Universidad de Granada – Fundación MEDINA

Doctoral Programme in Pharmacy

Genome Mining of the Strain *Streptomyces* sp. CA-170360: Identification and Heterologous Expression of Secondary Metabolite Biosynthetic Gene Clusters

Fernando Román Hurtado

PhD Thesis

2022

Thesis Directors:

Olga Genilloud Rodríguez

Marina Sánchez Hidalgo

Editor: Universidad de Granada. Tesis Doctorales
Autor: Fernando Román Hurtado
ISBN: 978-84-1117-370-4
URI: http://hdl.handle.net/10481/75457

Universidad de Granada – Fundación MEDINA

Programa de Doctorado en Farmacia

Análisis del Genoma de la Cepa *Streptomyces* sp. CA-170360: Identificación y Expresión Heteróloga de Rutas de Biosíntesis de Metabolitos Secundarios

Fernando Román Hurtado

Tesis Doctoral

2022

Dirigida por las Doctoras:

Olga Genilloud Rodríguez

Marina Sánchez Hidalgo

La presente Tesis Doctoral se ha llevado a cabo en los Departamentos de Microbiología, Química y Screening de Fundación MEDINA bajo la supervisión de las Doctoras Olga Genilloud Rodríguez y Marina Sánchez Hidalgo, en la línea de Investigación de Nuevas Dianas Terapéuticas del Programa de Doctorado en Farmacia (B15.56.1) de la Escuela de Doctorado de Ciencias de la Salud de la Universidad de Granada. Este trabajo ha sido financiado por el proyecto IIMENA (*Integrating Informatics and Metabolic Engineering for the Biosynthesis of Novel Antibiotics*) financiado por el Challenge Grant de la Novo Nordisk Foundation, liderado por el grupo del Dr. Tilmann Weber en la Universidad Técnica de Dinamarca (DTU) y en colaboración con el grupo del Dr. Sang Yup Lee del Instituto Avanzado de Ciencia y Tecnología de Corea (KAIST).

Durante el desarrollo de esta Tesis Doctoral, se realizó una estancia durante tres meses, del 10 de septiembre de 2018 al 7 de diciembre de 2018, en el Novo Nordisk Foundation Center for Biosustainability, de la Universidad Técnica de Dinamarca, bajo la supervisión de los Doctores Tilmann Weber y Yaojun Tong.

Los resultados de esta Tesis Doctoral han sido publicados en las siguientes revistas científicas:

- **Román-Hurtado, F., Sánchez-Hidalgo, M., Martín, J., Ortiz-López, F. J. and Genilloud, O.** (2021). Biosynthesis and Heterologous Expression of Cacaoidin, the First Member of the Lanthidin Family of RiPPs. *Antibiotics*. **10**:403. doi: 10.3390/antibiotics10040403.

- **Román-Hurtado, F., Sánchez-Hidalgo, M., Martín, J., Ortiz-López, F. J., Carretero-Molina, D., Reyes, F. and Genilloud, O.** (2021). One Pathway, Two Cyclic Non-Ribosomal Pentapeptides: Heterologous Expression of BE-18257 Antibiotics and Pentaminomycins from *Streptomyces cacaoi* CA-170360. *Microorganisms*. **9**:135. doi: 10.3390/microorganisms9010135.

Además, los resultados derivados de esta Tesis Doctoral han sido presentados en los siguientes congresos:

- Identificación de la ruta biosintética de un nuevo glicolantipéptido producido por *Streptomyces* sp. CA-170360 mediante *genome mining*. **Fernando Román-Hurtado**, Marina Sánchez-Hidalgo, Francisco Javier Ortiz-López y Olga Genilloud. XXVII Congreso de la Sociedad Española de Microbiología 2019 (SEM2019), en Málaga, 2019.

- Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in MEDINA microbial collection. Marina Sánchez-Hidalgo, Daniel Oves-Costales, **Fernando Román-Hurtado**, Olga Genilloud. I Congreso de Investigadores del PTS, en Granada, 2019.

- Identification and characterization of the biosynthetic gene cluster of the unprecedented glycosylated lanthipeptide MDN-0207. Marina Sánchez-Hidalgo, **Fernando Román-**

Hurtado, Francisco Javier Ortiz-López, Olga Genilloud. 14th International Symposium on the Genetics of Industrial Microorganisms (GIM 2019), Pisa (Italia), 2019.

- Identificación y expresión heteróloga de la ruta biosintética de cacaoidina, el primer lantipéptido de clase V. **Fernando Román-Hurtado**, Marina Sánchez-Hidalgo, Francisco Javier Ortiz-López, Jesús Martín y Olga Genilloud. II Congreso de Investigadores del PTS, en Granada, 2022.
- Genome Mining at Fundación MEDINA: Identification of the biosynthetic gene clusters of promising molecules. Marina Sánchez-Hidalgo, **Fernando Román-Hurtado**, Daniel Oves-Costales, José Miguel Quesada y Olga Genilloud. II Congreso de Investigadores del PTS, en Granada, 2022.

Además de las publicaciones directamente relacionadas con la Tesis Doctoral, el doctorando ha participado como coautor en las siguientes publicaciones:

- **Ortiz-López, F. J., Carretero-Molina, D., Sánchez-Hidalgo, M., Martín, J., González, I., Román-Hurtado, F., de la Cruz, M., García-Fernández, S., Reyes, F., Deisinger, J., Schneider, T. and Genilloud, O.** (2020). Cacaoidin, first member of the new lanthidin RiPP family. *Angew. Chem. Int.* **59**:12654-12658; 10.1002/anie.202005187.
- **Carretero-Molina, D., Ortiz-López, F. J., Martín, J., González, I., Sánchez-Hidalgo, M., Román-Hurtado, F., Díaz, C., de la Cruz, M., Genilloud, O. and Reyes, F.** (2021). Pentaminomycins F and G, Nonribosomal Peptides Containing 2-Pyridylalanine*. J. Nat. Prod.* **84**:1127-1134. doi: 10.1021/acs.jnatprod.0c01199.
- **Beck, C., Gren, T., Jorgensen, T. S., González, I., Román-Hurtado, F., Oves-Costales, D., Genilloud, O. and Weber, T.** (2021). Complete genome sequence of *Streptomyces* sp. strain CA-256286. *Microbiol. Resour. Announc.* **10**:e0029021. doi: 10.1128/MRA.00290-21.
- **Jorgensen, T. S., Gren, T., Kontou, E. E., González, I., Román-Hurtado, F., Oves-Costales, D., Thomsen, E., Charusanti, P., Genilloud, O. and Weber, T.** (2021). Complete genome sequence of the rare actinobacterium *Kutzneira* sp. strain CA-103260. *Microbiol. Resour. Announc.* **10**:e0049921. doi: 10.1128/MRA.00499-21.
- **Carretero-Molina, D., Ortiz-López, F. J., Gren, T., Oves-Costales, D., Martín, J., Román-Hurtado, F., Jorgensen, T. S., de la Cruz, M., Díaz, C., Vicente, F., Blin, K., Reyes, F., Weber, T. and Genilloud, O.** (2022). Discovery of gargantulides B and C, new 52-membered macrolactones from *Amycolatopsis* sp. Complete absolute stereochemistry of the gargantulide family. *Org. Chem. Front*. **9**:462-470. doi: 10.1039/D1QO01480C.

ABSTRACT

The emergence of antibiotic resistant pathogenic strains, especially Gram-negative bacteria, has seriously increased in the last decades, endangering the efficiency of the antibiotics used thus far. As a result, the discovery of new molecules with potential antimicrobial activity has become an essential matter. Historically, actinomycetes, particularly species from the genus *Streptomyces*, have been one of the most prolific sources of novel antibiotics.

The strain *Streptomyces cacaoi* CA-170360 from the microbial collection of Fundación MEDINA produces the novel cyclic pentapeptides pentaminomycins A-H, alongside the already known cyclic pentapeptides BE-18257 antibiotics, with a moderate antibacterial activity against *Acinetobacter baumannii*. This strain also produces cacaoidin, the first member of the new class V lanthipeptides (or lanthidins) RiPP family, with bioactivity against Gram-positive pathogens, such as Methicillin-Resistant *Staphylococcus aureus* (MRSA) and *Clostridium difficile*. The genome of this strain was sequenced and analyzed, and the biosynthetic gene clusters responsible of the production of these compounds were identified, together with other biosynthetic gene clusters involved in the production of different known secondary metabolites detected in the fermentations of this strain.

Pentaminomycins A-H and BE-18257 antibiotics are two different families of cyclopeptides synthesized by two independent non-ribosomal peptide synthetases encoded in tandem within the same biosynthetic gene cluster (*cpp*). The cluster lacks a thiosterase domain to release and cyclize the pentapeptides; however, it contains a stand-alone PBP-type protein with a beta-lactamase conserved domain (CppA) upstream the first NRPS gene. The heterologous expression of the *cpp* cluster in the host *Streptomyces albidoflavus* J1074 confirmed its implication in the biosynthesis of both pentaminomycins and BE-18257 antibiotics, while the generation of a knockout by genetic replacement of *cppA* demonstrated the involvement of this protein in the release and cyclization of the peptide chains of both cyclopeptide families.

Cacaoidin is a novel glycosylated lanthipeptide with remarkable unprecedented structural features, such as an unusually high number of D-amino acids, an *N,N*-dimethyl lanthionine system (NMe_2Lan), not previously found in known lanthipeptides, and an O-glycosylated Tyr residue with a not previously reported disaccharide formed by α-L-rhamnose and β-L-6-deoxygulose. The available predictive tools were not able to detect the biosynthetic gene cluster responsible of the production of cacaoidin, which was found using the C-terminal amino acid sequence of the structural peptide. The final *cao* cluster was determided after BLAST analysis and HHpred secondary structure prediction. The heterologous expression of the cluster in *S. albidoflavus* J1074 confirmed the involvement of the cluster in the biosynthesis of cacaoidin and led to the identification of a new variant of cacaoidin with a different disaccharide lacking two oxygen atoms (cacaoidin-2O). The generation of knockout strains in the heterologous host by genetic replacement allowed the assignment of different roles in several genes in the biosynthesis of cacaoidin: *cao4* is involved in the *N,N*-dimethylation of the N-terminal Ala residue and *cao8* and

Abstract

cao16 are two glycosyltransferases working cooperatively in the Tyr O-glycosylation. The generation of the aglycon of cacaoidin was achieved by the controlled acid hydrolysis of pure cacaoidin-2O.

