H., 2001. Identification, Cloning, Expression, and Characterization of the Extracellular Acarbose-Modifying Glycosyltransferase, AcbD, from *Actinoplanes* sp. Strain SE50. Journal of bacteriology 183 (15), 4484–4492.

Henderson, I.R., Owen, P., Nataro, J.P., 1999. Molecular switches--the ON and OFF of bacterial phase variation. Mol Microbiol 33 (5), 919–932.

Henkin, T.M., Grundy, F.J., Nicholson, W.L., Chambliss, G.H., 1991. Catabolite repression of alpha-amylase gene expression in *Bacillus subtilis* involves a trans-acting gene product homologous to the *Escherichia coli* lacI and galR repressors. Mol Microbiol 5 (3), 575–584.

Henkin, T.M., Yanofsky, C., 2002. Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. BioEssays : news and reviews in molecular, cellular and developmental biology 24 (8), 700–707.

Huffman, J.L., Brennan, R.G., 2002. Prokaryotic transcription regulators: More than just the helix-turn-helix motif. Current Opinion in Structural Biology 12 (1), 98–106.

International Diabetes Federation, 2014. Diabetes Atlas: Key Findings 2014. https://www.idf.org/diabetesatlas. Accessed 13 October 2015.

Ishihama, A., 1997. Promoter Selectivity Control of RNA Polymerase, in: Eckstein, F., Lilley, D.M.J. (Eds.), Mechanisms of Transcription, vol. 11. Nucleic Acids and Molecular Biology 11. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 53–70.

Jang, M.-S., Mouri, Y., Uchida, K., Aizawa, S.-I., Hayakawa, M., Fujita, N., Tezuka, T., Ohnishi, Y., 2016. Genetic and Transcriptional Analyses of the Flagellar Gene Cluster in *Actinoplanes missouriensis*. Journal of bacteriology 198 (16), 2219–2227.

Jensen, P.R., Dwight, R., Fenical, W., 1991. Distribution of actinomycetes in near-shore tropical marine sediments. Applied and Environmental Microbiology 57 (4), 1102–1108.

Jeong, Y., Kim, J.-N., Kim, M.W., Bucca, G., Cho, S., Yoon, Y.J., Kim, B.-G., Roe, J.-H., Kim, S.C., Smith, C.P., Cho, B.-K., 2016. The dynamic transcriptional and translational landscape of the model antibiotic producer *Streptomyces coelicolor* A3(2). Nature communications 7, 11605.

Jiang, X., Hall, A.B., Arthur, T.D., Plichta, D.R., Covington, C.T., Poyet, M., Crothers, J., Moses, P.L., Tolonen, A.C., Vlamakis, H., Alm, E.J., Xavier, R.J., 2019. Invertible promoters mediate bacterial phase variation, antibiotic resistance, and host adaptation in the gut. Science (New York, N.Y.) 363 (6423), 181–187.

Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 28 (1), 27–30.

Koussounadis, A., Langdon, S.P., Um, I.H., Harrison, D.J., Smith, V.A., 2015. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. Scientific reports 5, 10775.

Krause, J.P., Polen, T., Youn, J.-W., Emer, D., Eikmanns, B.J., Wendisch, V.F., 2012. Regulation of the malic enzyme gene malE by the transcriptional regulator MalR in *Corynebacterium glutamicum*. Journal of biotechnology 159 (3), 204–215.

Lee, D.J., Minchin, S.D., Busby, S.J.W., 2012. Activating transcription in bacteria. Annual review of microbiology 66, 125–152.

Lee, J.-S., 2002. Molekulare Analysen zur Oligosaccharidbildung in *Actinoplanes* sp. Inagural-Dissertation, Münster, 178 pp.

Leemhuis, H., Wehmeier, U.F., Dijkhuizen, L., 2004. Single amino acid mutations interchange the reaction specificities of cyclodextrin glycosyltransferase and the acarbose-modifying enzyme acarviosyl transferase. Biochemistry 43 (41), 13204–13213.

Lempp, M., Farke, N., Kuntz, M., Freibert, S.A., Lill, R., Link, H., 2019. Systematic identification of metabolites controlling gene expression in *E. coli.* ncomms 10 (1), 4463.

Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G., Lu, P., 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. Science (New York, N.Y.) 271 (5253), 1247–1254.

Licht, A., Bulut, H., Scheffel, F., Daumke, O., Wehmeier, U.F., Saenger, W., Schneider, E., Vahedi-Faridi, A., 2011. Crystal structures of the bacterial solute receptor AcbH displaying an exclusive substrate preference for β-D-galactopyranose. Journal of molecular biology 406 (1), 92–105.

Lloyd, G., Landini, P., Busby, S., 2001. Activation and repression of transcription initiation in bacteria. Essays in biochemistry 37, 17–31.

Ludwig, W., Euzéby, J., Whitman, W.B., 2012. Taxonomic outline of the phylum Actinobacteria, in: Goodfellow, M., Busse, H.-J., Kämpfer, P., Ludwig, W., Suzuki, K.-i., Trujillo, M.E., Whitman, W.B. (Eds.), Bergey's Manual® of Systematic Bacteriology. Volume Five The Actinobacteria, Part A, Second edition ed. Springer New York, New York, NY, pp. 29–31.

Luscombe, N.M., Austin, S.E., Berman, H.M., Thornton, J.M., 2000. An overview of the structures of protein-DNA complexes. Genome biology 1 (1), REVIEWS001.

Lyubetskaya, E.V., Leont'ev, L.A., Gelfand, M.S., Lyubetsky, V.A., 2003. Molecular Biology 37 (5), 707–715.

Madan Babu, M., Teichmann, S.A., 2003. Functional determinants of transcription factors in *Escherichia coli:* protein families and binding sites. Trends in genetics : TIG 19 (2), 75–79.

Maier, T., Schmidt, A., Güell, M., Kühner, S., Gavin, A.-C., Aebersold, R., Serrano, L., 2011. Quantification of mRNA and protein and integration with protein turnover in a bacterium. Molecular systems biology 7, 511.

Martin, J.F., 1992. Clusters of genes for the biosynthesis of antibiotics: regulatory genes and overproduction of pharmaceuticals. Journal of industrial microbiology & biotechnology (9.2), 73–90.

Martínez-Antonio, A., Collado-Vides, J., 2003. Identifying global regulators in transcriptional regulatory networks in bacteria. Current opinion in microbiology 6 (5), 482–489.

Matic, I., Taddei, F., Radman, M., 2004. Survival versus maintenance of genetic stability: a conflict of priorities during stress. Research in microbiology 155 (5), 337–341.

Merettig, N., 2009. Acarbose als Carbophor: Biochemische Charakterisierung Acarbose-resistenter α-Glucosidasen und ihre Bedeutung für die Produzentenstämme. Dissertation, Wuppertal, 172 pp.

Merino, E., Yanofsky, C., 2005. Transcription attenuation: a highly conserved regulatory strategy used by bacteria. Trends in genetics : TIG 21 (5), 260–264.

Monsalve, M., Mencía, M., Salas, M., Rojo, F., 1996. Protein p4 represses phage phi 29 A2c promoter by interacting with the alpha subunit of *Bacillus subtilis* RNA polymerase. Proc Natl Acad Sci U S A 93 (17), 8913–8918.

Mouri, Y., Konishi, K., Fujita, A., Tezuka, T., Ohnishi, Y., 2017. Regulation of Sporangium Formation by BldD in the Rare Actinomycete *Actinoplanes missouriensis*. Journal of bacteriology 199 (12).

Murakami, K.S., Darst, S.A., 2003. Bacterial RNA polymerases: the wholo story. Current Opinion in Structural Biology 13 (1), 31–39.

Nahoum, V., Roux, G., Anton, V., Rouge, P., Puigserver, A., Bischoff, H., Henrissat, B., Payan, F., 2000. Crystal structures of human pancreatic α-amylase in complex with carbohydrate and proteinaceous inhibitors. Biochemical Journal (346), 201–208.

Naville, M., Gautheret, D., 2009. Transcription attenuation in bacteria: theme and variations. Briefings in functional genomics & proteomics 8 (6), 482–492.

Nguyen, C.C., Saier, M.H., 1995. Phylogenetic, structural and functional analyses of the LacI-GalR family of bacterial transcription factors. FEBS Letters 377 (2), 98–102.

Nguyen, J., 1999. The regulatory protein Reg1 of *Streptomyces lividans* binds the promoter region of several genes repressed by glucose. FEMS Microbiology Letters 175 (1), 51–58.

Nguyen, J., Francou, F., Virolle, M.-J., Guérineau, M., 1997. Amylase and Chitinase Genes in *Streptomyces lividans* are Regulated by reg1, a Pleiotropic Regulatory Gene. Journal of bacteriology 179 (20), 6383–6390.

Nieselt, K., Battke, F., Herbig, A., Bruheim, P., Wentzel, A., Jakobsen, Ø.M., Sletta, H., Alam, M.T., Merlo, M.E., Moore, J., Omara, W.A.M., Morrissey, E.R., Juarez-Hermosillo, M.A., Rodríguez-García, A., Nentwich, M., Thomas, L., Iqbal, M., Legaie, R., Gaze, W.H., Challis, G.L., Jansen, R.C., Dijkhuizen, L., Rand, D.A., Wild, D.L., Bonin, M., Reuther, J., Wohlleben, W., Smith, M.C.M., Burroughs, N.J., Martín, J.F., Hodgson, D.A., Takano, E., Breitling, R., Ellingsen, T.E., Wellington, E.M.H., 2010. The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*. BMC genomics 11, 10.

Nudler, E., Mironov, A.S., 2004. The riboswitch control of bacterial metabolism. Trends in biochemical sciences 29 (1), 11–17.

Oehler, S., Eismann, E.R., Krämer, H., Müller-Hill, B., 1990. The three operators of the lac operon cooperate in repression. The EMBO Journal 9 (4), 973–979.

Oman, T.J., van der Donk, W.A., 2010. Follow the leader: the use of leader peptides to guide natural product biosynthesis. Nature chemical biology 6 (1), 9–18.

Ortseifen, V., 2016. Genombasierte Modellbildung zur Biosynthese von Acarviostatin-Metaboliten in drei *Actinoplanes* sp. SE50/110-Stämmen, Dissertation, Bielefeld, 160 pp.

Ortseifen, V., Kalinowski, J., Pühler, A., Rückert, C., 2017. The complete genome sequence of the actinobacterium *Streptomyces glaucescens* GLA.O (DSM 40922) carrying gene clusters for the biosynthesis of tetracenomycin C, 5`-hydroxy streptomycin, and acarbose. Journal of biotechnology 262, 84–88.

Pabo, C.O., Sauer, R.T., 1992. Transcription factors: structural families and principles of DNA recognition. Annual review of biochemistry 61, 1053–1095.

Paget, M.S., 2015. Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. Biomolecules 5 (3), 1245–1265.

Palleroni, N.J., 1976. Chemotaxis in *Actinoplanes*. Archives of microbiology 110 (1), 13–18.

Parenti, F., Coronelli, C., 1979. Members of the Genus *Actinoplanes* and their Antibiotics. Annual review of microbiology (33), 389–411.

Perez-Rueda, E., Collado-Vides, J., 2000. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. Nucleic acids research 28 (8), 1838–1847.

Philips, S.J., Canalizo-Hernandez, M., Yildirim, I., Schatz, G.C., Mondragón, A., O'Halloran, T.V., 2015. TRANSCRIPTION. Allosteric transcriptional regulation via changes in the overall topology of the core promoter. Science (New York, N.Y.) 349 (6250), 877–881.

Picard, F., Milhem, H., Loubière, P., Laurent, B., Cocaign-Bousquet, M., Girbal, L., 2012. Bacterial translational regulations: high diversity between all mRNAs and major role in gene expression. BMC genomics 13, 528.

Piepersberg, W., Distler, J., 2001. Aminoglycosides and Sugar Components in Other Secondary Metabolites, in: Rehm, H.-J., Reed, G. (Eds.), Biotechnology Set. Wiley-VCH Verlag GmbH, Weinheim, Germany, pp. 397–488.

Piepersberg, W., Uribe, P.M.D.-G., Stratmann, A., Thomas, H., Wehmeier, U.F., Zhang, C.-S., 2002. Developments in the Biosynthesis and Regulation of Aminoglycosides. Microbial Secondary Metabolites: Biosynthesis, Genetics and Regulation, 1–26.

Power, P.M., Sweetman, W.A., Gallacher, N.J., Woodhall, M.R., Kumar, G.A., Moxon, E.R., Hood, D.W., 2009. Simple sequence repeats in *Haemophilus influenzae*. Infection, Genetics and Evolution 9 (2), 216–228.

Pratelli, R., Pilot, G., 2014. Regulation of amino acid metabolic enzymes and transporters in plants. Journal of experimental botany 65 (19), 5535–5556.

Puls, W., Keup, U., Krause, H.P., Thomas, G., Hoffmeister, F., 1977. Glucosidase inhibition. Naturwissenschaften 64 (10), 536–537.

Rauenbusch, E., Schmidt, D., 1978. Verfahren zur Isolierung von (O {4,6-Dideoxy-4[[1S-(1,4,6/5)-4,5,6-trihydroxy-3-hydroxymethyl-2-cyclohexen-1-yl]-amino]-α-D-glucopyranosyl}-(1→4)-O-α-D-glucopyranosyl-(1→4)-D-glucopyranose) aus Kulturbrühen.

Rennie, M.J., Tipton, K.D., 2000. Protein and amino acid metabolism during and after exercise and the effects of nutrition. Annual review of nutrition 20, 457–483.

Reznikoff, W.S., Winter, R.B., Hurley, C.K., 1974. The location of the repressor binding sites in the lac operon. Proc Natl Acad Sci U S A 71 (6), 2314–2318.

Rockser, Y., Wehmeier, U.F., 2009. The gac-gene cluster for the production of acarbose from *Streptomyces glaucescens* GLA.O: identification, isolation and characterization. Journal of biotechnology 140 (1-2), 114–123.

Rodionov, D.A., 2007. Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. Chem. Rev. 107 (8), 3467–3497.

Rojo, F., 1999. Repression of Transcription Initiation in Bacteria. Journal of bacteriology 181 (10), 2987–2991.

Rutkauskas, D., Zhan, H., Matthews, K.S., Pavone, F.S., Vanzi, F., 2009. Tetramer opening in LacI-mediated DNA looping. Proc Natl Acad Sci U S A 106 (39), 16627–16632.

Ruwe, M., Persicke, M., Busche, T., Müller, B., Kalinowski, J., 2019. Physiology and Transcriptional Analysis of (p)ppGpp-Related Regulatory Effects in *Corynebacterium glutamicum*. Frontiers in microbiology 10, 2769.

Schaffert, L., März, C., Burkhardt, L., Droste, J., Brandt, D., Busche, T., Rosen, W., Schneiker-Bekel, S., Persicke, M., Pühler, A., Kalinowski, J., 2019a. Evaluation of vector systems and promoters for overexpression of the acarbose biosynthesis gene *acbC* in *Actinoplanes* sp. SE50/110. Microbial cell factories 18 (1), 114.

Schaffert, L., Schneiker-Bekel, S., Dymek, S., Droste, J., Persicke, M., Busche, T., Brandt, D., Pühler, A., Kalinowski, J., 2019b. Essentiality of the Maltase AmlE in Maltose Utilization and Its Transcriptional Regulation by the Repressor AmlR in the Acarbose-Producing Bacterium *Actinoplanes* sp. SE50/110. Frontiers in microbiology 10, 2448.

Schatz, H., 2006. Diabetologie kompakt: Grundlagen und Praxis, 4., erw. und aktualisierte Aufl ed. Thieme, Stuttgart, xviii, 405.

Schlösser, A., Weber, A., Schrempf, H., 2001. Synthesis of the *Streptomyces lividans* maltodextrin ABC transporter depends on the presence of the regulator MalR. FEMS Microbiology Letters 196, 77–83.

Schmidt, D., Frommer, W., Junge, B., Müller, L., Wingender, W., Truscheit, E., 1977. α-Glucosidase Inhibitors: New Complex Oligosaccharides of Microbial Origin. Naturwissenschaften (64), 535–536.

Schwientek, P., 2012. Genomics and Transcriptomics of the Industrial Acarbose Producer *Actinoplanes* sp. SE50/110. Dissertation, Bielefeld, 155 pp.

Schwientek, P., Szczepanowski, R., Rückert, C., Kalinowski, J., Klein, A., Selber, K., Wehmeier, U.F., Stoye, J., Pühler, A., 2012. The complete genome sequence of the acarbose producer *Actinoplanes* sp. SE50/110. BMC genomics 13, 112.

Schwientek, P., Wendler, S., Neshat, A., Eirich, C., Rückert, C., Klein, A., Wehmeier, U.F., Kalinowski, J., Stoye, J., Pühler, A., 2013. Comparative RNA-sequencing of the acarbose producer *Actinoplanes* sp. SE50/110 cultivated in different growth media. Journal of biotechnology 167 (2), 166–177.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics (Oxford, England) 30 (14), 2068–2069.

Seshasayee, A.S.N., Sivaraman, K., Luscombe, N.M., 2011. An overview of prokaryotic transcription factors : a summary of function and occurrence in bacterial genomes. Sub-cellular biochemistry 52, 7–23.

Stratmann, A., 1997. Identifizierung eines Acarbose-Biosynthesegenclusters in *Actinoplanes* sp. und Charakterisierung ausgewählter Enzyme des Acarbose-Stoffwechsels. Dissertation, Wuppertal.

Stratmann, A., Mahmud, T., Lee, S., Distler, J., Floss, H.G., Piepersberg, W., 1999. The AcbC Protein from *Actinoplanes* Species is a C7-cyclitol Synthase related to 3-Dehydroquinate Synthase and is involved in the biosynthesis of the α-Glucosidase Inhibitor Acarbose. Journal of Biological Chemistry 16 (274), 10889–10896.

Strauch, E., Takano, E., Baylis, H.A., Bibb, M.J., 1991. The stringent response in *Streptomyces coelicolor* A3(2). Molecular microbiology 5 (2), 289–298.

Swint-Kruse, L., Matthews, K.S., 2009. Allostery in the LacI/GalR family: variations on a theme. Current opinion in microbiology 12 (2), 129–137.

Szaniszlo, P., 1967. The nature of the intramycelial pigmentation of some *Actinoplanaceae*. J. Elisha Mitchell Sci. Soc (84), 24–26.

Thomas, H., 2001. Acarbose-Metabolismus in *Actinoplanes* sp. SE50/110. Dissertation, Wuppertal, 132 pp.

Truscheit, E., Frommer, W., Junge, B., Müller, L., Schmidt, D., Wingender, W., 1981. Chemistry and Biochemistry of Microbial α-Glucosidase Inhibitors. Angewandte Chemie (20), 744–761.

Uchida, K., Jang, M.-S., Ohnishi, Y., Horinouchi, S., Hayakawa, M., Fujita, N., Aizawa, S.-I., 2011. Characterization of *Actinoplanes missouriensis* spore flagella. Applied and Environmental Microbiology 77 (7), 2559–2562.

Uribe, P.M.D.-G., 2001. Untersuchungen zum Einbau des Stickstoffes in der Acarviose-Einheit der Acarbose bei *Actinoplanes* sp. 50/110: Die Aminotransferase AcbV. Dissertation, Wuppertal, 112 pp.

van der Woude, M.W., Henderson, I.R., 2008. Regulation and function of Ag43 (flu). Annual review of microbiology 62, 153–169.

van Wezel, G.P., White, J., Bibb, M.J., Postma, P.W., 1997a. The malEFG gene cluster of *Streptomyces coelicolor* A3(2): characterization, disruption and transcriptional analysis. Molecular & general genetics : MGG 254 (5), 604–608.

van Wezel, G.P., White, J., Young, P., Postma, P.W., Bibb, M.J., 1997b. Substrate induction and glucose repression of maltose utilization by *Streptomyces coelicolor* A3(2) is controlled by malR, a member of the lacI-galR family of regulatory genes. Molecular microbiology 23 (3), 537–549.

Virolle, M.J., Bibb, M.J., 1988. Cloning, characterization and regulation of an alpha-amylase gene from *Streptomyces limosus*. Mol Microbiol 2 (2), 197–208.

Virolle, M.-J., Gagnat, J., 1994. Sequences involved in growth-phase dependent expression and glucose repression of a *Streptomyces* α-amylase gene. Microbiology (Reading, England) (140), 1059–1067.

Vobis, G. (Ed.), 1989. *Actinoplanetes*. Section 28 IV. The Williams & Wilkins Co.

Vobis, G. (Ed.), 2006. The Genus *Actinoplanes* and Related Genera. Springer, New York, NY.

Vobis, G., Schäfer, J., Kämpfer, P., 2015. *Actinoplanes*, in: Whitman, W.B. (Ed.), Bergey's manual of systematics of archaea and bacteria, vol. 58. Wiley, [Hoboken, New Jersey], pp. 1–41.

Vogel, C., Marcotte, E.M., 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nature reviews. Genetics 13 (4), 227–232.

Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., Brown, P.O., 2002. Precision and functional specificity in mRNA decay. Proc Natl Acad Sci U S A 99 (9), 5860–5865.

Wehmeier, U.F., 2003. The Biosynthesis and Metabolism of Acarbose in *Actinoplanes* sp. SE 50/110: A Progress Report. Biocatalysis and Biotransformation 21 (4-5), 279–284.

Wehmeier, U.F., 2004. Acarbose, ein therapeutisch eingesetzter Wirkstoff: Biosynthese und Funktion. BIOspektrum 10 (1/04), 34–36.

Wehmeier, U.F., Piepersberg, W., 2004. Biotechnology and molecular biology of the alpha-glucosidase inhibitor acarbose. Applied microbiology and biotechnology 63 (6), 613–625.

Wehmeier, U.F., Piepersberg, W., 2009. Chapter 19 Enzymology of Aminoglycoside Biosynthesis—Deduction from Gene Clusters, in: , Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates, vol. 459. Methods in Enzymology. Elsevier, pp. 459–491.

Weickert, M.J., Adhya, S., 1992. A family of bacterial regulators homologous to Gal and Lac repressors. The Journal of biological chemistry 267 (22), 15869–15874.

Wendler, S., Hürtgen, D., Kalinowski, J., Klein, A., Niehaus, K., Schulte, F., Schwientek, P., Wehlmann, H., Wehmeier, U.F., Pühler, A., 2013. The cytosolic and extracellular proteomes of *Actinoplanes* sp. SE50/110 led to the identification of gene products involved in acarbose metabolism. Journal of biotechnology 167 (2), 178–189.

Wendler, S., Ortseifen, V., Persicke, M., Klein, A., Neshat, A., Niehaus, K., Schneiker-Bekel, S., Walter, F., Wehmeier, U.F., Kalinowski, J., Pühler, A., 2014. Carbon source dependent biosynthesis of acarviose metabolites in *Actinoplanes* sp. SE50/110. Journal of biotechnology 191, 113–120.

Wendler, S., Otto, A., Ortseifen, V., Bonn, F., Neshat, A., Schneiker-Bekel, S., Walter, F., Wolf, T., Zemke, T., Wehmeier, U.F., Hecker, M., Kalinowski, J., Becher, D., Pühler, A., 2015. Comprehensive proteome analysis of *Actinoplanes* sp. SE50/110 highlighting the location of proteins encoded by the acarbose and the pyochelin biosynthesis gene cluster. Journal of proteomics 125, 1–16.

Willoughby, L.G., 1971. Observations on some aquatic *Actinomycetes* of streams and rivers. Freshwater Biol 1 (1), 23–27.

Winkler, W.C., Breaker, R.R., 2005. Regulation of bacterial gene expression by riboswitches. Annual review of microbiology 59, 487–517.

Wolf, T., Droste, J., Gren, T., Ortseifen, V., Schneiker-Bekel, S., Zemke, T., Pühler, A., Kalinowski, J., 2017a. The MalR type regulator AcrC is a transcriptional repressor of acarbose biosynthetic genes in *Actinoplanes* sp. SE50/110. BMC genomics 18 (1), 562.

Wolf, T., Gren, T., Thieme, E., Wibberg, D., Zemke, T., Puhler, A., Kalinowski, J., 2016. Targeted genome editing in the rare actinomycete *Actinoplanes* sp. SE50/110 by using the CRISPR/Cas9 System. Journal of biotechnology 231, 122–128.

Wolf, T., Schneiker-Bekel, S., Neshat, A., Ortseifen, V., Wibberg, D., Zemke, T., Pühler, A., Kalinowski, J., 2017b. Genome improvement of the acarbose producer *Actinoplanes* sp. SE50/110 and annotation refinement based on RNA-seq analysis. Journal of biotechnology 251, 112–123.

Wong, O.K., Guthold, M., Erie, D.A., Gelles, J., 2008. Interconvertible lac repressor-DNA loops revealed by single-molecule experiments. PLoS biology 6 (9), e232.

Zhang, C.-S., Podeschwa, M., Altenbach, H.-J., Piepersberg, W., Wehmeier, U.F., 2003. The acarbose-biosynthetic enzyme AcbO from *Actinoplanes* sp. SE 50/110 is a 2-epi-5-epi-valiolone-7-phosphate 2-epimerase. FEBS Letters 540 (1-3), 47–52.

Zhang, C.-S., Stratmann, A., Block, O., Brückner, R., Podeschwa, M., Altenbach, H.-J., Wehmeier, U.F., Piepersberg, W., 2002. Biosynthesis of the C(7)-cyclitol moiety of acarbose in *Actinoplanes* species SE50/110. 7-O-phosphorylation of the initial cyclitol precursor leads to proposal of a new biosynthetic pathway. The Journal of biological chemistry 277 (25), 22853–22862.

Zhang, D., Zhao, Q., Jiang, M., Kang, Q., Bai, L., 2020. Biosynthetic pathway of deoxyaminosugar moiety in acarbose from *Actinoplanes* sp. SE50/110. Acta Microbiologica Sinica (60), 118–134.

Zhao, Q., Luo, Y., Zhang, X., Kang, Q., Zhang, D., Zhang, L., Bai, L., Deng, Z., 2020. A severe leakage of intermediates to shunt products in acarbose biosynthesis. ncomms 11 (1), 1468.

Zhou, Y., Kolb, A., Busby, S.J.W., Wang, Y.-P., 2014. Spacing requirements for Class I transcription activation in bacteria are set by promoter elements. Nucleic Acids Res 42 (14), 9209–9216.

Appendix

Danksagung

Zunächst möchte ich mich bei Jörn Kalinowski für die großartige Zeit in seiner Arbeitsgruppe bedanken: Hier sind vor allem die hervorragende wissenschaftliche Betreuung, die Gesprächsbereitschaft, die unvergesslichen Ausflüge und die (Ski-) Urlaube zu nennen. Darüber hinaus möchte ich Alfred Pühler für die konstruktiven Gespräche, seine wissenschaftliche Unterstützung und seinen Rat danken. Die Gespräche zur Vorbereitung von Publikationen waren stets sehr angenehm und konstruktiv.

Außerdem bedanke ich mich bei allen „Coryne and Friends“ für die tolle Gruppenatmosphäre, die konstruktiven Diskussionen und Seminare und für die gemeinsame Zeit. Mein besonderer Dank gilt dem gesamten *Actinoplanes*-Team um Susanne Schneiker-Bekel, Marcus Persicke, Lena Schaffert, Eric Thieme, Lars Wollenschläger und den ehemaligen Mitgliedern Timo Wolf, Vera Ortseifen, Tetiana Gren und Sergej Wendler.

Mein Dank gilt auch der Bayer AG und dem CLIB-Graduate Cluster Industrial Biotechnology für die finanzielle Unterstützung dieser Arbeit. Außerdem danke ich allen Co-Autoren und Unterstützern, die an den Publikationen mitgewirkt haben.

Ferner möchte ich mich bei allen iGEMern bedanken, die mir als Betreuer und Teammitglieder die tolle und unvergessliche iGEM-Zeit ermöglicht haben. Hier sind vor allem Boas Pucker und Nils Lübke zu nennen, denen ich für die fachlichen und oft spaßigen Debatten danke. Darüber hinaus möchte ich auch allen nicht namentlich genannten Kollegen, Weggefährten und Studenten des CeBiTec für die gemeinsame Zeit während meiner Promotion danken.

Zum Schluss möchte ich mich ganz herzlich bei meinen Eltern, meinen Freunden und speziell bei meiner Verlobten Vera und unserem Sohn Tim bedanken, die mich immer unterstützt und mir den Rücken freigehalten haben.

Erklärung

Hiermit erkläre ich, dass ich die hier vorliegende Dissertation eigenständig und ohne unerlaubte Hilfe angefertigt habe. Ich versichere, dass ich keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie alle aus der Literatur entnommenen Zitate als solche kenntlich gemacht habe. Die Dissertation wurde weder in der vorgelegten noch in ähnlicher Form bei einer anderen Institution eingereicht. Ich bewerbe mich hiermit erstmalig um den Doktorgrad der Naturwissenschaften.

Bielefeld, 19.10.2020

_____

Unterschrift: Julian Droste

Publications

Droste, J., Ortseifen, V., Schaffert, L., Persicke, M., Schneiker-Bekel, S., Pühler, A., Kalinowski, J., 2020. The expression dynamics of the *acb* gene cluster in *Actinoplanes* sp. SE50/110 is dependent on the growth phase. BMC genomics. DOI: 10.21203/rs.3.rs-41287/v1 [status: accepted for publication]

Authors contributions:

Julian Droste designed, planned, and interpreted the experimental work, performed data analysis and drafted the manuscript. Julian Droste and Dr. Vera Ortseifen performed the proteome studies. Julian Droste and Dr. Marcus Persicke performed the fermenter cultivation experiments. Dr. Vera Ortseifen, Dr. Lena Schaffert, Dr. Susanne Schneiker-Bekel, Prof. Dr. Alfred Pühler and Prof. Dr. Jörn Kalinowski assisted in interpreting the data and revised the manuscript. Prof. Dr. Jörn Kalinowski and Prof. Dr. Alfred Pühler coordinated the study. All authors read and approved the final manuscript.

The expression of the acarbose biosynthesis gene cluster in *Actinoplanes* sp. SE50/110 is dependent on the growth phase

Julian Droste[1], Vera Ortseifen[2], Lena Schaffert[1], Marcus Persicke[1], Susanne Schneiker-Bekel[2], Alfred Pühler[2], and Jörn Kalinowski[1§]

Affiliations

[1] Microbial Genomics and Biotechnology,

[2] Senior Research Group in Genome Research of Industrial Microorganisms,

Center for Biotechnology, Bielefeld University, Universitätsstraße 27, 33615 Bielefeld, Germany

§ Corresponding author:

Prof. Dr. Jörn Kalinowski

joern@CeBiTec.Uni-Bielefeld.DE

Abstract

Background: *Actinoplanes* sp. SE50/110 is the natural producer of the diabetes mellitus drug acarbose, which is highly produced during the growth phase and ceases during the stationary phase. In previous works, the growth-dependency of acarbose formation was assumed to be caused by a decreasing transcription of the acarbose biosynthesis genes during transition and stationary growth phase.

Results: In this study, transcriptomic data using RNA-seq and state-of-the-art proteomic data from seven time points of controlled bioreactor cultivations were used to analyze expression dynamics during growth of *Actinoplanes* sp. SE50/110. A hierarchical cluster analysis revealed co-regulated genes, which display similar transcription dynamics over the cultivation time. Aside from an expected metabolic switch from primary to secondary metabolism during transition phase, we observed a continuously decreasing transcript abundance of all acarbose biosynthetic genes from the early growth phase until stationary phase, with the strongest decrease for the monocistronically transcribed genes *acbA, acbB, acbD* and *acbE*. Our data confirm a similar trend for *acb* gene transcription and acarbose formation rate.

Surprisingly, the proteome dynamics does not follow the respective transcription for all *acb* genes. This suggests different protein stabilities or post-transcriptional regulation of the Acb proteins, which in turn could indicate bottlenecks in the acarbose biosynthesis. Furthermore, several genes are co-expressed with the *acb* gene cluster over the course of the cultivation, including eleven transcriptional regulators (e.g. ACSP50_0424), two sigma factors (ACSP50_0644, ACSP50_6006) and further genes, which have not previously been in focus of acarbose research in *Actinoplanes* sp. SE50/110.

Conclusion: In conclusion, we have demonstrated, that a genome wide transcriptome and proteome analysis in a high temporal resolution is well suited to study the acarbose biosynthesis and the transcriptional and post-transcriptional regulation thereof.

Keywords: *Actinoplanes*, Acarbose, Transcriptomic, Proteomic, Expression dynamics, co-regulation

1. Background

Actinoplanes sp. SE50/110 is a Gram-positive, aerobic bacterium belonging to the genus of *Actinoplanes*, within the family *Micromonosporaceae* [1, 2]. Members of the genus *Actinoplanes* can form sporangia, that contain motile spores, and typically grow in branched hyphae [1, 3]. *Actinoplanes* spp. are characterized by genomes with high G+C contents of 69-73 % [1, 3]. Several species are known for their potential to produce a variety of secondary metabolites, like antibiotics [4, 5]. Among them are more than 120 antibiotics, like actaplanin [6], teicoplanin [7], friulimicins [8] and ramoplanin [9]. *Actinoplanes* sp. SE50 strains are of special interest because of their ability to produce the pseudotetrasaccharide acarbose, which has an inhibitory effect on alpha-glucosidases and is therefore of special interest for pharmaceutical applications [10]. Due to its inhibitory effect, acarbose is used for the treatment of diabetes mellitus. The inhibition of the intestinal alpha-glucosidases decelerates the degradation of long-chain carbons and thus leads to a retarded resorption of monosaccharides into the blood system [11–13]. By this, the postprandial blood and serum sugar glucose is reduced, which is a risk factor for developing secondary complications, like cardiovascular diseases, diabetical retinopathies and diabetic food syndrome [14].

The strain *Actinoplanes* sp. SE50/110 is the best-studied acarbose producer and a high quality genome sequence is known [15]. Several biochemical studies of the enzymes of the acarbose biosynthesis (acb) gene cluster and genomic as well as proteomic studies were carried out to propose pathways for the biosynthesis of acarbose [13, 16–19]. Recently, tools for genome editing based on CRISPR/Cas9 [20], an overexpression system using different promoter strengths [21] and a protocol for conjugational plasmid transfer [22] were developed for *Actinoplanes* sp. SE50/110, which will further promote the acarbose research in this strain.

The transcriptional organization of the acb gene cluster, including transcription start sites, promoter elements and operon organization was recently elucidated [15]. The acb gene cluster in *Actinoplanes* sp. SE50/110 lacks genes coding for transcription factors. This is in contrast to the acarbose biosynthetic gene clusters of Streptomyces spp. [23, 24]. Only one study concerning a transcription factor involved in acarbose biosynthesis and its binding sites is known [25]. Since it is known that the formation of acarbose correlates with the course of cell growth and no acarbose is produced in the

stationary phase [26, 27], the expression dynamics of genes involved in the acarbose biosynthesis of *Actinoplanes* sp. SE50/110 was examined in this study. Bioreactor cultivations were conducted to maintain controlled cultivation conditions and both transcriptomic and proteomic analyses were used in a high temporal resolution to study whole genome expression dynamics during the course of cell growth. A hierarchical cluster analysis of the data was performed to elucidate co-expressed genes. Finally, acarbose biosynthesis (*acb*) genes were analyzed in detail regarding their respective transcript and protein dynamics. Furthermore, genes co-expressed to the acb gene cluster were elucidated.