RESUMEN

La aparición de cepas patógenas resistentes a antibióticos, especialmente bacterias Gram-negativas, ha aumentado enormemente en los últimos años, comprometiendo la eficacia de los antibióticos usados hasta el momento. Por ello, la búsqueda de nuevas moléculas con actividad antimicrobiana se ha convertido en una tarea indispensable. Históricamente, los actinomicetos, sobre todo las especies del género *Streptomyces*, han sido una de las grandes fuentes de nuevos antibióticos.

La cepa *Streptomyces cacaoi* CA-170360, perteneciente a la colección microbiana de Fundación MEDINA, produce una familia de nuevos ciclopentapéptidos, las pentaminomicinas A-H, junto con los ya conocidos antibióticos BE-18257, también ciclopentapéptidos, con una moderada actividad frente *Acinetobacter baumannii.* Esta cepa, además, produce cacaoidina, el primer lantipéptido de clase V (lantidina) descrito, con actividad frente a patógenos Gram-positivos, como *Staphylococcus aureus* resistente a meticilina (MRSA) y *Clostridium difficile*. El genoma de esta cepa fue secuenciado y analizado, llegando a identificar las rutas biosintéticas responsables de la producción de ambas familias de compuestos, así como otras de rutas de biosíntesis de metabolitos secundarios ya conocidos y detectados en las fermentaciones de la cepa.

Las pentaminomicinas A-H y los antibióticos BE-18257 son dos familias de ciclopentapéptidos sintetizados por dos sintetasas de péptidos no ribosomales (NRPS) independientes pero codificadas conjuntamente en la misma ruta biosintética (*cpp*). La ruta carece de un dominio tioesterasa necesario para la ciclación y liberación de los pentapéptidos, pero contiene una proteína aislada tipo PBP con un dominio beta-lactamasa conservado (CppA) aguas arriba de la primera NRPS. La expresión heteróloga de la ruta *cpp* en el hospedador *Streptomyces albidoflavus* J1074 confirmó que es la responsable de la biosíntesis tanto de las pentaminomicinas A-H como de los antibióticos BE-18257, y la generación del mutante por reemplazamiento génico del gen *cppA* demostró su implicación en la liberación y ciclación de ambas familias de pentapéptidos.

La cacaoidina es un nuevo lantipéptido glicosilado con unas características estructurales no descritas hasta el momento, como un número inusualmente elevado de D-aminoácidos, un anillo de lantionina dimetilado (NMe_2Lan), no encontrado en otros lantipéptidos, y una O-glicosilación de un residuo de tirosina con un disacárido no descrito hasta la fecha en ningún otro producto natural, formado por α-L-ramnosa y β-L-6-deoxigulosa. Las diferentes herramientas predictivas disponibles no fueron capaces de identificar la ruta biosintética responsable de la producción de cacaoidina, que se localizó usando la secuencia C-terminal de aminoácidos y encontrando el gen estructural en el genoma. Finalmente, la ruta *cao* se delimitó mediante análisis BLAST y de predicción de estructuras secundarias HHPred. Su expresión heteróloga en *S. albidoflavus* J1074 confirmó que es la ruta responsable de la producción de cacaoidina y llevó a la identificación de una nueva variante de cacaoidina, glicosilada con un disacárido diferente que tiene dos átomos menos de oxígeno (cacaoidina-2O). La generación de mutantes en el hospedador heterólogo nos

permitió asignar las funciones de diferentes genes en la biosíntesis de cacaoidina: el gen *cao4* participa en la doble metilación del residuo de Ala en el extremo amino-terminal, y los genes *cao8* y *cao16* son dos glicosiltransferasas que trabajan de forma cooperativa en la glicosilación del residuo de Tyr. Finalmente, la generación de aglicón se consiguió mediante la hidrólisis ácida controlada de cacaoidina-2O.

Index

Figures and Tables Index

A. INTRODUCTION

C. MATERIALS AND METHODS

D. RESULTS

A. Introduction

1. NATURAL PRODUCTS IN DRUG DISCOVERY

Natural products (NPs), commonly known as secondary metabolites, are a very broad and diverse group of non-essential bioactive molecules with low molecular weight produced by different organisms such as bacteria, fungi, lichens, marine invertebrates, plants, insects and mammals (Cannell, 1998). Among them, microorganisms are one the richest sources of bioactive natural products, derived from their secondary metabolism (Woodruff, 1980). Natural products show an enormous scaffold diversity and structural complexity which confer them a broad range of pharmacological activities and advantages for the drug discovery process, in comparison with synthetic small-molecule libraries (Lachance *et al.*, 2012). In fact, many natural products have reached the market without chemical modification, which highlights the ability of microorganisms to produce small, drug-like molecules.

Historically, natural products have been developed to be used in human health as antibiotics, antifungals, anticancer, immunosuppressants and antiparasitic drugs (Berdy *et al.,* 2012; Hwang *et al.*, 2014). Indeed, they have played such a remarkable role in the treatment of cancer and infectious diseases (Atanasov *et al.*, 2015), but also in other therapeutic areas, including cardiovascular diseases (Waltenberger *et al.*, 2016) or multiple sclerosis (Tintore *et al.*, 2019), just to name a few of them.

Since the discovery and production of the life-saving antibiotic penicillin by Alexander Fleming in 1928, a golden era ranged from the 1930s to 1960s and gave rise to many new antibiotics (Nathan and Cars, 2014), that revolutionized medicine and saved countless lives, preventing and treating bacterial infections (Piddock, 2012). Unfortunately, this rise of antibiotics was accompanied by the increasing appearance of pathogenic resistant strains, compromising their efficacy and making bacterial infections a serious threat (Spellberg and Gillbert, 2014). Numerous international organizations have declared antibiotic resistance as a global public health of concern (Spellberg *et al.*, 2016) as infectious diseases are now the second leading death cause in the world, reaching 4.95 million deaths associated with antimicrobial resistance in 2019 (Murray *et al*., 2022). This emerging antibiotic resistance crisis is mostly due to an increased overuse and misuse of antibiotics in the clinic and animal production, poor sanitation and hygiene, wildlife spread and enhanced global migration, in a context where the pharmaceutical industry has abandoned the research and development in new classes of antibiotics (Read and Woods, 2014; Singer et al, 2016).

All classes of antibiotics have been affected by the emergence of resistant bacteria, and cross-resistances have rapidly increased among antibiotics belonging to the same class of drugs (Cândido *et al*., 2019). Today, multidrug resistance represents one of the major challenges to treat infectious diseases, both in community and hospital-acquired infections, and limits the management of bacterial infections. Most research efforts were focused for many years on pathogenic Gram-positive bacteria, especially *Staphylococcus aureus* MRSA, but very limited options have been developed against multidrug resistant Gram-negative bacteria, frequently

associated with high mortality rates and fewer treatment options in hospital-acquired infections (Ayobami *et al*., 2022).

This situation makes critical the discovery of new molecules with potential antimicrobial activity. However, natural products research programs were widely abandoned by the pharmaceutical industry due to severe rediscovery issues, challenging isolation and low production concentrations of new producers (Hug *et al.*, 2018) and were replaced without much success by alternative approaches such as combinatorial chemical libraries (Silver, 2011). Though, this resulted unsuccessful as the chemistry was not able to create enough diversity of pharmacologically bioactive molecules (Berdy, 2012). Natural products represent an arsenal of compounds with unique structures that cannot be supplanted by the combinatorial chemistry. In fact, they are still an important source for the discovery of novel antibiotic drugs (Dai *et al.*, 2020; Newman and Cragg, 2020).

Among all the microorganisms, the filamentous bacteria actinomycetes are amazingly prolific in the number of natural products with antimicrobial activity which they can produce. About 75 % of known antibiotics are produced by actinomycetes and about 75 % of these are made by members of the genus *Streptomyces*. From all antibiotics used in medicine, more than 90 % were produced by actinomycetes (Martens and Demain, 2017), which makes them a promising source for the discovery of antibiotics.

2. ACTINOMYCETES

The actinomycetes are a heterogeneous group of Gram-positive, aerobic and filamenting bacteria with high G+C DNA content included in the class Actinobacteria (Ventura *et al*., 2007; Gao and Gupta, 2012). Their taxonomic description based on chemotaxonomic and micromorphological aspects underwent a significant rise before the 1980s (Labeda, 1987) and defined the criteria used for their classification until today. From the 1980s, the huge development of new molecular biology techniques allowed the improvement of phylogenetic analyses based on ribosomal rDNA sequences (Olsen *et al.*, 1986), resulting in an important progress in the study of the phylogenetic position of actinomycetes. The availability of the whole genome sequencing of several bacterial species led to a massive improvement in the study of actinomycetes.

The last published edition of the Bergey's Manual of Systematic Bacteriology contains a volume exclusively dedicated to the Phylum *Actinobacteria*, which is classified in 6 classes, 22 orders, 53 families and 222 genera (Goodfellow *et al*., 2012). Within the class Actinobacteria, 15 different orders are found: Actinomycetales, Actinopolysporales, Bifidobacteriales, Catenulisporales, Corynebacteriales, Frankiales, Glycomycetales, Jiangellales, Kineosporiales, Micrococcales, Micromonosporales, Propionibacteriales, Pseudonocardiales, Streptomycetales and Streptosporangiales.

The actinomycetes are widely distributed in soils, sediments, and mainly in plant-associated rhizospheres, and are ubiquitous in aquatic and terrestrial ecosystems, where they have been shown to play an important ecological role in the nutrients recycling. They are saprophytic microorganisms and can colonize all kinds of substrates, invertebrates, decomposing plant materials, from a broad range of marine and terrestrial, desert and extreme ecosystems. Actinomycetes share their habitat with other bacteria and fungi, competing for the nutrients and producing a great diversity of compounds, in response to different stimuli and stresses. Several actinomycete species can form parasitic or symbiotic associations with animals or plants (van der Meij *et al*., 2017).

These bacteria are distinguished from the rest by their complex biological cycle, with an initial formation of a vegetative mycelium and the subsequent differentiation with the formation of an aerial mycelium and sporulation, whereby morphological development and secondary metabolism often coincide (Yagüe *et al.*, 2013).

Their extraordinary metabolism and production of secondary metabolites have made them one of the most prolific sources of novel antibiotics, including among others the most important antimicrobial drug classes of β-lactams, tetracyclines, rifamycins, aminoglycosides, macrolides and glycopeptides. The actinomycetes, and particularly the species from the genus *Streptomyces*, have proved to be a tremendous high-impact source of valuable chemicals. Approximately two-thirds of all known antibiotics are predominantly produced by *Streptomyces* (Barka *et al.*, 2015).

However, in response to the growing needs of drug discovery programs for new bioactive compounds, the focus in microbial natural products discovery has shifted towards minor taxa within the actinomycetes, frequently named as rare genera or non-*Streptomyces*, or novel *Streptomyces* isolates found in underexplored habitats, especially the marine and extreme environments, which have already afforded the discovery of novel antimicrobials with unique chemical moieties (Hozzein *et al.*, 2011; Tiwari *et al.*, 2014). It is believed that around 16 % of the total number of bioactive compounds are produced by rare actinomycetes, mostly from members of the family *Micromonosporaceae*, with additional smaller contributions from the *Pseudonocardiaceae* and *Thermomonosporaceae* (Tiwari and Gupta, 2012). In particular, actinomycetes belonging to some of these minor genera such as *Micromonospora*, *Nocardia*, *Actinomadura*, *Actinoplanes*, and *Saccharopolyspora* have been found to produce chemically unique antibiotics featuring potent activities (Tiwari and Gupta, 2012). The discovery of gentamycin from *Micromonospora purpurea* in 1963 (Weinstein *et al.*, 1963) triggered the search for new compounds from the so-called “rare actinomycetes”, which are obtained in lower frequencies than other species of the genus *Streptomyces*. Relevant compounds produced by these rare actinomycetes include erythromycin (from *Sacchapolyspora erythraea* in 1952) (McGuire *et al*., 1952), vancomycin (from *Amycolatopsis orientalis* in 1955) (McCormick *et al*., 1955), teicoplanin (from *Actinoplanes teichomyceticus* in 1978) (Parenti *et al*., 1978), or the potent lantibiotic microbisporicin, also named NAI107 (from *Microbispora corallina* in 2005) (Castiglione *et al*., 2008). Furthermore, salinosporamide A, which exhibits anticancer activity and

is currently in a phase III clinical trial under the name of marizomib for patients with gliobastoma (Roth *et al*., 2019), is produced by a strain of the genus *Salinispora*, an obligate marine actinomycete isolated from a heat-treated marine sediment sample (Feling *et al.*, 2003), and abyssomicin (Bister *et al.*, 2004) and proximicins (Fiedler *et al.*, 2008) are produced by *Verrucosispora* strains. These results support that these rare actinomycetes are a valuable source of novel compounds, and that improved isolation strategies are required to increase the frequency in which they are isolated (Takahashi and Nakashima, 2018).

2.1. The genus *Streptomyces*

The first reference to the *Streptomyces* genus was given in 1875 by the polish botanic and bacteriologist Ferdinand Julius Cohn (Cohn, 1875). Based on its filamentous aspect, he named it *Streptothrix* (rolled hair), but in 1943, this denomination changed to the currently used, *Streptomyces* (rolled fungi) (Waksman and Henrici, 1943).

The current taxonomy of the *Streptomyces* genus is: Domain Bacteria; Phylum XXVI. Actinobacteria; Class I. Actinobacteria; Order XIV. Streptomycetales; Family I. *Streptomycetaceae*; Genus I. *Streptomyces*. According to the List of Prokaryotic names with Standing in Nomenclature (LPSN), 834 type strains belonging to the *Streptomyces* genus have been validly published under the ICNP (International Code of Nomenclature of Prokaryotes) to date (Parte *et al.,* 2020).

The species of the genus *Streptomyces* are distributed mainly in diverse habitat of soil, with concentrations between 10^5 and 10^8 cfu/g. Their ability to colonize the soils is due to its capability to produce a wide variety of organic compounds using different carbon and nitrogen sources. In fact, their “moist soil smell” is caused by the production of the volatile metabolite geosmine. Also, their biological cycle, where spores constitute a semi-lethargy state, allows their survival in the soils for very long periods of nutrients and water shortage (McCarthy and Williams, 1992).

2.1.1. Streptomyces *biological cycle*

The complex biological cycle of *Streptomyces* species presents several steps characterized by changes in both its morphology and physiology. The life cycle begins with the germination of a spore under favorable conditions of O_2, nutrients and water. The energy for this process is obtained from the degradation of the endogenous trehalose disaccharide spore reservoir (Hirsch and Ensign, 1978). From this germinated spore, the non-septate germ tube grows and generates a truss of branched multinucleated hyphae which are inserted into the substrate to be nurtured. The synthesis of RNA and proteins begins immediately with the germination of the spore and the synthesis of DNA starts with the growing of the germinative tube (Hardisson *et al.*, 1978). This classic model was based in the idea that the differentiation is produced vertically (from bottom to top) and implies the formation of two different types of mycelia. First, the vegetative or

substrate mycelium that grows inside and on the agar until it undergoes a death process, after which the hyphae start to differentiate to form a reproductive mycelium (aerial) that grows in the air to finally result in the formation of individual spores (Fernández and Sánchez, 2002) (Figure A1). This aerial mycelium metabolizes the material arising from the substrate mycelium as nutrients source (Braña *et al.*, 1986). When the growth of the multinucleated hyphae ceases, septa are formed in the ends of the hyphae, producing mononucleated compartments or pre-spores. These pre-spores are coated by a fibrous layer, originating individual, pigmented and hydrophobic spores (Flärdh, 2003). The spores constitute a lethargy state and have the capability to survive for long periods of time until the environmental conditions provide the humidity and optimal nutrient concentrations for their germination. The substrate and aerial mycelia are multinucleated, and the spores are the only uninucleate stage of the *Streptomyces* life cycle (Claessen *et al.*, 2006). Traditionally, it was assumed that the growth was homogenous in the whole surface of the plate, but subsequent studies have modified this accepted classical model (Manteca and Sánchez, 2009).

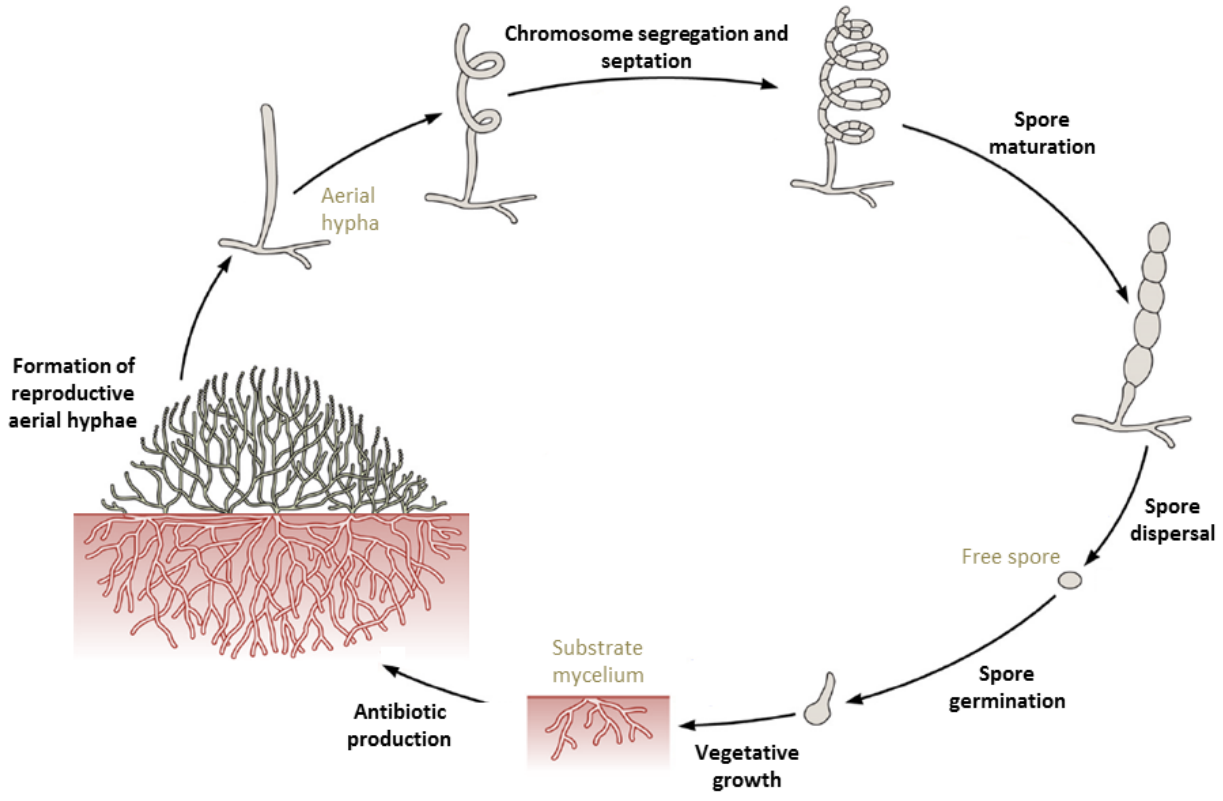


Figure A1. Schematic representation of the life cycle of sporulating actinomycetes. Figure adapted from Barka et al. *(2016).*

Transcriptomic and proteomic studies proved that there are previously unknown steps in the process of differentiation (Manteca *et al*., 2007; Yagüe *et al*., 2013): once the spores are germinated, a primary young and septate mycelium (MI) is developed, where alive and dead cells coexist due to a process of programmed cellular death. This compartmentalized MI is the real vegetative mycelium of *Streptomyces*, where the primary metabolism is developed and conjugation, recombination and mutagenesis mechanisms are activated. A programmed cell death program occurs from internal to the external parts of a colony, and the peripheral cells make way for the secondary multinucleated mycelium (MII), in which two stages are identified. In the first stage, defined as early MII, the mycelium grows inside the culture medium (this corresponds with the traditional substrate mycelium) until it starts to generate an hydrophobic enveloping layer which allows the mycelium to grow outside the culture medium, forming the late MII (aerial mycelium) (Manteca *et al.*, 2007). Both early and late MII are the reproductive mycelia involved in the sporulation and production of secondary metabolites and are the stages when the mechanisms of genetic variability are activated, transposons and insertion sequences, previous to the sporulation stage. Before sporulation takes place, a second programmed cell death happens which affects both the substrate and the aerial mycelia. During the MII stage many antibiotics (secondary metabolites) are produced to ensure the survival of the bacteria as defense against other microorganisms (Yagüe *et al.*, 2013) (Figure A2). All the studies of this cell cycle have been carried out in solid media, as most of the *Streptomyces* species do not sporulate in liquid media, although *S. griseus*, *S. acrymicini* or *S. albus* are exceptions (Novella *et al.*, 1992).

2.1.2. Secondary metabolites from Streptomyces

Streptomyces species produce about 75 % of the secondary metabolites found in actinomycetes (Barka *et al.*, 2015). Many of these secondary metabolites have antibiotic activity and more than 50 % of all known natural antibiotics are synthesized by bacteria from the *Streptomyces* genus (Berdy, 2005). After the discovery of streptomycin from *S. griseus* in 1944, a long list of examples of the success of this genus as a source of antimicrobials has been generated.