2. Results and discussion

2.1 Acarbose production of *Actinoplanes* sp. SE50/110 steadily decreases during the growth phase and almost ceases in stationary phase

In this study, the changes of acarbose production during the growth of *Actinoplanes* sp. SE50/110 were analyzed. Therefore, bioreactor cultivations were used to achieve controlled cultivation parameters. *Actinoplanes* sp. SE50/110 was cultivated in maltose minimal medium in three biological replicates. Spores were generated for inoculation by first growing *Actinoplanes* sp. SE50/110 in NBS complex medium and afterward plating the cells on SFM agar plates to generate spores, which in turn served as inoculum. Samples were taken at regular intervals to monitor the course of growth and acarbose formation.

Within the controlled conditions of reactor cultivations, a correlation between acarbose formation and the course of biomass production over time was shown (Figure 1A) as it was observed in previous studies [26]. Acarbose was produced, starting in the lag phase (24.0 h) and continuing during growth (47.8, 72.3, 96.5 h), until the cultivations reached the transition phase (120.0 to 144.3 h). The acarbose concentration in the supernatant remains almost constant during the stationary phase (144.3 and 168.0 h). The specific product formation rate, defined as produced acarbose normalized to the mean cell dry weight and to cultivation time, increased during the first 48 h and then decreased steadily (Figure 1B). The specific product formation is a direct indicator for acarbose production of the mycelial growing strain during a defined period and is not biased by hitherto formed acarbose. The findings of an acarbose production by *Actinoplanes* sp. SE50/110 in a growth-dependent manner is in good accordance to

shake flask cultivations reported in the literature [25, 26]. For further analyses of the growth dependency of acarbose formation transcriptome and proteome dynamics were examined over the whole fermentation process.

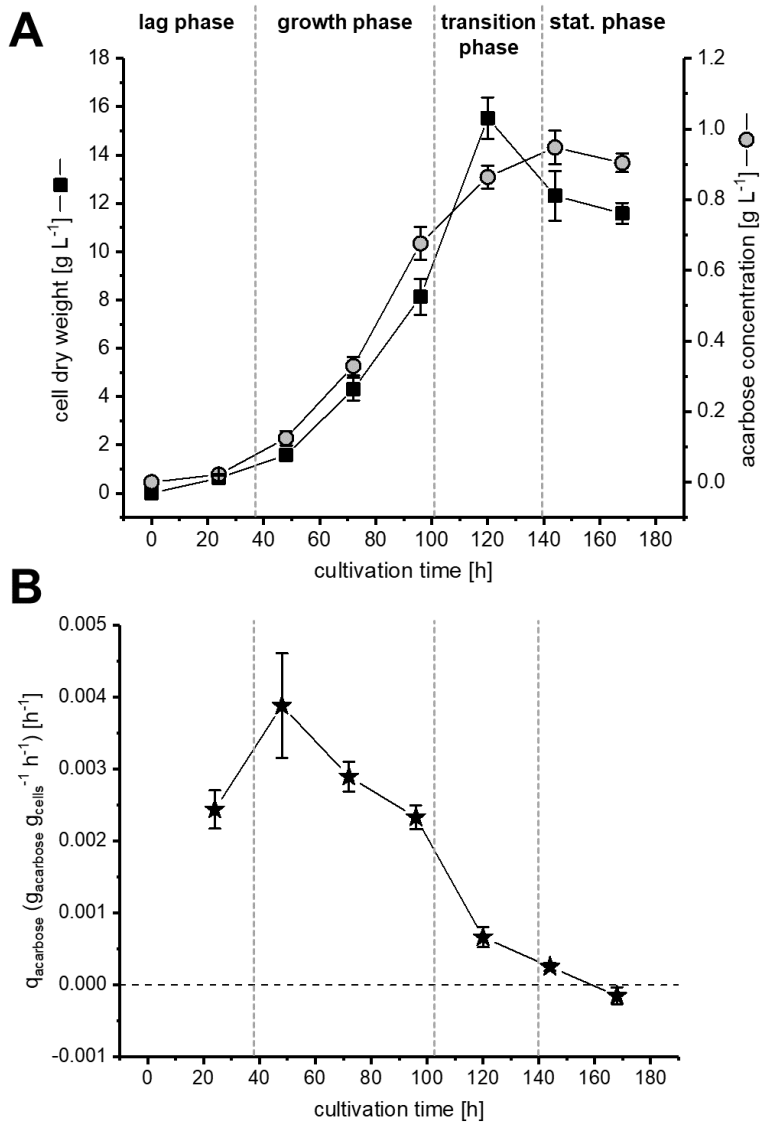

Figure 1: Characterization of growth and acarbose production of *Actinoplanes* sp. SE50/110 in controlled fermenter conditions. Different growth phases (lag, growth, transition and stationary phase) were indicated by vertical dashed lines. (A) Cell dry weight (black boxes) and acarbose concentration (grey circles) over the cultivation course. Plotted are the means and standard deviations of three biological replicates, each of which were measured in three technical replicates. (B) Specific product formation rates ($q_{Acarbose}$) defined as produced acarbose normalized on the mean cell dry weight and cultivation time difference.

2.2 Analysis of whole transcriptome data of *Actinoplanes* sp. SE50/110

2.2.1 Processing and filtering of transcriptomic data

Whole transcriptome analysis using RNA-seq was subsequently carried out, in which seven time points in three biological replicates were compared to RNA pooled from all analyzed time points for each replicate. Thereby, a normalized analysis over the entire course of cultivation is possible, minimizing technical and biological variances. Consequently, a relative transcript abundance of one and a $\log_2$(fold change) of zero correspond to the average amount of transcript over all time points. For 8,364 of all 8,402 annotated features (99.5%), reads could be found for all analyzed time points. A principal component analysis (PCA) was performed to determine the differences of each time point to the pooled sample (Supplementary Figure 1). For cluster analyses genes were ruled out, if the transcription shows no significant difference (p_{adj}-value > 0.05) at all time points compared to the mean value of the respective transcript. This filtering results in 6,770 genes with a significant different transcription for at least one time point. A schematic overview of processing and filtering steps can be found in Supplementary Figure 2.

2.2.2 Overview of temporal transcriptome dynamics

To gain a first overview, the number of genes was evaluated for which significantly increased or decreased transcript amounts were measured (p_{adj}-value <0.05) (Figure. 2). The highest number of genes with a significant difference in transcript amount compared to the respective average amount over the whole cultivation time was observed during the lag phase (24 h) and the late stationary phase (168 h). The transcript amount was significantly increased for 1,421 (17.0 %) genes and decreased for 1,246 (14.9 %) genes in the lag phase (24 h). In the late stationary phase 2,491 (29.8 %) of all genes show an increased and 2,531 (30.3 %) a decreased transcript amount. The minimal differences regarding transcription was observed in the mid growth phase (96.5 h).

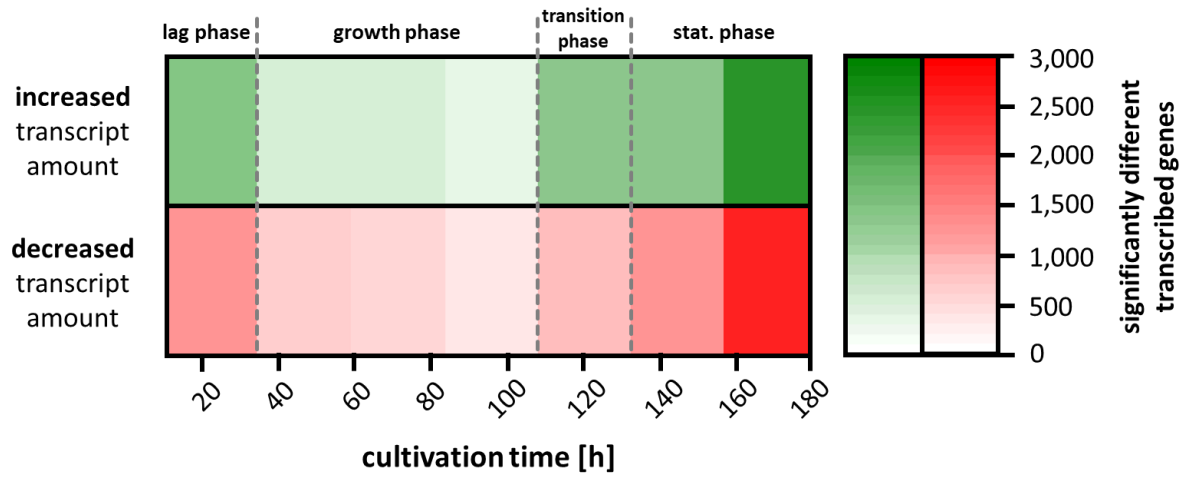

Figure 2: Overview about transcriptome dynamics in *Actinoplanes* sp. SE50/110. Number of genes with significantly (padj value < 0.05) increased (green) and decreased (red) transcript abundances during cultivation at the given time points. Growth phases are indicated with dashed lines.

Most genes are transcribed during filamentous growth and show their mean transcription level in the mid growth phase. The observed trend of differences in transcription is in good accordance to the expectation as a minimal number of differentially transcribed genes is expected during filamentous growth [28]. In contrast to that, the greatest difference regarding transcription was observed for the late stationary phase (168 h). This could be a hint for a typical switch from primary to secondary metabolism [28, 29].

However, it should be noted that this first global analysis highlights only genes with significantly differential transcript amounts at single time points and ignores trends in temporal transcriptome dynamics of single genes. To analyze these trends and identify co-regulated genes on a transcriptional level, a hierarchical cluster analysis was implemented, as it is described below.

2.2.3 Identification of operon structures by combining whole transcriptome data sets of different time points

Two or more genes, that are transcribed from a single promoter, form an operon. The analysis of the operon structure of *Actinoplanes* sp. SE50/110 is an important step to investigate the co-regulation of single genes and large operons. The operon detection

was performed using the software ReadXplorer [30]. The data of all 21 RNA-seq experiments were combined to increase the number of reads in regions with low coverage. The identified primary operons were checked for experimental validation using the TSS determined from the data of sequenced 5´-end enriched libraries [15]. If an operon has an assigned TSS, it is experimentally validated. If not, it was specified as predicted operon. The class of sub-operons consists of operons which show a TSS for a posterior gene in a primary operon. All other genes, which could not be connected to an operon, were assigned to be monocistronically transcribed.

Table 2: Number of monocistronic genes, primary operons and sub-operons of *Actinoplanes* sp. SE50/110 obtained from operon analysis using ReadXplorer [30] obtained from RNA-seq data of this study.

Genes per transcript	Primary operons	Sub-operons	Monocistronic genes
1	-	604	4,757
2	689	63	-
3	181	32	-
4	95	17	-
≥ 5	64	19	-
Total	**1,029**	**735**	**4,757**

Under the studied conditions 1,029 primary operons containing 2,751 genes could be detected by combining the whole transcriptome data sets of all analyzed time points (Table 1, Supplementary Figure 2).

408 (39.7 %) of all primary operons could be experimentally validated, as a TSS could be assigned to the first gene of the corresponding operon. By analyzing the internal TSS, 735 sub-operons could be determined inside the 1,029 primary operons. The majority (604) of the sub-operons consists of a single gene (Table 1, Supplementary Figure 3).

The largest primary operon contains 16 genes, which encodes mainly genes with no annotated function (“hypothetical protein”) [15] (Supplementary Table 1).

The number of monocistronically transcribed genes was determined to be 4,757 (56.6 % of all CDS), of which 1,789 genes (37.6 %) were associated with a TSS (Table 1).

2.2.4 Global identification of transcription start sites (TSS), 5´-UTR lengths and promoter consensus motifs in the *Actinoplanes* sp. SE50/110 genome sequence

For the analysis of growth-dependent transcription, a fundamental knowledge about the transcriptional landscape of the *Actinoplanes* sp. SE50/110 genome is required. The identification of transcription start sites (TSS) and corresponding promoters, which are only active in specific growth phases is useful for understanding regulatory processes and networks in *Actinoplanes* sp. SE50/110.

Based on the 5´ enriched library data from [15] and the whole transcriptome profile from this study, the positions of TSS was determined using the software ReadXplorer [30]. The automated prediction revealed 7,937 TSS. Filtering and manual curation resulted in 4,228 primary TSS, which could be assigned to 2,787 CDS (33.2 % of all annotated features) (Supplementary Table 2). This is a 3-fold increase to previous studies (1,427 TSS assigned to 799 CDS) by [31].

The 5´-untranslated region (5´-UTR) was determined as distance of TSS to the corresponding translation start site (TLS). Transcripts with a distance ≤3 nt were classified as leaderless transcripts. This results in 1,179 TSS (14.03% of all CDS), which belong to leaderless transcripts, whereas 3,049 TSS were assigned to transcripts carrying a 5´-UTR. The 5´-UTR length ranges between 4 and 494 nt, but 90 % of all 5´-UTRs are less than 200 nt in length (Supplementary Figure 4).

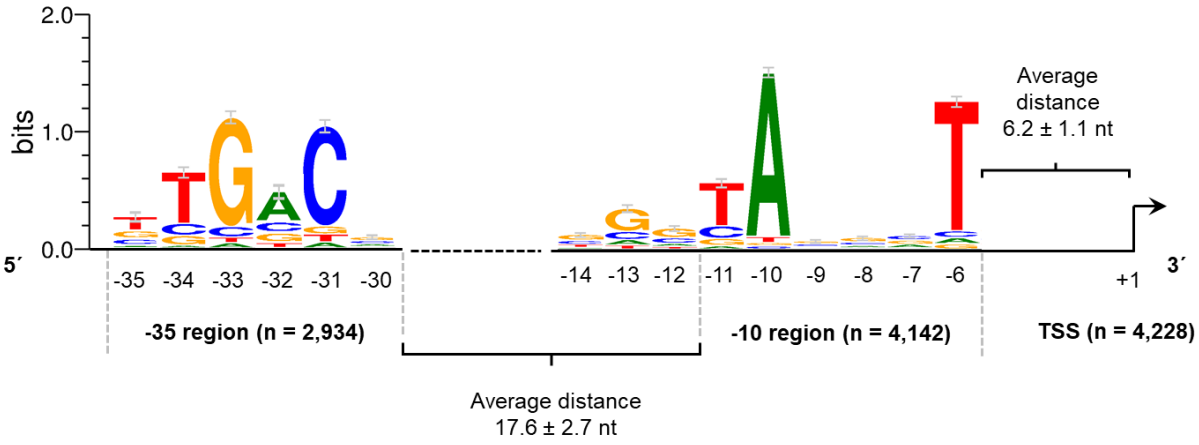

Figure 3: Conserved -10 and -35 regions identified in the promoter regions of the *Actinoplanes* sp. SE50/110 genome. The motifs were identified using Improbizer [32] searching upstream of 4,228 TSS, resulting in 4,142 putative -10 regions and 2,934

putative -35 regions. The sequence logos were created using the software WebLogo [35].

Upstream of the identified TSS, promoter motifs could be found, such as the -10 region (Pribnow box) and the -35 region. Therefore, 50 bp upstream of each identified primary TSS were searched with the tool Improbizer [32]. For the -10 region a conserved hexamer motif represented by TAnnnT was found in 4,143 (98 %) of all sequences examined (Supplementary Table 1). This result is in line with the findings from [31], analyzing the upstream sequences of 318 TSS in *Actinoplanes* sp. SE50/110.

In this study, the T on the first position of the identified hexamer was found in 63.6 % of the analyzed sequences. For the A on second position within the -10 motif, a frequency of 90.8 % was determined. In the last position of the -10 hexamer a T is present in 85.7 % in the considered sequences in *Actinoplanes* sp. SE50/110. Therefore, the identified -10 region perfectly matches the most highly conserved bases of the -10 motif in the model organisms *Escherichia coli* [33] and *Streptomyces coelicolor* A3(2) [28]. The slightly overrepresented G at position -13 indicates that some promoters feature an extended -10 region [34].

The average distance of the -10 hexamer to the corresponding TSS was found to be 6.2 ± 1.1 nt , whereas 82% of all spacer lengths range between 5 and 7 nt (Figure 3). The TSS itself is a purine in 75.4 % of the cases (24.2 % A and 51.2 % G).

In the -35 region the consensus hexamer nTGACn was determined in 2,934 of all 4,228 TSS upstream sequences (69.4 %) using the software Improbizer [32], whereas the highest frequencies for G at position three (83.7 %) and C at position five (82.0 %) were found. However, the T at position two (67.9 %) and the A at position four (53.0 %) are less conserved in the *Actinoplanes* sp. SE50/110 -35 promoter region. This motif resembles the -35 consensus motif of *E. coli*, which is TTGACA [33]. The average distance between the -10 and the -35 region was found to be 17.6 ± 2.5 nt for 2,906 promoters, which contain both a -10 and -35 region. This spacer length is in common with the average distance of 17 nt described as optimal in *E. coli* consensus promoters [33].

In general, the promoter analysis is in accordance to the results described in the literature [15, 31]. However, in this study a much higher amount of data was used to determine the consensus motifs. Especially the consensus sequence regarding the -35 region could be improved, as it is more related to the motifs described in the

literature, e.g. for *E. coli* [33] or *S. coelicolor* A3(2) [28]. In addition to promoter analysis, for 93.2% of all analyzed leadered transcripts with a 5´-UTR-length > 10 nt a conserved ribosome binding site (RBS, Shine-Dalgarno sequence) could be found. The detected consensus motif is nGGAGn (Supplementary Figure 5).

2.2.5 Identification of co-transcribed genes by hierarchical cluster analysis of transcription dynamics data

In order to identify co-regulated genes, a hierarchical cluster analysis was performed using transcriptome data determined for each time point (Supplementary Table 3). The hierarchical cluster analysis was performed within the software Omics Fusion [36] and resulted in an optimal cluster amount of 36 (Figure 4; Supplementary Figures 6 and 7). The clusters contain 45 to 645 genes.

An overview about predominant functions and pathways in a group of genes can be achieved by classification according to COG (clusters of orthologous groups) database [37] and KEGG (Kyoto Encyclopedia of Genes and Genomes) database [38]. Therefore, the cluster obtained from hierarchical cluster analysis can be examined for enriched gene functions and overrepresented pathways (Supplementary Table 3). Furthermore, strong changes in the course of transcription could be a hint for metabolic switches or different stages of the *Actinoplanes* sp. SE50/110 life cycle [39].

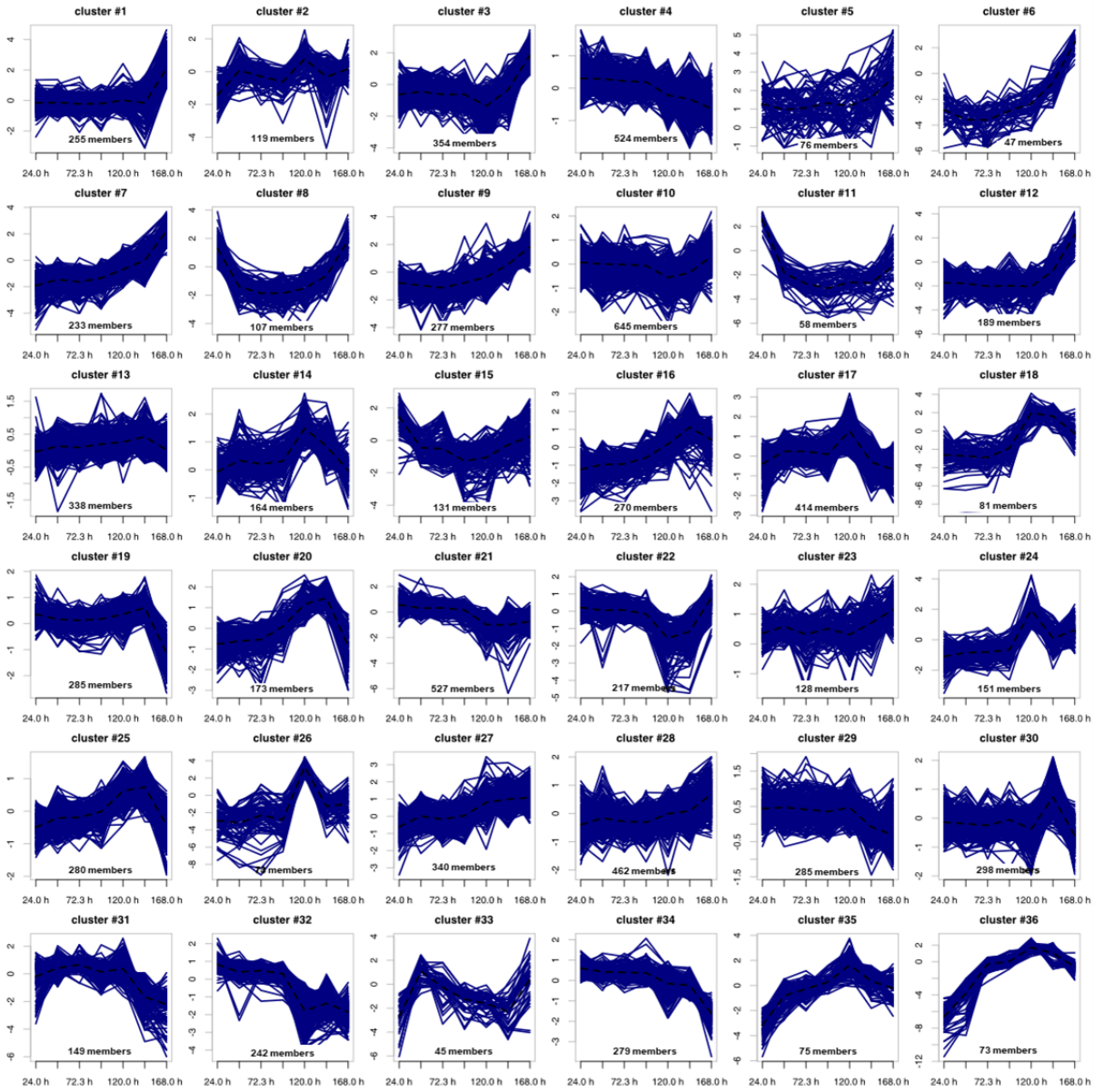

Figure 4: Hierarchical cluster analysis cluster analysis of transcription dynamics with calculation of optimal cluster size and cluster grouping using the tool Omics Fusion [36]. The log2(fold-changes) values for each time point and gene were used as input and are shown in the cluster graphs. The respective mean value of each cluster is visualized by a dashed line. For better visualization the trend of each cluster is scaled differently.

The earliest transcriptional change can be observed for genes, which are grouped in clusters 8, 11 and 15. These genes are highly transcribed in the lag phase, but almost no longer present during filamentous growth. Additionally, an increasing transcription in the late stationary phase could be observed for several genes of these clusters.

Functional analysis of these three clusters revealed an enrichment of genes encoding chemotaxis, motility and flagellum associated proteins (COG class N; cell motility). More than 65 % of the genes encoding proteins of these COG class in the *Actinoplanes* sp. SE50/110 genome are grouped in these three clusters. Since the cultivation was inoculated with spores generated on SFM agar plates, cell motility proteins are necessary and therefore highly transcribed in this stage of life cycle [39, 40]. Genes involved in flagellar development have been identified and described to be highly expressed in spores in *Actinoplanes* spp. [41, 42]. Due to the fact, that these genes are only expressed for a short time after inoculation, spore formation and cell motility seem to play a minor role in the further course of cultivation under the tested conditions. Interestingly, transcription of many of these genes increases in the late stationary phase. Therefore, it can be assumed that spores are formed at the end of the cultivation. So far, sporulation in liquid media could not be shown for *Actinoplanes* sp. SE50/110, yet. However, it was described for *Bacillus subtilis* [43] and *Streptomyces* spp. [44].

The clusters 4, 21, 32 and 34 with continuously decreasing transcript abundance are dominated by genes encoding ribosomal proteins and other proteins involved in protein biosynthesis (COG class J; translation, ribosomal structure and biogenesis). More than 60% of these features are located in these clusters. Considering only the 30S and 50S ribosomal proteins, these are almost only distributed among clusters 32 and 34. The corresponding profiles closely match the pattern of the stringent response in other close related actinobacteria, like *S. coelicolor* A3(2) [45] or *Corynebacterium glutamicum* [46, 47]. A continuously decreasing transcript level of ribosomal proteins and other proteins with functions related to the protein biosynthesis fits well to the expectations of an enhanced translation machinery for boosting cell growth at the beginning of the growth phase. The transcription of these genes decreases during cultivation reaching the lowest level in the stationary phase at which growth stops. This effect was previously described in *S. coelicolor* A3(2) [29].

Typical for actinomycetes is an increased production of secondary metabolites in the stationary phase, which is reflected by the transcription of the respective genes [28, 29]. The genome of *Actinoplanes* sp. SE50/110 harbors 20 predicted biosynthetic gene clusters for secondary metabolites, including the acarbose biosynthetic gene cluster [15]. Genes which are associated with one of these gene clusters were

identified with antiSMASH 5.0 [48]. Most of these predicted secondary metabolite gene clusters are highly transcribed in the transition phase (clusters 17, 20 and 30; Figure 4) and in the stationary phase (cluster 1, 3, 7, 10, 12 and 16; Figure 4). Therefore, for 14 of all 20 predicted secondary metabolite gene clusters, an increased transcription during the late growth phase and stationary phase could be shown. They encode for terpene (carotenoid), NRPS, PKS, lassopeptide, lantipeptide, bacteriocin, or melanin biosynthesis. These findings indicate a typical switch from primary to secondary metabolism described for most organisms [49, 50]. In *S. coelicolor* A3(2) similar effects were observed by analyzing the transcription of secondary metabolite gene clusters in a growth-dependent manner [29].

The six remaining secondary metabolite gene cluster encode two siderophore, a terpene, a pyochelin and the acarbose biosynthesis and display different transcription dynamics. The two siderophore biosynthesis gene cluster differ from each other regarding their transcriptional course: For the first one, an increased transcription both during lag and stationary phase (cluster 15; Figure 4) was observed, whereas the second one shows a slight increase of transcription during the stationary phase (cluster 3; Figure 4). This could indicate different needs for iron in the growth versus stationary phase. In contrast to the beforementioned carotenoid biosynthesis gene cluster, the second terpene cluster was found to be highly transcribed during lag and stationary growth phase showing a similar transcription course as several cell motility and spore formation genes (cluster 8; Figure 4). However, further investigation has to be made regarding the metabolic product of this gene cluster to determine a potential connection to e.g. sporulation.

The pyochelin biosynthesis gene cluster was found to be transcribed similar to the growth curve (cluster 36; Figure 4). An increasing transcription was observed during growth, but no further increase could be found during transition and stationary phase. The gene products of pyochelin biosynthesis were analyzed in previous studies regarding their sub-cellular localization. The results revealed a close connection to the bacterial cell membrane [51]. The same localization was identified for the products of the *acb* gene cluster. However, transcription of the *acb* genes was found to be increased in the early growth phase and decreases until cells reaching the stationary phase (clusters 31 and 32; Figure 4).

Due to the fact, that the transcription is increased in the early growth phase, the *acb* gene products seem to be more important for cell metabolism and therefore do not qualify as genes involved in secondary metabolite biosynthesis. The expression dynamics of the *acb* genes over the course of the cultivation will be discussed in detail in chapter 4.1.3.

2.3 Analysis of proteome data during the whole cultivation process

2.3.1 Processing and filtering of proteome data

To investigate the expression dynamics of *Actinoplanes* sp. SE50/110 proteome analysis was performed at all seven time points. Proteins were isolated from *Actinoplanes* cells (cytosolic fraction) and from the supernatant (extracellular fraction) and proteins were measured using state-of-the-art mass spectrometry (QExactive mass spectrometer). This resulted in a total number of 2,675 proteins (32.3 % of all annotated CDS), whereas 2,496 were identified in the cellular fraction and 878 were found in the extracellular fraction. Principal component analyses (PCA) were performed to check the quality of the proteome data (Supplementary Figure 8 and 9).

Out of 878 proteins identified in the extracellular fraction 699 (79.6 %) could also be detected in the cellular fraction. According to previous protein localization predictions [51] of these 699 proteins identified in both fractions 534 could be assigned as cytosolic, 103 are membrane associated or located at the inner membrane and 53 are previously predicted as extracellular proteins since a signal peptide could be identified in the amino acid sequence of the respective protein. Furthermore, 36 of the 179 proteins exclusively identified in the extracellular fraction were predicted as cytosolic, 64 were predicted as membrane associated or inner membrane proteins and 71 proteins were predicted as extracellular proteins [51]. The identification of proteins predicted as cytosolic in the extracellular fraction was reported previously for several bacteria [52, 53]. Additionally, it could be shown for *Bacillus subtilis* that only 21 % of over 900 extracellular identified proteins show a signal peptide [53]. In this study, for 124 (14.1 %) of the 878 proteins identified in the extracellular fraction, a signal peptide was predicted, which is in good accordance to the literature. Proteins which were identified in both extracellular and cytosolic fraction as well as predicted as extracellular proteins [51], were excluded from the cellular fraction data set and kept in

the extracellular data. These proteins were assumed to be genuine extracellular proteins due to their predicted signal peptide.

The different filter steps resulted in 2,234 proteins, of which 1,654 were identified in the cellular, 183 in the extracellular and 397 proteins were found in both fractions. In this way, for the cellular fraction for 1,468 proteins (71.6 %) data could be obtained for at least 6 of 7 time points. In the extracellular fraction for 240 proteins (41.4 %) data are available for at least 6 time points (Supplementary Table 4). Differential expression analysis was performed according to processing of transcriptomic data. A schematic overview on processing and filtering steps can be found in Supplementary Figure 10.

2.3.2 Overview of proteome dynamics in *Actinoplanes* sp. SE50/110

The whole proteome analysis revealed 2,234 different proteins, which could be detected in the cellular and extracellular fraction of *Actinoplanes* sp. SE50/110. Applying differential expression analysis 1,441 proteins (1,374 cytosolic and 67 extracellular) could be identified, which show a significant difference (p_{adj} <0.05) for at least one time point. The number of significant different protein amounts compared to the respective mean value (protein fold-change) changes during the cultivation process. Figure 5 shows the total amount of genes with a significant difference regarding the respective protein amount.

Strikingly, the extracellular proteome fraction shows only a small number of significant different protein levels. This shows higher stability of extracellular proteins compared to cytosolic proteins due to absence of proteases or other influences [54].

For the cellular fraction clearly changes of the proteome repertoire during growth could be observed. The highest number of significant different protein amounts was observed in the late stationary phase after 168 h with 428 increased (29.7 %) and 278 decreased (19.3 %) proteins. Furthermore, a high number of significantly different protein abundancies was detected at the beginning of the cultivation in the lag phase (190 increased and 89 decreased proteins) and in the early growth phase (292 increased and 150 decreased protein amounts). These findings are in common with the overview on transcript level (Figure 2), in which highest numbers were also observed in the lag and stationary phase.

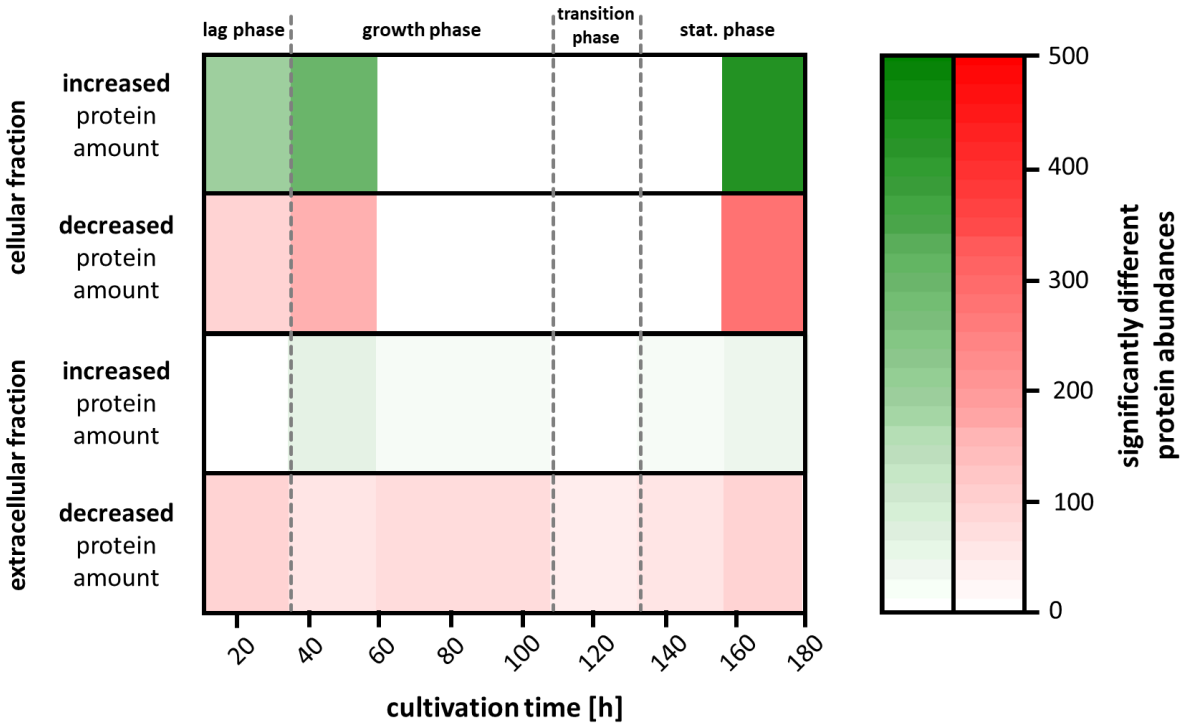

Figure 5: Overview about proteome dynamics in *Actinoplanes* sp. SE50/110. Number of proteins with significantly (padj-value < 0.05) increased (green) and decreased (red) protein abundances during cultivation at the given time points in the cellular and extracellular fraction. Growth phases are indicated with dashed lines.

Interestingly, nearly no significant differences on proteome level were observed during filamentous growth and transition phase. This shows a stable protein repertoire of the cells during filamentous growth, which is in common with the transcription profile (Figure 2). Both on transcriptional and proteome level strongest changes could be observed in the lag phase and in the stationary phase, where secondary metabolism could be observed. A slight offset between transcriptional and proteomic changes can be explained by the time of translation, since changed transcript level is necessary before changes of protein level can occur [55]. Minimal number of significantly different protein abundancies was observed between 72.3 h and 144 h. In accordance to that, the minimal number of differentially transcribed genes was found in the time between 47.3 h and 96 h, which reflects the offset of one time point (24 h) between transcription and translation.

2.3.3 Identification of different post-translational modifications by comprehensive proteome analysis

Post-translational modification is an important mechanism for regulation of protein activity, localization or stability. To get an overview on different protein modifications in *Actinoplanes* sp. SE50/110, peptide sequences were analyzed for all known modification types using MaxQuant software [56]. This resulted in the following modification types: Oxidation, acetylation, phosphorylation and glutamine (Gln) converted into pyroglutamic acid (pyro-Glu). All other modifications were low abundant or not significant. It has to be noted, that phosphorylation sites could not be determined in detail since phospho-proteome analysis requires specific sample preparation. However, 55 proteins were found to be phosphorylated at serine, threonine or tyrosine residues.

The whole proteome analysis of *Actinoplanes* sp. SE50/110 revealed a number of 821 (30.7 % of all detected proteins) proteins which are post-translationally modified at different positions. This finding matches the results for other bacteria obtained from the literature, e.g. *Leptospirillum* spp. [57]. However, under stress or nutrient-limiting conditions, bacterial proteomes were found to be more often modified on a post-translational level [58, 59].

In total, 176 proteins were found to be acetylated at their respective N-terminus. 380 proteins containing oxidations at methionine residues. This modification was identified up to three times per peptide. However, most of the modifications occur only once per protein. Finally, conversion of glutamine to pyroglutamic acid was observed in 415 proteins. All post-translational modifications and their respective positions can be found in Supplementary Table 4.

Interestingly, several proteins encoded by the *acb* gene cluster were shown to be modified. An overview on Acb protein modifications is given in Table 2.

Table 2: Overview on post-translational modifications (without oxidations) of Acb proteins of *Actinoplanes* sp. SE50/110 during growth.

Gene	Annotated function	Modification(s)	Time Point(s)
AcbZ	acarbose-resistant alpha-amylase	Gln → pyro-Glu	T1; T3; T4; T6
AcbW	ABC-type transporter; ATPase	Gln → pyro-Glu	T4; T6
AcbU	1-epi-valienol-7-phosphate 1-kinase	Gln → pyro-Glu	T3

AcbR	1-epi-valienol-1,7-bisphosphate-1-adenylyltransferase	Gln → pyro-Glu	T4
AcbQ	acarbose 4-alpha-glucanotransferase; amlyomaltase	Acetylation Gln → pyro-Glu	T3 T5
AcbO	2-epi-5-epi-valiolone-7-phosphate 2-epimerase	Gln → pyro-Glu	T3
AcbE	acarbose-resistant alpha-amylase	Gln → pyro-Glu	T3; T4
AcbD	acarviose transferase	Gln → pyro-Glu	T2; T5; T7

Interestingly, the only N-acetylation among Acb proteins was identified in the putative acarbose 4-alpha-glucanotransferase AcbQ during the middle of the growth phase (T3; 72.3 h). Since N-terminal acetylation can affect the protein stability in both directions [65-69], it can be assumed that stability of AcbQ is post-translational influenced. Interestingly, AcbQ shows one of the most stable protein abundances among Acb proteins over the cultivation process (Figure 8). This could indicate, that AcbQ possibly plays an important role in the physiology of *Actinoplanes* sp. SE50/110, e.g. within the acarbose metabolism. However, the specific function of AcbQ in the acarbose biosynthesis pathway has not yet been proven [13, 18]. Nevertheless, if AcbQ is an important enzyme in *Actinoplanes* sp. SE50/110 preventing its degradation is a possible action to increase production by the cell [63, 64]. It is notably, that most of the glutamine to pyroglutamic acid modifications of the Acb proteins were identified during the filamentous growth phase (72.3 h and 96.5 h). This could be a hint for altered enzymatic activity during filamentous growth caused by this modification. Nevertheless, this has to be proven by further experiments.

2.4 Combining transcriptome and proteome data to elucidate expression dynamics of *Actinoplanes* sp. SE50/110 using a combined clustering approach

The expression of genes in bacteria is regulated on transcriptional, post-transcriptional, translational and post-translational level. By combining transcriptome and proteome data of each gene, correlation of transcription and translation could be performed. However, if transcriptome and proteome data do not correlate in an expected manner

different regulatory stages could be responsible for that, such as protein degradation [55].

Pearson correlation of each available transcript/protein data pair was obtained. The overall Pearson coefficient was found to range from 0.10 to 0.63. In previous studies Pearson correlation coefficients of about 0.4 to 0.5 were reported for correlation of transcriptome and proteome data in bacteria [55, 65, 66], and between 0.66 and 0.76 for *Saccharomyces cerevisiae* [67]. Compared to this, the transcript/protein data pairs of *Actinoplanes* sp. SE50/110 display a broad range of correlation. Weak correlations can be referred to technical and methodological constrains, but also to translational and post-translational regulation processes [68, 69]. Therefore, correlation of transcription and protein abundance is often poor [55, 70].

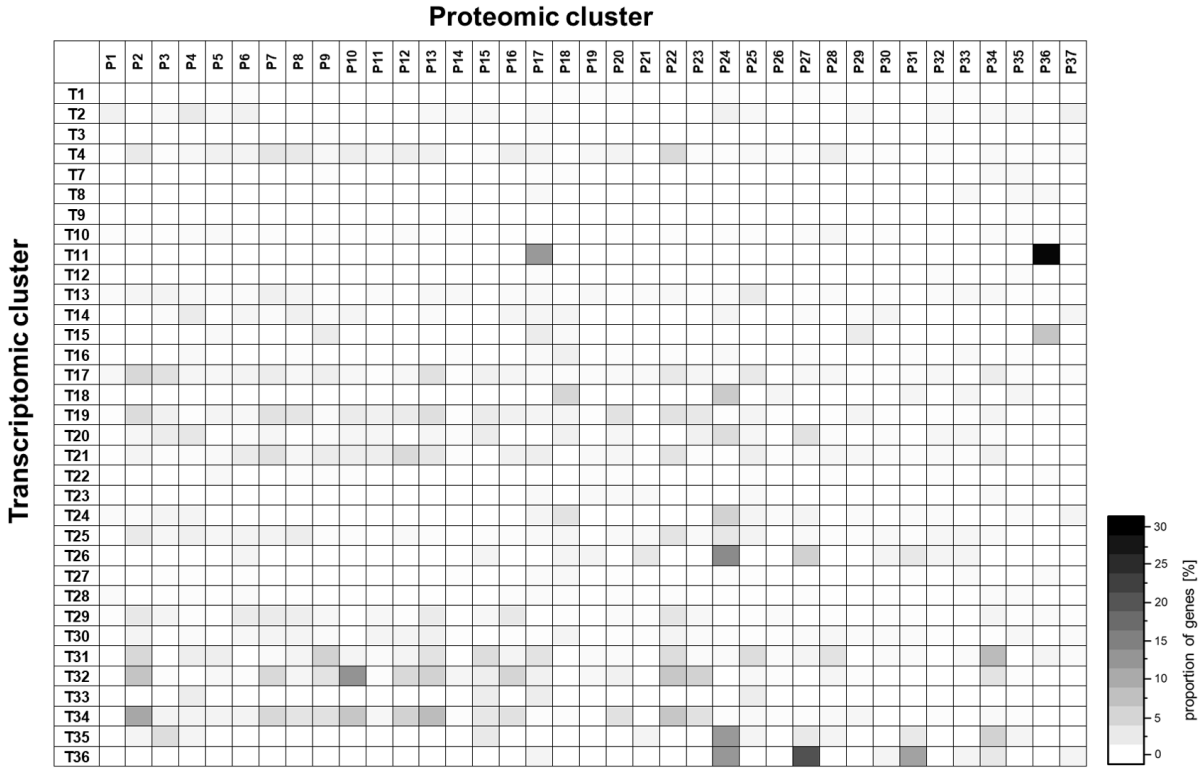

Figure 6: Connected heatmap of clustered transcriptome and proteome dynamics in *Actinoplanes* sp. SE50/110. Transcriptome clusters (T) are arranged vertically and proteome clusters (P) are arranged horizontally. Strong color indicates a high proportion of genes of the corresponding transcriptomic cluster which are present in the respective protein cluster.

For the different growth phases the differences for the respective transcriptome and proteome data were obtained. In the lag phase (24.0 h) a correlation of 0.44 was observed. Strongest correlation between transcriptome and proteome data was found in the early growth phase (47.8 h and 72.3 h) with a Pearson correlation coefficient of 0.63 and 0.48 respectively. Lowest correlation was observed when the cells entered the early stationary phase (120.0 h) with a Pearson coefficient of 0.10. Interestingly, the difference between transcriptome and proteome data is less when comparing transcriptomic data from time point X to proteome data from time point X+1 meaning to compare transcriptome data from 24.0 h with proteome data from 47.8 h, and so on. This results in Pearson correlation coefficients of 0.61, 0.61, 0.50, 0.24, 0.21 and 0.20 respectively. These findings indicate the offset between transcription and translation caused by protein folding and processing. Furthermore, proteins are more stable, and their half-life time is much longer compared to the corresponding mRNA [71, 72].

Transcriptome and proteome data were compared using a combined clustering approach via connected heatmaps. Therefore, only genes of which both transcriptome and proteome data are available were considered (2,050 genes). The cluster analysis of the proteome data resulted in an optimal cluster amount of 37 proteome clusters (Supplementary Figures 11 and 12) connected to 34 of the 36 transcriptome clusters identified previously (Figure 6). The transcriptomic cluster 5 and 6 are not included in this analysis, since no proteome data were obtained for genes inside this cluster, which contains of several hypothetical proteins, a few transcriptional regulator gene as well as tRNAs and rRNAs. Since tRNAs and rRNAs of course have not protein data and regulators are often low expressed, it is not surprising that no protein data are available for the genes in transcriptome cluster 5.

The resulting transcript and protein clusters were compared to each other regarding co-occurrences. This way, genes with both the same transcription pattern (co-transcribed) and the same protein dynamic can be identified. However, the respective transcription and protein trend can differ. The results are given in percentage of the respective transcriptome cluster size (Figure 6).

It is striking, that several transcriptomic clusters are distributed over a lot of protein clusters (e.g. clusters T4, T17, T19, T21, T32 and T34), since only a few co-occurrences could be found. A reason for that could be the large cluster size and the functional diversity of the genes inside these clusters.

The highest similarity was observed for genes located in transcription cluster T11 and proteome cluster P36. In this cluster pair predominantly genes of sporulation, chemotaxis and motility can be found. The remaining genes of cluster 11 of which a protein could be detected are grouped in protein cluster P17. Both protein clusters show a strong protein signal in the lag phase (cluster P36) or early growth phase (cluster P17). This is in common with the transcription dynamics of these genes grouped in cluster T11, which show an increased transcription in the lag phase. This shows a close connection of transcript and protein abundance of genes involved in chemotaxis, sporulation, flagellar biosynthesis and motility (COG class N).

Further co-occurrences were observed in clusters T32 and P10 (Figure 6). These clusters mainly consist of ribosomal proteins and other translation related genes and proteins (COG class J). Both clusters show a continuously decreasing transcript and protein abundance during the cultivation process. This shows, that changes on transcript level have an immediate effect on proteome level. Ribosomal proteins seem to be mainly regulated on transcriptomic level, since protein dynamics is highly similar to transcriptional changes. On transcriptional as well as on proteome level decreasing signals were observed for about 50 % of all annotated ribosomal proteins in the *Actinoplanes* sp. SE50/110 genome. However, some of the co-transcribed ribosomal proteins could be found in other protein clusters, like P23, which show a more constant protein level during filamentous growth. This shows that some ribosomal proteins are more stable than others.

Interestingly, the proteins of the acarbose biosynthesis gene cluster are distributed over 5 different cluster (P5, P13, P15, P28 and P34), although the *acb* genes were found to be highly co-transcribed in transcriptomic clusters T31 and T32. This indicates a regulation of *acb* gene expression on a post-transcriptional level and at least different protein half-live times. These findings will be discussed in the next chapter more detailed.

2.5 The genes of the acarbose biosynthetic gene cluster are transcriptionally and post-transcriptionally regulated during filamentous growth

When analyzing trends over the time course for the differentially transcribed genes, those of the acarbose biosynthetic gene cluster were particularly striking. As shown above, the temporal transcription dynamics of the *acb* genes (Figure 7B) seem to be highly similar to the specific acarbose formation rate (Figure 1B). Especially the genes *acbZ*, *acbB* and *acbA* follow the course of acarbose formation rate with an increase during the first 48 h and continuous decrease afterwards. This trend was observed for all further *acb* genes as well, but less strong.

The genes *acbZ, acbB, acbA, acbE* and *acbD*, which represent the monocistronically transcribed genes in the *acb* gene cluster, were grouped within a cluster of genes with substantial decrease of transcript amounts over the growth curve (Figure 4, cluster 32). All other *acb* genes cluster with genes showing a slight increase until late growth phase followed by a decrease of transcript amounts (Figure 4, cluster 31). Consequently, lower transcript levels were measured in the stationary phase compared to the filamentous growth phase for all *acb* genes (Figure 7). Remarkably, the most distinct differences were detected for the genes coding for the extracellular proteins AcbE (acarbose-resistant alpha-amylase) and AcbD (acarviose transferase) with fold changes of 32.7 and 60.9 on transcriptomic level, when comparing the filamentous growth (72.3 h) and the stationary phase (168 h). The genes *acbB* (coding for dTDP-4-keto-6-deoxy-glucose dehydratase) and *acbA* (encoding dTDP-glucose synthase) show fold changes of 17.4 and 15.2 comparing the filamentous growth phase and the stationary phase.

For all other *acb* genes, fold changes between 4.9 and 9.7 were found on transcript level. These *acb* genes code for the proteins of acarbose biosynthesis, an exporter and the extracellular alpha-amylase (*acbZ*). The fold change of *acbZ* was determined as 8.0 comparing the growth and the stationary phase. However, it was grouped in cluster 32 together with *acbB*, *acbA*, *acbE* and *acbD*.

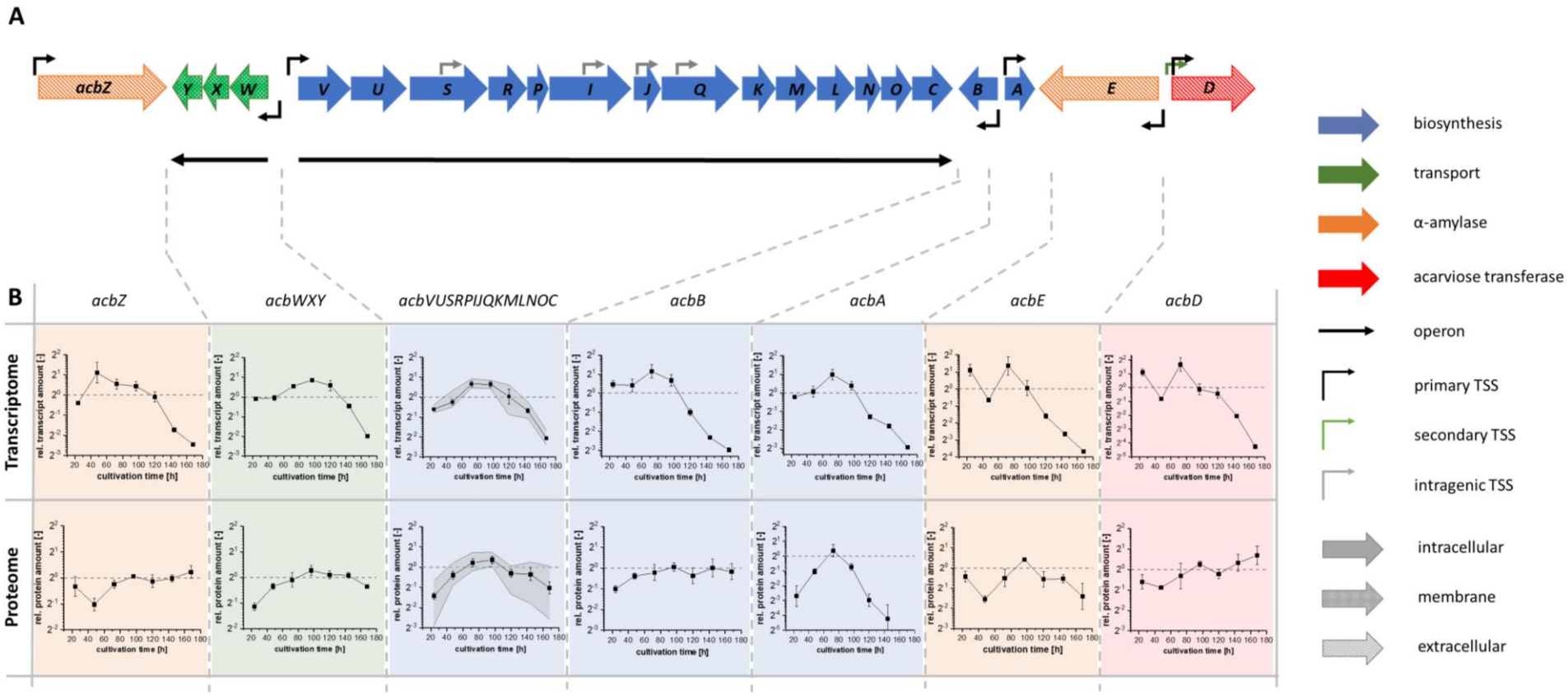

Figure 7: Overview about the expression dynamics of the acb gene cluster in *Actinoplanes* sp. SE50/110. (A) The *acb* gene cluster with its transcriptional landscape including operon structure and TSS [15]. The function of all genes and operons are color-coded. The sub-cellular localization (according to [51]) of the corresponding gene products are encoded by filled, dotted and striped arrow content. (B) Dynamics of the relative transcript abundances and the relative protein amounts of the acb genes and Acb proteins. A relative abundance of one corresponds to the average amount of RNA or protein over all time points. Mean values and standard deviation of three biological replicates are shown for each time point. For the operon *acbWXY* only protein abundancies of AcbW are shown. For the operon *acbVUSRPIJQKMLNOC* the maximum and minimum values are shown by grey area.

The similar course of transcription during growth and similar fold changes between growth and stationary phase indicate a co-regulation of these genes. Although the genes *acbZ*, *acbB*, *acbA*, *acbE* and *acbD* are transcribed monocistronically, co-regulation was reported for the pair *acbD* and *acbE* [25]. The genes *acbE* and *acbD* as well as *acbB* and *acbA* are located in opposite directions to each other in the genome sharing an intergenic region (Figure 7). It was assumed that the intergenic regions between the pairs *acbA* & *acbB* and *acbD* & *acbE* harbor binding sites for transcriptional regulators, which explains the co-regulation of these genes [13, 73]. For *acbE* and *acbD* it was shown, that the MalR type transcriptional regulator AcrC (ACSP50_6387) is the repressor of these two genes in *Actinoplanes* sp. SE50/110 [25].

All other *acb* genes, encoding for intracellular acarbose metabolism and acarbose export, are transcribed in the two operons *acbVUSRPIJQKMLNOC* and *acbWXY* (2.2.3). These two operons, which are also located in a head-to-head arrangement, seem to be co-regulated as well. So far, no regulators of these two operons are described.

Noticeably, the course of the transcript abundance for all *acb* genes (Figure 7) are more or less in accordance with the course of the specific product formation rate (Figure 1B). This might be an indication that the transcription of these genes has a direct influence on the acarbose production. However, protein abundances are not in correlation with the specific product formation rate for all Acb proteins. Especially the alpha-amylase AcbZ, the dTDP-4-keto-6-deoxy-glucose dehydratase AcbB and the acarviose transferase AcbD are not following their respective transcript signals on protein level. Protein levels for these enzymes seem to be almost constant during cultivation process. This could be a hint for a post-transcriptional regulation or high protein stability due to the secretion of these proteins, since less protease activity is expected in the extracellular space.

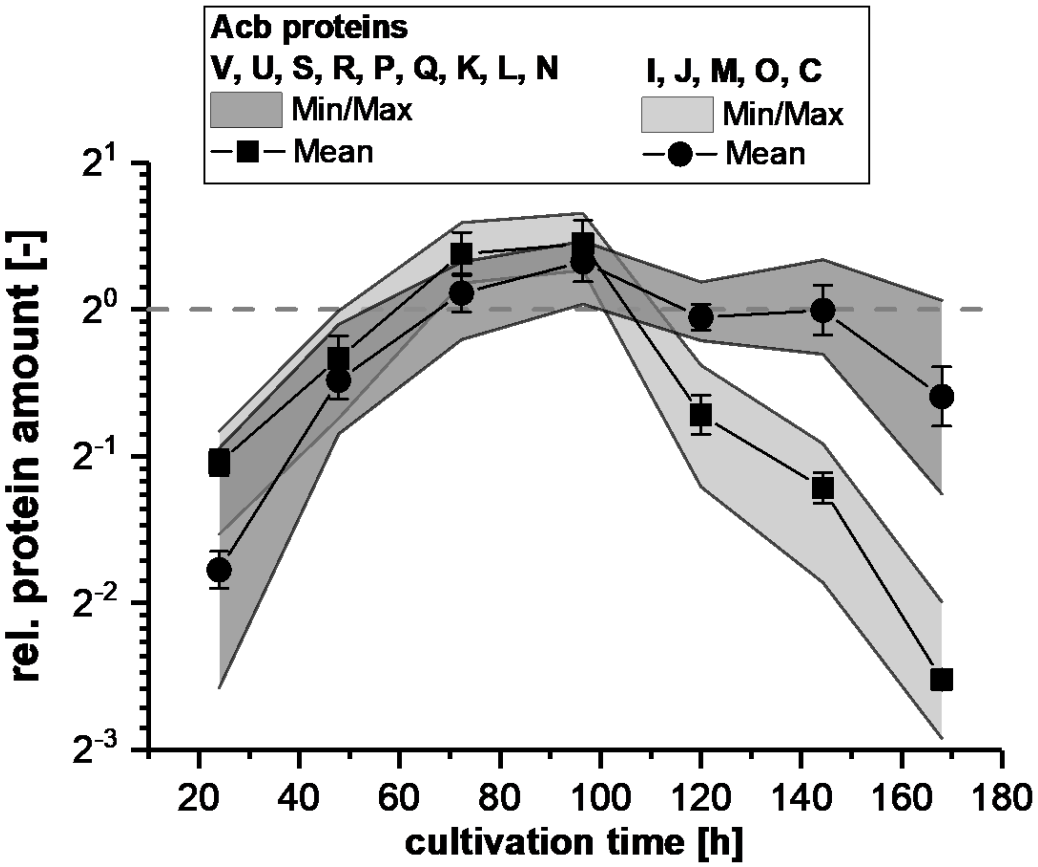

Figure 8: Dynamic of the protein abundancies of the Acb proteins encoded by the operon *acbVUSRPIJQKMLNOC*. Proteins were divided into two groups according to their behavior in the transition and stationary phase.

Based on the literature [18] and current models [26] acarbose is formed intracellularly and the secreted proteins AcbZ, AcbE and AcbD are not essential for acarbose formation [26]. The acarbose-resistant alpha-amylases AcbE and AcbZ degrade starch and maltodextrins to maltose and maltotriose or higher malto-oligosaccharides in the extracellular space [74]. The gene *acbD* encodes an acarviose transferase, which is supposed to catalyze the transfer of acarviosyl moieties from acarviose metabolites to the hydroxyl group of various sugars [75, 76]. Therefore, a direct correlation of the expression of the genes *acbZ*, *acbE* and *acbD* with the acarbose formation was not expected. However, it could be shown in previous studies, that AcbD is essential for acarbose formation in *Actinoplanes* sp. SE50/110 since an Δ*acbD* deletion mutant shows no acarbose formation [77]. As AcbD is proposed to transfer sugar moieties onto acarbose it can be assumed, that AcbD expression is important during the whole cultivation process and therefore should be expressed constantly. However, protein abundance of AcbD was found to be even slightly increased during growth. In contrast,

acbE, which is transcribed highly similar to *acbD*, shows a decreasing protein abundance in parallel to its transcription. This difference of protein abundances of the transcriptionally co-regulated genes *acbE* and *acbD* could be explained by the fact, that *acbD* seem to be transcribed from two or even three different TSS with different leader transcripts [15], which can influence AcbD translation efficiency (Figure 7).

Interestingly, AcbB, which is involved in the synthesis of dTDP-4-amino-4,6-dideoxy-D-glucose, is the only intracellular acarbose biosynthesis enzyme whose protein dynamics highly differs from its transcription profile during growth. The transcription of *acbB* strongly decreases during cultivation whereas its protein abundance stays on a constant level. In contrast, *acbA,* which seem to be highly co-regulated with *acbB* on a transcript level, shows a different protein dynamic which seems to be coupled to the corresponding transcription signal. It is striking that the expression pattern of AcbB and AcbA differ, although they are involved in the same part of acarbose biosynthesis [19]. Differences in expression strength could not be explained by differences in ribosome binding sites, since both genes are transcribed leaderless [15]. Therefore, an almost constant protein level could be due to a higher protein stability of AcbB or a regulatory effect on protein level. Protein modifications were not found for neither AcbB nor AcbA.

Since AcbA shows strongest decrease on proteome level, it would be an interesting target for overexpression as low AcbA amounts might be a bottleneck in the pathway operated by AcbA, AcbB and AcbV.

Strikingly, the proteins encoded by the large operon *acbVUSRPIJQKMLNOC* show diverse abundancies on protein level until the transition phase (Figure 8) at which acarbose formation decreases (Figure 1A), whereas their respective transcription seem to be similar (Figure 7).

The proteins AcbI, AcbJ, AcbM, AcbO and AcbC show a stronger decreasing protein abundance between 96 h and 168 h, whereas the other proteins encoded by the operon *acbVUSRPIJQKMLNOC* only slightly decrease. This could be due to different protein half-life times. Nevertheless, this could be an indication, that the five proteins AcbI, AcbJ, AcbM, AcbO and AcbC might be responsible for the decreasing acarbose formation during transition and stationary phase. Since AcbC, AcbO and AcbM catalyze the first steps in acarbose biosynthesis, down-regulation of the acarbose formation due to decreasing protein amounts might be beneficial to save energy and resources. This might also give explanation for the decrease of AcbA protein

abundance during the transition and stationary phase, since AcbA catalyzes the first step of the second synthesis branch of the acarbose biosynthesis [19]. Additionally, this step is in competition to reactions of central metabolism in *Actinoplanes* sp. SE50/110, as D-glucose-1-phosphate serves as a substrate in other cellular processes. Furthermore, a homologous gene (*ACSP50_3024*) was identified in the genome of *Actinoplanes* sp. SE50/110 [78]. This gene was found to be transcribed constantly over the whole cultivation process with a slight increase in the stationary phase (cluster 10). Therefore, it can be assumed, that available D-glucose-1-phosphate is consumed by ACSP50_3024 and no substrate is available for further acarbose biosynthesis.

Due to this, it can be assumed that acarbose formation is blocked due to absence of the first steps of both branches of the acarbose biosynthesis pathway.

2.6 Identification of genes co-expressed to the *acb* gene cluster

When analyzing genes co-transcribed to the *acb* genes located in one of the operons *acbWXY* and *acbVUSRPIJQKMLNOC* (cluster 31), it is noticeable that in addition to the *acb* genes only 120 other genes are located in cluster 31. Since most other clusters harbor more members, this indicates that the transcriptional dynamic of the *acb* genes is not rare but also not common in *Actinoplanes* sp. SE50/110. Analyzing the genes co-transcribed to these *acb* genes in cluster 31, also the genes *galG* and *galF* are among the genes with this transcription dynamics. The ABC-transporter GalHFG was formerly proposed as an acarbose importer, but it was shown that acarbose binds with low affinities to GalH. GalH has a high binding affinity to galactose, wherefore GalHFG is now suggested as a putative galactose importer [79]. The observation that these genes are co-regulated with the *acb* gene cluster and the direct genomic proximity might be an indication that these genes are after all involved in the acarbose metabolism. However, further experiments and proofs, like deletion mutants of these genes are needed to question the current assumptions about these genes.

Of the 120 co-transcribed genes, 30 genes are annotated as “hypothetical proteins” and 22 as “uncharacterized proteins”. Furthermore, two transcriptional regulators (*ACSP50_0424*, *ACSP50_8200*), two two-component regulator systems (*ACSP50_2300*, *ACSP50_5226*) and a sigma factor (*ACSP50_7877*) could be found inside of cluster 31 indicating a co-transcription to the acarbose biosynthesis operons.

These regulators could be interesting targets for gene deletions analyzing the effect on acarbose biosynthesis.

A further example is the operon *ACSP50_6408* to *ACSP50_6411,* which was found to be transcribed in the same course as the two operons in *acb* gene cluster. This operon encodes enzymes involved in the histidine metabolism (formation of ergothioneine from L-histidine). Ergothioneine has been described to be synthesized in many actinomycetes, cyanobacteria, methylobacteria and some fungi. It is described to be resistant to autooxidation and therefore enable survive of microbes under oxidative stress [80].

Interestingly, the gene *ACSP50_2474* encoding a maltose degrading enzyme (AmlE), which was previously identified to be essential for maltose utilization in *Actinoplanes* sp. SE50/110 [81], was identified to be co-transcribed to the two *acb* operons. In the related species *S. glaucescens* GLA.O the *amlE*-homolog was even identified as part of the *gac* acarbose biosynthesis gene cluster, which suggests a co-evolution in this species [81]. This shows the close connection between maltose and acarbose metabolism, since maltose seem to be essential for production of acarbose in *Actinoplanes* sp. SE50/110.

Furthermore, the gene *cgt* (*ACSP50_5024*) was also found to be transcribed parallel to the *acb* operons. The function of the gene product is unclear. Several functional analyses were carried out on the extracellular protein Cgt, but no enzymatic activity could be determined [82]. However, the deletion of *cgt* lead to an increase of acarbose formation in *Actinoplanes* sp. SE50/110 [82]. This effect is supposed to be caused by the reduced metabolic burden, since Cgt was identified to be one of the highest abundant proteins in *Actinoplanes* sp. SE50/110 [83-85]. The similar expression patterns shown in this study, support the suggestion, that by deletion of genes co-expressed with the *acb* gene cluster, the acarbose formation might be improved. To further reduce the metabolic burden in order to improve the acarbose formation, this study suggests deletion of the 52 genes without functional annotation (see above).

In contrast to cluster 31, in cluster 32 there are 237 genes beside the monocistronically transcribed *acb* genes (*acbA*, *acbB, acbD, acbE* and *acbZ*) showing the same transcription pattern. However, 51 of these 237 genes are annotated as “hypothetical” or “uncharacterized proteins”. A high number of ribosomal proteins is located in this

cluster. This indicates a close connection of these *acb* genes to the primary metabolism of *Actinoplanes* sp. SE50/110.

Furthermore, 9 transcriptional regulators (*ACSP50_1631*, *ACSP50_2235, ACSP50_4697*, *ACSP50_5005*, *ACSP50_6401*, *ACSP50_6463, ACSP50_8007, ACSP50_8120* and *ACSP50_8287*), a two-component regulator system (*ACSP50_3744*, *ACSP50_3745*) and 2 sigma factor genes (*ACSP50_0644*, *ACSP50_6006*) were determined to show the same transcription dynamics as the monocistronic *acb* genes located in cluster 32. Several genes located in cluster 32 are involved in amino acid transport and metabolism (31), nucleotide transport and metabolism (19) and carbohydrate transport and metabolism (19). This shows, that most of the genes co-transcribed with the monocistronically transcribed *acb* genes belong to the central metabolism of *Actinoplanes* sp. SE50/110.

Using the data from combined clustering approach (2.4) 21 different genes were found to be clustered regarding both transcript and protein abundance with the *acb* genes (Figure 6). Of these genes, 6 are annotated as “uncharacterized proteins”. Among the remaining 15 genes, a glycosyl transferase gene (ACSP50_7756) was identified. It needs to be elucidated in future, whether this enzyme is involved in the acarbose biosynthesis or by-component formation. In addition, one transcriptional regulator (*ACSP50_0424*) was found to show similar transcript as well as protein profile. This regulator gene seems to be widespread in the family Micromonosporaceae but no specific function was reported, yet. Therefore, it would be interesting to further analyze these transcriptional regulators regarding a potential effect on acarbose biosynthesis since they are highly co-expressed with several *acb* genes.

3. Conclusions

The combination of robust and controlled cultivation conditions with state-of-the-art transcriptomics and proteomics in a high temporal resolution is well suited to answer a variety of biological questions. The close connection of acarbose biosynthesis and growth could be elucidated.

Using the transcriptomic data, comprehensive analyses of the transcriptional landscape of *Actinoplanes* sp. SE50/110 were performed. Using high-quality RNA-seq data of different growth phases more than 99% of all genomic features were covered

in the analyses. This way, the operon structure with 1,029 primary operons, 4,228 transcription start sites and a consensus promoter sequences (-10 motif: TAnnnT; -35 motif: nTGACn) of the *Actinoplanes* sp. SE50/110 genome were obtained.

Through high-accurate proteome studies, 1,441 proteins could be identified under the tested conditions, of which 1,374 were found in the cellular fraction and 67 in the supernatant of *Actinoplanes* sp. SE50/110.

In this study, co-regulations of genes during different growth phases and in correlation to their respective protein dynamics were shown. Especially for acarbose biosynthesis genes striking results regarding transcription and protein dynamics could be achieved. It could be shown, that transcription of the acarbose biosynthesis gene cluster is in close correlation to the specific product formation rate regarding acarbose. However, on protein level several differences were found. Unlike to the other Acb proteins, AcbZ, AcbB and AcbD show a protein dynamic which differ from their respective transcription pattern suggesting that these proteins are more stable or post-transcriptional regulated. AcbB in particular seems to play an important role in acarbose biosynthesis since protein level was found to be constant during whole cultivation process. However, AcbA which catalyzes the step directly before AcbB seems to be a limiting factor in this branch of acarbose biosynthesis as protein level of AcbA strongly decreases since the middle of the filamentous growth phase (72.3 h).

Finally, genes could be identified, which beforehand were not in the focus of acarbose research in *Actinoplanes* sp. SE50/110. The combined clustering approach revealed several genes which are strictly co-expressed with the *acb* gene cluster. In this context, the transcriptional regulator genes *ACSP50_0424* was described as interesting target for further analyses regarding a potential effect on acarbose biosynthesis, because strong co-expression to *acb* genes was found.

This approach of analyzing the expression dynamics can be applied to other strains and organisms as well as experimental settings, like the spike-in of nutrients, stress factors or substances enhancing product formation. Thus, the temporal transcriptional response upon induced changes in environmental conditions could be elucidated with the here established methodology.

4. Methods

4.1 Strains, media, and cultivation conditions

The bacterial strain *Actinoplanes* sp. SE50/110 (ATCC 31044) was grown on soy flour medium (SFM; 20 g L^{-1} soy flour, 20 g L^{-1} mannitol, 20 g L^{-1} agar, pH 8, tap water), in NBS complex medium and maltose-containing minimal medium as described elsewhere [51]. Bioreactor cultivations of *Actinoplanes* sp. SE50/110 was carried out in three biological replicates. NBS medium was inoculated from glycerol stocks prepared according to 81. After 2 days of cultivation 300 µL was plated on SFM agar plates and incubated for 5 days at 28 °C to generate spores. The spores were harvested by adding 2 mL ddH_2O on the plates and carefully detaching them with a cotton swab. One plate resulted in 1 mL of spore suspension. Spores from 30 plates were pooled and used for the inoculation of each bioreactor cultivation. For the cultivation 3 L DASGIP® Bioblock (Eppendorf, Hamburg, Germany) bioreactors with a working volume of 1.6 L were used. Cells were grown in maltose minimal medium at 28 °C. The pH was set to 6.7 and automatically controlled by the addition of 10 % H_3PO_4 or 2 N NaOH respectively. Dissolved oxygen level was set to 30 % controlled by stirrer speed and oxygen partial pressure in the air supply in a two-step cascade.