From a structural point of view, the antibiotics produced by *Streptomyces* belong to different classes of antibiotics, such as β-lactams, aminoglycosides, macrolides, peptides, tetracyclines, lincosamides, ansamycins, epoxides, aminocoumarines and nucleosides (De Simeis and Serra, 2021). (Table A1).

Furthermore, other secondary metabolites produced by *Streptomyces* have found different applications as antitumoral agents (bleomycin, actinomycin D, doxorubicin), immunosuppressors (tacrolimus, rapamycin, ascomycin), antifungals (amphotericin B, nystatin) and antihelmintic agents (avermectin, milbemycin), herbicides (phosphotricin) and growing promoters in plants (pteridic acids A and B) and siderophores (desferroxiamine B) (Weber *et al.*, 2003; Martín, 2004; Igarashi *et al*., 2002). This wide variety of natural products produced by *Streptomyces* shows a

huge structural diversity including polyketides, alkaloids, terpenoids and peptides, both ribosomally and non-ribosomally synthesized.

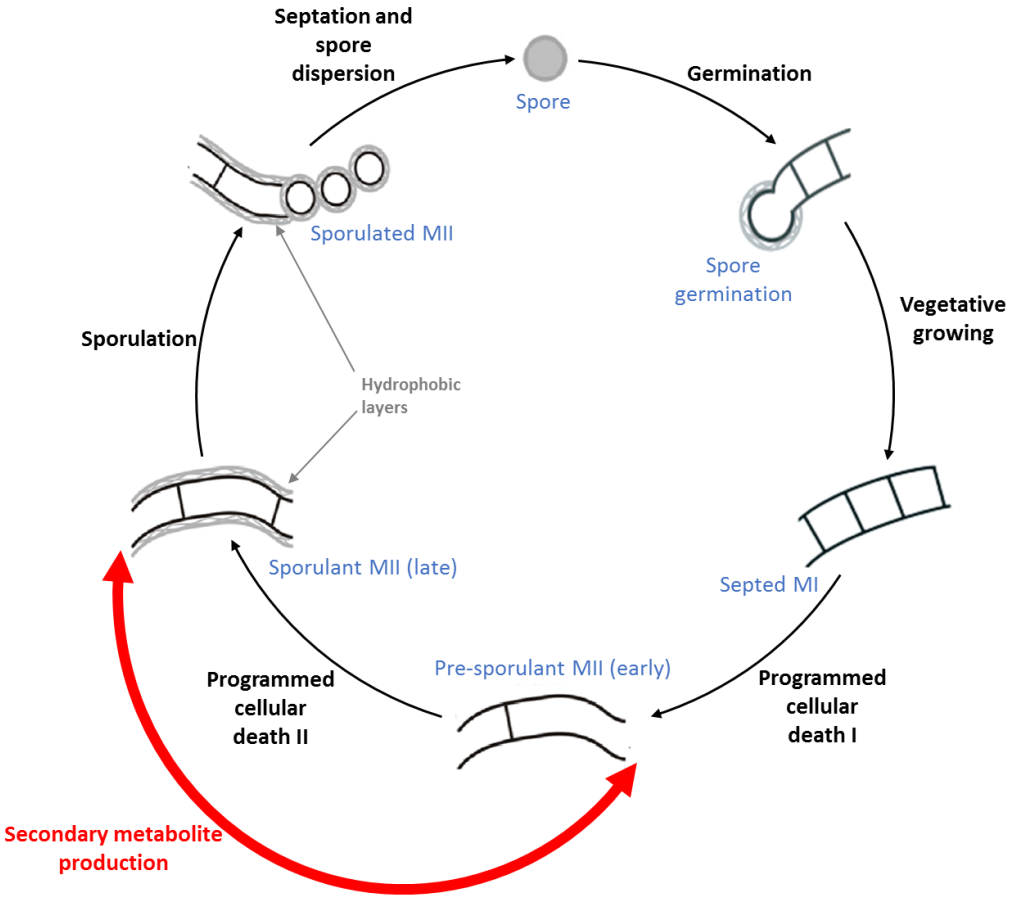


Figure A2. Updated Streptomyces *biological cycle proposed by Manteca and Sánchez (2009). Figure modified from Yagüe* et al. *(2013).*

Class	Subclass	Example	Producer *Streptomyces* species
β-Lactams	Penicillin	Penicillin N	*S. lipmanii; S. microflavus; S. griseus; S. cattleya*
	Cephalosporine	Cephalosporin C	*S. clavuligerus; S. microflavus; S. lipmanii, S. hygroscopus*
	Carbapenems	Thienamycins	*S. cattleya; S. griseus; S. penemifaciens; S. flavogriseus; S. olivaceus; S. cremeus; S. flavoviridis*
	Monobactam	Nocardicin	*S. alcalophilus*
	PBP inhibitors	Clavulanic acid	*S. clavuligerus; S. jumonjinesis; S. katsuharamanus*
		Olivanic acid	*S. olivaceus; S. griseus*
Macrolides		Oleandomycin	*S. antibioticus*
		Spiramycin	*S. ambofaciens*
		Erythromycin A	*S. erythraeus*
		Josamycin	*S. narbonensis*
		Midecamycin	*S. mycarofaciens*
		Rapamycin	*S. hygroscopicus*
		Leucomycin	*S. kitasatoensis*
Tetracyclines		Tetracycline	*S. aureofaciens; S. rimosus; S. avellaneus; S. lusitanus; S. viridifaciens*
		Chlortetracycline	*S. aureofaciens; S. lusitanus; S. lividans*
		Oxytetracycline	*S. alboflavus; S. albofaciens; A. erumpens; S. griseus; S. platensis; S. rimosus; S. varsoviensis*
		6-demethyltetracycline	*S. aureofaciens*
Aminoglycosides		Streptomycin	*S. griseus; S. bikiniensis; S. streptomycinii; S. ornatus; S. humidus; S. subrutilus; S. mashuensis; S. glaucescens; S. griseocarnus; S. rimosus; S. albus*
		Neomycin	*S. fradiae; S. catenulae; S. chrestomyceticus; S. albogriseolus*
		Kanamycin	*S. kanamyceticus*
		Paromomycym	*S. rimosus; S. catenulae; S. chrestomyceticus; S. fradiae*
		Tobramycin	*S. tenebrarius; S cremeus*
Peptides	Glycopeptide	Mannopeptimycin	*S. hygroscopicus*
	Lipopeptides	Daptomycin	*S. roseosporus; S. lividans*
	Streptogramins	Streptogramins A and B	*S. virginiae; S. halstedii; S. pristinaespiralis*
Aminocumarines		Novobiocin	*S. spheroids; S niveus; S. caeruleus*
Aminonucleosides		Puromycin	*S. alboniger*
Amphenicol		Chloramphenicol	*S. venezuelae*
Lincosamides		Lincomycin	*S. lincolnensis*
Epoxide		Fosfomycin	*S. fradiae; S. viridochromogenes; S. wedmorensis; S. lividans*

Table A1. Examples of antibiotics produced by different species of Streptomyces*. Table adapted from De Simeis and Serra, 2021.*

3. BIOSYNTHESIS OF PEPTIDIC NATURAL PRODUCTS IN ACTINOMYCETES

The genes for the biosynthesis of secondary metabolites are organized in biosynthetic gene clusters (BGCs) that encode the biosynthetic enzymes, regulatory genes, resistance genes to ensure self-protection from bioactive metabolites, and transporter genes for the release of the metabolite outside of the cell (Martín and Liras, 1989). The production of secondary metabolites is highly regulated and frequently associated with the processes of cell differentiation and growth arrest, as a consequence of the depletion of essential nutrients such as carbon, nitrogen or phosphate sources. In liquid fermentations they are frequently produced in the late exponential phase or the stationary growth phase. Additionally, stress factors, like thermal stress, radiation, acidification or variations in oxygen concentration can have an important influence on their production (Martín and Demain, 1980; von Döhren and Gräfe, 1997).

Among the natural products produced by actinomycetes, many important therapeutic agents, ranging from antibiotics, anti-inflammatory, anti-tumor or anti-depressants, correspond to peptidic molecules that and can be divided in two types according to their mechanism of biosynthesis (Dang and Süssmuth, 2017; Neethu *et al.,* 2022) (Figure A3). The first type include the Ribosomally Synthesized and Post-translationally modified Peptides (RiPPs), that have been attracting interest as a prolific source of untapped antimicrobials (Montalbán-López, 2021), such as duramycin, a class II lanthipeptide produced by *Streptomyces cinnamoneus* used as a potent plant pathogen controlling agent (Goethals *et al*., 2001). Duramycin has also proved to be useful in cystic fibrosis clearing phase II clinical trials, and has a wide range of efficacy as an anti-viral and anti-tumor agent (Richard *et al*., 2015; Yates *et al*., 2012). On the other hand, the non-ribosomal peptides (NRP) are synthesized by modular Non-Ribosomal Peptide Synthetases (NRPS) (Süssmuth and Mainz, 2017), like daptomycin, a lipopeptide produced by *Streptomyces roseosporus* used for the treatment of infections caused by Gram-positive pathogens like MRSA (Robbel and Marahiel, 2010).

Peptide natural product research also faces the challenge to discover new structural diversity and bioactivity against novel molecular targets, helped by the full understanding and engineering of the peptide biosynthetic machineries (Dang and Süssmuth, 2017).