The cultivations were examined by determination of the cell dry weight and quantification of acarbose as described previously [25]. For both transcriptome and proteome analysis, 1 mL of the cell suspension was centrifuged for 15 sec at 16.000 *g* and immediately frozen in liquid nitrogen. Cell pellets were stored at -80 °C until further processed for RNA or protein isolation. For the extraction of extracellular proteins 10 mL of the cell suspension was centrifuged for 2 minutes at 4,000 *g*. The supernatant was frozen and lyophilized thereafter.

4.2 Acarbose quantification using HPLC measurement

Acarbose was quantified from the supernatant of *Actinoplanes* sp. SE50/110 by high-performance liquid chromatography. Therefore, 1 mL of the culture supernatant was centrifuged (20,000 *g*, 2 min) to remove residual biomass and other particles. Next, 200 µL of the supernatant was mixed with 800 µL methanol by vortexing and centrifuged again (20,000 *g*, 2 min) to remove precipitate. Afterwards, the samples were transferred to HPLC vials and analyzed in an Agilent 1100 HPLC system

(G1312A Binary Pump, G1329A ALS autosampler, G1315A diode-array detector (DAD)) using a Thermo Fisher Scientific (Waltham, MA, USA) Hypersil APS-2 column (125 x 4 mm, particle size: 3 µm) heated to 40 °C. As a mobile phase 32 % phosphate buffer (0.62 g L^{-1} KH_2PO_4 and 0.38 g L^{-1} K_2HPO_4 x H_2O) (solvent A) mixed with 68 % acetonitrile (solvent B) was used with an isocratic flow of 1 mL min^{-1}. 40 µL of each sample was injected and separated for 10 minutes. Acarbose was detected at 210 nm (reference of 360 nm) with a DAD detector and quantified from the peak areas with a calibration curve.

4.3 Isolation of total RNA and RNA-seq

RNA was isolated with a Macherey-Nagel NucleoSpin® RNA Plus Kit (MACHEREY-NAGEL, Düren, Germany) in combination with an rDNase Set (MACHEREY-NAGEL, Düren, Germany) as described previously [21]. In brief, frozen cell pellets were resuspended in 500 µL LB-buffer and transferred to 2 mL lysing matrix tubes. Cell disruption was carried out in a homogenizer three times for 20 s at 6.5 m s^{-1} and cooled on ice between steps. Subsequently, the cell suspension was centrifuged, and the RNA was extracted from the supernatant using the NucleoSpin® RNA Plus Kit in combination with rDNase Set for an on-column DNA digestion. After the clean-up the DNA-digestion was repeated in-solution. Residual DNA was tested negatively with two primer pairs binding to genomic DNA of *Actinoplanes* sp. SE50/110. Quality and quantity of the RNA were analyzed with an Xpose® spectrophotometer (Unchained Labs, Pleasanton, CA, USA) and an Agilent RNA 6000 Pico kit run on an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA).

For each time point the RNA of three fermenters was isolated and used for cDNA library preparation. The preparation of cDNA libraries were performed according to [15]. Additionally, a pool was generated for each fermenter of all time points using equimolar amounts of RNA.

All cDNA libraries were sequenced using TruSeq kits (Illumina, San Diego, CA, USA) on an HiSeq1500 sequencer system (Illumina, San Diego, CA, USA) using a paired-end mode with 2 x 70 nt read length.

RNA-seq yielded about 1.39 to 4.96 million read pairs for the 21 libraries (t_1 to t_7 in 3 biological replicates) and 1.91 to 2.79 million read pairs for the library of the pooled RNA samples over all time points. The reads were mapped to the reference sequence

(GenBank: LT827010.1) using bowtie2 v2.3.2 in the paired-end mode [86], resulting in 90.3 % to 98.9 % mapped reads. The coverage of each annotated feature (protein-coding genes and RNA genes) was determined using featureCounts [87]. Afterwards, the software ReadXplorer was used for visualization of the transcriptomic data [30, 88]. After coverage analysis, differential expression analysis was carried out using the tool DESeq2 [89], to compare the differences in transcription for each gene at each time point. As a reference a pooled RNA sample from all time points was used as a mean value of transcription over the whole cultivation process. Thereby, $\log_2$ (fold-changes) (M-values) for each transcript compared to the mean transcript amount for all time points was determined in three biological replicates. RNA-seq data have been deposited in the ArrayExpress database at EMBL-EBI [90] under accession number E-MTAB-8857.

4.4 Extraction of proteins from *Actinoplanes* sp. SE50/110 cell pellets

For protein isolation, the freeze-dried cell pellet was dissolved in 500 µl of 100 mM ammonium bicarbonate (Honeywell Fluka) and transferred to 2 mL lysing matrix tubes (0.1 mm spherical silica beads, MP Biomedicals, Santa Ana, California, USA). The cell suspension was disrupted in a homogenizer (FastPrep FP120, Thermo Fisher Scientific, Waltham, MA, USA) three times for 20 s at 6.5 m s^{-1} and cooled on ice between steps.

Next, the organic solvent trifluoroethanol (TFE) was used for isolation of the cytosolic proteins [91]. Therefore, the supernatant of the ribolysed cell pellets was transferred in low protein binding collection tubes (Thermo Fisher Scientific, MA, USA). 1 % RapiGest (Waters Corporation, Milford, MA, USA), 100 µl of TFE (Honeywell Fluka, Morristown, NJ, USA) and 5 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) was added to the cell suspension. The sample was incubated for 60 minutes at 60 °C and inverted several times during that time. In a second incubation step 20 µL of 200 mM Chloroacetamid (CAA) was added and left for 90 minutes in the dark. Afterwards, 5 µL of 200 mM TCEP was mixed to the solution. The sample was inverted and incubated at room temperature for additional 60 minutes. In the next step, the proteins within the cell suspension were digested.

4.5 Extraction of proteins from the supernatant of *Actinoplanes* sp. SE50/110

For extraction of extracellular proteins, the method developed by [83] was used. In brief, the proteins of the freeze-dried supernatant were isolated by phenol extraction with subsequent methanol precipitation and several washing steps with 70% ethanol.

After drying the extracellular proteins were resuspended in 200µl TE-Buffer and 5 mM dithiothreitol (DTT) and incubated for 60 minutes at 60 °C. 20 µL of 200 mM iodacetamid (IAA) was added and left for 90 minutes in the dark. Next, 5 µL of 200 mM DTT was mixed to the solution and incubated for additional 60 minutes. The isolated proteins of the extracellular space were digested afterwards according to the proteins of the cellular fraction.

4.6 Protein digestion and mass spectrometry measurements and data analysis

For protein digestion, the solutions were diluted with 435 µL of 100mM ammonium bicarbonate and 435 µL of bidistilled water. In the next step, 5 µL Trypsin Gold (1µg μL^{-1}) (Mass Spectrometry Grade, Promega, WI, USA) was added and the solution was incubated at 37 °C overnight. The digest was purified on the next day with Sep Pak C18 cartridges (Waters Corporation, Milford, MA, USA). The Sep Pak C18 cartridges was rinsed with 1 mL of solution B (65 % Acetonitril, 35 % bidistilled water; 0.1 % TFA). The column was equilibrated with 1 mL of solution A (98 % Acetonitril, 2 % bidistilled water; 0.1 % TFA). Next the digest mixed with 1 mL of solution A was added and run through the column slowly. Subsequently, the cartridges were washed with 1 mL of solution A. The proteome was eluted with 100 µL solution B in low protein binding collection tubes and dried in a vacufuge concentrator (Eppendorf, Hamburg, Germany). Next the dried peptide mixture was resuspended in 15 µL of solution A and measured with a Nano-Drop™ 2000/2000c (Thermo Fisher, MA, USA).

LC-MS/MS measurement were carried out using a QExactive mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) online coupled to the LC system. The peptides were separated on a 25 cm steel column Acclaim™ PepMap™ 100 C18-LC-column with a particle size of 2 µm and a diameter of 75 µm (Thermo Fisher, MA, USA).

Identification and label-free quantification (LFQ) analysis was performed using the software MaxQuant with default settings and a false discovery rate of p_{adj} <0.05 [56]. For identification, ms spectra were searched with Andromeda against the target-decoy protein sequence database of *Actinoplanes* sp. SE50/110 based on the annotation data from [15]. Only unique peptides were used for the quantification. An oxidation of methionine (15.99 Da) were allowed up to three times per peptide. As static modification a carbamidomethylation of cysteine (57.02 Da) and the dynamic modification of the N terminus with an acetylation (42.01 Da) was allowed.

The statistical analysis of LFQ data obtained from MaxQuant was performed with Perseus 1.6.10.43 [92]. In total, 565,792 MS/MS spectra were recorded for the cytosolic fraction resulting in 210,637 identified peptide sequences corresponding to 2,663 proteins. For the extracellular fraction, 243,903 MS/MS spectra were recorded resulting in 35,000 identified peptide sequences, which could be associated to 911 proteins. As a first filtering step, proteins which were identified with less than two unique peptides were excluded from further analyses. Next, all time points of a protein were ruled out, of which the respective protein could not be identified in all three replicates. LFQ intensities were normalized by z-normalization and afterwards $\log_2$ (fold-change) was calculated by comparing each time point against the mean protein intensity over all analyzed time points. Finally, a two-sample *t*-test was performed to identify significant differences in protein level for each time point for the respective protein. The false discovery rate was set to p_{adj} <0.05.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [93] via the PRIDE partner repository [94] with the dataset identifier PXD017973.

4.7 Hierarchical cluster analysis

For hierarchical cluster analysis, the multi-omics data integration tool set *Omics Fusion* (https://fusion.cebitec.uni-bielefeld.de/) was used [36]. Genes were considered for analysis, when at least two third of the time points had numerical values. Missing values for single time points were replaced by means of the earlier and later time point. Hierarchical cluster analysis with grouping of clusters were performed with the Wards method for linkage and Euclidean distances, respectively. The maximum number of

clusters was set to 50 and the optimal cluster size was calculated using the Krzanowski-Lai index [95, 96].

First, both transcriptomic and proteomic data sets were clustered individually. In the second step a combined clustering approach was used. Therefore, only features of which both transcriptomic and proteomic data are available were used. Again, Wards linkage method with Euclidean distance was used. The maximal cluster number was set to 50. As a result, a combined heatmap was generated indicating co-occurrences of features in the respective proteomic and transcriptomic cluster.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

The transcriptome datasets generated for this study can be found in the Array Express database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-8857.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD017973.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

JD acknowledges support from the CLIB-Graduate Cluster Industrial Biotechnology at Bielefeld University, Germany, which is supported by the Ministry of Innovation, Science and Research (MIWF) of the federal state North Rhine-Westphalia, Germany and Bielefeld University, Germany. We gratefully thank our cooperation partner Bayer AG (Leverkusen, Germany) for their continuous support and funding. We acknowledge the support for the Article Processing Charge by the Deutsche Forschungsgemeinschaft and the Open Access Publication Fund of Bielefeld University.

Authors' Contribution

JD designed, planned, and interpreted the experimental work, performed data analysis and drafted the manuscript. JD and VO performed the proteome studies. JD and MP performed the fermenter cultivation experiments. VO, LS, SS-B, AP and JK assisted in interpreting the data and revised the manuscript. JK and AP coordinated the study. All authors read and approved the manuscript.

Acknowledgments

We gratefully thank Julia Voß for support regarding proteome measurements and proteome data analysis. We thank Dr. Tobias Busche and the IIT Biotech GmbH for support in RNA sequencing. We cordially thank Dr. Stefan Albaum for support with using the tool *Omics Fusion*.

Supplementary Material

Supplementary Figures (.docx): Includes all supplementary figures.

Supplementary Table 1 (.xlsx): Operon structure of the *Actinoplanes* sp. SE50/110 genome.

Supplementary Table 2 (.xlsx): Transcription start sites (TSS) and promoter motifs in the *Actinoplanes* sp. SE50/110 genome.

Supplementary Table 3 (.xlsx): Transcription dynamics of *Actinoplanes* sp. SE50/110

Supplementary Table 4 (.xlsx): Proteome dynamics of *Actinoplanes* sp. SE50/110.

References

1. Parenti F, Coronelli C. Members of the genus *Actinoplanes* and their antibiotics. Annu Rev Microbiol. 1979;33:389–411. doi:10.1146/annurev.mi.33.100179.002133.

2. Ludwig W, Euzéby J, Whitman WB. Taxonomic outline of the phylum Actinobacteria. In: Goodfellow M, Busse H-J, Kämpfer P, Ludwig W, Suzuki K-i, Trujillo ME, Whitman WB, editors. Bergey's Manual® of Systematic Bacteriology: Volume Five The Actinobacteria, Part A. New York, NY: Springer New York; 2012. p. 29–31. doi:10.1007/978-0-387-68233-4_2.

3. Vobis G, Schäfer J, Kämpfer P. *Actinoplanes*. In: Whitman WB, editor. Bergey's manual of systematics of archaea and bacteria. [Hoboken, New Jersey]: Wiley; 2015. p. 1–41. doi:10.1002/9781118960608.gbm00139.

4. Okami Y, Hotta K. Search and Discovery of New Antibiotics. In: Goodfellow M, Williams ST, Mordarski M, editors. Actinomycetes in biotechnology. San Diego: Academic Press; 1988. p. 33–67. doi:10.1016/B978-0-12-289673-6.50007-5.

5. Vobis G. The Genus *Actinoplanes* and Related Genera. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. The Prokaryotes. New York, NY: Springer New York; 2006. p. 623–653. doi:10.1007/0-387-30743-5_24.

6. Debono M, Merkel KE, Molloy RM, Barnhart M, Presti E, Hunt AH, Hamill RL. Actaplanin, new glycopeptide antibiotics produced by *Actinoplanes missouriensis*. The isolation and preliminary chemical characterization of actaplanin. J Antibiot. 1984;37:85–95. doi:10.7164/antibiotics.37.85.

7. Bardone MR, Paternoster M, Coronelli C. Teichomycins, new antibiotics from *Actinoplanes teichomyceticus* nov. sp. II. Extraction and chemical characterization. J Antibiot. 1978;31:170–7. doi:10.7164/antibiotics.31.170.

8. Aretz W, Meiwes J, Seibert G, Vobis G, Wink J. Friulimicins: novel lipopeptide antibiotics with peptidoglycan synthesis inhibiting activity from *Actinoplanes friuliensis* sp. nov. I. Taxonomic studies of the producing microorganism and fermentation. J Antibiot. 2000;53:807–15. doi:10.7164/antibiotics.53.807.

9. Ciabatti R, Kettenring JK, Winters G, Tuan G, Zerilli L, Cavalleri B. Ramoplanin (A-16686), a new glycolipodepsipeptide antibiotic. III. Structure elucidation. J Antibiot. 1989;42:254–67. doi:10.7164/antibiotics.42.254.

10. Truscheit E, Frommer W, Junge B, Müller L, Schmidt DD, Wingender W. Chemistry and Biochemistry of Microbialα-Glucosidase Inhibitors. Angew. Chem. Int. Ed. Engl. 1981;20:744–61. doi:10.1002/anie.198107441.

11. Creutzfeldt W. Acarbose for the Treatment of Diabetes Mellitus. Berlin, Heidelberg: Springer Berlin Heidelberg; 1988.

12. Bischoff H. Pharmacology of alpha-glucosidase inhibition. Eur J Clin Invest. 1994;24 Suppl 3:3–10. doi:10.1111/j.1365-2362.1994.tb02249.x.

13. Wehmeier UF, Piepersberg W. Biotechnology and molecular biology of the alpha-glucosidase inhibitor acarbose. Appl Microbiol Biotechnol. 2004;63:613–25. doi:10.1007/s00253-003-1477-2.

14. Li C, Hung Y-J, Qamruddin K, Aziz MFA, Stein H, Schmidt B. International noninterventional study of acarbose treatment in patients with type 2 diabetes mellitus. Diabetes Res Clin Pract. 2011;92:57–64. doi:10.1016/j.diabres.2010.12.033.

15. Wolf T, Schneiker-Bekel S, Neshat A, Ortseifen V, Wibberg D, Zemke T, et al. Genome improvement of the acarbose producer *Actinoplanes* sp. SE50/110 and annotation refinement based on RNA-seq analysis. J Biotechnol. 2017;251:112–23. doi:10.1016/j.jbiotec.2017.04.013.

16. Zhang C-S. Genomic Analysis of Secondary Metabolite Producing Actinomycetes: AcbM is a 2-epi-5-epi-valiolone 7-kinase [Doctoral dissertation]. Wuppertal: Bergische Universität Wuppertal; 2002.

17. Zhao Q, Luo Y, Zhang X, Kang Q, Zhang D, Zhang L, et al. A severe leakage of intermediates to shunt products in acarbose biosynthesis. ncomms. 2020;11:613. doi:10.1038/s41467-020-15234-8.

18. Zhang C-S, Stratmann A, Block O, Brückner R, Podeschwa M, Altenbach H-J, et al. Biosynthesis of the C(7)-cyclitol moiety of acarbose in *Actinoplanes* species SE50/110. 7-O-phosphorylation of the initial cyclitol precursor leads to proposal

of a new biosynthetic pathway. J Biol Chem. 2002;277:22853–62. doi:10.1074/jbc.M202375200.

19. Zhang D, Zhao Q, Jiang M, Kang Q, Bai L. Biosynthetic pathway of deoxyaminosugar moiety in acarbose from *Actinoplanes* sp. SE50/110. Acta Microbiologica Sinica. 2020:118–34. doi:10.13343/j.cnki.wsxb.20190109.

20. Wolf T, Gren T, Thieme E, Wibberg D, Zemke T, Pühler A, Kalinowski J. Targeted genome editing in the rare actinomycete *Actinoplanes* sp. SE50/110 by using the CRISPR/Cas9 System. J Biotechnol. 2016;231:122–8. doi:10.1016/j.jbiotec.2016.05.039.

21. Schaffert L, März C, Burkhardt L, Droste J, Brandt D, Busche T, et al. Evaluation of vector systems and promoters for overexpression of the acarbose biosynthesis gene acbC in *Actinoplanes* sp. SE50/110. Microb Cell Fact. 2019;18:114. doi:10.1186/s12934-019-1162-5.

22. Gren T, Ortseifen V, Wibberg D, Schneiker-Bekel S, Bednarz H, Niehaus K, et al. Genetic engineering in *Actinoplanes* sp. SE50/110 - development of an intergeneric conjugation system for the introduction of actinophage-based integrative vectors. J Biotechnol. 2016;232:79–88. doi:10.1016/j.jbiotec.2016.05.012.

23. Rockser Y, Wehmeier UF. The gac-gene cluster for the production of acarbose from *Streptomyces glaucescens* GLA.O — Identification, isolation and characterization. J Biotechnol. 2009;140:114–23. doi:10.1016/j.jbiotec.2008.10.016.

24. Guo X, Geng P, Bai F, Bai G, Sun T, Li X, et al. Draft genome sequence of *Streptomyces coelicoflavus* ZG0656 reveals the putative biosynthetic gene cluster of acarviostatin family α-amylase inhibitors. Lett Appl Microbiol. 2012;55:162–9. doi:10.1111/j.1472-765X.2012.03274.x.

25. Wolf T, Droste J, Gren T, Ortseifen V, Schneiker-Bekel S, Zemke T, et al. The MalR type regulator AcrC is a transcriptional repressor of acarbose biosynthetic genes in *Actinoplanes* sp. SE50/110. BMC Genomics. 2017;18:562. doi:10.1186/s12864-017-3941-x.

26. Wendler S, Ortseifen V, Persicke M, Klein A, Neshat A, Niehaus K, et al. Carbon source dependent biosynthesis of acarviose metabolites in *Actinoplanes* sp. SE50/110. J Biotechnol. 2014;191:113–20. doi:10.1016/j.jbiotec.2014.08.019.

27. Wolf T. Transcriptional regulation of acarbose biosynthesis in *Actinoplanes* sp. SE50/110 analyzed by next-generation sequencing, transcriptomics and genome editing [Doctoral dissertation]. Bielefeld: Bielefeld University; 2017.

28. Jeong Y, Kim J-N, Kim MW, Bucca G, Cho S, Yoon YJ, et al. The dynamic transcriptional and translational landscape of the model antibiotic producer *Streptomyces coelicolor* A3(2). Nat Commun. 2016;7:11605. doi:10.1038/ncomms11605.

29. Nieselt K, Battke F, Herbig A, Bruheim P, Wentzel A, Jakobsen ØM, et al. The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*. BMC Genomics. 2010;11:10. doi:10.1186/1471-2164-11-10.

30. Hilker R, Stadermann KB, Doppmeier D, Kalinowski J, Stoye J, Straube J, et al. ReadXplorer--visualization and analysis of mapped sequences. Bioinformatics. 2014;30:2247–54. doi:10.1093/bioinformatics/btu205.

31. Schwientek P, Neshat A, Kalinowski J, Klein A, Rückert C, Schneiker-Bekel S, et al. Improving the genome annotation of the acarbose producer *Actinoplanes* sp. SE50/110 by sequencing enriched 5'-ends of primary transcripts. J Biotechnol. 2014;190:85–95. doi:10.1016/j.jbiotec.2014.03.013.

32. Ao W, Gaudet J, Kent WJ, Muttumu S, Mango SE. Environmentally induced foregut remodeling by PHA-4/FoxA and DAF-12/NHR. Science. 2004;305:1743–6. doi:10.1126/science.1102216.

33. Singh SS, Typas A, Hengge R, Grainger DC. *Escherichia coli* σ^{70} senses sequence and conformation of the promoter spacer region. Nucleic Acids Res. 2011;39:5109–18. doi:10.1093/nar/gkr080.

34. Mitchell JE, Zheng D, Busby SJW, Minchin SD. Identification and analysis of 'extended -10' promoters in *Escherichia coli.* Nucleic Acids Res. 2003;31:4689–95. doi:10.1093/nar/gkg694.

35. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14:1188–90. doi:10.1101/gr.849004.

36. Brink BG, Seidel A, Kleinbölting N, Nattkemper TW, Albaum SP. Omics Fusion - A Platform for Integrative Analysis of Omics Data. J Integr Bioinform. 2016;13:296. doi:10.2390/biecoll-jib-2016-296.

37. Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 2015;43:D261-9. doi:10.1093/nar/gku1223.

38. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30. doi:10.1093/nar/28.1.27.

39. Palleroni NJ. Chemotaxis in *Actinoplanes*. Arch. Microbiol. 1976;110:13–8. doi:10.1007/BF00416963.

40. Uchida K, Jang M-S, Ohnishi Y, Horinouchi S, Hayakawa M, Fujita N, Aizawa S-I. Characterization of *Actinoplanes missouriensis* spore flagella. Appl Environ Microbiol. 2011;77:2559–62. doi:10.1128/AEM.02061-10.

41. Mouri Y, Konishi K, Fujita A, Tezuka T, Ohnishi Y. Regulation of Sporangium Formation by BldD in the Rare Actinomycete *Actinoplanes missouriensis*. J Bacteriol 2017. doi:10.1128/JB.00840-16.

42. Jang M-S, Mouri Y, Uchida K, Aizawa S-I, Hayakawa M, Fujita N, et al. Genetic and Transcriptional Analyses of the Flagellar Gene Cluster in *Actinoplanes missouriensis*. J Bacteriol. 2016;198:2219–27. doi:10.1128/JB.00306-16.

43. Phillips ZEV, Strauch MA. *Bacillus subtilis* sporulation and stationary phase gene expression. Cell Mol Life Sci. 2002;59:392–402. doi:10.1007/s00018-002-8431-9.

44. Chandra G, Chater KF. Developmental biology of *Streptomyces* from the perspective of 100 actinobacterial genome sequences. FEMS Microbiol Rev. 2014;38:345–79. doi:10.1111/1574-6976.12047.

45. Strauch E, Takano E, Baylis HA, Bibb MJ. The stringent response in *Streptomyces coelicolor* A3(2). Mol Microbiol. 1991;5:289–98. doi:10.1111/j.1365-2958.1991.tb02109.x.

46. Brockmann-Gretza O, Kalinowski J. Global gene expression during stringent response in *Corynebacterium glutamicum* in presence and absence of the rel

gene encoding (p)ppGpp synthase. BMC Genomics. 2006;7:230. doi:10.1186/1471-2164-7-230.

47. Ruwe M, Persicke M, Busche T, Müller B, Kalinowski J. Physiology and Transcriptional Analysis of (p)ppGpp-Related Regulatory Effects in *Corynebacterium glutamicum*. Front Microbiol. 2019;10:2769. doi:10.3389/fmicb.2019.02769.

48. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 2019;47:W81-W87. doi:10.1093/nar/gkz310.

49. DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science. 1997;278:680–6. doi:10.1126/science.278.5338.680.

50. Kolter R, Siegele DA, Tormo A. The stationary phase of the bacterial life cycle. Annu Rev Microbiol. 1993;47:855–74. doi:10.1146/annurev.mi.47.100193.004231.

51. Wendler S, Otto A, Ortseifen V, Bonn F, Neshat A, Schneiker-Bekel S, et al. Comprehensive proteome analysis of *Actinoplanes* sp. SE50/110 highlighting the location of proteins encoded by the acarbose and the pyochelin biosynthesis gene cluster. J Proteomics. 2015;125:1–16. doi:10.1016/j.jprot.2015.04.013.

52. Wendler S, Otto A, Ortseifen V, Bonn F, Neshat A, Schneiker-Bekel S, et al. Comparative proteome analysis of *Actinoplanes* sp. SE50/110 grown with maltose or glucose shows minor differences for acarbose biosynthesis proteins but major differences for saccharide transporters. J Proteomics. 2016;131:140–8. doi:10.1016/j.jprot.2015.10.023.

53. Otto A, Bernhardt J, Meyer H, Schaffer M, Herbst F-A, Siebourg J, et al. Systems-wide temporal proteomic profiling in glucose-starved *Bacillus subtilis*. Nat Commun. 2010;1:137. doi:10.1038/ncomms1137.

54. Armengaud J, Christie-Oleza JA, Clair G, Malard V, Duport C. Exoproteomics: exploring the world around biological systems. Expert Rev Proteomics. 2012;9:561–75. doi:10.1586/epr.12.52.

55. Koussounadis A, Langdon SP, Um IH, Harrison DJ, Smith VA. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. Sci Rep. 2015;5:10775. doi:10.1038/srep10775.

56. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26:1367–72. doi:10.1038/nbt.1511.

57. Li Z, Wang Y, Yao Q, Justice NB, Ahn T-H, Xu D, et al. Diverse and divergent protein post-translational modifications in two growth stages of a natural microbial community. Nat Commun. 2014;5:4405. doi:10.1038/ncomms5405.

58. Grangeasse C, Stülke J, Mijakovic I. Regulatory potential of post-translational modifications in bacteria. Front Microbiol. 2015;6:500. doi:10.3389/fmicb.2015.00500.

59. Brown CW, Sridhara V, Boutz DR, Person MD, Marcotte EM, Barrick JE, Wilke CO. Large-scale analysis of post-translational modifications in *E. coli* under glucose-limiting conditions. BMC Genomics. 2017;18:301. doi:10.1186/s12864-017-3676-8.

60. Hollebeke J, van Damme P, Gevaert K. N-terminal acetylation and other functions of Nα-acetyltransferases. Biol Chem. 2012;393:291–8. doi:10.1515/hsz-2011-0228.

61. Hu LI, Lima BP, Wolfe AJ. Bacterial protein acetylation: the dawning of a new age. Mol Microbiol. 2010;77:15–21. doi:10.1111/j.1365-2958.2010.07204.x.

62. Hwang C-S, Shemorry A, Varshavsky A. N-terminal acetylation of cellular proteins creates specific degradation signals. Science. 2010;327:973–7. doi:10.1126/science.1183147.

63. VanDrisse CM, Escalante-Semerena JC. Protein Acetylation in Bacteria. Annu Rev Microbiol. 2019;73:111–32. doi:10.1146/annurev-micro-020518-115526.

64. Drazic A, Myklebust LM, Ree R, Arnesen T. The world of protein acetylation. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics. 2016;1864:1372–401. doi:10.1016/j.bbapap.2016.06.007.

65. Bathke J, Konzer A, Remes B, McIntosh M, Klug G. Comparative analyses of the variation of the transcriptome and proteome of *Rhodobacter sphaeroides*

throughout growth. BMC Genomics. 2019;20:358. doi:10.1186/s12864-019-5749-3.

66. Picard F, Milhem H, Loubière P, Laurent B, Cocaign-Bousquet M, Girbal L. Bacterial translational regulations: high diversity between all mRNAs and major role in gene expression. BMC Genomics. 2012;13:528. doi:10.1186/1471-2164-13-528.

67. Greenbaum D, Colangelo C, Williams K, Gerstein M. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol. 2003;4:117. doi:10.1186/gb-2003-4-9-117.

68. Berghoff BA, Konzer A, Mank NN, Looso M, Rische T, Förstner KU, et al. Integrative "omics"-approach discovers dynamic and regulatory features of bacterial stress responses. PLoS Genet. 2013;9:e1003576. doi:10.1371/journal.pgen.1003576.

69. Maier T, Schmidt A, Güell M, Kühner S, Gavin A-C, Aebersold R, Serrano L. Quantification of mRNA and protein and integration with protein turnover in a bacterium. Mol Syst Biol. 2011;7:511. doi:10.1038/msb.2011.38.

70. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2012;13:227–32. doi:10.1038/nrg3185.

71. Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D, Brown PO. Precision and functional specificity in mRNA decay. Proc Natl Acad Sci U S A. 2002;99:5860–5. doi:10.1073/pnas.092538799.

72. Christiano R, Nagaraj N, Fröhlich F, Walther TC. Global proteome turnover analyses of the Yeasts *S. cerevisiae* and *S. pombe*. Cell Rep. 2014;9:1959–65. doi:10.1016/j.celrep.2014.10.065.

73. Stratmann A. Identifizierung eines Acarbose-Biosynthesegenclusters in *Actinoplanes* sp. und Charakterisierung ausgewählter Enzyme des Acarbose-Stoffwechsels.; 1997.

74. Wehmeier UF, Piepersberg W. Chapter 19 Enzymology of Aminoglycoside Biosynthesis—Deduction from Gene Clusters. In: Hopwood DA, editor. Complex enzymes in microbial natural product biosynthesis: Polyketides, aminocoumarins

and carbohydrates. London, San Diego, Calif: Academic; Academic Press/Elsevier; 2009. p. 459–491. doi:10.1016/S0076-6879(09)04619-9.

75. Hemker M, Stratmann A, Goeke K, Schröder W, Lenz J, Piepersberg W, Pape H. Identification, cloning, expression, and characterization of the extracellular acarbose-modifying glycosyltransferase, AcbD, from *Actinoplanes* sp. strain SE50. J Bacteriol. 2001;183:4484–92. doi:10.1128/JB.183.15.4484-4492.2001.

76. Leemhuis H, Wehmeier UF, Dijkhuizen L. Single amino acid mutations interchange the reaction specificities of cyclodextrin glycosyltransferase and the acarbose-modifying enzyme acarviosyl transferase. Biochemistry. 2004;43:13204–13. doi:10.1021/bi049015q.

77. Gren T. Development and application of genetic engineering methods for *Actinoplanes* sp. SE50/110; 2017.

78. Schwientek P, Szczepanowski R, Rückert C, Kalinowski J, Klein A, Selber K, et al. The complete genome sequence of the acarbose producer *Actinoplanes* sp. SE50/110. BMC Genomics. 2012;13:112. doi:10.1186/1471-2164-13-112.

79. Licht A, Bulut H, Scheffel F, Daumke O, Wehmeier UF, Saenger W, et al. Crystal structures of the bacterial solute receptor AcbH displaying an exclusive substrate preference for β-D-galactopyranose. J Mol Biol. 2011;406:92–105. doi:10.1016/j.jmb.2010.11.048.

80. Cumming BM, Chinta KC, Reddy VP, Steyn AJC. Role of Ergothioneine in Microbial Physiology and Pathogenesis. Antioxid Redox Signal. 2018;28:431–44. doi:10.1089/ars.2017.7300.

81. Schaffert L, Schneiker-Bekel S, Dymek S, Droste J, Persicke M, Busche T, et al. Essentiality of the Maltase AmlE in Maltose Utilization and Its Transcriptional Regulation by the Repressor AmlR in the Acarbose-Producing Bacterium *Actinoplanes* sp. SE50/110. Front. Microbiol. 2019;10:403. doi:10.3389/fmicb.2019.02448.

82. Schaffert L, Schneiker-Bekel S, Gierhake J, Droste J, Persicke M, Rosen W, et al. Absence of the highly expressed small carbohydrate binding protein Cgt improves the acarbose formation in *Actinoplanes* sp. SE50/110. Appl Microbiol Biotechnol 2020. doi:10.1007/s00253-020-10584-1.

83. Wendler S, Hürtgen D, Kalinowski J, Klein A, Niehaus K, Schulte F, et al. The cytosolic and extracellular proteomes of *Actinoplanes* sp. SE50/110 led to the identification of gene products involved in acarbose metabolism. J Biotechnol. 2013;167:178–89. doi:10.1016/j.jbiotec.2012.08.011.

84. Ortseifen V. Genombasierte Modellbildung zur Biosynthese von Acarviostatin-Metaboliten in drei *Actinoplanes* sp. SE50/110-Stämmen; 2016.

85. Schwientek P, Wendler S, Neshat A, Eirich C, Rückert C, Klein A, et al. Comparative RNA-sequencing of the acarbose producer *Actinoplanes* sp. SE50/110 cultivated in different growth media. J Biotechnol. 2013;167:166–77. doi:10.1016/j.jbiotec.2012.10.019.

86. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9. doi:10.1038/nmeth.1923.

87. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30. doi:10.1093/bioinformatics/btt656.

88. Hilker R, Stadermann KB, Schwengers O, Anisiforov E, Jaenicke S, Weisshaar B, et al. ReadXplorer 2-detailed read mapping analysis and visualization from one single source. Bioinformatics. 2016;32:3702–8. doi:10.1093/bioinformatics/btw541.

89. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. doi:10.1186/s13059-014-0550-8.

90. Athar A, Füllgrabe A, George N, Iqbal H, Huerta L, Ali A, et al. ArrayExpress update - from bulk to single-cell expression data. Nucleic Acids Res. 2019;47:D711-D715. doi:10.1093/nar/gky964.

91. Wang H, Qian W-J, Mottaz HM, Clauss TRW, Anderson DJ, Moore RJ, et al. Development and evaluation of a micro- and nanoscale proteomic sample preparation method. J Proteome Res. 2005;4:2397–403. doi:10.1021/pr050160f.

92. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods. 2016;13:731–40. doi:10.1038/nmeth.3901.