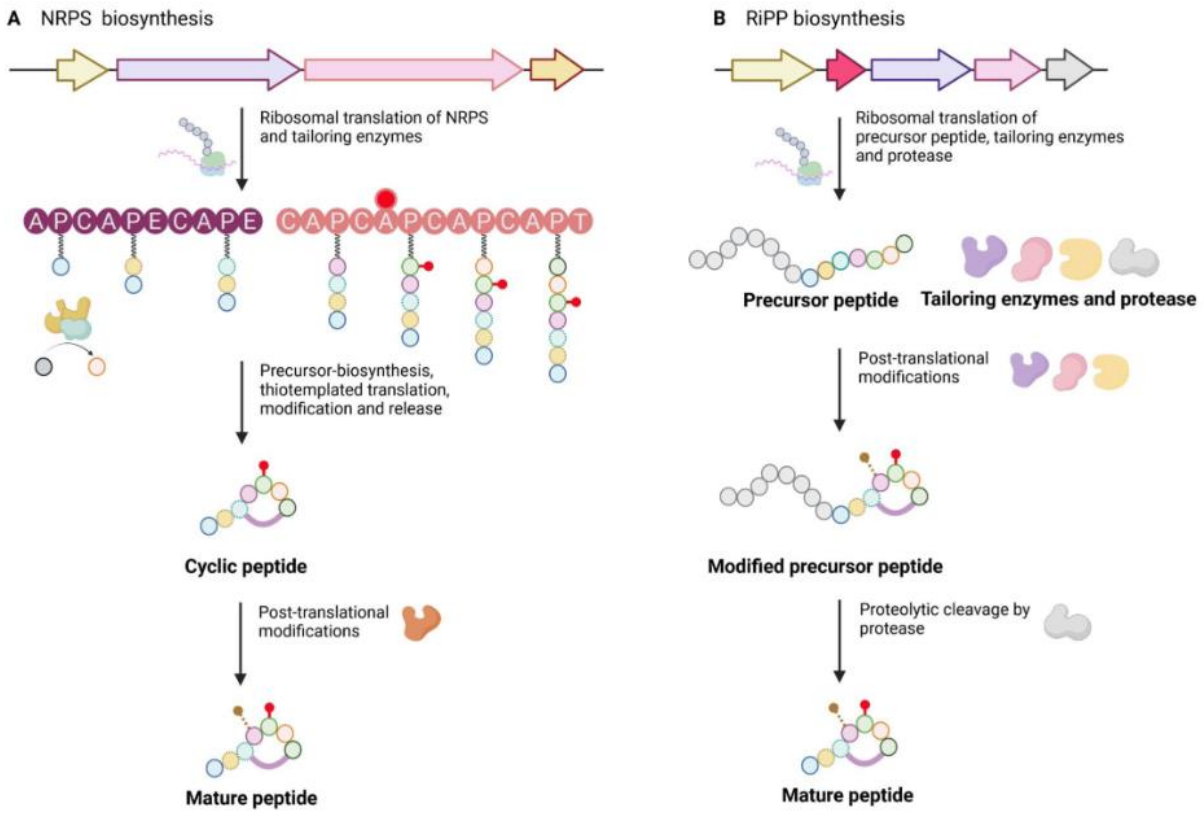


Figure A3. Schematic overview of NRP (A) and RiPP (B) biosynthesis. Figure taken from Wenski et al.*, 2022.*

3.1. Non-ribosomal peptide synthetases (NRPS)

Non-ribosomal peptides can include in their sequence the 20 proteinogenic amino acids and a large variety of more than 300 non-proteinogenic amino acids and aryl acids that are used as the monomer building blocks for oligomerization and diversification during chain elongation and after chain termination (Schwarzer *et al.*, 2003). The NRP chains can be modified at their N- or C-termini, such as the long-chain fatty acid attached in the N-terminus of daptomycin (Miao *et al.*, 2005) or the amines found in the C-terminus of bleomycin (Du *et al.*, 2000).

Non-ribosomal peptide synthetases are multi-modular enzymatic assembly lines that generally contain one module for each condensed amino acid corresponding to a programmed peptide biosynthesis (Figure A4). A common NRPS assembly line includes initiation, elongation, and termination modules. The minimal NRPS biosynthetic module contains three domains that ensure the elongation step on a growing peptidyl chain: a thiolation or peptidyl carrier protein domain (T or PCP) and two catalytic domains, the adenylation (A) domain that determines the selectivity of the amino acid, and the condensation (C) domain, responsible for the peptide bond-forming chain elongation (Süssmuth and Mainz, 2017). Each PCP domain must be primed by post-

translational modification of a serine side chain with phosphopantetheine, catalyzed by dedicated phosphopantetheinyl transferases (PPTases). During NRP assembly, peptidyl chains are linked to the terminal thiolate of the phosphopantetheine prosthetic group in a thioester linkage.

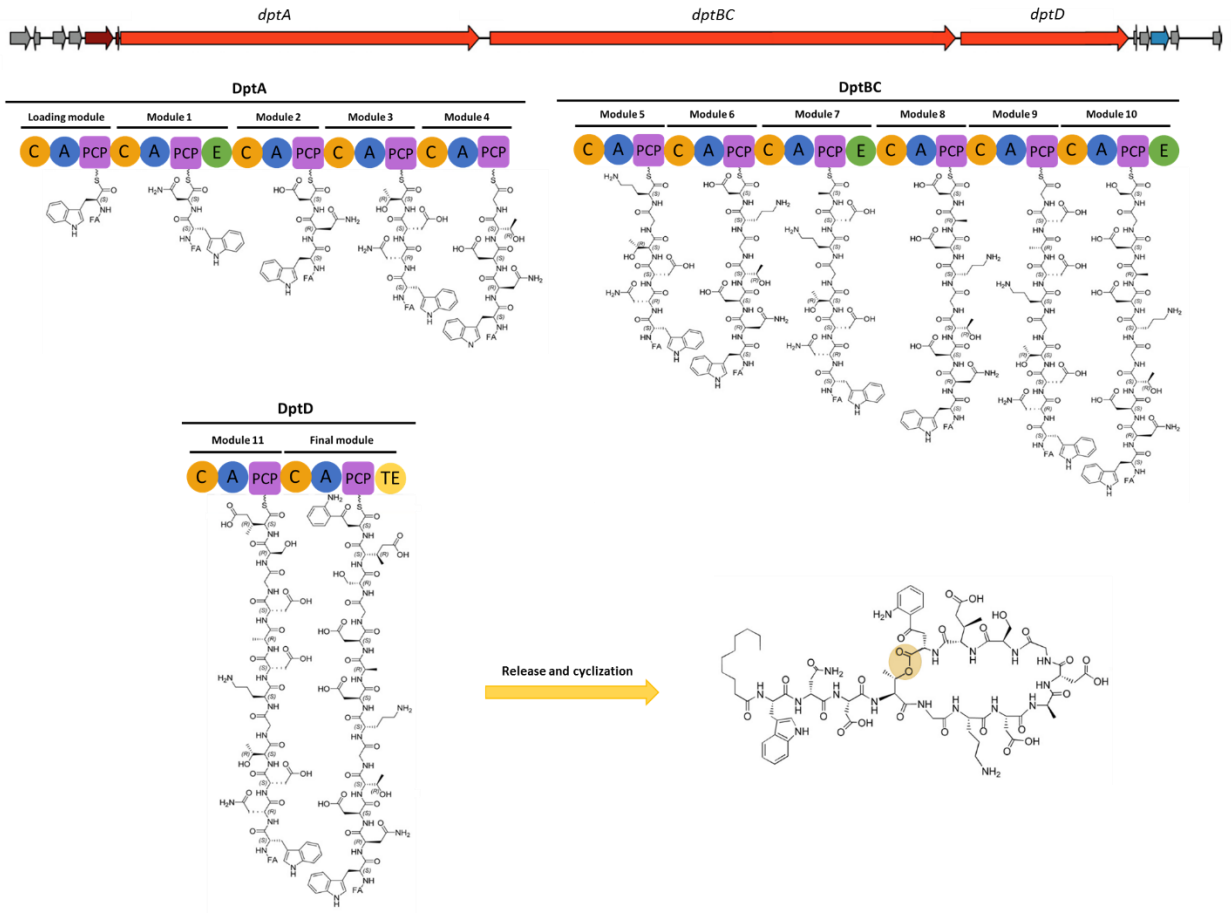


Figure A4. NRPS assembly line and the resulting chemical structure of daptomycin (abbreviations: C, condensation; A, adenylation; PCP, peptidyl-carrier protein; E, epimerase; TE, thioesterase). Picture modified from Robbel and Marahiel, 2010.

The two catalytic domains A and C are involved in the following different functions in the NRPS elongation. The A domain selects the amino acid, activating the carboxylate with ATP to generate the aminoacyl-AMP and then binding the aminoacyl group to the thiolate of the adjacent PCP domain. The C domain from the n module catalyzes the condensation of the upstream peptidyl-S-PCP_{n-1} to the downstream aminoacyl-S-PCP_n during the chain elongation.

The initiation or loading module in an NRPS assembly line is often a two-domain A-PCP module that selects the first amino acid and links it covalently to PCP_1. Then the peptidyl chain elongates leaving free the amino group of the N-terminal residue. Some initiation modules can have a three domain C-A-PCP organization. This usually happens in NRPS that generate N-acetyl peptides, such as plipastatin (Steller *et al.*, 1999), as the first C domain catalyzes this N-acylation of the initiating amino acid while it is installed on PCP_1 (Schmoock *et al.*, 2005).

One characteristic feature of NRPs is the high content of D-amino acids residues in their peptide chains. Since D-amino acids, especially non-proteinogenic D-amino acids, are not commonly present in microbial producer cells, L- to D-amino acids epimerization is catalyzed by an epimerization (E) domain found in some elongation modules (Stachelhaus and Walsh, 2000). The domain organization in such modules is C-A-PCP-E.

There are additional C domains known as cyclization domains (Cy) that catalyze the heterocyclization of cysteine and serine/threonine to thiazoline and oxazoline rings, respectively (Gehring *et al.*, 1998). NRPS modules can also contain additional domains involved in oxidation, N-methylation or transamination of the elongating peptide. Oxidase (Ox) domains, can be found in modules that contain Cy domains and are involved in the oxidation of the thiazoline ring to a thiazole (Cy-A-PCP-Ox) (Schneider *et al.*, 2003). This organization can be found in the bleomycin and epothilone synthases. Many non-ribosomal peptides may contain N-methylated amino acids, like cyclosporine, whose 7 out to the 11 residues are N-methylated. This N-methylation is catalyzed by a methyltransferase (MT) domain that transfers the methyl group from S-adenosyl methionine (SAM) to the amino group of the aminoacyl-S-PCP intermediate in C-A-MT-PCP modules (Hacker *et al.*, 2000).

The cleavage of the full-length peptide product from the enzyme to complete the biosynthesis is usually ensured by a thioesterase (TE) domain located at the termination or final module of the NRPS assembly line. Following the release from the assembly lines, many NRPs can be chemically modified by specialized enzymes encoded in their biosynthetic gene clusters. Among these modifications, which are often required for their biological activity, hydroxylations, oxidative cyclizations, halogenations, acylations or glycosylations can be found (Süssmuth and Mainz, 2017).

3.1.1. Non-TE-mediated release mechanisms in NRPS

In recent years, there have been increasing examples of alternative NRPS product release mechanisms mediated by enzymes distinct from the thioesterases (Little and Hertweck, 2022).

One of these mechanisms is mediated by the reductase (R) domain located on the C-termini of some NRPS enzymes. This R domain can catalyze the NAD(P)H-dependent reductive and non-redox type releases of peptidyl-S-NRPs. These R domains belong to the short-chain dehydrogenases (SDR) family of NAD(P)H dependent oxidoreductases. The first biochemically

characterized R domain to perform this reaction was described by Ehmann *et al.* (1999) and is involved in the biosynthesis of L-lysine in *Saccharomyces cerevisiae*. This reductive release mechanism has been also proposed in the biosynthesis of the tetrahydroisoquinoline NRP antibiotic saframycin and its analogs by *Streptomyces lavendulae* (Fu *et al.*, 2009).