93. Deutsch EW, Csordas A, Sun Z, Jarnuczak A, Perez-Riverol Y, Ternent T, et al. The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. Nucleic Acids Res. 2017;45:D1100-D1106. doi:10.1093/nar/gkw936.

94. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. 2019;47:D442-D450. doi:10.1093/nar/gky1106.

95. Krzanowski WJ, Lai YT. A Criterion for Determining the Number of Groups in a Data Set Using Sum-of-Squares Clustering. Biometrics. 1988;44:23. doi:10.2307/2531893.

96. Albaum SP, Hahne H, Otto A, Haußmann U, Becher D, Poetsch A, et al. A guide through the computational analysis of isotope-labeled mass spectrometry-based quantitative proteomics data: an application study. Proteome Sci. 2011;9:30. doi:10.1186/1477-5956-9-30.

Droste, J.*, Wolf, T.*, Gren, T., Ortseifen, V., Schneiker-Bekel, S., Zemke, T., Pühler, A., Kalinowski, J., 2017. The MalR type regulator AcrC is a transcriptional repressor of acarbose biosynthetic genes in *Actinoplanes* sp. SE50/110. BMC genomics 18 (1), 562. DOI: 10.1186/s12864-017-3941-x [status: published] *Authors contributed equally to this study

Authors contributions:

Julian Droste and Dr. Timo Wolf designed, planed and interpreted the experimental work of this study. Dr. Tetiana Gren developed PCR targeting for *Actinoplanes* sp. SE50/110. Dr. Tetiana Gren, Dr. Vera Ortseifen and Dr. Timo Wolf constructed the deletion mutant. Julian Droste and Dr. Timo Wolf carried out the transcriptomic experiments. Julian Droste carried out all other experimental work and revised the manuscript. Dr. Timo Wolf drafted the manuscript. Prof. Dr. Jörn Kalinowski, Prof. Dr. Alfred Pühler, Dr. Susanne Schneiker-Bekel and Dr. Vera Ortseifen assisted in interpreting the data and revised the manuscript. Till Zemke advised the project and was involved in revising the manuscript. Prof. Dr. Jörn Kalinowski and Prof. Dr. Alfred Pühler coordinated this study. All authors read and approved the final manuscript.

Wolf et al. BMC Genomics (2017) 18:562
DOI 10.1186/s12864-017-3941-x

BMC Genomics

RESEARCH ARTICLE Open Access

The MalR type regulator AcrC is a transcriptional repressor of acarbose biosynthetic genes in *Actinoplanes* sp. SE50/110

Timo Wolf[1†], Julian Droste[1†], Tetiana Gren[1], Vera Ortseifen[2], Susanne Schneiker-Bekel[2], Till Zemke[3], Alfred Pühler[2] and Jörn Kalinowski[1*]

Abstract

Background: Acarbose is used in the treatment of diabetes mellitus type II and is produced by *Actinoplanes* sp. SE50/110. Although the biosynthesis of acarbose has been intensively studied, profound knowledge about transcription factors involved in acarbose biosynthesis and their binding sites has been missing until now. In contrast to acarbose biosynthetic gene clusters in *Streptomyces* spp., the corresponding gene cluster of *Actinoplanes* sp. SE50/110 lacks genes for transcriptional regulators.

Results: The acarbose regulator C (AcrC) was identified through an in silico approach by aligning the LacI family regulators of acarbose biosynthetic gene clusters in *Streptomyces* spp. with the *Actinoplanes* sp. SE50/110 genome. The gene for *acrC*, located in a head-to-head arrangement with the maltose/maltodextrin ABC transporter *malEFG* operon, was deleted by introducing PCR targeting for *Actinoplanes* sp. SE50/110. Characterization was carried out through cultivation experiments, genome-wide microarray hybridizations, and RT-qPCR as well as electrophoretic mobility shift assays for the elucidation of binding motifs. The results show that AcrC binds to the intergenic region between *acbE* and *acbD* in *Actinoplanes* sp. SE50/110 and acts as a transcriptional repressor on these genes. The transcriptomic profile of the wild type was reconstituted through a complementation of the deleted *acrC* gene. Additionally, regulatory sequence motifs for the binding of AcrC were identified in the intergenic region of *acbE* and *acbD*. It was shown that AcrC expression influences acarbose formation in the early growth phase. Interestingly, AcrC does not regulate the *malEFG* operon.

Conclusions: This study characterizes the first known transcription factor of the acarbose biosynthetic gene cluster in *Actinoplanes* sp. SE50/110. It therefore represents an important step for understanding the regulatory network of this organism. Based on this work, rational strain design for improving the biotechnological production of acarbose can now be implemented.

Keywords: *Actinoplanes*, Acarbose, MalR, AcrC, Transcriptional regulation, Actinomycetes

* Correspondence: Joern.Kalinowski@cebitec.uni-bielefeld.de
†Equal contributors
[1]Microbial Genomics and Biotechnology, Center for Biotechnology, Bielefeld University, Universitätsstraße 27, 33615 Bielefeld, Germany
Full list of author information is available at the end of the article

Background

Acarbose (acarviosyl-1,4-maltose) is used for the treatment of diabetes mellitus type II, as it supports the reduction of blood sugar levels, due to its inhibitory effect on alpha-glucosidases in the human intestine [1–3]. The Gram-positive actinobacterium *Actinoplanes* sp. SE50/110 is a natural producer of the pseudotetrasaccharide acarbose and the genome includes the acarbose biosynthetic (*acb*) gene cluster [4, 5]. Therefore, *Actinoplanes* sp. SE50 strains are used for the biotechnological production of acarbose [6]. *Actinoplanes* species are characterized by genomes with high G + C contents of 69–73%, can produce motile spores and typically grow in branched hyphae [7, 8].

Based on biochemical studies of the enzymes encoded by the *acb* gene cluster as well as genome-wide omics studies, models for the enzymatic pathways of acarbose biosynthesis have been proposed and targets for metabolic engineering have been suggested [3, 9–11]. However, functional studies concerning the acarbose biosynthesis based on genetic engineering of *Actinoplanes* sp. SE50/110 or rational strain designs have not been carried out until now. Recently, tools for genetic engineering of *Actinoplanes* sp. SE50 strains were developed [12, 13]. Combined with the high quality genome sequence and annotation of *Actinoplanes* sp. SE50/110 [14], targeted mutagenesis will facilitate the validation of acarbose biosynthesis and its regulation.

The transcriptional organization of the *acb* gene cluster, including transcription start sites, promoter elements and operon organization, was recently elucidated [14]. The cluster is divided into seven transcription units, with most of the genes coding for biosynthetic enzymes organized in one operon. The genes *acbZ*, *acbD* and *acbE* are transcribed monocistronically and encode proteins of the extracellular carbohydrate and acarbose metabolism. The genes *acbE* and *acbD* are located adjacently and oriented divergently [14]. However, profound knowledge about transcription factors involved in acarbose biosynthesis and their binding sites is missing until now. In contrast to acarbose biosynthetic gene clusters in *Streptomyces* spp. [15, 16], the *acb* gene cluster in *Actinoplanes* sp. SE50/110 lacks genes coding for transcription factors.

Nevertheless, it is known that expression of the genes *acbD* and *acbE* is inducible by maltotriose, when expressed heterologously in *Streptomyces lividans* [4]. It was suggested that dyadic symmetry element boxes (DSE) in the intergenic regions of the oppositely oriented genes *acbA* and *acbB* as well as *acbE* and *acbD,* might be possible operator sites for carbohydrate dependent transcriptional regulators [3]. Similar DSE boxes associated with maltose/maltotriose induction and glucose repression were identified upstream of alpha-amylase genes in several *Streptomyces* spp. [17, 18].

In this study, we expanded the toolbox for genetic engineering of *Actinoplanes* sp. SE50/110 through the successful application of PCR targeting ("ReDirect" technology), and applied this technology for the functional characterization of the MalR type transcription factor acarbose regulator C (AcrC). The rationale for classifying this transcription factor as a regulator of *acb* genes is shown by an in silico approach, cultivation experiments, transcriptomics as well as electrophoretic mobility shift assays for the elucidation of its DNA-binding motifs.

Results

In silico analysis for the identification of a transcriptional regulator of the acarbose biosynthetic gene cluster and construction of a deletion mutant

Recently, the transcriptional organization of the acarbose biosynthetic gene cluster (*acb* gene cluster), including transcription start sites, promoter elements and operon organization was elucidated [14]. However, profound knowledge about transcription factors involved in acarbose biosynthesis and their binding sites was missing until now. The *acb* gene cluster in *Actinoplanes* sp. SE50/110 lacks genes coding for transcriptional regulators. Interestingly, two other gene clusters for the production of acarviostatins have been identified in *Streptomyces* spp.. These are the *gac* gene cluster from *Streptomyces glaucescens* GLA.O [15, 19] and the *sct* gene cluster from *Streptomyces coelicoflavus* ZG0656 [16], which each include two LacI-type regulators (*garC1*, *garC2*, and *scarC1*, *scarC2*, respectively). When using protein alignment tools such as BLASTP [20] with the protein sequences of these regulators as an input and the *Actinoplanes* sp. SE50/110 genome for searching, the LacI family transcriptional regulator ACSP50_6387 was the best hit in all four cases. The pairwise identity of the regulators GarC1 and GarC2 from *S. glaucescens* GLA.O and ScarC1 as well as ScarC2 from *S. coelicoflavus* ZG0656 with ACSP50_6387 was between 59.7 and 63.4%, as determined through alignments using MUSCLE [21] (Fig. 1). These observations lead to the conclusion that ACSP50_6387 is a possible transcriptional regulator of the *acb* gene cluster. The *ACSP50_6387* gene was originally named *malR* and is located head to head to the maltose/maltodextrin ABC transporter gene cluster *malEFG* [11]. As this regulator also shows high similarities to MalR regulators, binding to the upstream region of the *malEFG* operon in other *Actinobacteria*, it was assumed that this regulator has a similar function in *Actinoplanes* sp. SE50/110 [22, 23]. In this study, it was shown that the LacI family regulator ACSP50_6387 is not the repressor of the *malEFG* operon, but is the first identified transcriptional regulator of the *acb* gene cluster, which is why it was named acarbose regulator C (AcrC). Conclusive evidence for this is given in the following.

A deletion mutant of the MalR-type regulator gene *acrC* was constructed using PCR targeting [24]. For this technology, also called "ReDirect" technology, a cosmid

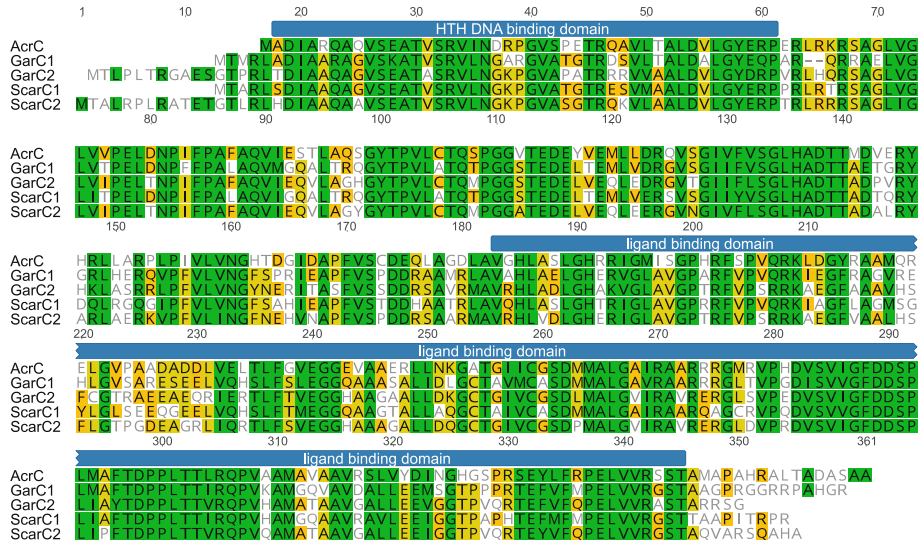

Fig. 1 AcrC was identified through alignment with transcriptional regulators from acarbose biosynthetic gene clusters of streptomycetes. The protein alignment of AcrC from *Actinoplanes* sp. SE50/110, GarC1 and GarC2 from *S. glaucescens* as well as ScarC1 and ScarC2 from *S. coelicoflavus* is shown. The protein domains were determined with Pfam [62] and refer to the exact amino acid positions of AcrC. The alignment was performed with MUSCLE [21] in Geneious [63]

containing the chromosomal region surrounding *acrC* and the *malEFG* operon was modified by applying λ RED-mediated recombination [25]. The complete coding region of *acrC* was replaced with the selection marker *aac(3)IV*, conferring apramycin resistance and an *oriT* (RK2) for conjugational transfer of the cosmid. The gene disruption of *acrC* in *Actinoplanes* sp. SE50/110 was verified by PCR on isolated DNA and by sequencing of the PCR products. These results proved the successful application of the so-called "ReDirect" technology in *Actinoplanes* sp. SE50/110 for the first time.

Establishment of whole genome microarrays for *Actinoplanes* sp. SE50/110 and application on a Δ*acrC* deletion mutant

In order to characterize the transcriptional regulator AcrC, comparative genome wide transcriptome analyses were conducted. Therefore, the wild type *Actinoplanes* sp. SE50/110 and the mutant Δ*acrC* were each cultivated in triplicates in minimal medium supplemented with maltose or glucose as single carbon source. Maltose minimal medium was used, as it is known as an acarbose production medium [26]. It was assumed that maltose or a metabolic product of maltose is an effector of AcrC, due to its similarity to MalR-like regulators. Therefore, maltose or a derivative might prevent the repressor AcrC from binding to its operator sites and consequently might lower the effect of a deletion mutant on the transcript levels of relevant genes. To better analyze the effect of the deletion mutant Δ*acrC* on the transcriptome, glucose minimal medium was used in parallel.

RNA samples from the biological replicates were taken in the middle of the growth phase of both strains in each maltose and glucose minimal medium, respectively. RNA was isolated and the three replicates were combined for each strain and condition. Subsequently, whole genome microarrays were used to identify genes regulated by AcrC. Agilent oligonucleotide microarrays were constructed, consisting of a total of 43,803 features and representing 8238 genes of *Actinoplanes* sp. SE50/110. Furthermore, the arrays contained 1417 control spots. The standard protocol for microarray hybridization was adapted due to the high G + C content of *Actinoplanes* sp. SE50/110. Additionally, the technical variance was determined in a "yellow experiment" (data not shown). The $\log_2$(fold change) cut-off (M-value) for a false discovery rate of 0.01 was determined as 1.1 and −1.1, respectively.

Whole transcriptome analysisallowed the identification of several genes for which different transcript abundances were measured when comparing the mutant Δ*acrC* with the *Actinoplanes* sp. SE50/110 wild type (Fig. 2). For each cultivation condition, the data from two arrays (dye swap) were combined to make statistically reliable conclusions. When using the RNA from the strains grown in maltose minimal medium, 23 genes with a $\log_2$(fold change) greater than 1.1 were determined indicating significantly higher transcript levels of these genes in the mutant (t-test $p < 0.05$). For 54 genes, an $\log_2$(fold change) less than −1.1 was determined and thus the transcript abundances were significantly lower in the mutant (t-test $p < 0.05$, Fig. 2a). In glucose minimal medium, the $\log_2$(fold change) was above 1.1 for 73

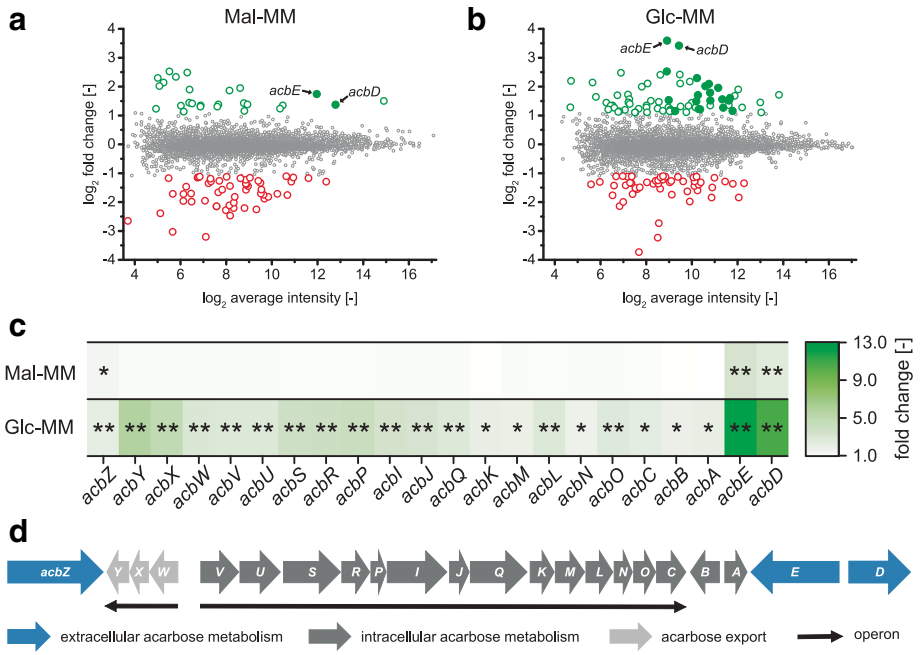

Fig. 2 Differential transcriptional analysis of the deletion mutant Δ*acrC* compared to the wild type. **a** Ratio/intensity plot from whole genome microarrays of the strain *Actinoplanes* sp. SE50/110 Δ*acrC* compared to the *Actinoplanes* sp. SE50/110 wild type grown in maltose minimal medium (Mal-MM). Green and red dots represent genes with significantly different transcript levels in the Δ*acrC* strain. Filled dots show *acb* genes. **b** Ratio/intensity plot from whole genome microarrays of the strain Δ*acrC* compared to the wild type grown in glucose minimal medium (Glc-MM). **c** Heatmap of the fold change of transcript abundance for the genes of the *acb* gene cluster, derived from the microarray data shown in 2A and 2B. Significance of $p < 0.05$ is marked with a single asterisk, significance of $p < 0.01$ with two asterisks (t-test, two-sample, Holm). **d** Transcriptional organization of the *acb* gene cluster with protein localizations depicted by coloring

genes and below −1.1 (t-test $p < 0.05$) for 51 genes, when comparing the strain Δ*acrC* to the wild type (Fig. 2b). This data provides the first evidence for genes transcriptionally regulated by AcrC (full list of genes with significantly differential transcript abundancies in Additional file 1).

In total, significantly higher transcript amounts were detected for seven genes in the strain Δ*acrC* in both maltose and glucose minimal medium. Among them were uncharacterized (ACSP50_2985 and ACSP50_6701) and hypothetical proteins (ACSP50_6700), a predicted extracellular protein with unknown function (ACSP50_6253) and the gene *dapE2*, putatively coding for a succinyl-diaminopimelate desuccinylase. The *dapE2* gene is highly similar to the *dapE1* gene, but since the latter is located together with *dapC* in the *Actinoplanes* sp. SE 50/110 genome it is a possible paralog. *DapE2* is located downstream of *acrC*, which is why polar effects through the replacement of *acrC* with the highly transcribed apramycin resistance cassette cannot be ruled out. Apart from the gene *acrC* itself, only two additional genes were identified with significantly reduced transcript amounts in the Δ*acrC* strain in both maltose and glucose minimal medium. These included *ACSP50_2217*, coding for a NADPH:quinone reductase and *ACSP50_4307*, coding for an oxidoreductase.

Most striking when analyzing the genes with significantly different transcript amounts in both cultivation conditions, were two of the genes of the *acb* gene cluster. For *acbE* (fold change of 3.4 in maltose, 12.1 in glucose medium) and *acbD* (fold change of 2.6 in maltose, 10.7 in glucose medium) significantly elevated transcript levels were measured in the strain Δ*acrC* (Fig. 2). In glucose minimal medium, these represented the genes with the overall largest differences in the transcript amount. *acbE* and *acbD* are genes encoding proteins of the extracellular acarbose metabolism [27]. For the other *acb* genes, which code for proteins of the acarbose biosynthesis or the export of acarbose, no significant differences in RNA amounts were measured in maltose minimal medium. However, in glucose minimal medium an increased transcript level was detected for all *acb* genes in the transcription factor knockout strain (Fig. 2c). For *acbM*, *acbN* and *acbB*, the fold change was just below the cut-off of 2.1 (M-value 1.1) but above 1.9. For the remaining *acb* genes, the fold changes were between 2.2 and 5.7.

Strikingly, no significant differences in the transcript abundance for genes of the operon *malEFG* were measured with the microarrays. This is surprising, as the gene for AcrC is located in direct proximity to this operon on the opposite DNA strand. To validate this unexpected

result, reverse transcription quantitative PCR (RT-qPCR) measurements were performed with RNA from cultivations in different carbon sources (data shown in Additional file 2). This way it was also possible to rule out that the lack of differences in the transcript levels for *malEFG* originate from maltose being the effector molecule and glucose acting through carbon catabolite repression. When comparing the strain Δ*acrC* with the wild type, no differences in the transcript amounts of *malE* could be detected with glucose, maltose, a mixture of glucose and maltose, glycerol, or mannitol as carbon source. However, with all tested carbon sources the transcript amounts of *acbE* were elevated in the Δ*acrC* strain compared to the wild type. The observations described here, are the first indications, that AcrC is a repressor of at least two *acb* genes and does not regulate the *malEFG* operon.

The transcription of the genes *acbD* and *acbE* is regulated by the repressor AcrC

A complementation of *acrC* in the deletion mutant Δ*acrC* was conducted to rule out polar effects of the gene replacement and to prove that the transcriptomic as well as phenotypic effects of the Δ*acrC* mutant can be attributed to the repressor effects of the transcriptional regulator. For the complementation of *acrC* in the deletion mutant Δ*acrC*, the ϕC31-based integrative vector pSET152 was used, for which the integration site in *Actinoplanes* sp. SE50/110 is known [12].

The complementation of *acrC* and the effect on the transcription of the genes *malE, acbD* and *acbE* was analyzed through RT-qPCR (Fig. 3). Therefore, RNA isolated from the middle of the growth phase of strains grown in glucose minimal medium was used. The transcript levels of the single genes in the Δ*acrC* deletion strain as well as the complementation strain, were compared to the levels of the wild type. The complementation of *acrC* was validated, as only a slightly reduced relative transcript amount compared to the wild type was measured (fold change 0.45), but no transcripts were detected in the Δ*acrC* deletion strain. The results of the RT-qPCR analysis for the *malE* gene are in line with the data from the microarray, confirming that the transcription of *malE* is not influenced by AcrC. The relative transcript amounts for the genes *acbD* and *acbE* in the deletion strain Δ*acrC* were significantly elevated compared to the wild type strain and therefore validated the results of the microarrays (fold change 39.5 for *acbE* and 63.3 for *acbD*). In the complementation strain, the transcript amounts for these genes were only moderately elevated, showing the nearly successful reconstitution of the transcriptomic profile of the wild type (fold change 1.5 for *acbE* and 5.1 for *acbD*). It should be noted that the transcription of the genes *acbD* and *acbE* is highly regulated during growth and dependent on the growth phase of *Actinoplanes* sp. SE50/110 (our unpublished results). This can have a strong impact on the variance of biological replicates.

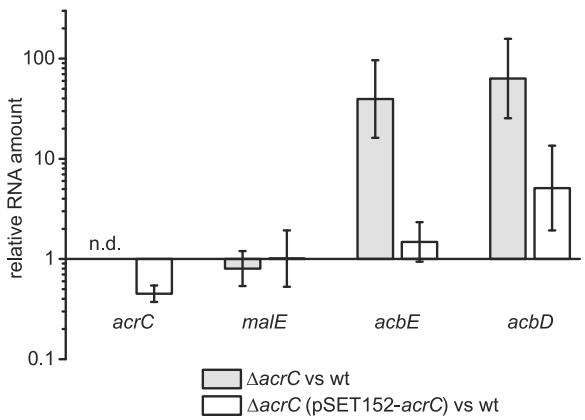

Fig. 3 Relative RNA amounts of single genes in the deletion and complementation strain compared to the wild type. Relative transcript abundances of the deletion strain *Actinoplanes* sp. SE50/110 Δ*acrC* and the complementation strain *Actinoplanes* sp. SE50/110 Δ*acrC* (pSET152-*acrC*) were compared with the wild type *Actinoplanes* sp. SE50/110 (wt). The means and standard derivations of three biological replicates are shown. RNA was isolated from the growth phase of shake flask cultivations in glucose minimal medium and analyzed by RT-qPCR

AcrC has an effect on the acarbose production

Comparative cultivations of the *Actinoplanes* sp. SE50/110 wild type, the mutant Δ*acrC* and the complementation strain Δ*acrC* (pSET152-*acrC*) were carried out to examine differences in growth and acarbose production. When comparing the three strains with respect to the cell dry weight, no significant differences were detected in growth behavior (Fig. 4). The production of different acarviose metabolites by *Actinoplanes* sp. SE50/110 is dependent on the available carbon source. When supplying glucose as carbon source, mainly acarviosyl-glucose is formed, which is why no production of acarbose is expected under these conditions [26]. Therefore, the acarbose concentration was determined solely for the cultivation in minimal medium with maltose, since acarviosyl-maltose (acarbose) is formed under these conditions [6, 26].

For cultivations of *Actinoplanes* sp. SE50/110 in maltose minimal medium in shake flasks, an acarbose concentration of up to 0.98 g L^{-1} for the wild type, 0.93 g L^{-1} for the deletion mutant and 0.75 g L^{-1} for the complementation strain was achieved. This corresponds to the expected product titer between 0.7 g L^{-1} and 1.0 g L^{-1} described in the literature for these conditions [3, 26]. In the early growth phase of the cultivations, a maximum of the specific product formation rate was obtained for all strains (Fig. 4). This shows that acarbose is produced during growth and not in the stationary phase and confirms the hypothesis of biomass-associated acarbose production of *Actinoplanes* sp. SE50/110 [26]. However, the strains

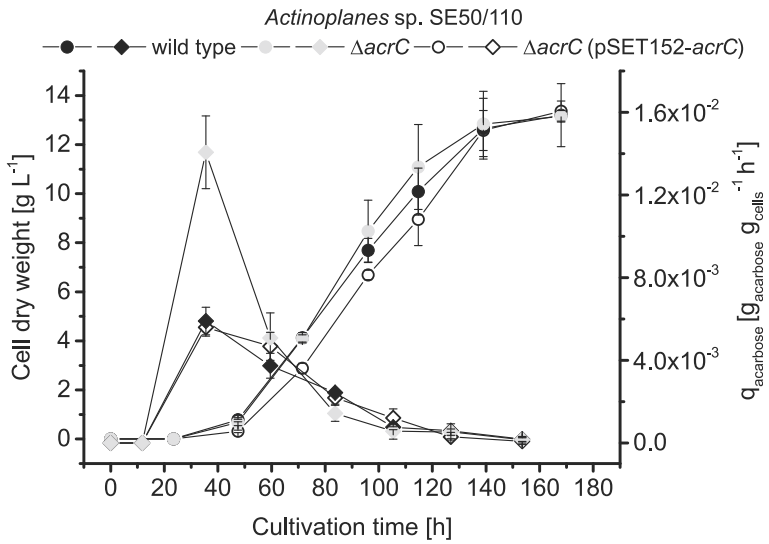

Fig. 4 Growth and product formation of the wild type, deletion and complementation stain. Cell dry weight (*circles*) and specific product formation rates ($q_{Acarbose}$, *diamonds*) of the *Actinoplanes* sp. SE50/110 wild type, the deletion strain Δ*acrC* and the complementation Δ*acrC* (pSET152-*acrC*). Samples were taken from shake flask cultivation in maltose minimal medium inoculated with spores. The means and standard derivations of five biological and two technical replicates are shown

differed with respect to the specific product formation level, defined as produced acarbose normalized to the mean cell dry weight and cultivation time. A significantly higher maximal specific product formation rate was achieved after 47.5 h in the *ΔacrC* strain ($1.4 \times 10^{-2} \pm 0.2 \times 10^{-2}$ h^{-1}) compared to the wild type ($5.9 \times 10^{-3} \pm 0.7 \times 10^{-3}$ h^{-1}) and the complementation strain ($5.5 \times 10^{-3} \pm 0.4 \times 10^{-3}$ h^{-1}). Thus, there is an effect of AcrC expression on the product formation of acarbose in the early growth phase.

The intergenic region between *acbE* and *acbD* features a binding site for AcrC

For the identification of precise binding sites of AcrC, band shift assays were carried out. Therefore, the AcrC protein was expressed in *Streptomyces lividans* TK23 and purified through a C-terminal hexa-histidine tag. The successful expression and purification of AcrC-His_6 was verified by SDS page and a tryptic peptide fingerprint analysis using MALDI-ToF-MS/MS (Data not shown). Electrophoretic mobility shift assays (EMSA) were carried out with the purified protein and Cy3 labeled PCR fragments.

When using the intergenic region of *acbE* and *acbD* as well as the intergenic region of *malE*/*acrC* together with AcrC, a retardation of the DNA was observed. No band-shift was detected when using the upstream region of *dapE2* as a control (Fig. 5a). Therefore, AcrC binds to the promoter regions of *acbE* and *acbD* as well as of *acrC* itself but not to the promoter region of *dapE2*.

An analysis of the intergenic region of *acbD* and *acbE* revealed two potential DNA binding sites with inverted repeat sequences, which are typical for the specific binding of transcriptional regulators [28, 29]. Upstream of the translation start of *acbE*, the motif 5′-CTTGCTG-3 bp-TAGCAAG-3′ (O1) is found at a distance of 60 bp. The TSS of *acbE* is located 40 bp downstream of this palindromic motif. A secondary TSS of *acbD* is located 21 bp upstream of this motif. Upstream of the start codon of *acbD* (50 bp) the motif 5′-CTGCAAG-2 bp-CTTGCAG-3′ (O2) can be identified. The primary TSS of *acbD* can be found 15 bp upstream of this motif (Fig. 5b). A similar inverted repeat motif is also located in the intergenic region of *malE* and *acrC,* but with a weaker consensus sequence in the second repeat of 5′-CTTGCAG-3 bp -ATGGAAG-3′. The repeat is found downstream of two *acrC* TSS as well as upstream of one *malE* and a third *acrC* TSS (Fig. 5c). When unlabeled double-stranded oligonucleotides covering only these motifs were added to the EMSAs as competitive DNA in excess amounts, the binding of AcrC to the DNA was reversed (Fig. 5a). A complete displacement was observed starting at a 50 fold molar excess of the double-stranded displacement oligonucleotide over the labeled PCR fragment. When using a 25 fold excess, the displacement was partial (data not shown). This is a proof that the identified DNA regions are required AcrC binding. The identified motifs were used to build a position weight matrix and the *Actinoplanes* sp. SE50/110 genome was scanned for additional motifs. However, the motif was not identified upstream of other genes with significantly different transcript amounts when comparing the Δ*acrC* with the wild type through microarrays.

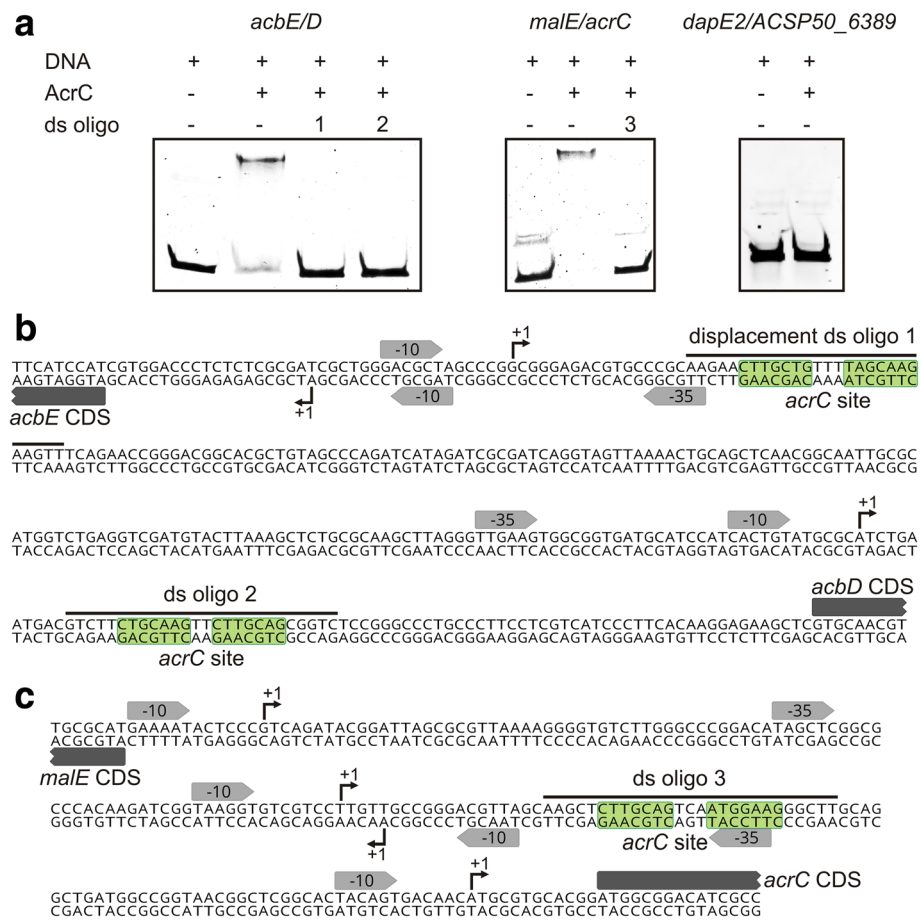

Fig. 5 Electrophoretic mobility shift assays with AcrC protein and the intergenic region of *acbE* and *acbD*. **a** EMSAs with the 342 bp fragment of the intergenic region of *acbE/acbD*, the 217 bp intergenic region *malE/acrC* as well as the 203 bp region *dapE/ACSP50_6389*. 0.05 pmol Cy3 labeled PCR fragments were incubated with 80 pmol purified AcrC protein, 0.05 μg herring sperm DNA for blocking of unspecific binding, and 100 mg BSA. 12.5 pmol unlabeled double-stranded oligonucleotides (ds oligo) covering the *acrC* site plus 5 bp up- and downstream were added as indicated. Separation was carried out with 10% native polyacrylamide (TBE) gels and visualized by fluorescence imaging. **b** Intergenic region of *acbE* and *acbD* used for the EMSAs with the promoter motives described in [14] and the *acrC* binding sites. **c** Intergenic region of *malE* and *acrC* used for the EMSA with promoter motives

To identify a possible effector of AcrC, which interacts with the ligand-binding domain and causes its dissociation from the DNA-binding site by a conformational change, different sugars were added to the protein-DNA mix. However, a retardation of the DNA was still observed, when adding glucose, galactose, maltose, maltotriose or acarbose in a range of 1 to 20 mM (data not shown). Therefore, an effector could not be identified yet.

Discussion

Genetic engineering technologies and whole genome microarrays were established to characterize the transcription factor AcrC

The transcription factor AcrC was identified through an in silico approach by comparing the regulators of the acarbose biosynthetic gene clusters from *S. glaucescens* GLA.O [15] and *S. coelicoflavus* ZG0656 [16] with the genome of *Actinoplanes* sp. SE50/110. AcrC is a member of the LacI/GalR family of transcriptional regulators, which is mainly composed of repressor proteins of genes involved in carbohydrate and nucleotide metabolism [30, 31].