A recently described type of chain release is performed by homologs of penicillin binding proteins (PBP). In *Streptomyces* sp. JAMM992, a PBP-like enzyme present in the surugamides A-F BGC, SurE, has been shown to catalyze a chain release lactamization reaction in surugamides biosynthesis (Zhou *et al.*, 2019). SurE homologues are also encoded in other NRPS BGCs from *Actinobacteria*, even as embedded domains within a NRPS enzyme instead of the typical TE domains (Zhou *et al.,* 2019).

Other NRPS release mechanisms have been described in fungi, such as those performed by R* domains or C_T domains, and in myxobacteria, such as those catalyzed by oxygenases (Little and Hertweck, 2022). However, none of these alternative mechanisms have been reported in actinomycetes.

Although different mechanisms can be suggested for product release in the majority of NRPSs that have been studied so far, there are a number of NRPSs that do not appear to have an obvious release enzyme, and a nonenzymatic release mechanism has been proposed (McErlean *et al.,* 2019). One such example is the biosynthesis of capuramycin-type nucleoside antibiotics by different strains of *Streptomyces* (Biecker *et al.,* 2019). Since two TE-lacking NRPS genes are present in the BGC (*capU* and *capV*), it was hypothesized that the C domain of CapV catalyzed the intramolecular lactamization. However, the C domains seemed to be non-functional, leading to a spontaneous non-enzymatic lactamization and product release.

3.2. Ribosomally synthesized and post-translationally modified peptides (RiPPs)

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a major family of natural products with a wide variety of structures and biological activities. They are normally produced as longer precursor peptides, typically around 20-110 residues in length, encoded by a structural gene. The precursor peptide comprises a core peptide, where several post-translational modifications (PTMs), such as amino acid dehydration, macrocyclization, glycosylation or epimerization, take place to increase metabolic stability of the peptide, and a leader peptide, appended to the N-terminus of the core peptide, which is usually important for the recognition of many enzymes and ensures that PTMs take place in the appropriate order. The leader peptide is also important for the peptide export and is removed to generate the mature peptide (Ortega and van der Donk, 2016; Montalbán-López *et al.*, 2021) (Figure A5). In most RiPPs, the recognition of the leader peptide by some PTM enzymes is carried out by the RiPP recognition element (RRE), a conserved PqqD-like domain containing an antiparallel β-sheet with a β-strand (2-6) conformation (Burkhart *et al.,* 2015; Lee and van der Donk, 2022). In some cases, a follower sequence in the C-terminus of the core peptide is needed for the biosynthesis and

proteolyzed with the leader peptide to release the mature peptide (Arnison *et al.*, 2013). The core peptide of some RiPPs is flanked by short regions known as recognition sequences (RS) that are identified by the PTM enzymes and are required for the chemical transformation on the core peptide. In the majority of these systems, the leader peptide is followed by multiple copies of the core peptide, which leads to the biosynthesis of a family of different RiPPs. (Montalbán-López *et al.*, 2021) (Figure A5).

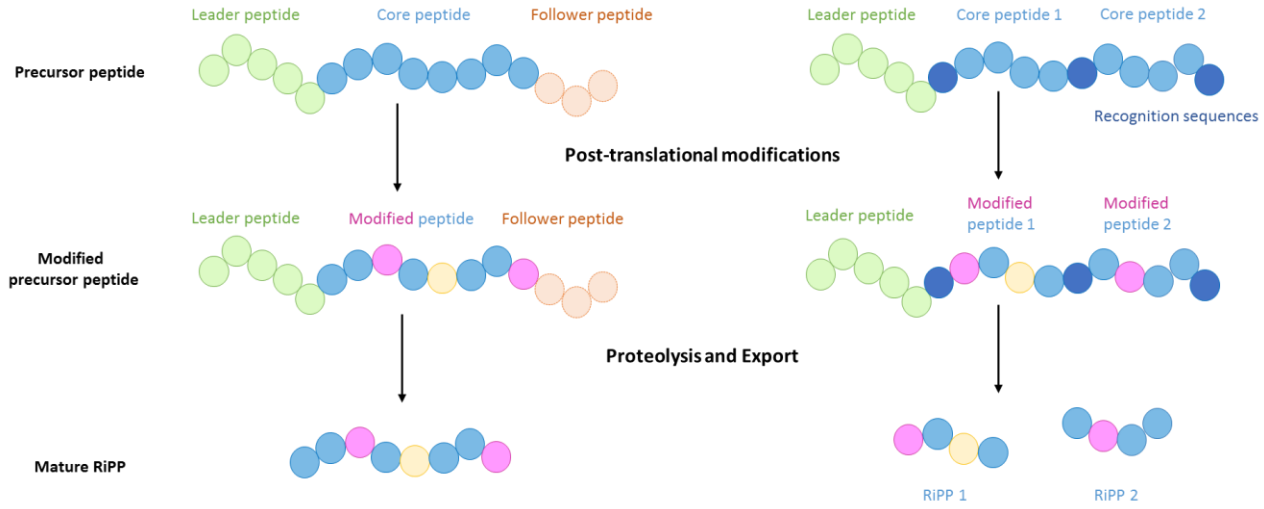


Figure A5. General biosynthetic pathways for RiPPs. Figure adapted from Lee and van der Donk et al.*, 2022.*

A wide variety of different RiPP classes have been defined based on their PTMs and structural features (Table A2). Nevertheless, the discovery and characterization of novel RiPP biosynthetic gene clusters is providing key information to identify new RiPPs showing new or improved activity or unusual chemical structures (Hetrick and van der Donk, 2017). Thus, RiPPs represent a promising group to discover novel chemical scaffolds and drug leads.

RiPP Class	Class-defining PTM or structural feature	Example (Producer bacteria)
Atropitides	Aromatic amino acids crosslinked giving a non-canonical atropisomer	Tryptorubin (*Streptomyces* sp. CLI2509)
Autoinducing peptides	Cyclic ester or thioester	AIP-I (*Staphylococcus aureus*)
Bacterial head-to-tail cyclized peptides	N-to-C cyclization differing from cyanobactins in the biosynthetic macrocyclization machinery	Enterocin AS-48 (*Enterococcus faecalis*)
Bottromycins	N-terminal macrocylic amidine C-terminal follower peptide instead of N-terminal leader peptide	Bottromycin A1 (*Streptomyces bottropensis*)
Guanidinotides	α-guanidino acid containing peptides	Pheganomycin (*Streptomyces cirratus*)
Lanthipeptides	(Methyl)lanthionine Labionin	Nisin (*Lactococcus lactis/Streptoccus*)
Lasso peptides	N-terminal macrolactam with threaded C-terminal tail through a ring	Humidimycin (*Streptomyces humidus*)
Linaridins	Dehydroamino acids (Dhb) but lacking Lan/MeLan	Cypemycin (*Streptomyces* sp. OH-4156)
Linear azol(in)e-containing peptides (LAPs)	(Methyl)-oxazol(in)e/thiazol(in)e heterocycles	Azolemycin (*Streptomyces* sp. FXJ1.264)
Lipolanthines	C-terminal labionin/avionin N-terminal FAS/PKS segment	Microvionin (*Microbacterium arborescens*)
Mycofactocin	Val-Tyr crosslink	Mycofactocin (*Mycobacteria* sp.)
Ryptides	Arg-Tyr crosslink	RRR (*Streptococcus* sp.)
Sactipeptides	Intramolecular thioether linkages between Cys side chains and α-carbons of other amino acids (sactionine crosslink)	Streptosactin (*Streptococcus* sp.)
Streptide	Trp-to-Lys carbon-carbon cross link	Streptide (*Streptoccoccus* sp.)
Thioamitides	Thioamide linkages	Thioviridamide (*Streptomyces olivoviridis*)
Thiopeptides	A central sin-membered nitrogen-containing ring Additional dehydrations and cyclodehydrations	Thiostrepton (*Streptomyces azureus* ATCC 14921)

Table A2. Currently known RiPP classes and representative examples, along their characteristic structural features, found in Gram-positive bacteria. Table adapted from Montalbán-López et al., 2021.

One of the best characterized RiPP classes is the lanthipeptides (Repka *et al.,* 2017), which are characterized by the presence of the thioether-bridged amino acids lanthionine (Lan) and methyllanthionine (MeLan). Because of their potent antimicrobial activities, many lanthipeptides are also named lantibiotics.

On the other hand, linaridins, defined as linear, dehydrated (arid) peptides, are a growing class of RiPP formed by a few members. However, *in silico* analysis show that this class is widespread in nature and contains a high structural diversity (Ma and Zhang, 2020).

3.2.1. Lanthipeptides

Lanthipeptides are polycyclic peptides characterized by the presence of the thioether-cross-linked amino acids meso-lanthionine (Lan) and/or 3-methyllanthionine (MeLan) in their structures. Lanthipeptides with antimicrobial activity, also known as lantibiotics, are small peptides containing 19-38 amino acids that often show a potent activity against Gram-positive bacteria such as methicillin-resistant *Staphylococcus aureus* (MRSA) and *Clostridium difficile* by disrupting the integrity of their bacterial cell walls (Pokhrel *et al*., 2022). The most extensively studied lantibiotic is nisin, produced in *Lactococcus lactis*. Nisin binds to the pyrophosphate of bacterial peptidoglycan precursor lipid II to inhibit cell wall biosynthesis and is able also to flip and insert into the bacteria membrane starting pore formation after binding to lipid II, resulting in a multi-pronged attack on the integrity of the Gram-positive bacterial cell wall and membrane (van der Donk and Nair, 2014). This unique mode of action lessens the frequency to develop resistant pathogens and makes lantibiotics promising candidates as new antibiotics. Other than nisin, there are several lantibiotics that share the same dual mode of action, such as mutacin 1140 (MU1140) produced by *Streptococcus mutans*, gallidermin synthesized by *Staphylococcus gallinarum*, NVB302 produced by *Actinoplanes liguriae* and NAI107, also named microbisporicin, isolated from the *Microbispora* (Pokhrel *et al*., 2022).

Lanthipeptides usually undergo extensive post-translational modifications and in all cases a lanthionine residue is generated. A lanthionine ring is a bis-amino-bis acid where two alanine residues are linked by a thioether group connecting their β-carbons. This crosslink is generated by the enzymatic dehydration of Ser or Thr residues to form 2,3-dehydroalanine (Dha) or (Z)-2,3-dehydrobutyrine (Dhb), respectively, followed by the enzymatic Michael-type stereoselective addition of the thiol of Cys to the unsaturated amino acid to give the thioether bridge. The resultant enolate can be protonated to produce either a lanthionine (Lan) from Ser or a methyllanthionine (MeLan) from Thr. These cross-links are essential not only for the activity of the mature compound but also for the stability against proteolysis and heat denaturation (Knerr and van der Donk, 2012).

Depending on the way that the dehydration takes place, lanthipeptides are divided into four different classes (I-IV) (Figure A6). A new lanthipeptide class V has been recently proposed by Ortiz-López *et al.* (2020) and Xu *et al.* (2020). The precursor peptides of the class I

lanthipeptide family (generally named LanA) are post-translationally modified by two distinct modification enzymes: a dehydratase that generates the moiety Dha/Dhb, named LanB, and a cyclase that catalyzes the addition of the Cys thiol, known as LanC (Knerr and van der Donk, 2012). Nisin, discovered in 1928 as a substance that inhibited the growth of *Lactobacillus bulgaricus,* is the representative member of class I (Rogers, 1928). In contrast, a single modification enzyme, called LanM, is responsible for processing the precursor peptides of class II lanthipeptides such as mersadicin and cinnamycin (Widdick *et al.,* 2003). LanM contains an amino-terminal dehydratase domain and a carboxy-terminal LanC-like domain (Willey and van der Donk, 2007). The III and IV classes of lanthipeptides, like SapT or labyrinthopeptin and venezuelin, respectively, are biosynthesized by a single trifunctional enzyme, known as LanKC for class III and LanL for class IV. These enzymes contain an amino-terminal phosphor-Ser/phosphor-Thr lyase domain, a central kinase-like domain and a carboxy-terminal LanC-like cyclase domain (Goto *et al.*, 2010; Krawczyk *et al.*, 2012). The carboxy-terminal LanC-like domains of class III enzymes are unique as they lack the zinc-binding site that is believed to activate the Cys thiol for the nucleophilic attack in LanC proteins (Wang and van der Donk, 2012). Class III lanthipeptides are also structurally characterized by tri-amino acid cross-links named labionin (Lab), which enables the formation of two ring systems within a peptide chain (Müller *et al.*, 2010). The novel class V lanthipeptides includes only three members: lexapeptide (Xu *et al.,* 2020), pristinin A3 (Kloosterman *et al.,* 2020) and cacaoidin (Ortiz-López *et al.,* 2020), considered the first member of this new class and subject of study in this thesis. Xu *et al*. (2020), proved that this class V lanthipeptides are post-translationally modified by a three-component lanthionine synthetase containing three monofunctional proteins: LxmK, LxmX and LxmY. LxmK is a kinase that phosphorylates Ser and Thr residues. Then, LxmX is a phosphotransferase homolog that removes the phosphate groups, generating D-amino acids. And, finally, LxmY is a cyclase that generates the lanthionine ring (Xu *et al*., 2020). This mechanism is similar to that described in class II-IV lanthipeptides, including phosphorylation, elimination and cyclization processes, with the difference than in this new class V three proteins are required. Moreover, these three monofunctional proteins share no similarities with the multifunctional synthetases LanM, LanKC or LanL from class II-IV lanthipeptides.

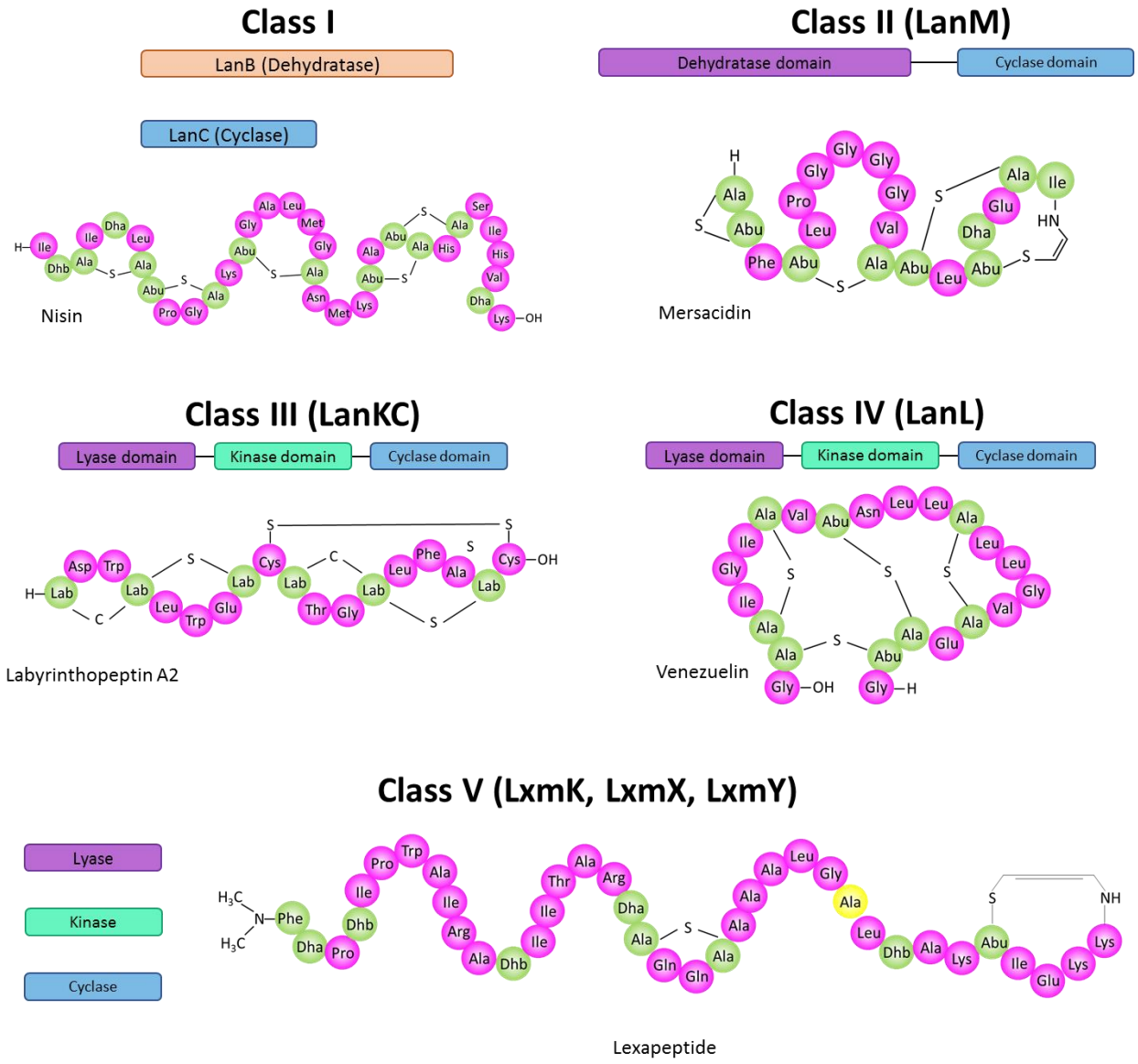


Figure A6. Structures of the canonical post-translational modifications in lanthipeptides, and representative members of each class. The amino acid sequences are given with pink, green and yellow representing proteinogenic, post-translationally modified and D-amino acids, respectively. Figure adapted from van der Donk and Nair, 2014.

In addition to the lanthionine generation, a number of additional post-translational modification enzymes involved in lanthipeptide maturation have been characterized. Many of these modifications happen at the N- or C-terminal ends of the core sequence to increase the stability or to change the overall charge of the peptides, after which the leader sequence is

removed to afford the final mature natural product. Conceptually, the leader peptide may facilitate the biosynthesis by interacting with the post-translational modification enzymes, by transporting the peptide product and by contributing to immunity within the producing strain. However, the leader peptide has no evidence to play a direct role in the chemical transformations in the core peptide. In fact, most tailoring reactions do not require the presence of the leader peptide or may take place after leader peptide removal (Repka *et al.*, 2017). Following the post-transcriptional modifications, the amino-terminal leader peptide is removed to generate the bioactive molecule. For class I lanthipeptides, removal of the amino-terminal leader peptide is carried out by a subtilisin-like serine protease named LanP (van der Meer *et al.*, 1993). During the biosynthesis of class II lanthipeptides, the leader peptide removal occurs simultaneously with the transport of the modified precursor peptide, and is catalyzed by the papain-like Cys protease domain of $LanT_p$, a transmembrane ATP-binding cassette (ABC) transporter superfamily used to export the modified peptide from the producing cell (Havarstein *et al.*, 1995). A few class II lanthipeptides proteases employ a combination of $LanT_p$ and LanP proteases to complete leader processing (Caetano *et al.*, 2011). The mechanism of leader peptide removal for class III lanthipeptides remains unclear, posing a major obstacle to their production and bioengineering. Völler *et al.* (2013) reported a putative class III prolyl oligopeptidases (POP), which is capable of cleaving the leader peptide of modified precursor peptides C-terminal of a Pro residue. These POP peptidases only process substrates that contain a fully modified core peptide, with a specificity mechanism designed to select against premature substrate cleavage. However, they are not found in all class III BGCs and it is unclear if other proteases would be required for leader processing in class III lanthipeptides. More investigation is still needed to understand the key biosynthetic step of leader peptide removal.

Several lanthipeptides, such as epidermin, gallidermin or NAI-107 contain the amino acid S-[(Z)-2-aminovinyl]-D-cysteine (AviCys) or S-[(Z)-2-aminovinyl]-D-3-methylcysteine (AviMeCys) rings at their C-termini. The enzymes responsible for the formation of these rings are named LanD proteins and carry out the oxidative decarboxylation of the C-terminal Cys to generate the enethiolate intermediate (Chatterjee *et al.*, 2005). This family of LanD proteins contains a stoichiometric noncovalently bound flavin mononucleotide (FMN) and has been named the homo-oligomeric flavin containing Cys decarboxylase (HFCD) family. These Avi(Me)Cys structures provide desirable properties such as stability, high target specificity and resistance to proteases (Grant-Mackie *et al.*, 2021).

D-amino acids have been frequently found in ribosomally synthesized products contributing to proteolytic stability, structural conformation and bioactivity. As only L-amino acids are genetically encoded, access to the D stereochemistry requires a post-translational modification. For non-lanthipeptide natural products, D-amino acids are incorporated by epimerization of the corresponding L-amino acid, either by deprotonation-protonation mechanisms or by a radical mechanism. In lanthipeptides, based on the mechanism of (Me)Lan formation, the conversion of L-serine to D-alanine is suggested to happen by initial dehydration of L-Ser to Dha. Instead of lanthionine formation, the diastereoselective hydrogenation of Dha is

carried out to give D-Ala. This reaction is catalyzed by the generic enzyme LanJ (Yang and van der Donk, 2015).