After the in silico identification of AcrC as a possible transcription factor of the *acb* gene cluster, methods for creating deletion mutants as well as a cost effective genome wide transcriptomics method with a relatively fast data evaluation pipeline were needed. When the work on AcrC was conducted, both elements were missing for *Actinoplanes* sp. SE50/110, and therefore PCR targeting ("ReDirect" technology) and genome wide microarrays were established for this organism.

The gene of *acrC* was replaced with an apramycin resistance cassette by applying PCR targeting [24], which proved the successful application of this technology in *Actinoplanes* sp. SE50/110 for the first time. This expands the toolbox for genetic engineering of *Actinoplanes* sp. SE50/110 additionally to the application of integrative

vectors [12] and the meanwhile adapted genome editing using CRISPR/Cas9 [13].

The applicatiin of microarrays and RT-qPCR showned that the transcript levels of the genes *acbE* and *acbD* were elevated in the Δ*acrC* strain. This effect, caused by the deletion of *acrC,* was reversed by a complementation of *acrC*, confirming the successful reconstitution of the transcriptomic profile of the wild type. Although a clear effect of the complementation was shown, the transcript amount of *acrC* was only half as large as the transcript amount of the wild type, possibly resulting in slightly increased transcript amounts for *acbE* and *acbD* in the complementation strain compared to the wild type. An explanation for this could be possible polar effects at the integration site on the transcription of *acrC*. Another reason for the slight variances between the transcript levels of the wild type and Δ*acrC* strain could be that the transcription of the genes *acbD* and *acbE* is highly regulated during growth and dependent on the growth phase of *Actinoplanes* sp. SE50/110. This can lead to variances on the transcript levels between the strains, as it was observed for the comparison of the relative RNA amount of *acbD* in the complementation strain with the wild type.

Polar effects on neighboring genes were also observed through the replacement of *acrC* with the highly transcribed antibiotic resistance cassette. The gene *dapE2*, located directly downstream of *acrC*, is transcribed stronger in the strain Δ*acrC* and this effect was not reversed through the complementation (data not shown). Additionally it was shown that AcrC does not bind to the upstream region of *dapE2*, leading to the conclusion that the increased transcription of this gene in the mutant strain is caused by polar effects. Such effects on neighboring genes are unavoidable when applying PCR-targeting by replacing a target gene with a resistance marker cassette [32, 33]. This method can be expanded and improved by removing the antibiotic resistance cassette through site-specific recombination systems [34, 35], but this has not yet been applied to *Actinoplanes* sp. SE50/110. Alternatively, the recently established CRISPR/Cas9 technology enables scar-free and resistance marker-free deletions in the genome of *Actinoplanes* sp. SE50/110 with a single conjugation [13].

Well-functioning and reliable genetic engineering technologies in combination with fast and easily applicable whole genome transcriptomic methods will be indispensable for the clarification of regulatory networks in *Actinoplanes* sp. SE50/110. Although RNA-Seq has several advantages over microarrays, such as its single-nucleotide resolution and a much greater (log-linear) dynamic range [36, 37], the latter still have a legitimacy, as they can be used to simultaneously screen multiple samples in a cost-effective manner. The genome of *Actinoplanes* sp. SE50/110 harbors about 500 genes, which contain predicted DNA binding domains and might function as transcriptional regulators, of which now only the first one is functionally characterized. The methods established here will be helpful for the screening of many more transcription factors and understanding their biological functions. This knowledge will be of high value for metabolic engineering of this biotechnologically important organism.

AcrC is the missing repressor of the acarbose biosynthetic gene cluster

When comparing the whole transcriptome of the deletion mutant Δ*acrC* with the *Actinoplanes* sp. SE50/110 wild type, it was noticeable that no significant differences in the transcript abundance for genes of the maltose/maltodextrin ABC transporter operon *malEFG* were detected. This was not expected, as the gene coding for AcrC is located adjacently and divergently oriented to *malEFG*. Furthermore, its function was predicted as a MalR-type regulator and AcrC shows high similarities to MalR regulators, acting as a repressor of the *malEFG* operon in other *Actinobacteria* [22, 23]. The deletion of *malR* in *S. coelicolor* results in a glucose-insensitive transcription of *malE* [22, 38]. The transcriptional repressor MalR from *S. lividans* was shown to not only bind to regulatory sequences upstream of *malEFG*, but also to operator sites upstream of alpha-amylase genes [23, 39, 40]. Glucose repression of alpha-amylase genes mediated through LacI/GalR type transcription factors was also reported for other Gram-positive bacteria [41–43]. In this study, it was shown that AcrC binds to the intergenic region of *acbE* and *acbD* in *Actinoplanes* sp. SE50/110 and acts as a transcriptional repressor on these genes. AcbE is an acarbose-resistant alpha-amylase, which degrades starch and maltodextrins to maltose and maltotriose or higher malto-oligosaccharides [44]. The gene *acbD* encodes an acarviose transferase, which is proposed to catalyze the transfer of acarviosyl moieties from acarbose to the hydroxyl group of various sugars [45, 46]. The architecture of the catalytic site of AcbD is similar to other enzymes of the alpha-amylase family [44, 46]. Although the MalR type regulator AcrC of *Actinoplanes* sp. SE50/110 does not influence the transcription of the *malEFG* operon, it still binds upstream of similar genes as MalR does in *Streptomyces* spp.

Two binding sites for AcrC, each composed of a palindromic 7 bp repeat (5′-CTTGC(A/T)G-3′) where identified in the intergenic region of *acbE* and *acbD*. The regulatory motif resembles the core binding site of MalR in *S. lividans*, which is described as 5′-CTTGCAG-3′, occurring as an inverted and a direct repeat upstream of *malE* but downstream of the promoter site [23]. Additional motifs were identified upstream of amylase and

chitinase genes as direct or inverted repeats with a spacer of 3–15 bp [40]. In *Actinoplanes* sp. SE50/110 one of the operator sites is located downstream of the *acbD* transcription start sites and therefore blocks the RNA polymerase, but the other operator is located upstream of the promoter of *acbE*. However, the binding motif is located three base pairs upstream of the −35 region of this promoter, possibly acting by sterically blocking the RNA-polymerase from binding to the promoter. The close proximity of the two operator sites (182 bp) hints towards a possible tetrameric protein assembled of two homodimers, similar to the *E. coli* lactose repressor protein LacI [47, 48]. The repressor function of LacI is strengthened by DNA-looping with two operator sequences [49–51]. AcrC could form a similar structure, causing nearly the complete intergenic region between *acbE* and *acbD* to from a loop, thereby blocking all three promoters and increasing the repression effect. In *Actinoplanes* sp. SE50/110 the consensus-binding motif for AcrC also occurs as an inverted repeat with in the intergenic region between *malE* and *acrC*. Binding of AcrC to this region was shown with band shift assays. However, the potential binding site is located upstream of the *malE* TSS and downstream of two out of three *acrC* TSS. Together, with the observation that the transcription of *malE* is unchanged when deleting *acrC*, it can be assumed that only a transcriptional autoregulation of *acrC* occurs.

The consensus binding motif of AcrC was not identified upstream of the six additional genes with significantly different transcript amounts in both carbon source conditions. Although not consistently differentially transcribed in both conditions, transcriptional regulators were among the genes with significantly different transcript amounts in each condition. Therefore, indirect effects through changed metabolite concentrations or affected regulatory networks cannot be ruled out as cause for the differential transcript levels of these genes.

An effector molecule interacting with the ligand-binding domain of AcrC and thereby leading to a detachment of the repressor from the operator site was not detected through in vitro band shift assays. Nevertheless, the effect of the *acrC* deletion on the transcription of the *acb* genes, in particular *acbE* and *acbD*, was stronger in glucose containing medium compared to maltose minimal medium. This could indicate a detached repressor from the operator in maltose conditions. Combined, this could lead to the conclusion that maltose itself is not the effector of AcrC but a metabolic product directly derived from it. Maltodextrins can be built up intracellularly from maltose [52, 53] and are therefore promising candidates to be the effectors of AcrC, as it was also shown for MalR in *S. lividans* [23].

During the early growth phase, a significantly higher maximal specific product formation rate was achieved in the *ΔacrC* strain compared to the wild type and the complementation strain. Thus, there is a clear effect of AcrC expression on the acarbose formation in the early growth phase. This supports the assumption that AcrC is responsible for the repression of genes of the *acb* gene cluster in vivo, as the acarbose production is directly influenced by the deletion of the transcriptional regulator *acrC*. Based on literature and current models, acarbose is formed intracellularly and the extracellular proteins AcbE and AcbD are not directly involved in acarbose biosynthesis, when growing *Actinoplanes* sp. SE50/110 in maltose minimal medium [3, 26]. Therefore, a direct correlation of the transcription of the genes *acbE* and *acbD* with acarbose formation is not expected. It could be suspected that the gene products of *acbD* and *acbE* have additional enzymatic functions or that indirect effects such as feedback inhibition might influence acarbose formation.

The ABC transporter MalEFG was suggested as a possible acarbose-metabolite re-importer and AglEFG might be an additional maltose/maltodextrin importer [10, 54]. The proteins MalE, MalF and MalG were detected in high abundancies in both maltose and glucose-grown *Actinoplanes* sp. SE50/110 cultures [11]. This could lead to the conclusion that MalEFG imports acarviosyl metabolites independently from the available carbon source and could explain a possible evolutionary change of the AcrC regulon in *Actinoplanes* sp. SE50/110, dependent on the presence of the functional acarbose cluster. However, it could be beneficial to regulate the energy consuming expression and secretion of AcbE and AcbD, depending on the available carbon source. This function is implemented by AcrC in *Actinoplanes* sp. SE50/110 and could explain the special interaction between AcrC and the transcription of *acb* genes.

Conclusions

The identification of AcrC as a repressor of genes of the acarbose biosynthetic gene cluster is an important step towards understanding the transcriptional regulation of the acarbose biosynthesis. This study not only describes the first documented transcription factor of the *acb* gene cluster in *Actinoplanes* sp. SE50/110 but is also the first functional study of genetic engineering that influences acarbose production in the biotechnologically important rare actinomycete *Actinoplanes* sp. SE50/110. Genetic engineering technologies were developed and can be used in combination with the described microarrays as well as RNA-Seq, to further elucidate the complex regulatory network of *Actinoplanes* sp. SE50/110. Based on this work, rational strain design for the improvement of acarbose production can be carried out.

Methods

Strains, media and reagents

All standard cloning procedures were carried out with *Escherichia coli* DH5αMCR [55]. *E. coli* BW25113/pIJ790 [24] was used for λ RED recombineering of cosmids. *E. coli* ET12567/pUZ8002 [56] was used as a conjugation host for the target organism *Actinoplanes* sp.SE50/110 (ATCC 31044) to generate mutant strains (this study). *Streptomyces lividans* TK23 [57] was used for overexpression of AcrC.

Actinoplanes sp. SE50/110 was grown on soy flour medium agar (SFM; 20 g L^{-1} soy flour, 20 g L^{-1} mannitol, 20 g L^{-1} agar, pH 8, tap water) and in NBS medium for molecular cloning procedures as well as strain maintenance. Minimal medium was supplemented with 2.4 C-mole of the respective sugar as carbon source. The composition of the liquid media is described elsewhere [10]. When needed, chloramphenicol (25 μg mL^{-1}), kanamycin (50 μg mL^{-1}), apramycin (50 μg mL^{-1}) or hygromycin (100 μg/mL) was added to the media.

Soy flour (full fat) was used from Sobo Naturkost (Cologne, Germany) and purchased at a local store. For all PCRs, Phusion High-Fidelity PCR Master Mix with GC Buffer (NEB, Ipswich, MA, USA) was used. Gibson assembly master mix was prepared with Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA, USA), T5 Exonuclease (Epicentre, Madison, WI, USA) and TaqDNA Ligase (NEB, Ipswich, MA, USA).

Cultivation of *Actinoplanes* sp. SE50/110 and quantification of acarbose

For the cultivation of *Actinoplanes* sp. SE50/110 50 mL of medium were inoculated with 1 mL of spore suspension. Spores were harvested from freshly grown SFM agar plates with cultures grown for 6–7 days at 28 °C after uniformly plating 300 μL of a glycerol stock. Spores were washed off by adding 2 mL ddH_20 and carefully detaching them with a cotton swab. One plate resulted in roughly 1 mL spore suspension. The suspension of all plates for one strain was mixed before inoculation.

Cell dry weights were determined by harvesting 1 mL of cell suspension in weighed reaction tubes (20,000 *g*, 5 min). The supernatant was stored at −20 °C for acarbose quantification. The cell pellets were washed twice with ddH_20, dried at 70 °C for 48 h and weighed. For subsequent RNA isolation, 1 mL of cell suspension was centrifuged for 15 s at 16,000 *g* and immediately frozen in liquid nitrogen. Cell pellets were stored at −80 °C until further processed for RNA isolation.

Acarbose in the supernatant of *Actinoplanes* sp. SE50/110 cultivations was quantified by HPLC. Therefore, the supernatant was centrifuged (20,000 *g*, 2 min) to remove residual particles. Afterwards, 200 μL supernatant were mixed with 800 μL methanol, vortexed and centrifuged again (20,000 *g*, 2 min) to remove the resulting precipitate. The supernatant was transferred to HPLC vials and analyzed in a HPLC system (Finnigan Mat P4000 pump, AS3000 autosampler and UV6000LP detector, Thermo Fisher Scientific, Waltham, MA, USA). A flow of 1 mL min^{-1} of a mixture of 68% acetonitrile and 32% phosphate buffer (0.62 g L^{-1} KH_2PO_4 and 0.38 g L^{-1} $K_2HPO_4{\cdot}2H_2O$) was applied on a Hypersil APS-2 amino LC column (125 × 4 mm and 3 μm particle size, Thermo Fisher Scientific, Waltham, MA, USA) heated to 40 °C. The detection of acarbose was carried out with an UV detector at 210 nm. The acarbose concentration calculated with from the peak area and with a calibration curve.

Construction of *Actinoplanes sp.* SE50/110 mutants

The regulator gene *acrC* was disrupted from start to stop codon in *Actinoplanes* sp. SE50/110 by applying PCR targeting, also called ReDirect. The ReDirect protocol (version 1.4) was carried out as described in detail elsewhere [25]. All primers used in this study are listed in Additional file 3. The plasmid pIJ773 [24] (received from B. Ostash, Ivan Franko National University of Lviv, Ukraine was used as template for the disruption cassette containing an apramycin resistance (*aac(3)IV*) and an *oriT* (RK2). The chromosomal sequence of *acrC* on a pcc2FOS based fosmid, containing the genomic region 12,914 bp downstream to 24,255 bp upstream of *acrC*, was replaced with the disruption cassette. The chloramphenicol resistance cassette on the pcc2FOS vector was replaced by a hygromycin resistance gene (received from L. Horbal, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Germany) as a second selection marker. Conjugation of the cosmid was carried out as described previously [12]. After purification of exconjugants from *E. coli*, successful double-crossovers were verified by apramycin resistance and recovery of hygromycin sensitivity.

In order to complement the disrupted gene in the Δ*acrC* strain, a modified version of the integrative vector pSET152 [58] was used. The apramycin resistance gene of pSET152 was exchanged for a hygromycin resistance gene and the *acrC* gene including the 5′-UTR and promoter region (determined with data from [14]) was cloned in the multiple cloning site by isothermal Gibson assembly [59].

DNA of *Actinoplanes* sp. SE50/110 strains was isolated as described before [13]. PCR was used to confirm the constructed cosmids and plasmids as well as the genotype of all *Actinoplanes* sp. SE50/110 strains. PCR fragments were purified and Sanger sequencing was carried out by the in-house sequencing core facility.

Transcriptomic analyses

RNA isolation

For RNA isolation frozen cell pellets were suspended in 800 μL RLT buffer (RNeasy mini kit, Qiagen, Hilden,

Germany) and transferred to 2 mL lysing matrix tubes (0.1 mm spherical silica beads, MP Biomedicals, Santa Ana, California, USA). Cell disruption was carried out in a homogenizer (FastPrep FP120, Thermo Fisher Scientific, Waltham, MA, USA) for two times 20 s at speed setting 6.5 and 1 min on ice in between. Subsequently, the cell suspension was centrifuged for 3 min at 13,000 *g* and 4 °C. The supernatant was used for RNA extraction using a Qiagen RNeasy mini kit in combination with an RNase-free DNase kit (Qiagen, Hilden, Germany) for on-column and off-column DNA digestion. PCR with primers binding to genomic *Actinoplanes* sp. SE50/110 DNA was used to verify complete removal of residual DNA. Quality and quantity of the RNA was analyzed with a NanoDrop 1000 spectrometer (Peqlab, Erlangen, Germany) and an Agilent RNA 6000 Pico kit run on an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA).

Whole genome oligonucleotide microarrays

Custom whole genome oligonucleotide microarrays representing the coding sequence of *Actinoplanes* sp. SE50/110 were designed with eArray (Agilent Technologies, Santa Clara, CA, USA) and ordered in the 4x44K format (Agilent Technologies, Santa Clara, CA, USA). These consist of 43,803 features representing 8238 genes and 1417 control spots. All experimental procedures, including sample preparation, cDNA synthesis and labeling, microarray hybridization and washing as well as scanning and feature extraction, were carried out as described by the manufacturer. The kit Two-Color Microarray-Based Prokaryote Analysis FairPlay III Labeling (Version 1.4, Agilent Technologies, Santa Clara, CA, USA) was used with the following adjustments, which were optimized and tested in previous experiments. The quantities and volumes of the components of the hybridization samples were adjusted to fit the 4x44K array format. The mix was prepared with 330 ng of each labeled cDNA and 11 μL gene expression blocking agent. The cDNA blocking mix was filled up to 55 μL with H_2O and mixed with 55 μL Hi-RPM hybridization buffer. 100 μL of the hybridization mix were used for the hybridization of one array. Washing of the microarrays was carried out including stabilization and drying solution. The number and length of the washing steps was increased (two wash cycles, with 5 min wash buffer 1 and 1 min wash buffer 2) to reduce signal artifacts due to the high G + C content of *Actinoplanes* sp. SE50/110. Amersham CyDye mono-reactive dye packs were used from GE Healthcare (Little Chalfont, UK). All other microarray specific reagents as well as the hybridization oven and the microarray scanner were used from Agilent Technologies (Santa Clara, CA, USA).

Feature extraction was performed with the Agilent Feature Extraction Software Version 10.7.3.1 (Agilent Technologies, Santa Clara, CA, USA), applying the protocol GE2_107_Sep09. Subsequent data analysis, including LOWESS normalization and statistical analysis was performed with EMMA2 [60]. A *p*-value of 0.05 was used as a cut-off for significance and the M-value cut-offs for a false discovery rate of 0.01 were determined as 1.1 and −1.1, respectively.

Reverse transcription quantitative PCR

RT-qPCR was applied for relative mRNA quantification of single genes. Primers were designed to amplify 75 to 150 bps of intragenic regions (list of primers in Additional file 3). A SensiFast SYBR No-Rox One-Step Kit (Bioline, London, UK) and 96 well lightcycler plates (Sarstedt, Nümbrecht, Germany) were used for measurements in a LightCycler 96 System (Roche, Mannheim, Germany). 1 μL of template RNA, adjusted to 200 ng μL^{-1}, was mixed with 19 μL master mix containing 1 μL of specific primers (10 μM each), 0.2 μL reverse transcriptase, 0.4 μL RNase inhibitor, 10 μL reaction mix and 7.4 μL 5 M betain. A minimum of three biological replicates in each technical duplicates was included for every measurement. Two negative controls with 1 μL H_2O as template were included for each analyzed gene. Reverse transcription was performed at 45 °C for 20 min, followed by 2 min at 95 °C, a three step amplification (95 °C 5 s, 60 °C 10 s, 72 °C 10 s, 60 cycles) and a melting profile. The LightCycler 96 V1.1 software was used for inspection of control measurements and melting curve analysis. The relative RNA amount was normalized on total RNA (200 ng) and calculated as $2^{-\Delta Cq}$. ΔCq was calculated as the difference of the mean Cq in the mutant strain compared to the control strain.

Heterologous expression and purification of AcrC in *Streptomyces lividans*

For the heterologous expression and purification of the AcrC protein, the *acrC* gene was cloned by Gibson assembly [59] into the multiple cloning site of the pGM1202 expression vector (G. Muth, unpubl. Data, available through Addgene # 69615) which includes a pSG5 origin of replication [61], the P_{tipA} promoter and a C-terminal His_6-tag. The expression vector was transferred into *Streptomyces lividans* TK23 by conjugation. The strain was grown in 50 mL yeast extract-malt extract (YEME) medium with 50 μg mL^{-1} apramycin in a 250 mL flask at 28 °C and 180 rpm. After 3 days, 15 mL of the culture were transferred to 200 mL fresh YEME medium supplemented with 25 μg mL^{-1} thiostrepton to induce gene expression. The cells were cultivated for further 3 to 4 days at 28 °C and 180 rpm. Afterwards, the cells were harvested by centrifugation at 5000 *g* for 20 min at 4 °C. The pellet was resuspended in ice-cold lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8). Cell disruption was carried out with a French press for three times. Cell debris were separated

from the soluble fraction by centrifugation (5000 *g*, 1 h) at 4 °C. The protein was purified from the supernatant using Protino® Ni-TED 1000 Packed Columns as described by the manufacturer (Macherey-Nagel, Düren, Germany) and stored in 30 mM Tris-HCl, 300 mM NaCl, pH 8.5 buffer at 4 °C.

Electrophoretic mobility shift assays

DNA band shift assays were performed with Cy3-labeled PCR fragments and ds oligos for displacements (list of primers in Additional file 3). Cy3-labeled primers (Metabion, Steinkirchen, Germany) were used to produce PCR fragments, which were then purified by using a PCR Clean Up Kit (Macherey Nagel, Düren, Germany). The oligonucleotides were annealed by heating 5 min to 95 °C and then ramp to 4 °C at 0.1 °C s^{-1}.

The binding assay was performed in a final reaction volume of 20 μL containing 80 pmol His-tagged AcrC protein, 4 μL of 5× EMSA binding buffer (100 mM Na_2HPO_4, 375 mM KCl, 25% Glycerin, pH 8), 2.5 mM $MgCl_2$ and 0.1 mM EDTA. In addition, 0.05 μg of herring sperm DNA and 0.1 μg BSA (bovine serum albumin) was added to each reaction to block unspecific protein-DNA interactions. After incubation for 20 min at room temperature the samples were separated on a 10% native polyacrylamide gel (Biorad, Hercules, CA, USA) at 170 V using TBE (89 mM Tris base, 89 mM boric acid, 2 mM EDTA) as running buffer. The gel was scanned on a Typhoon 8600 Variable Mode Imager (GE Healthcare, Little Chalfont, UK).

Additional files

Additional file 1: List of genes with significantly differential transcript abundancies in the mutant strain Δ*acrC* compared to the wild type in maltose and glucose minimal medium. (XLSX 122 kb)

Additional file 2: Relative RNA amounts of *malE* and *acbE* in the deletion strain compared to the wild type in different carbon sources. (PDF 206 kb)

Additional file 3: List of primers used in this study. (PDF 150 kb)

Abbreviations

AcrC: Acarbose regulator C; ds oligo: Double-stranded oligonucleotides; DSE: Dyadic symmetry element; EMSA: Electrophoretic mobility shift assays; Glc-MM: Glucose minimal medium; HTH: Helix-turn-helix; Mal-MM: Maltose minimal medium; nAcb: Acarbose biosynthesis; RT-qPCR: Reverse transcription quantitative PCR; wt: Wild type

Acknowledgments

We thank Bohdan Ostash and Liliya Horbal for providing strains and plasmids and Robin Wördenweber for proofreading this manuscript. JD, TG, VO and TW acknowledge support from the CLIB-Graduate Cluster Industrial Biotechnology at Bielefeld University, Germany, which is supported by the Ministry of Innovation, Science and Research (MIWF) of the federal state North Rhine-Westphalia, Germany and Bielefeld University, Germany. We gratefully thank our cooperation partner Bayer AG (Leverkusen, Germany) for their continuous support and funding.

Funding

This study was funded by the Bayer AG and CLIB-Graduate Cluster Industrial Biotechnology. We acknowledge support of the publication fee by Deutsche Forschungsgemeinschaft and the Open Access Publication Funds of Bielefeld University.

Availability of data and materials

The microarray datasets generated during the current study are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-5459. All other datasets supporting the conclusions of this article are included within the article and its additional files.

Authors' contributions

JD and TW designed, planed and interpreted the experimental work of this study. TG developed PCR targeting for *Actinoplanes* sp. SE50/110. TG, VO and TW constructed the deletion mutant. JD and TW carried out the transcriptomic experiments. JD carried out all other experimental work and revised the manuscript. TW drafted the manuscript. JK, AP, SSB and VO assisted in interpreting the data and revised the manuscript. TZ advised the project and was involved in revising the manuscript. JK and AP coordinated this study. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Microbial Genomics and Biotechnology, Center for Biotechnology, Bielefeld University, Universitätsstraße 27, 33615 Bielefeld, Germany. [2]Senior Research Group in Genome Research of Industrial Microorganisms, Center for Biotechnology, Bielefeld University, Universitätsstraße 27, 33615 Bielefeld, Germany. [3]Product Supply, Bayer Pharma AG, Friedrich Ebert Str. 217-475, 42117 Wuppertal, Germany.

Received: 26 January 2017 Accepted: 13 July 2017
Published online: 25 July 2017

References

1. Creutzfeldt W. Acarbose for the Treatment of Diabetes Mellitus. Berlin, Heidelberg: Springer Berlin Heidelberg; 1988.
2. Bischoff H. Pharmacology of α-glucosidase inhibition. Eur J Clin Investig. 1994;24:3–10. doi:10.1111/j.1365-2362.1994.tb02249.x.
3. Wehmeier UF, Piepersberg W. Biotechnology and molecular biology of the α-glucosidase inhibitor acarbose. Appl Microbiol Biotechnol. 2004;63:613–25. doi:10.1007/s00253-003-1477-2.
4. Wehmeier UF. The biosynthesis and metabolism of Acarbose in *Actinoplanes* sp. SE 50/110: a progress report. Biocatal Biotransformation. 2003;21:279–84. doi:10.1080/10242412031000161438.
5. Schwientek P, Szczepanowski R, Ruckert C, Kalinowski J, Klein A, Selber K, et al. The complete genome sequence of the acarbose producer *Actinoplanes* sp. SE50/110. BMC Genomics. 2012;13:112. doi:10.1186/1471-2164-13-112.
6. Truscheit E, Frommer W, Junge B, Müller L, Schmidt DD, Wingender W. Chemistry and biochemistry of microbial α-Glucosidase inhibitors. Angew Chem Int Ed Engl. 1981;20:744–61. doi:10.1002/anie.198107441.
7. Parenti F, Coronelli C. Members of the genus *Actinoplanes* and their antibiotics. Annu Rev Microbiol. 1979;33:389–411. doi:10.1146/annurev.mi.33.100179.002133.
8. Vobis G, Schäfer J, Kämpfer P. Actinoplanes. In: Bergey DH, Whitman WB, Goodfellow M, Kämpfer P, Busse H-J, editors. Bergey's manual of systematic bacteriology. 2nd ed. New York: Springer; 2012. p. 1–41. doi:10.1002/9781118960608.gbm00139.
9. Zhang C-S, Stratmann A, Block O, Brückner R, Podeschwa M, Altenbach H-J, et al. Biosynthesis of the C(7)-cyclitol moiety of acarbose in *Actinoplanes* species SE50/110. 7-O-phosphorylation of the initial cyclitol precursor leads

to proposal of a new biosynthetic pathway. J Biol Chem. 2002;277:22853–62. doi:10.1074/jbc.M202375200.
10. Wendler S, Hurtgen D, Kalinowski J, Klein A, Niehaus K, Schulte F, et al. The cytosolic and extracellular proteomes of *Actinoplanes* sp. SE50/110 led to the identification of gene products involved in acarbose metabolism. J Biotechnol. 2013;167:178–89. doi:10.1016/j.jbiotec.2012.08.011.
11. Wendler S, Otto A, Ortseifen V, Bonn F, Neshat A, Schneiker-Bekel S, et al. Comparative proteome analysis of *Actinoplanes* sp. SE50/110 grown with maltose or glucose shows minor differences for acarbose biosynthesis proteins but major differences for saccharide transporters. J Proteome. 2016;131:140–8. doi:10.1016/j.jprot.2015.10.023.
12. Gren T, Ortseifen V, Wibberg D, Schneiker-Bekel S, Bednarz H, Niehaus K, et al. Genetic engineering in *Actinoplanes* sp. SE50/110 - development of an intergeneric conjugation system for the introduction of actinophage-based integrative vectors. J Biotechnol. 2016;232:79–88. doi:10.1016/j.jbiotec.2016.05.012.
13. Wolf T, Gren T, Thieme E, Wibberg D, Zemke T, Puhler A, Kalinowski J. Targeted genome editing in the rare actinomycete *Actinoplanes* sp. SE50/110 by using the CRISPR/Cas9 system. J Biotechnol. 2016;231:122–8. doi:10.1016/j.jbiotec.2016.05.039.
14. Wolf T, Schneiker-Bekel S, Neshat A, Ortseifen V, Wibberg D, Zemke T, et al. Genome improvement of the acarbose producer *Actinoplanes* sp. SE50/110 and annotation refinement based on RNA-seq analysis. J Biotechnol. 2017;251:112–23. doi:10.1016/j.jbiotec.2017.04.013.
15. Rockser Y, Wehmeier UF. The gac-gene cluster for the production of acarbose from *Streptomyces glaucescens* GLA.O—identification, isolation and characterization. J Biotechnol. 2009;140:114–23. doi:10.1016/j.jbiotec.2008.10.016.
16. Guo X, Geng P, Bai F, Bai G, Sun T, Li X, et al. Draft genome sequence of *Streptomyces coelicoflavus* ZG0656 reveals the putative biosynthetic gene cluster of acarviostatin family alpha-amylase inhibitors. Lett Appl Microbiol. 2012;55:162–9. doi:10.1111/j.1472-765X.2012.03274.x.
17. Virolle M-J, Long CM, Shing C, Bibb MJ. Cloning, characterisation and regulation of an α-amylase gene from *Streptomyces venezuelae*. Gene. 1988;74:321–34. doi:10.1016/0378-1119(88)90166-7.
18. Virolle MJ, Gagnat J. Sequences involved in growth-phase-dependent expression and glucose repression of a *Streptomyces* α-amylase gene. Microbiology. 1994;140(Pt 5):1059–67. doi:10.1099/13500872-140-5-1059.
19. Ortseifen V, Winkler A, Albersmeier A, Wendler S, Puhler A, Kalinowski J, Ruckert C. Complete genome sequence of the actinobacterium *Streptomyces glaucescens* GLA.O (DSM 40922) consisting of a linear chromosome and one linear plasmid. J Biotechnol. 2015;194:81–3. doi:10.1016/j.jbiotec.2014.11.036.
20. Altschul S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402. doi:10.1093/nar/25.17.3389.
21. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7. doi:10.1093/nar/gkh340.
22. van Wezel GP, White J, Young P, Postma PW, Bibb MJ. Substrate induction and glucose repression of maltose utilization by *Streptomyces coelicolor* A3(2) is controlled by malR, a member of the lacI-galR family of regulatory genes. Mol Microbiol. 1997;23:537–49. doi:10.1046/j.1365-2958.1997.d01-1878.x.
23. Schlösser A, Weber A, Schrempf H. Synthesis of the *Streptomyces lividans* maltodextrin ABC transporter depends on the presence of the regulator MalR. FEMS Microbiol Lett. 2001;196:77–83. doi:10.1111/j.1574-6968.2001.tb10544.x.
24. Gust B, Challis GL, Fowler K, Kieser T, Chater KF. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc Natl Acad Sci U S A. 2003;100:1541–6. doi:10.1073/pnas.0337542100.
25. Gust B, Kieser T, Chater K. REDIRECT technology: PCR targeting system in *Streptomyces coelicolor* A3(2) 2002. Norwich: John Innes Centre.
26. Wendler S, Ortseifen V, Persicke M, Klein A, Neshat A, Niehaus K, et al. Carbon source dependent biosynthesis of acarviose metabolites in *Actinoplanes* sp. SE50/110. J Biotechnol. 2014;191:113–20. doi:10.1016/j.jbiotec.2014.08.019.
27. Wendler S, Otto A, Ortseifen V, Bonn F, Neshat A, Schneiker-Bekel S, et al. Comprehensive proteome analysis of *Actinoplanes* sp. SE50/110 highlighting the location of proteins encoded by the acarbose and the pyochelin biosynthesis gene cluster. J Proteome. 2015;125:1–16. doi:10.1016/j.jprot.2015.04.013.
28. Rhodes D, Schwabe JW, Chapman L, Fairall L. Towards an understanding of protein-DNA recognition. Philos Trans R Soc Lond Ser B Biol Sci. 1996;351:501–9. doi:10.1098/rstb.1996.0048.
29. Huffman JL, Brennan RG. Prokaryotic transcription regulators: more than just the helix-turn-helix motif. Curr Opin Struct Biol. 2002;12:98–106. doi:10.1016/S0959-440X(02)00295-6.
30. Weickert MJ, Adhya S. A family of bacterial regulators homologous to gal and Lac repressors. J Biol Chem. 1992;267:15869–74.
31. Swint-Kruse L, Matthews KS. Allostery in the LacI/GalR family: variations on a theme. Curr Opin Microbiol. 2009;12:129–37. doi:10.1016/j.mib.2009.01.009.
32. Herrmann S, Siegl T, Luzhetska M, Petzke L, Jilg C, Welle E, et al. Site-specific recombination strategies for engineering actinomycete genomes. Appl Environ Microbiol. 2012;78:1804–12. doi:10.1128/AEM.06054-11.
33. Myronovskyi M, Luzhetskyy A. Genome engineering in actinomycetes using site-specific recombinases. Appl Microbiol Biotechnol. 2013;97:4701–12. doi:10.1007/s00253-013-4866-1.
34. Fedoryshyn M, Petzke L, Welle E, Bechthold A, Luzhetskyy A. Marker removal from actinomycetes genome using Flp recombinase. Gene. 2008;419:43–7. doi:10.1016/j.gene.2008.04.011.
35. Zelyas N, Tahlan K, Jensen SE. Use of the native *flp* gene to generate in-frame unmarked mutations in *Streptomyces* spp. Gene. 2009;443:48–54. doi:10.1016/j.gene.2009.03.022.
36. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5:621–8. doi:10.1038/nmeth.1226.
37. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10:57–63. doi:10.1038/nrg2484.
38. van Wezel GP, White J, Bibb MJ, Postma PW. The malEFG gene cluster of *Streptomyces coelicolor* A3(2): characterization, disruption and transcriptional analysis. Mol Gen Genet. 1997;254:604–8. doi:10.1007/s004380050458.
39. Nguyen J, Francou F, Virolle MJ, Guérineau M. Amylase and chitinase genes in *Streptomyces lividans* are regulated by reg1, a pleiotropic regulatory gene. J Bacteriol. 1997;179:6383–90. doi:10.1128/jb.179.20.6383-6390.1997.
40. Nguyen J. The regulatory protein Reg1 of *Streptomyces lividans* binds the promoter region of several genes repressed by glucose. FEMS Microbiol Lett. 1999;175:51–8. doi:10.1111/j.1574-6968.1999.tb13601.x.
41. Virolle M-J, Bibb MJ. Cloning, characterization and regulation of an α-amylase gene from *Streptomyces limosus*. Mol Microbiol. 1988;2:197–208. doi:10.1111/j.1365-2958.1988.tb00021.x.
42. Henkin TM, Grundy FJ, Nicholson WL, Chambliss GH. Catabolite repression of α-amylase gene expression in *Bacillus subtilis* involves a trans-acting gene product homologous to the *Escherichia coli lacI* and *galR* repressors. Mol Microbiol. 1991;5:575–84. doi:10.1111/j.1365-2958.1991.tb00728.x.
43. Afzal M, Shafeeq S, Manzoor I, Kuipers OP. Maltose-dependent transcriptional regulation of the mal Regulon by MalR in *Streptococcus pneumoniae*. PLoS One. 2015;10:e0127579. doi:10.1371/journal.pone.0127579.
44. Wehmeier UF, Piepersberg W. Chapter 19 Enzymology of Aminoglycoside Biosynthesis—Deduction from Gene Clusters. In: Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates: Elsevier; 2009. p. 459–91. doi:10.1016/S0076-6879(09)04619-9.
45. Hemker M, Stratmann A, Goeke K, Schroder W, Lenz J, Piepersberg W, Pape H. Identification, cloning, expression, and characterization of the extracellular acarbose-modifying glycosyltransferase, AcbD, from Actinoplanes sp. strain SE50. J Bacteriol. 2001;183:4484–92. doi:10.1128/JB.183.15.4484-4492.2001.
46. Leemhuis H, Wehmeier UF, Dijkhuizen L. Single amino acid mutations interchange the reaction specificities of Cyclodextrin Glycosyltransferase and the Acarbose-modifying enzyme Acarviosyl Transferase. Biochemistry. 2004;43:13204–13. doi:10.1021/bi049015q.
47. Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, et al. Crystal structure of the lactose Operon repressor and its complexes with DNA and inducer. Science. 1996;271:1247–54. doi:10.1126/science.271.5253.1247.
48. Matthews KS, Nichols JC. Lactose Repressor Protein: Functional Properties and Structure. In: Progress in nucleic acid research and molecular biology. New York: Academic Press; 1998. p. 127–64. doi:10.1016/S0079-6603(08)60035-5.
49. Oehler S, Eismann ER, Kramer H, Muller-Hill B. The three operators of the *lac* operon cooperate in repression. EMBO J. 1990;9:973–9.
50. Wong OK, Guthold M, Erie DA, Gelles J. Interconvertible lac repressor-DNA loops revealed by single-molecule experiments. PLoS Biol. 2008;6:e232. doi:10.1371/journal.pbio.0060232.

51. Rutkauskas D, Zhan H, Matthews KS, Pavone FS, Vanzi F. Tetramer opening in LacI-mediated DNA looping. Proc Natl Acad Sci U S A. 2009;106:16627–32. doi:10.1073/pnas.0904617106.
52. Boos W, Shuman H. Maltose/maltodextrin system of *Escherichia coli*: transport, metabolism, and regulation. Microbiol Mol Biol Rev. 1998;62:204–29.
53. Seibold GM, Wurst M, Eikmanns BJ. Roles of maltodextrin and glycogen phosphorylases in maltose utilization and glycogen metabolism in *Corynebacterium glutamicum*. Microbiology. 2009;155:347–58. doi:10.1099/mic.0.023614-0.
54. Brunkhorst C, Schneider E. Characterization of maltose and maltotriose transport in the acarbose-producing bacterium *Actinoplanes* sp. Res Microbiol. 2005;156:851–7. doi:10.1016/j.resmic.2005.03.008.
55. Grant SG, Jessee J, Bloom FR, Hanahan D. Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. Proc Natl Acad Sci U S A. 1990;87:4645–9. doi:10.1073/pnas.87.12.4645.
56. Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. Practical *Streptomyces* genetics. Norwich: John Innes Foundation; 2000.
57. Hopwood DA, Hintermann G, Kieser T, Wright HM. Integrated DNA sequences in three streptomycetes form related autonomous plasmids after transfer to *Streptomyces lividans*. Plasmid. 1984;11:1–16. doi:10.1016/0147-619X(84)90002-7.
58. Bierman M, Logan R, O'Brien K, Seno ET, Nagaraja Rao R, Schoner BE. Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. Gene. 1992;116:43–9. doi:10.1016/0378-1119(92)90627-2.
59. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009;6:343–5. doi:10.1038/nmeth.1318.
60. Dondrup M, Albaum SP, Griebel T, Henckel K, Junemann S, Kahlke T, et al. EMMA 2–a MAGE-compliant system for the collaborative analysis and integration of microarray data. BMC Bioinformatics. 2009;10:50. doi:10.1186/1471-2105-10-50.
61. Muth G, Nußbaumer B, Wohlleben W, Pühler A. A vector system with temperature-sensitive replication for gene disruption and mutational cloning in streptomycetes. Mol Gen Genet. 1989;219:341–8. doi:10.1007/BF00259605.
62. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44:D279–85. doi:10.1093/nar/gkv1344.
63. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–9. doi:10.1093/bioinformatics/bts199.

Droste, J., Kulisch, M., Wolf. T., Schaffert, L., Schneiker-Bekel, S., Pühler, A., Kalinowski, J., 2020. A maltose-regulated large genomic region is activated by MalT in *Actinoplanes* sp. SE50/110. Applied Microbiology and Biotechnology. DOI: 10.1007/s00253-020-10923-2 [status: published]

Authors contributions:

Julian Droste designed, planned, and interpreted the experimental work, performed data analysis and drafted the manuscript. Julian Droste and Martin Kulisch performed cultivation experiments and transcriptome analyses. Dr. Timo Wolf performed comparative transcriptome analysis of *Actinoplanes* sp. SE50/110 on maltose compared to glucose. Dr. Lena Schaffert assisted in analysis of maltose metabolism in *Actinoplanes* spp. Prof. Dr. Jörn Kalinowski, Prof. Dr. Alfred Pühler and Dr. Susanne Schneiker-Bekel assisted in interpreting the data and revised the manuscript. Prof. Dr. Jörn Kalinowski and Prof. Dr. Alfred Pühler coordinated the study. All authors read and approved the manuscript.

Applied Microbiology and Biotechnology
https://doi.org/10.1007/s00253-020-10923-2

APPLIED GENETICS AND MOLECULAR BIOTECHNOLOGY

A maltose-regulated large genomic region is activated by the transcriptional regulator MalT in *Actinoplanes* sp. SE50/110

Julian Droste[1] · Martin Kulisch[1] · Timo Wolf[1] · Lena Schaffert[1] · Susanne Schneiker-Bekel[2] · Alfred Pühler[2] · Jörn Kalinowski[1]

Received: 23 July 2020 / Revised: 15 September 2020 / Accepted: 21 September 2020

Abstract

Actinoplanes sp. SE50/110 is the industrially relevant producer of acarbose, which is used in the treatment of diabetes mellitus. Recent studies elucidated the expression dynamics in *Actinoplanes* sp. SE50/110 during growth. From these data, we obtained a large genomic region (*ACSP50_3900* to *ACSP50_3950*) containing 51 genes, of which 39 are transcribed in the same manner. These co-regulated genes were found to be stronger transcribed on maltose compared with glucose as a carbon source. The transcriptional regulator MalT was identified as an activator of this maltose-regulated large genomic region (MRLGR). Since most of the genes are poorly annotated, the function of this region is farther unclear. However, comprehensive BLAST analyses indicate similarities to enzymes involved in amino acid metabolism. We determined a conserved binding motif of MalT overlapping the -35 promoter region of 17 transcription start sites inside the MRLGR. The corresponding sequence motif 5′-TCATCC-5nt-GGATGA-3′ displays high similarities to reported MalT binding sites in *Escherichia coli* and *Klebsiella pneumoniae*, in which MalT is the activator of *mal* genes. A *malT* deletion and an overexpression mutant were constructed. Differential transcriptome analyses revealed an activating effect of MalT on 40 of the 51 genes. Surprisingly, no gene of the maltose metabolism is affected. In contrast to many other bacteria, MalT is not the activator of *mal* genes in *Actinoplanes* sp. SE50/110. Finally, the MRLGR was found partly in other closely related bacteria of the family Micromonosporaceae. Even the conserved MalT binding site was found upstream of several genes inside of the corresponding regions.

Key points

- *MalT is the maltose-dependent activator of a large genomic region in ACSP50_WT.*
- *The consensus binding motif is similar to MalT binding sites in other bacteria.*
- *MalT is not the regulator of genes involved in maltose metabolism in ACSP50_WT.*

Keywords *Actinoplanes* · Transcriptomic · Transcriptional regulation · MalT

Introduction

Actinoplanes sp. SE50/110 is the natural producer of the pseudotetrasaccharide acarbose (acarviosyl-1,4-maltose), which functions as an α-glucosidase inhibitor and is used in the treatment of diabetes mellitus (Truscheit et al. 1981). It is a Gram-positive, aerobic bacterium, which grows in branched hyphae and can form sporangia and motile spores (Vobis et al. 2015). The genome of *Actinoplanes* sp. SE50/100 has a high G+C content of 71.32%, which was first sequenced by Schwientek et al. (2012). Today, a refined high-quality genome of *Actinoplanes* sp. SE50/110 is available (Wolf et al. 2017b).

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00253-020-10923-2) contains supplementary material, which is available to authorized users.

✉ Jörn Kalinowski
joern@CeBiTec.Uni-Bielefeld.DE

[1] Microbial Genomics and Biotechnology, Center for Biotechnology, Bielefeld University, Universitätsstraße 27, 33615 Bielefeld, Germany

[2] Senior Research Group in Genome Research of Industrial Microorganisms, Center for Biotechnology, Bielefeld University, Universitätsstraße 27, 33615 Bielefeld, Germany

Published online: 28 September 2020

Actinoplanes spp. are known for their potential to produce a variety of secondary metabolites and antibiotics, like actaplanin (Debono et al. 1984), friulimicins (Aretz et al. 2000), moenomycin (Horbal et al. 2016), ramoplanin (Ciabatti et al. 1989), and teicoplanin (Bardone et al. 1978). Mining the genomes of actinomycetes, gene clusters for the production of several industrially relevant products could be identified. Also, the genome of *Actinoplanes* sp. SE50/110 harbors about 20 gene clusters, which are potentially responsible for the biosynthesis of secondary metabolites (Wolf et al. 2017b).

Actinoplanes sp. SE50 strains are industrially relevant producer of acarbose (Wehmeier and Piepersberg 2004). Therefore, understanding of the metabolism and the regulatory processes of this bacterium is an important step to optimize acarbose-producing conditions and to identify potential targets for metabolic engineering in order to increase acarbose productivity in the future.

Recent studies analyzed expression dynamics of all genes and operons during growth (Droste et al. 2020). Many co-regulated genes were identified by hierarchical cluster analyses, such as the *acb* gene cluster responsible for acarbose biosynthesis. A total of 71 genes were found to be transcribed coordinately, showing an increasing transcript amount during growth (Cluster 36, Droste et al. 2020). Interestingly, 41 genes were found to be located in close proximity in a region comprised of 51 genes (*ACSP50_3900* to *ACSP50_3950*). Differential transcriptome analyses revealed an increased transcription of this genomic region on maltose compared with glucose as a carbon source (Supplemental Fig. S1 and Supplemental Table S1).

In this study, we analyzed this maltose-regulated large genomic region (MRLGR) and its transcriptional regulation. A conserved sequence motif analysis was applied to prove co-regulation of these genes. Interestingly, only two transcriptional regulator genes (*ACSP50_3915* and *ACSP50_3917*) were found inside the MRLGR. We investigated the effects of the transcriptional regulator MalT (*ACSP50_3915*) on the surrounding genes by deletion and overexpression. Furthermore, we analyzed the potential function of the corresponding gene products.

Materials and methods

Strains, media, and cultivation conditions

All cloning procedures were carried out with *Escherichia coli* DH5αMCR (Grant et al. 1990). For the conjugational transfer of plasmids into *Actinoplanes* sp. SE50/110 (ATCC21044), the strain *E. coli* ET12567 (pUZ8002) (Kieser et al. 2004) was used as a conjugation host to generate the mutated strains of *Actinoplanes* in this study.

For *malT* (*ACSP50_3915*), the gene deletion CRISPR/Cas9 technique based on the plasmid pCRISPomyces-2 was used as described by Wolf et al. (2016). Spacer for the generation of the guide RNA (gRNA) and primer for amplification and cloning of up- and downstream flanking sequences are listed in Supplemental Table S2. Cloning procedures were carried out according to Cobb et al. (2014) and Wolf et al. (2016). The deletion plasmid was transferred into *Actinoplanes* sp. SE50/110 by conjugation as described before (Gren et al. 2016). The successful deletion of *malT* was proven by PCR and Sanger sequencing with primers listed in Supplemental Table S2. Gene deletion resulted in the strain *Actinoplanes* sp. SE50/110 Δ*malT* (referred to as ACSP50_Δ*malT* in this study).

A *malT* overexpression plasmid was constructed based on the integrative vector pSET152 (Gren et al. 2016) using the strong promoter P_{gapDH} from *Eggerthella lenta* (Schaffert et al. 2019a), resulting in the strain *Actinoplanes* sp. SE50/110 pSET152::P_{gapDH}::*malT* (referred to as ACSP50_OE*malT* in this study). The strain *Actinoplanes* sp. SE50/110 pSET152 (referred to as ACSP50_pSET in this study) containing the plasmid pSET152 was used as an empty vector control.

The *Actinoplanes* sp. SE50/110 wild type strain (referred to as ACSP50_WT in this study), and the mutants (ACSP50_pSET, ACSP50_Δ*malT*, ACSP50_OE*malT*) derived from this strain were grown on soy flour medium (SFM; 20-g L^{-1} soy, 20-g L^{-1} mannitol, 20-g L^{-1} agar, pH 8.0, tap water) agar plates and in NBS (11 g L^{-1} glucose × 1 H_2O, 4 g L^{-1} peptone, 4 g L^{-1} yeast extract, 1 g L^{-1} $MgSO_4$ × 7 H_2O, 2 g L^{-1} KH_2PO_4, 4 g L^{-1} K_2HPO_4) complex medium. For shake flask cultivations, minimal medium supplemented with maltose or glucose as a carbon source was used as described elsewhere (Wendler et al. 2015).

Shake flask cultivations were carried out in five biological replicates in 250-mL Corning® Erlenmeyer baffled cell culture flasks. Therefore, 50 mL of minimal medium was inoculated with spore suspension obtained from bacterial strains grown on SFM agar plates for 6 to 7 days at 28 °C and harvested with 1 mL ddH_2O. Cell growth was examined by the determination of cell dry weight. For RNA isolation and subsequent transcriptome analyses, 1 mL cell suspension was centrifuged for 15 s at maximum speed and immediately frozen in liquid

nitrogen. Cell pellets were stored at − 80 °C until RNA isolation (Wolf et al. 2017a).

RNA isolation and transcriptome analysis

RNA isolation

For the transcriptome analysis, RNA was isolated using a Macherey-Nagel NucleoSpin® RNA Plus kit in combination with Macherey-Nagel rDNase Set (Macherey-Nagel, Düren, Germany). Therefore, cell pellets were resuspended in 500 μL LBP buffer (NucleoSpin® RNA Plus kit, Macherey-Nagel) and transferred into 2-mL lysing matrix tubes (0.1-mm spherical silica beads, MP Biomedicals, Santa Ana, CA, USA). Cell disruption was carried out in a homogenizer (FastPrep FP120, Thermo Fisher Scientific, Waltham, MA, USA) two times for 30 s at speed setting 6.5 and 1 min on ice in between. Following this, cell debris were centrifuged for 2 min at maximum speed at 4 °C. The supernatant was used for RNA isolation according to the manufacturer's protocol. To verify the complete removal of residual DNA in the samples, PCR with primers binding to genomic *Actinoplanes* sp. SE50/110 DNA was performed. Quality and quantity of the RNA were analyzed with a NanoDrop 1000 spectrometer (Peqlab, Erlangen, Germany) and an Agilent RNA 6000 Pico kit run on an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA).

Whole-genome oligonucleotide microarray

Custom whole-genome oligonucleotide microarrays representing nearly all coding sequences of *Actinoplanes* sp. SE50/110 were used as described previously (Wolf et al. 2017a). Summarized, Agilent custom microarrays in the 4x44K format were used with a Two-Color Microarray-Based Prokaryote Analysis FairPlay III Labeling kit (version 1.4, Agilent Technologies, Santa Clara, CA, USA). After feature extraction using the manufacturer's software package, data analysis was performed with the software EMMA2 (Dondrup et al. 2009). The data was normalized (LOWESS) and a *t* test (one-sample, Holm) was applied. A *p* value of 0.05 was used as a cutoff for significance, and the log2 (ratio) cutoffs for a false discovery rate of 0.01 were experimentally determined as 1.1 and − 1.1 (Wolf et al. 2017a).

Reverse transcription quantitative PCR

Reverse transcription quantitative PCR (RT-qPCR) was carried out using a Bioline SensiFast SYBR No-Rox One-Step Kit (Bioline, London, UK) in 96-well LightCycler plates (Sarstedt, Nümbrecht, Germany) and measured in a Roche LightCycler 96 System (Roche, Mannheim, Germany). Samples were prepared according to the protocol described by Wolf et al. (2017a). The relative transcript amount was calculated as $2^{-\Delta Cq}$, whereas ΔCq was determined as difference of the mean Cq in the mutated strain compared with the respective control strain. Primers used for the RT-qPCR are listed in Supplemental Table S2.

Results

Forty-one genes of the maltose-regulated large genomic region are co-regulated in *Actinoplanes* sp. SE50/110

In recent studies, several co-expressed genes were identified by transcriptome and proteome analyses during growth of *Actinoplanes* sp. SE50/110 (Droste et al. 2020). These genes might belong to the same regulons. A large genomic region of 51 genes (ACSP50_3900 to ACSP50_3950) was found to be transcribed coordinately during growth. For 41 of the 51 genes, a highly similar continuously increasing transcript amount over the course of the cultivation was determined by hierarchical cluster analysis (Droste et al. 2020). Interestingly, we also identified this genomic region by comparative transcriptome analysis of *Actinoplanes* sp. SE50/110 wild type strain grown on maltose compared with glucose minimal medium (Supplemental Fig. S1 and Supplemental Table S1). The aim of this experiment was to identify genes with an increased transcript amount on maltose as a carbon source compared with glucose. Cells were cultivated in minimal medium, and samples for transcriptome analysis were taken after 72 h (Supplemental Fig. S1). By analyzing the top scorer of this experiment (genes which are highly transcribed on maltose compared with glucose), it could be shown that the genes of the region ACSP50_3900 to ACSP50_3950 are among the genes with the highest *M* values. This way, it was shown that these genes are stronger transcribed on maltose compared with glucose. The similar transcription pattern, the close proximity of these genes, and the maltose-dependent expression indicate a maltose-dependent co-regulation of this genomic region during growth. Therefore, this genomic region (Fig. 1) was named maltose-regulated large genomic region.

Analyzing the annotation of these genes, no gene products involved in the maltose metabolism of *Actinoplanes* sp. SE50/110 could be identified (Schaffert et al. 2019b). Twenty-four of the 51 genes have no annotated function ("hypothetical protein," "uncharacterized protein") according to the NCBI database (GenBank: LT827010). Interestingly, 10 genes were

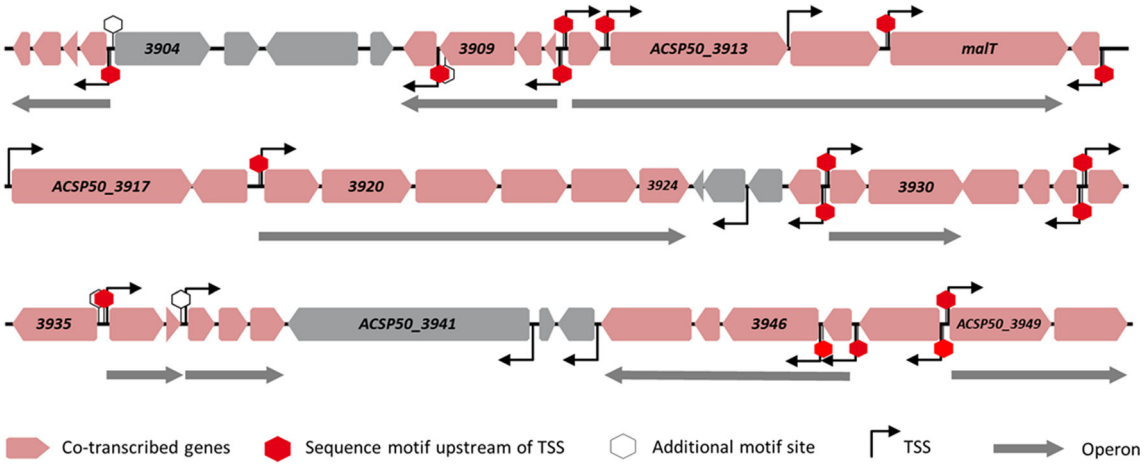

Fig. 1 Maltose-regulated large genomic region (MRLGR) ranging from ACSP50_3900 to ACSP50_3950 in *Actinoplanes* sp. SE50/110. Genes found to be co-regulated during growth are marked in light red. Data are obtained from Droste et al. (2020). Transcription start sites (TSS) and operon structure are indicated by black and gray arrows respectively. Additionally, locations of the sequence motif shown in Fig. 2 are marked with hexagons. If the sequence motif was found upstream of a TSS, the position is marked with a red hexagon, whereas additional locations of the motif are visualized in white, black-edged hexagons. The annotations of these genes are listed in Supplemental Table S3

annotated as membrane or transport related. Further 17 genes were annotated as peptidases, transferases, glucosidases, and other enzymes. Finally, two transcriptional regulator genes could be found inside the MRLGR (ACSP50_3915 and ACSP50_3917). Interestingly, both are annotated as LuxR family (MalT-like) transcriptional regulators.

A conserved palindromic sequence motif was identified in the promoter region of 17 genes of the MRLGR in *Actinoplanes* sp. SE50/110

The strict co-regulation of genes in bacteria is likely to be caused by a transcriptional regulator. Therefore, binding sites for transcription factors or alternative sigma factors might be conserved upstream of the transcription start sites (TSS) of these genes. Since many genes are organized in operons, a TSS was not identified upstream of every gene (Droste et al. 2020). Therefore, the transcription is initiated at the same sequence position for several genes. For 23 genes of this MRLGR, at least one TSS could be identified using the dataset of Droste et al. (2020). The tool MEME (Bailey et al. 2009) was used to identify motifs within the sequences 71 bp upstream of the TSS (− 70 to + 1) of these genes. A palindromic hexanucleotide sequence (5′-TCATCC-5 nt-GGATGA-3′) was identified in 17 sequences with an *e* value of 5.4×10^{-29} (Fig. 2). The distance to the upstream TSS was determined as 34.4 ± 1.0 bases from the 3′ end of the conserved motif and therefore overlaps with the -35 region of the corresponding promoters. This type of motif hints toward a characteristic binding site for transcription factors (Rhodes et al. 1996; Huffman and Brennan 2002). The identification of the transcriptional regulator responsible for maltose-dependent regulation of the MRLGR in *Actinoplanes* sp. SE50/110 is discussed in the chapter after next.

Functional analysis of the proteins encoded by the MRLGR in *Actinoplanes* sp. SE50/110

It could be shown that at least 41 of 51 genes of the MRLGR are strictly co-regulated dependent on the presence of maltose. However, the function of most of these genes belonging to this regulon is unclear, since they were poorly annotated by the automated annotation software pipeline Prokka, version 1.11 (Seemann 2014). The annotated functions of these genes according to the NCBI database (GenBank: LT827010) are listed in Supplemental Table S3. Therefore, we used the tools KEGG mapper (Kanehisa and Sato 2020; Kanehisa et al. 2016a) and GhostKOALA (Kanehisa et al. 2016b) in this work to further analyze the potential function of the genes of the MRLGR in *Actinoplanes* sp. SE50/110 (Supplemental Table S3). Interestingly, no common pathway or metabolism could be identified containing a significant number of proteins encoded by the MRLGR, although co-regulation of these genes was observed. However, the strongest commonality was found for 10 proteins, which contain domains that are similar to enzymes of the amino acid metabolism (Table 1). Therefore, we assume that parts of the MRLGR products are involved in the amino acid metabolism, such as arginine biosynthesis. For most of these proteins, at least one homologous gene/protein was identified in the genome of *Actinoplanes* sp. SE50/110 (Table 1).

The enzymes ACSP50_3921, ACSP50_3922, and ACSP50_3923 are potentially involved in the arginine biosynthesis. By in silico analysis of the respective enzymatic

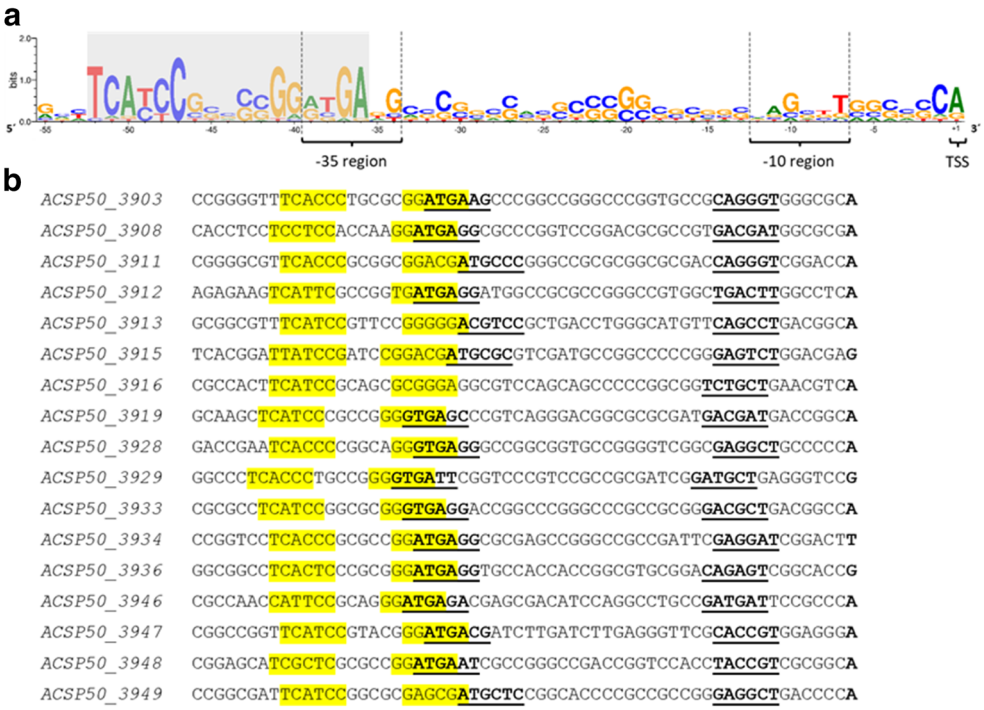

Fig. 2 TSS upstream sequences of genes co-regulated in the MRLGR (ACSP50_3900 to ACSP50_3950) in *Actinoplanes* sp. SE50/110 with an assigned TSS. The TSS are assigned according to Wolf et al. (2017b) and Droste et al. (2020) (submitted to BMC Genomics). **a** Consensus sequence of the promoter region of 17 genes of the MRLGR in *Actinoplanes* sp. SE50/110. The promoter motifs (-10 and -35 region) are marked with dashed lines. A conserved palindromic sequence motif overlapping the -35 region is highlighted in gray. **b** TSS upstream sequences used for consensus sequence shown in **a**. The corresponding TSS and promoter elements are shown in bold letters. The -10 and -35 regions are underlined. The palindromic sequence motif site is marked in yellow

reactions, a flux toward citrulline from ornithine and arginine could be observed.

A comprehensive BLAST analysis by the algorithms BLASTP and tBLASTn (Altschul et al. 1997) of the genomic region *ACSP50_3900* to *ACSP50_3950* was performed using respective protein sequences as input data. The full list of BLAST analysis can be found in Supplemental Table S3.

The results of the BLAST analyses revealed high similarities of the proteins to three different organisms: *Pseudosporangium ferrugineum*, *Couchioplanes caeruleus*, and *Krasilnikovia cinnamomea* (Fig. 3). The genomes of these three bacteria were searched for a similar genomic region compared with the MRLGR of *Actinoplanes* sp. SE50/110. In addition, the corresponding genomic region of the close

Table 1 Annotated function of 10 genes inside the MRLGR. The putative metabolic pathway and homologous genes in the genome of ACSP50_WT were listed

Locus tag	Annotated function (Wolf et al. 2017b)	Metabolic pathway	Homologous genes[1] in ACSP50_WT
ACSP50_3919	Class II glutamine amidotransferase	Amino acid metabolism	*ACSP50_6409*
ACSP50_3920	Amino acid permease	Amino acid metabolism	*ACSP50_2706*; *ACSP50_3876*
ACSP50_3921	Arginine deiminase	Arginine biosynthesis	*ACSP50_8316*
ACSP50_3922	Ornithine carbamoyltransferase	Arginine biosynthesis	*ACSP50_4060*
ACSP50_3923	Carbamate kinase	Arginine biosynthesis	*ACSP50_6398*
ACSP50_3924	Cyclic nucleotide–binding protein (phosphodiesterase)	Put. serine/threonine biosynthesis	
ACSP50_3944	Beta-Ala-His dipeptidase	Amino acid metabolism	*ACSP50_1214*
ACSP50_3946	Amino acid permease	Amino acid metabolism	
ACSP50_3948	Threonine/serine exporter family protein	Serine/threonine biosynthesis	
ACSP50_3950	Aminopeptidase P family protein	Amino acid metabolism	*ACSP50_1832*

[1] Revealed by BLASTP analysis, *e* value < $7e^{-14}$

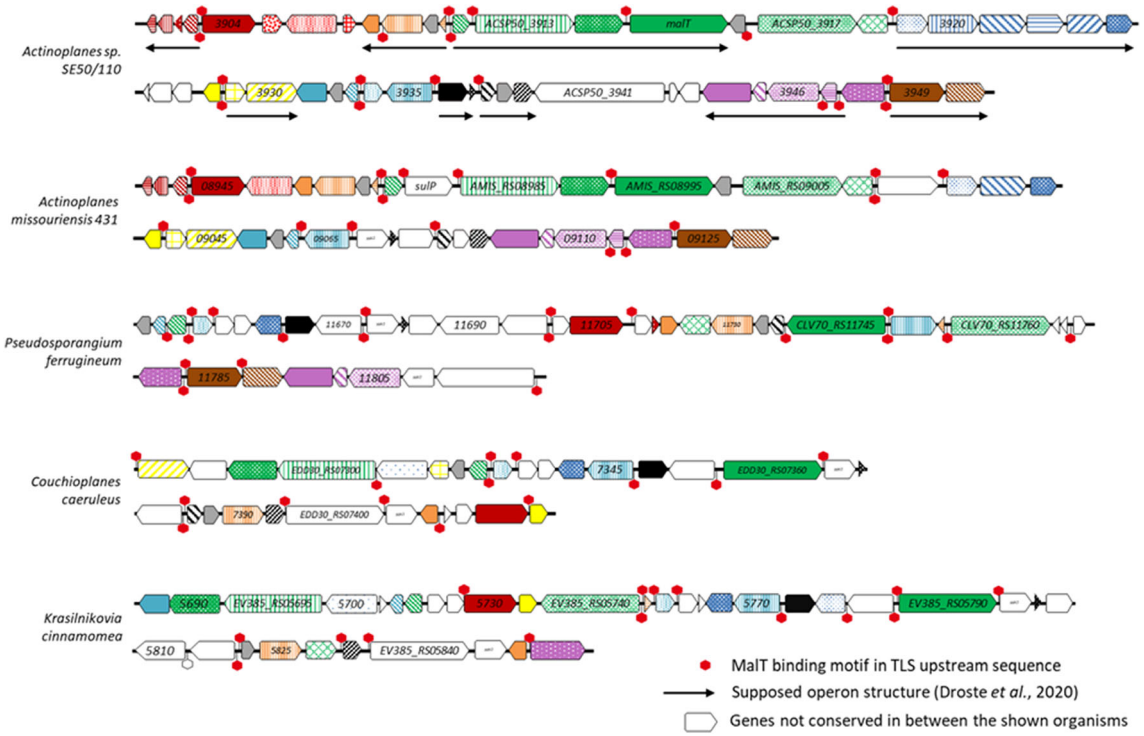

Fig. 3 Comparison of the MRLGR with similar genomic regions of *Actinoplanes missouriensis*, *Pseudosporangium ferrugineum*, *Couchioplanes caeruleus*, and *Krasilnikovia cinnamomea* containing homologous gene products identified by BLAST analysis. Genes of homologous proteins are marked in the same color code. The positions of the conserved sequence motif (5′-TCATCC-5 bp-GGATGA-3′) in all strains are marked with red (upstream of ORF) and white (additional sites) hexagons. A detailed list of all shown genes and their annotated function is given in Supplemental Table S4

relative *Actinoplanes missouriensis* was analyzed for comparison. For all strains, at least 23 genes homologous to genes from *ACSP50_3900* to *ACSP50_3950* were found to be located in close proximity to each other. However, not all genes were found in the same order and direction. Genes, which seem to be organized in operons in *Actinoplanes* sp. SE50/110, are rearranged in the other strains (Fig. 3). Even between the two *Actinoplanes* spp., clear differences were identified regarding this genomic region.

For the regulator gene *malT* (ACSP50_3915), a homologous gene could be identified in all analyzed genomes (Fig. 3), whereas ACSP50_3917 homologs were only identified in three of four species. By this, ACSP50_3915 is more conserved compared with ACSP50_3917. We assume that ACSP50_3915 is the key regulator of the surrounding genes.

Interestingly, several genes which were not found to be co-regulated in *Actinoplanes* sp. SE50/110 (Fig. 1), like *ACSP50_3904* to *ACSP50_3907*, *ACSP50_3925* to *ACSP50_3927*, or *ACSP50_3941* to *ACSP50_3943*, are not conserved between the analyzed strains (Supplemental Table S4), except for *ACSP50_3904*. No homologous genes could be identified in the compared strains, not even in the close relative *A. missouriensis*. Furthermore, the operon *ACSP50_3920* to *ACSP50_3924* is lacking in the four analyzed bacterial genomes, except for *ACSP50_3924* encoding a cyclic nucleotide–binding protein, which was identified in all strains. Additionally, *A. missouriensis* contains an *ACSP50_3921* homolog coding for an arginine deiminase. Strikingly, genes encoding a polyphosphate kinase (*ppk2*) were found in one or even two copies in the corresponding genomic regions of the analyzed bacteria but lack in the MRLGR of *Actinoplanes* sp. SE50/110.

Finally, the palindromic sequence motif identified in the -35 region of the MRLGR genes could also be identified upstream of several open reading frames (ORFs) in the analyzed genomic regions of *A. missouriensis*, *P. ferrugineum*, *C. caeruleus*, and *K. cinnamomea* (Fig. 3). This confirms the close relation of these genomic regions.

On the one hand, several genes of this region seem to be highly conserved as well as the identified palindromic sequence motif upstream of the ORFs. On the other hand, the arrangement and order of these genes are highly diverse comparing different bacterial strains. Therefore, it can be assumed that this genomic region was passed on by horizontal gene transfer in several related species of the family Micromonosporaceae. We assume that the gene

products of this region are important, but not all are essential for each respective strain. Especially for growth on glucose, most of the genes seem to be low or not transcribed at all in *Actinoplanes* sp. SE50/110.

The transcriptional regulator MalT (ACSP50_3915) is the activator of the MRLGR in *Actinoplanes* sp. SE50/110

The observations above lead to the conclusion that the genes of the MRLGR are strictly regulated by one common transcriptional regulator. Since only two transcriptional regulator genes (*ACSP50_3915* and *ACSP50_3917*) could be identified in the MRLGR, it was assumed that at least one of these regulators is responsible for regulation of the MRLGR.

However, only *ACSP50_3915* (*M* value of 2.23) but not *ACSP50_3917* (*M* value of 0.24) was found to be transcriptionally "upregulated" on maltose compared with glucose in our transcriptome analysis (Supplemental Fig. S1). In addition, a higher protein similarity was found for ACSP50_3915 (42% similarity) to MalT in *E. coli* (Supplemental Fig. S2). Therefore, it was assumed that MalT (ACSP50_3915) might be the transcriptional regulator of the MRLGR.

In order to prove this regulatory function, the corresponding gene *ACSP50_3915* was deleted using CRISPR/Cas9 (Wolf et al. 2016), resulting in an *Actinoplanes* sp. SE50/110 Δ*malT* deletion mutant (ACSP50_Δ*malT*). In addition, *malT* was overexpressed in *Actinoplanes* sp. SE50/110 using the strong promoter P_{gapDH} from *Eggerthella lenta* (Schaffert et al. 2019a) combined with the integrative vector pSET152 (Gren et al. 2016), resulting in the *malT* overexpression strain ACSP50_OE*malT*.

Both constructed mutant strains ACSP50_Δ*malT* and ACSP50_OE*malT* were cultivated in comparison with the *Actinoplanes* sp. SE50/110 wild type strain (ACSP50_WT) and an empty vector control strain (ACSP50_pSET) in a shake flask cultivation in minimal medium supplemented with maltose and glucose as a carbon source (Fig. 4). It could be shown that the regulator deletion mutant ACSP50_Δ*malT* grows slightly slower under both conditions (maltose and glucose) compared with the wild type (Fig. 4a). The *malT* overexpression strain (ACSP50_OE*malT*) shows significantly reduced growth on both glucose and maltose as a carbon source compared with an empty vector control (Fig. 4b).

Samples for transcriptome analysis were taken in the middle of the growth phase (after 96 h) of all strains on maltose and glucose each (Fig. 4), except for ACSP50_WT grown on maltose (transcriptome samples after 72 h) and ACSP50_Δ*malT* on glucose (transcriptome samples after 110 h) (Fig. 4a). The RNA was isolated and pooled from three biological replicates. Transcriptome analysis was carried out using whole-genome microarrays as described elsewhere (Wolf et al. 2017a).

In total, 141 genes were found to be significantly differentially transcribed on glucose, of which 28 genes show an increased and 113 genes a decreased transcript amount in ACSP50_Δ*malT* compared with the wild type strain (Fig. 5). On maltose as a carbon source, 247 genes with significant differential transcription were identified (101 increased and 146 decreased transcript amount). Strikingly, only 69 (11 increased and 58 decreased) differentially transcribed genes were found under both conditions (Supplemental Table S5). In addition to three genes annotated as hypothetical or uncharacterized proteins, two genes with a membrane-associated gene product (*ACSP50_0484*, *ACSP50_2520*), two RNA polymerase sigma-24 subunits (*ACSP50_3334*, *ACSP50_3840*), a polyhydroxyalkanoate depolymerase (*ACSP50_3332*), a NAD-dependent deacetylase (*ACSP50_4603*), an epimerase (*ACSP50_4604*), and a serine hydrolase (*ACSP50_8214*) were found to be "transcriptionally upregulated" under both conditions in ACSP50_Δ*malT*. Among the 58 genes, which show a significantly decreased transcript amount under both conditions, 18 genes with no annotated function, two glutathione-dependent formaldehyde dehydrogenases (*ACSP50_1264*, *ACSP50_4381*), 8 genes annotated as membrane proteins or transporters, an anti-sigma factor (*ACSP50_0205*), a glycosyl transferase (*ACSP50_2948*), a transglycosylase (*ACSP50_1322*), and a trehalose synthase (*ACSP50_7524*) were identified. All

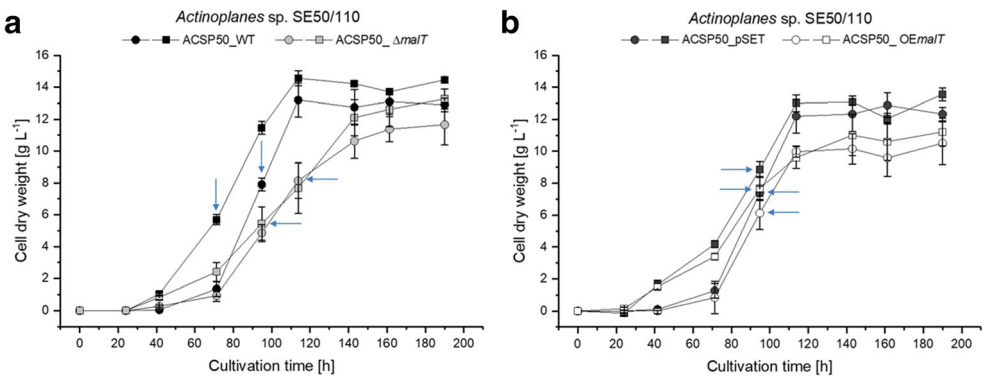

Fig. 4 Growth of ACSP50_WT (black), ACSP50_Δ*malT* (gray), and ACSP50_OE*malT* (white). Cell dry weight for cells grown in minimal medium inoculated with spores supplemented with glucose (circles) and maltose (squares) as a carbon source. The means and standard deviations of five biological and two technical replicates are shown. Sampling points for transcriptome analysis are indicated with blue arrows

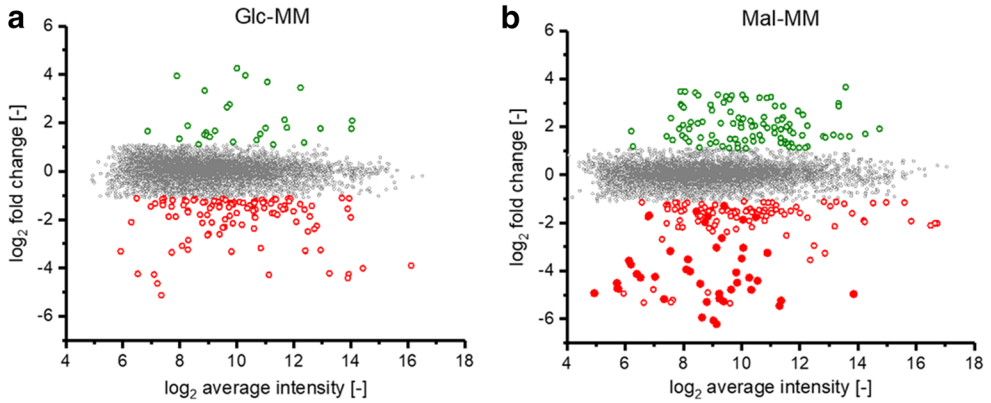

Fig. 5 Differential transcriptional analysis of ACSP50_Δ*malT* compared with ACSP50_WT. **a** Ratio/intensity plot from whole-genome microarrays of the Δ*malT* mutant compared with the wild type grown in glucose minimal medium (Glc-MM). **b** Ratio/intensity plot from whole-genome microarrays of the Δ*malT* mutant compared with the wild type grown in maltose minimal medium (Mal-MM). Green and red dots represent genes with significantly different transcript levels in the Δ*malT* strain (*M* value > 1.1 or < − 1.1 respectively; p_{adj} value > 0.05). Filled dots show genes of the MRLGR

results of the differential transcriptome analysis can be found in Supplemental Table S5.

Interestingly, the transcription of the MRLGR was strongly influenced by the deletion of the *malT* gene on maltose minimal medium but not on glucose (Fig. 5). This can be explained by the fact that the genes are not transcribed on glucose at all in the wild type, which is why no effect on the respective genes is visible on glucose (Supplemental Fig. S1 and Supplemental Table S1). A total of 32 of all 51 genes of the MRLGR are significantly less transcribed ($p_{adj} < 0.05$; *M* value < 1.1) in ACSP50_Δ*malT* on maltose as a carbon source, whereas 41 were previously described to be co-regulated in *Actinoplanes* sp. SE50/110 (Fig. 1, Droste et al. 2020). However, all genes of the MRLGR, which were identified to be less transcribed in ACSP50_Δ*malT*, were also previously described as co-regulated, except for *ACSP50_3907*. Additionally, 23 of the 31 genes, which were both described as co-regulated as well as significantly downregulated in ACSP50_Δ*malT*, were found to be significantly upregulated on maltose compared with glucose (Supplemental Table S6). These matches indicate that MalT is the maltose-dependent transcriptional activator of these genes. Strikingly, no genes of the maltose metabolism were found to be significantly different transcribed in ACSP50_Δ*malT* compared with ACSP50_WT. The maltase AmlE (*ACSP50_2474*), which was described to be essential for maltose degradation (Schaffert et al. 2019a, 2019b), shows an *M* value of − 0.393 (p_{adj} value > 0.5), or the operon *malEFG*, which was described to encode the maltose import system of *Actinoplanes* sp. SE50/110 (Wendler et al. 2016), exhibits *M* values of − 0.162 to 0.213 (p_{adj} values > 0.5) in the mutant strain compared with the wild type both grown on maltose minimal medium (Supplemental Table S5).

The results of the whole-genome microarrays were confirmed with RT-qPCR for the genes of the MRLGR, since RT-qPCR is more sensitive compared with the microarray technique. The genes, which were found to be less transcribed in the microarray data (Fig. 5), were confirmed to be downregulated by RT-qPCR data (Fig. 6). Strikingly, for 10 further genes, a significantly decreased transcription was determined. Thereby, all genes, which were previously described as co-regulated as well as upregulated on maltose, could be identified to be significantly less transcribed in ACSP50_Δ*malT* using RT-qPCR. Only *ACSP50_3948* ("hypothetical protein") was found to be co-regulated along with the other genes of the MRLGR, but not influenced by the deletion of *malT*. In conclusion, 42 of 51 genes inside the MRLGR were found to be significantly less transcribed in the deletion mutant ACSP50_Δ*malT*.

In addition, the transcript levels of the MRLGR genes were measured for *malT* overexpression (ACSP50_OE*malT*) using the strong promoter P_{gapDH} from *Eggerthella lenta* (Schaffert et al. 2019a) compared with the empty vector control (Fig. 6 and Supplemental Table S7).

In total, 39 of the 51 genes were identified to be significantly upregulated in the *malT* overexpression strain ACSP50_OE*malT*. All of these 39 genes were also identified to be downregulated in ACSP50_Δ*malT*, except for *ACSP50_3906* and *ACSP50_3926* (Fig. 6), which are downregulated in ACSP50_Δ*malT* but show no significant transcriptional changes through *malT* overexpression (Fig. 6).

Looking at all datasets, the results match the proposed operon structure of the MRLGR. Genes transcribed in the same operon show the same transcription trends under the different tested conditions. The genes, which are transcribed in the same operon together with *malT* (*ACSP50_3912* to *ACSP50_3915*), show 10.000-fold to 3.000-fold decreased transcription compared with the wild type strain both cultivated on maltose minimal medium. In contrast to that, an overexpression of *malT* leads to an increased transcription of all

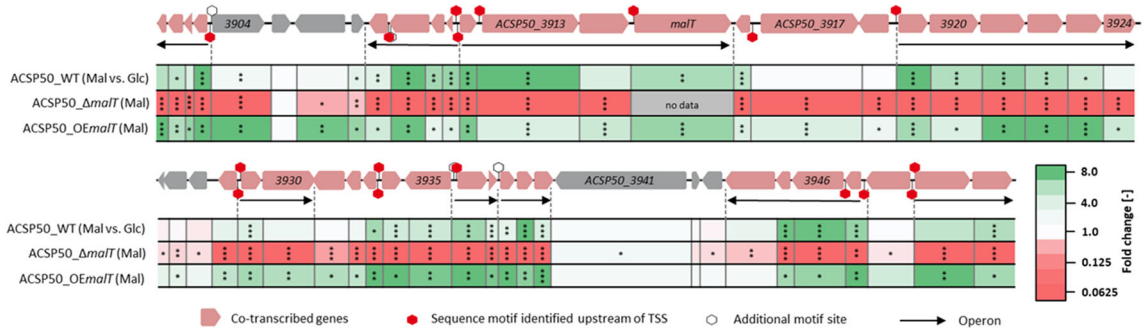

Fig. 6 Differential transcriptional analysis of the MRLGR of ACSP50_WT under different expression levels of the transcriptional activator MalT (ACSP50_3915) and on different carbon sources. The values for ACSP50_WT on maltose (Mal) compared with glucose (Glc) (green color indicates increased transcription on maltose) as well as the deletion mutant ACSP50_Δ*malT* and the *malT* overexpression strain ACSP50_OE*malT* on maltose both compared with ACSP50_WT are shown. In ACSP50_Δ*malT*, no *malT* gene is present, whereas ACSP50_OE*malT* contains two copies of *malT*, one in the genome and one on the integrated plasmid. Heatmap of the fold changes of transcript abundance for the genes in the genomic region surrounding *malT* was derived from whole-genome microarray (Mal vs. Glc) and reverse transcription quantitative PCR (RT-qPCR) data (ACSP50_Δ*malT* (Mal) and ACSP50_OE*malT* (Mal)). Green color indicates increased transcription (for "Mal vs. Glc": green = increased on maltose). Significance value of $p < 0.05$ is marked with a single asterisk, $p < 0.01$ with two asterisks, and $p < 0.001$ with three asterisks (*t* test, two-sample, Holm). The results of the RT-qPCR are listed in Supplemental Table S7

genes in the operon (fold changes 2.53 to 6.92). The genes of the operon *ACSP50_3919* to *ACSP50_3924* are significantly "downregulated" in ACSP50_Δ*malT* compared with the wild type strain (fold changes 0.001 to 0.011) on maltose minimal medium, whereas an overexpression of *malT* (ACSP50_OE*malT*) leads to an increased transcription of this operon of 5.02 to 10.22 fold.

The same effect was found for operons *ACSP50_3900* to *ACSP50_3903*, *ACSP50_3908* to *ACSP50_3911*, *ACSP50_3929* to *ACSP50_3930*, *ACSP50_3936* to *ACSP50_3937*, *ACSP50_3938* to *ACSP50_3940*, and *ACSP50_3949* to *ACSP50_3950*. In ACSP50_Δ*malT*, a decreased transcription was observed, whereas an overexpression of *malT* leads to an increased transcription of these operons.

The operon *ACSP50_3944* to *ACSP50_3947* shows only partly this effect. Since deletion of *malT* leads to a decreased transcription of all genes in the operon, an overexpression does not affect transcription of the last gene inside the operon *ACSP50_3944*. However, since this gene is the last gene in this operon, this effect could be explained by less transcription of operon's last genes due to shortened transcripts.

Interestingly, most of the genes which were not observed to be transcribed coordinately with the transcriptional activator gene *malT* were found to be less influenced regarding their respective transcription in ACSP50_Δ*malT* and ACSP50_OE*malT* compared with ACSP50_WT. These genes are *ACSP50_3905*, *ACSP50_3925* to *ACSP50_3927*, *ACSP50_3941* to *ACSP50_3943*, and *ACSP50_3948*, which show mostly no or a less strong effect regarding the *malT* expression level (Supplemental Table S6). This trend also correlates with the respective transcription level on maltose compared with glucose. Genes in this genomic region which seem to be transcribed coordinately and affected by the *malT* expression level show an increased transcription on maltose compared with glucose, whereas the abovementioned genes do not show any difference in transcription on maltose compared with glucose. An overview about all transcriptomic studies regarding the MRLGR is given in Supplemental Table S6. In conclusion, 37 genes were identified to be influenced by the expression level of MalT.

Discussion

A MRLGR was identified by expression dynamics analysis. A total of 41 of 51 genes inside this MRLGR were found to be transcribed coordinately, showing a continuously increasing transcription during growth (Droste et al. 2020). Therefore, it can be assumed that these genes are co-regulated on a transcriptional level. A conserved palindromic sequence motif (5′-TCATCC-5 nt-GGATGA-3′) overlapping the -35 region of the corresponding promoter was identified upstream of 17 TSS of the MRLGR genes. This sequence motif partly matches the binding motif of the transcriptional activator MalT in *E. coli* and *Klebsiella pneumoniae* described as a repeat of a 5′-GGA(T/G)GA core hexanucleotide, bordered by two G residues on both sides 5′ GGGGA(T/G)GAGG (Richet and Raibaud 1989; Vidal-Ingigliardi et al. 1991; Boos and Shuman 1998). In *E. coli*, the 5′ end of this so-called MalT box was identified at position -34.5 to -35.5 in relation to the TSS, which overlaps with the -35 region of the corresponding promoters (Boos and Shuman 1998). The distance to the TSS of the potential regulatory sequence identified in *Actinoplanes* sp. SE50/110 is in good accordance with that 34.4 ± 1.0 nt. MalT is the ATP-dependent transcriptional activator of the maltose regulon in *E. coli* (Richet and Raibaud 1989). MalT was found to be maltotriose-dependent in *E. coli*.

The genes of the MRLGR seem to be transcribed dependent on carbon source, activated on maltose, and repressed on glucose. Protein similarity of MalT in *E. coli* and in *Actinoplanes* sp. SE50/110 suggests that ACSP50_3915 also contains a maltose- or maltotriose-binding domain, as it was described for *E. coli*.

MalT-like regulators are widespread over different bacteria (Supplemental Fig. S3). Strikingly, two MalT-like regulators (*ACSP50_3915* and *ACSP50_3917*) were found in the MRLGR in *Actinoplanes* sp. SE50/110, of which *ACSP50_3915* shows the highest similarity to the *malT* gene of *E. coli* (Supplemental Fig. S2). Except for these two proteins, no other homologs (compared with MalT in *E. coli*) were identified in the *Actinoplanes* sp. SE50/110 genome. Furthermore, for gene deletion of *ACSP50_3917* in *Actinoplanes* sp. SE50/110, only slight effects on genes of the MRLGR were found (unpublished data). This could be due to an indirect effect of this regulator on the MRLGR genes. Therefore, it can be assumed that *ACSP50_3915* is the main transcriptional activator of the MRLGR and binds to the identified motifs in the -35 promoter region of these genes. Interestingly both regulators seem to have no effect on the genes of the maltose metabolism.

This could be confirmed by different transcriptomic studies on ACSP50_WT as well as deletion and overexpression mutants of *malT*. The deletion of *malT* leads to a significantly decreased transcription of 42 of these 51 genes, whereas overexpression of *malT* leads to a significantly increased transcription of at least 39 genes on maltose minimal medium. This results in a number of at least 37 genes, which are regulated by MalT, since both deletion and overexpression lead to significantly different transcription levels of these genes. In general, genes which were not affected by MalT seem to be less conserved in the MRLGR, as these genes could not be identified in similar genomic regions in other bacteria (Fig. 3 and Supplemental Table S4) and they were found to be not co-regulated or increased transcribed on maltose compared with glucose (Supplemental Table S6). This confirms the maltose dependency of the MalT regulon.

Interestingly, the MRLGR contains no genes encoding enzymes or proteins involved in maltose metabolism. Furthermore, none of the genes described for maltose utilization or transport (Schaffert et al. 2019a, 2019b) was found to be affected by deletion or overexpression of *malT* in *Actinoplanes* sp. SE50/110, and no similar sequence to the described MalT binding site was found upstream of these genes (data not shown).

Therefore, it can be assumed that transcription of genes involved in the maltose utilization is regulated MalT independently, although transcription of *malT* itself and therefore the genomic region *ACSP50_3900* to *ACSP50_3950* shows increased transcription on maltose (Fig. 6; Schaffert et al. 2019a, 2019b). Previous studies showed that transcription of the supposed maltose importer MalEFG is not regulated by a MalR homolog as it is described in *E. coli* (Wolf et al. 2017a). Regulation of maltose metabolism seems to be different in *Actinoplanes* sp. SE50/110 compared with other bacteria like *E. coli*. Therefore, the expression of several genes involved in maltose metabolism seems to be constitutive, since transcripts and proteins could be identified independently on supplied carbon source (Schaffert et al. 2019a, 2019b; Wolf et al. 2017a; Wendler et al. 2016). In contrast to that, the expression of the MRLGR, which in turn seems to be not involved in carbon metabolism, is highly influenced by the presence of maltose. Since maltose metabolism is closely connected to acarbose biosynthesis (Schaffert et al. 2019a, 2019b; Wendler et al. 2016), it could be assumed that gene products of the MRLGR are involved in the biosynthesis of acarbose precursors or related pathways.

However, the annotated function of most of the genes inside of this genomic region is unclear. Interestingly, several genes located in this genomic region were also found to be located in close proximity to each other in other organisms, like *P. ferrugineum*, *C. caeruleus*, and *K. cinnamomea*. Indeed, the arrangement and order of the homologous genes in these soil bacteria differ from *Actinoplanes* sp. SE50/110. Furthermore, several genes present in the MRLGR are not present in the other analyzed genomes, not even in close relatives like *A. missouriensis*. However, it was shown that the majority of these non-conserved genes are not regulated by MalT. Since these genes are strictly regulated dependent on maltose, it can be assumed that the corresponding proteins are needed especially on maltose. Indeed, the analyses using BLAST revealed protein functions for 10 gene products connected to the amino acid metabolism, such as arginine biosynthesis. Nevertheless, the presence of several homologs in the *Actinoplanes* sp. SE50/110 genome suggests that these genes are not mainly responsible for this biosynthetic pathway. A maltose-dependent regulation of these amino acid biosynthesis genes could not be explained. No common metabolic pathway could be identified for the annotated function of the conserved gene products of this genomic region. Most of the encoded proteins are enzymes or transport-related proteins involved in amino acid biosynthesis. However, also several genes annotated as hypothetical or uncharacterized proteins were reported. Since all strains containing parts of this genomic region were found in similar habitats, it can be assumed that this genomic region is involved in the metabolism of substrate specific for their respective soil habitat. As a soil bacterium, isolated from coffee plantation in Kenia (Frommer et al. 1975), a special nutrient supply of *Actinoplanes* sp. SE50/110 could be a reason for a sugar-dependent regulation of genes involved in amino acid uptake, peptide degradation, and amino acid biosynthesis. A close connection of sugar

and amino acid metabolism has been reported for prokaryotes (Gänzle et al. 2007), as well as eukaryotes (Binder 2010; Rennie and Tipton 2000). In plants, several regulatory effects of sugar on specific parts of the amino acid metabolisms have been shown (Pratelli and Pilot 2014). A further explanation could be that gene products of the MRLGR are involved in biosynthesis of a metabolite, which is not essential. Therefore, it could be regulated dependent on availability of maltose as an indicator of good environmental conditions. However, the analysis of the MRLGR for secondary metabolite genes using the web tool antiSMASH 5.0 (Blin et al. 2019) revealed no hits (data not shown).

In conclusion, it can be assumed that this genomic region harbors genes important for specific habitats of *Actinoplanes* sp. SE50/110. As it can be found partly in other soil bacteria, which occur in similar environments, the proteins encoded in this genomic region could be involved in uptake and degradation of specific nutrients or in production of an optional metabolite.

Code availability Not applicable.

Authors' contributions JD designed, planned, and interpreted the experimental work; performed data analysis; and drafted the manuscript. JD and MK performed cultivation experiments and transcriptome analyses. TW performed comparative transcriptome analysis of *Actinoplanes* sp. SE50/110 on maltose compared with glucose. LS assisted in analysis of maltose metabolism in *Actinoplanes* spp. JK, AP, and SS-B assisted in interpreting the data and revised the manuscript. JK and AP coordinated the study. All authors read and approved the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. JD acknowledges support from the CLIB-Graduate Cluster Industrial Biotechnology at Bielefeld University, Germany, which is supported by the Ministry of Innovation, Science and Research (MIWF) of the federal state North Rhine-Westphalia, Germany, and Bielefeld University, Germany. We gratefully thank our cooperation partner Bayer AG (Leverkusen, Germany) for their continuous support and funding. We acknowledge the support for the Article Processing Charge by the Deutsche Forschungsgemeinschaft and the Open Access Publication Fund of Bielefeld University.

Data availability The microarray datasets generated in this study can be found in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession E-MTAB-8815. All other data supporting the conclusion of this study are included in the article and its additional files.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Aretz W, Meiwes J, Seibert G, Vobis G, Wink J (2000) Friulimicins. Novel lipopeptide antibiotics with peptidoglycan synthesis inhibiting activity from *Actinoplanes friuliensis* sp. nov. I. Taxonomic studies of the producing microorganism and fermentation. J Antibiot 53:807–815. https://doi.org/10.7164/antibiotics.53.807

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:W202–W208. https://doi.org/10.1093/nar/gkp335

Bardone MR, Paternoster M, Coronelli C (1978) Teichomycins, new antibiotics from *Actinoplanes teichomyceticus* nov. sp. II. Extraction and chemical characterization. J Antibiot 31:170–177. https://doi.org/10.7164/antibiotics.31.170

Binder S (2010) Branched-chain amino acid metabolism in *Arabidopsis thaliana*. Arabidopsis Book 8:e0137. https://doi.org/10.1199/tab.0137

Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. https://doi.org/10.1093/nar/gkz310

Boos W, Shuman H (1998) Maltose/maltodextrin system of *Escherichia coli*. Transport, metabolism, and regulation. Microbiol Mol Biol Rev 62:204–229

Ciabatti R, Kettenring JK, Winters G, Tuan G, Zerilli L, Cavalleri B (1989) Ramoplanin (A-16686), a new glycolipodepsipeptide antibiotic. III. Structure elucidation. J Antibiot 42:254–267. https://doi.org/10.7164/antibiotics.42.254

Cobb RE, Wang Y, Zhao H (2014) High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. ACS Synth Biol 4:723–728. https://doi.org/10.1021/sb500351f

Debono M, Merkel KE, Molloy RM, Barnhart M, Presti E, Hunt AHNN, Hamill RL (1984) Actaplanin, new glycopeptide antibiotics produced by *Actinoplanes missouriensis*. The isolation and preliminary chemical characterization of actaplanin. J Antibiot 37:85–95. https://doi.org/10.7164/antibiotics.37.85

Dondrup M, Albaum SP, Griebel T, Henckel K, Jünemann S, Kahlke T, Kleindt CK, Küster H, Linke B, Mertens D, Mittard-Runte V, Neuweger H, Runte KJ, Tauch A, Tille F, Pühler A, Goesmann A (2009) EMMA 2-a MAGE-compliant system for the collaborative analysis and integration of microarray data. BMC Bioinf 10:50. https://doi.org/10.1186/1471-2105-10-50

Droste J, Ortseifen V, Persicke M, Schneiker-Bekel S, Pühler A, Kalinowski J (2020) The expression of the acarbose biosynthesis gene cluster in *Actinoplanes* sp. SE50/110 is dependent on the growth phase. BMC Genomics. https://doi.org/10.21203/rs.3.rs-41287/v1

Frommer W, Puls W, Schäfer D, Schmidt D. (1975) Glycoside-hydrolase enzyme inhibitors. German patent.

Gänzle MG, Vermeulen N, Vogel RF (2007) Carbohydrate, peptide and lipid metabolism of lactic acid bacteria in sourdough. Food Microbiol 24:128–138. https://doi.org/10.1016/j.fm.2006.07.006

Grant SG, Jessee J, Bloom FR, Hanahan D (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. PNAS 87:4645–4649. https://doi.org/10.1073/pnas.87.12.4645

Gren T, Ortseifen V, Wibberg D, Schneiker-Bekel S, Bednarz H, Niehaus K, Zemke T, Persicke M, Pühler A, Kalinowski J (2016) Genetic engineering in *Actinoplanes* sp. SE50/110-development of an inter-generic conjugation system for the introduction of actinophage-based integrative vectors. J Biotechnol 232:79–88. https://doi.org/10.1016/j.jbiotec.2016.05.012

Horbal L, Ostash B, Luzhetskyy A, Walker S, Kalinowski J, Fedorenko V (2016) A gene cluster for the biosynthesis of moenomycin family antibiotics in the genome of teicoplanin producer *Actinoplanes teichomyceticus*. Appl Microbiol Biotechnol 100:7629–7638. https://doi.org/10.1007/s00253-016-7685-3

Huffman JL, Brennan RG (2002) Prokaryotic transcription regulators. More than just the helix-turn-helix motif. Curr Opin Struct Biol 12:98–106. https://doi.org/10.1016/S0959-440X(02)00295-6

Kanehisa M, Sato Y (2020) KEGG Mapper for inferring cellular functions from protein sequences. Protein Sci 29:28–35. https://doi.org/10.1002/pro.3711

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016a) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:D457–D462. https://doi.org/10.1093/nar/gkv1070

Kanehisa M, Sato Y, Morishima K (2016b) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Kieser T, Bibb M, Buttner MJ, Chater KF, Hopwood DA (2004) Practical *Streptomyces* genetics. John Innes Centre, Norwich

Pratelli R, Pilot G (2014) Regulation of amino acid metabolic enzymes and transporters in plants. J Exp Bot 65:5535–5556. https://doi.org/10.1093/jxb/eru320

Rennie MJ, Tipton KD (2000) Protein and amino acid metabolism during and after exercise and the effects of nutrition. Annu Rev Nutr 20:457–483. https://doi.org/10.1146/annurev.nutr.20.1.457

Rhodes D, Schwabe JW, Chapman L, Fairall L (1996) Towards an understanding of protein-DNA recognition. Philos Trans R Soc Lond B Biol Sci 351:501–509. https://doi.org/10.1098/rstb.1996.0048

Richet E, Raibaud O (1989) MalT, the regulatory protein of the *Escherichia coli* maltose system, is an ATP-dependent transcriptional activator. EMBO J 8:981–987

Schaffert L, März C, Burkhardt L, Droste J, Brandt D, Busche T, Rosen W, Schneiker-Bekel S, Persicke M, Pühler A, Kalinowski J (2019a) Evaluation of vector systems and promoters for overexpression of the acarbose biosynthesis gene *acbC* in *Actinoplanes* sp. SE50/110. Microb Cell Factories 18:114. https://doi.org/10.1186/s12934-019-1162-5

Schaffert L, Schneiker-Bekel S, Dymek S, Droste J, Persicke M, Busche T, Brandt D, Pühler A, Kalinowski J (2019b) Essentiality of the maltase AmlE in maltose utilization and its transcriptional regulation by the repressor AmlR in the acarbose-producing bacterium *Actinoplanes* sp. SE50/110. Front Microbiol 10:2448. https://doi.org/10.3389/fmicb.2019.02448

Schwientek P, Szczepanowski R, Rückert C, Kalinowski J, Klein A, Selber K, Wehmeier UF, Stoye J, Pühler A (2012) The complete genome sequence of the acarbose producer *Actinoplanes* sp. SE50/110. BMC Genomics 13:112. https://doi.org/10.1186/1471-2164-13-112

Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Truscheit E, Frommer W, Junge B, Müller L, Schmidt DD, Wingender W (1981) Chemistry and biochemistry of microbial α-glucosidase inhibitors. Angew Chem Int Ed Eng 20:744–761. https://doi.org/10.1002/anie.198107441

Vidal-Ingigliardi D, Ricbet E, Raibaud O (1991) Two MalT binding sites in direct repeat. J Mol Biol 218:323–334. https://doi.org/10.1016/0022-2836(91)90715-I

Vobis G, Schäfer J, Kämpfer P (2015) *Actinoplanes*. In: Whitman WB, Rainey F, Kämpfer P, Trujillo M, Chun J, DeVos P, Hedlund B, Dedysh S (eds) Bergey's manual of systematics of archaea and bacteria. Wiley, Chichester, pp 1–41

Wehmeier UF, Piepersberg W (2004) Biotechnology and molecular biology of the alpha-glucosidase inhibitor acarbose. Appl Microbiol Biotechnol 63:613–625. https://doi.org/10.1007/s00253-003-1477-2

Wendler S, Otto A, Ortseifen V, Bonn F, Neshat A, Schneiker-Bekel S, Walter F, Wolf T, Zemke T, Wehmeier UF, Hecker M, Kalinowski J, Becher D, Pühler A (2015) Comprehensive proteome analysis of *Actinoplanes* sp. SE50/110 highlighting the location of proteins encoded by the acarbose and the pyochelin biosynthesis gene cluster. J Proteome 125:1–16. https://doi.org/10.1016/j.jprot.2015.04.013

Wendler S, Otto A, Ortseifen V, Bonn F, Neshat A, Schneiker-Bekel S, Wolf T, Zemke T, Wehmeier UF, Hecker M, Kalinowski J, Becher D, Pühler A (2016) Comparative proteome analysis of *Actinoplanes* sp. SE50/110 grown with maltose or glucose shows minor differences for acarbose biosynthesis proteins but major differences for saccharide transporters. J Proteome 131:140–148. https://doi.org/10.1016/j.jprot.2015.10.023

Wolf T, Gren T, Thieme E, Wibberg D, Zemke T, Pühler A, Kalinowski J (2016) Targeted genome editing in the rare actinomycete *Actinoplanes* sp. SE50/110 by using the CRISPR/Cas9 system. J Biotechnol 231:122–128. https://doi.org/10.1016/j.jbiotec.2016.05.039

Wolf T, Droste J, Gren T, Ortseifen V, Schneiker-Bekel S, Zemke T, Pühler A, Kalinowski J (2017a) The MalR type regulator AcrC is a transcriptional repressor of acarbose biosynthetic genes in *Actinoplanes* sp. SE50/110. BMC Genomics 18:562. https://doi.org/10.1186/s12864-017-3941-x

Wolf T, Schneiker-Bekel S, Neshat A, Ortseifen V, Wibberg D, Zemke T, Pühler A, Kalinowski J (2017b) Genome improvement of the acarbose producer *Actinoplanes* sp. SE50/110 and annotation refinement based on RNA-seq analysis. J Biotechnol 251:112–123. https://doi.org/10.1016/j.jbiotec.2017.04.013