The precursor peptides generally contain a Ser residue immediately following the leader sequence, which is converted to Dha. Upon removal of the leader peptide, the Dha is hydrolyzed to the corresponding pyruvyl group, and the resultant ketone of this pyruvyl group can be reduced to generate an amino-terminal lactyl group in an NADPH-dependent manner (Ortega *et al.*, 2014). This reaction is catalyzed by the generally named LanO dehydrogenase. This post-translational modification has been seen in the class I lanthipeptides epilancin 15X, epicidin 280 or gallidermin.

One of the most potent lantibiotics to date is microbisporicin (also known as NAI-107), which shows potent activity against several clinically relevant Gram-positive pathogens and contains two modifications not previously observed in lanthipeptides: a halogenation on a Trp residue and dihydroxylation of a Pro residue (Repka et al, 2017). The halogenation of the Trp residue is carried out by the LanH enzyme, a flavin-dependent halogenase that catalyzes the halogenation in an $FADH_2$-dependent manner. The dihydroxylation of the Pro residue is likely catalyzed by the LanO enzyme, which shows similarities to cytochrome P450.

Lanthipeptides producing strains must protect themselves against the bactericidal activity of the fully produced peptide. Such producing strains encode a self-immunity system consisting in two different components: a transmembrane ABC transporter system from the LanFEG gene products and a small single-pass immunity protein named LanI. Both components usually work synergistically for the complete self-immunity against the lantibiotic toxicity in producing strains, with LanI protein intercepting the lanthipeptide and the LanFEG complex transporting the bioactive product from the membrane to the extracellular space (van der Donk and Nair, 2014).

Lanthipeptide biosynthesis is very well distributed among microorganisms, being this family the most abundant of all RiPPs. Even though the enzymology of lanthipeptide biosynthesis is much better understood that a decade ago, many questions remain, and new questions have been raised. While challenging to study, the lanthipeptide biosynthetic enzymes are very promising with respect to the new bioengineering and synthetic biology applications and continuing investigations are warranted (Repka *et al.*, 2017).

3.2.2. Linaridins

Linaridins (linear arid peptides) are a small but complex group of linear dehydrated RiPPs, characterized by the presence of dehydrobutyrine (Dhb). Linaridins share the presence of thioether crosslinks with the lanthipeptide family but these crosslinks are generated by a different pathway. Cypemycin is the founding member of this family, though it was first considered a lanthipeptide because of its C-terminal aminovinyl cysteine. However, Claesen and Bibb (2010) later showed that cypemycin biosynthesis does not involve any lanthionine

synthetase-like enzyme and the sequence of the core peptide illustrated that the AviCys structure was formed from two Cys residues. Based on these differences from the biosynthetic pathway to AviCys in lanthipeptides, cypemycin was classified as belonging to a separate group, the linaridins (Arnison *et al.*, 2013).

Thus far, ten linaridins have been characterized: cypemycin (Minami *et al.*, 1994), grisemycin (Claesen and Bibb, 2011), legonaridin (Rateb *et al.*, 2015), salinipeptins A-D (Shang *et al.*, 2019), pegvadins A-B (Georgiou *et al.*, 2020) and mononaridin (Wang *et al.*, 2021), with features such as Dhb, AviCys moiety in the C-terminus, an *N,N*-dimethylalanine at the N-terminus and dehydrothreonine moieties, L-*allo*-isoleucine, D-amino acids and dimethylimidazolidin-a-one (Figure A7). Linaridins are a rare structurally diverse family of RiPPs whose members only possess one common structural feature, the dehydration of Thr to Dhb (Mo *et al.*, 2017), which is the class-defining modification in linaridins (Claesen and Bibb, 2010). Linaridins are still a relatively underexplored family of RiPPs and the mechanism of this Dhb installation remains unclear. The most likely candidates are homologs of CypH or the enzymes LegH and LegE, encoded in the legonaridin BGC, that are homologous, respectively, to the N- and C-terminal domains of CypH, which consists an N-terminal horizontally transferred transmembrane helix (HTTM) domain and a C terminal α/β hydrolase fold and participates in processing the precursor peptide CypA in cypemycin maturation (Rateb *et al.*, 2015).

Cypemycin, the prototypical member of the linaridin family, is linear and has only the AviCys ring system, thus it was an ideal system to investigate AviCys formation in linaridins. The AviCys moiety is also found in some lanthipeptides where is formed by Michael addition of an enethiolate, resulting from the oxidative decarboxylation of the C-terminal Cys residue, to a Dha resulting from Ser dehydration. The *in vitro* activity of the flavoprotein CypD was reconstituted by Claesen and Bibb (2010), showing that CypD catalyzes an oxidative decarboxylation of the C-terminal Cys of the precursor peptide CypA. However, Ding *et al.* (2018b) showed that CypD only catalyzes this Cys decarboxylation but it is not responsible for the subsequent AviCys ring formation, which is achieved by the dethiolation of the Cys residue at the C-terminal fourth position in the CypA, producing Dha. They suggest that this thiolation is likely prior to the CypD catalyzed decarboxylation of the C-terminal Cys, leading to the formation of the AviCys ring.

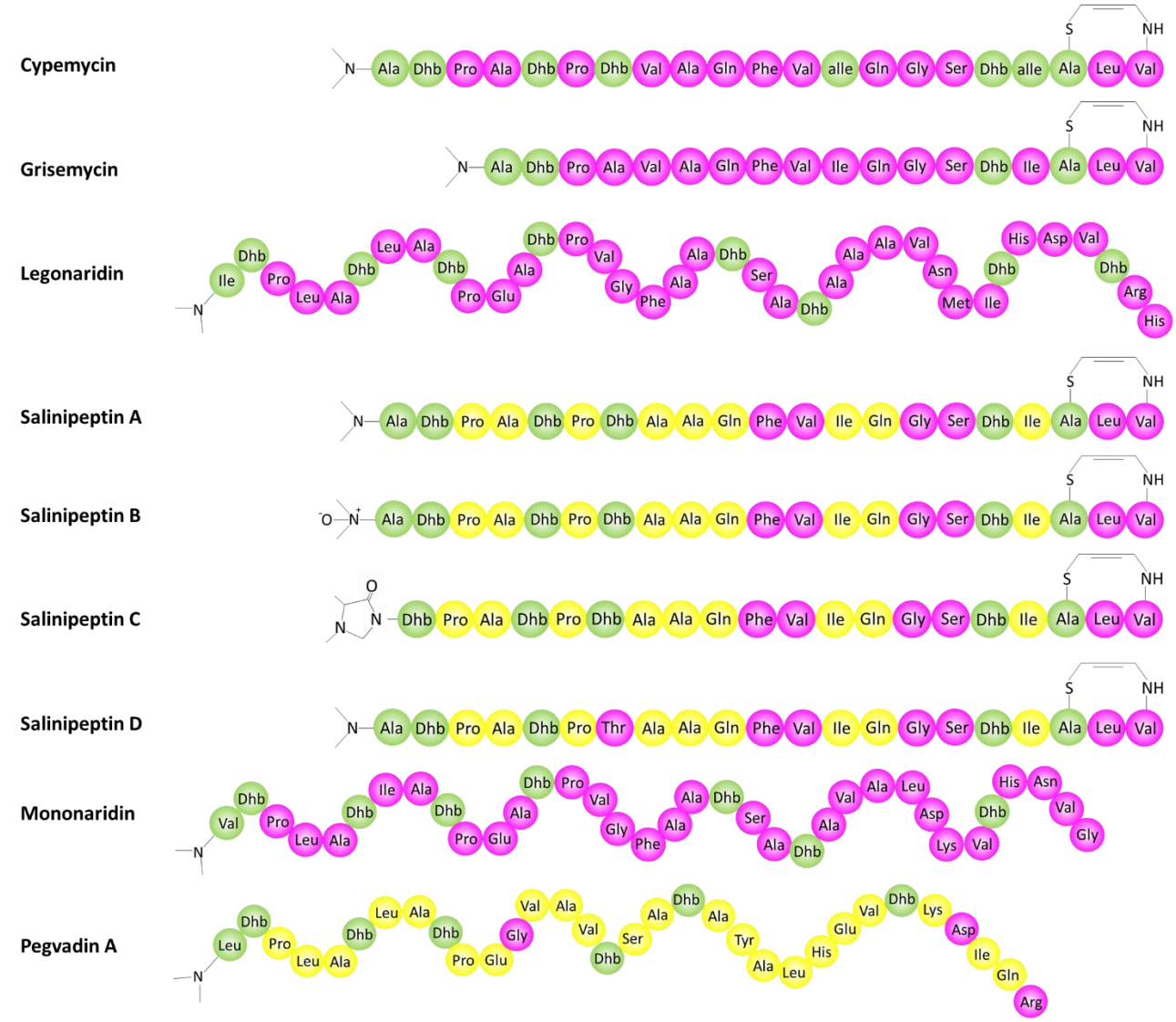


Figure A7. Structures of RiPPs belonging to the linaridin family. The amino acid sequences are given with pink, green and yellow representing proteinogenic, post-translationally modified and D-amino acids, respectively.

4. GENOME MINING IN *STREPTOMYCES*

Microbial genome mining is an alternative approach to traditional culture-based methods for the discovery of novel natural products. The concept of genome mining appeared in 2002 when the first whole genome from the model actinomycete strain *Streptomyces coelicolor* A3(2) was sequenced (Bentley *et al.*, 2002). This whole genome sequence revealed that *S. coelicolor* had a silent biosynthetic potential as it only produced four metabolites while harbored 22 secondary metabolite biosynthetic gene clusters.

The efficiency improvement and the decreasing costs in the Next Generation Sequencing (NGS) techniques have led to an increasing number of available microbial genomes in public databases. NGS technologies can provide both short (50-400 bp) and long reads (1-100 Kb) (Besser *et al*., 2018). Among the most commonly used NGS platforms (Illumina, PacBio and Nanopore), Illumina is based on the bridge amplification process, consisting in the repeated amplification of DNA molecules (about 500 bp), containing ligated adapters in their ends, on a solid support with oligonucleotide sequences complementary to the ligated adapter. After repeated rounds of amplification, clonal clusters are created. During the synthesis reactions, detectable fluorescent terminator nucleotides are incorporated. The terminators are removed after detection for the next round of synthesis, providing a high number of short reads (Quail *et al*., 2008).

The sequencing technology developed by Pacific Biosciences (Menlo Park, CA, USA), known as PacBio or SMRT -Single Molecule Real Time- sequencing, allows the sequencing of up to 30–50 Kb fragments. The SMRT method consists in the immobilization of an engineered DNA polymerase to the bottom of a well in a SMRT flow cell. The DNA polymerase binds to the DNA and incorporates nucleotides labeled with different phospho-linked fluorophores in a growing chain. After incorporation, the phosphate-linked fluorescent moiety is released, allowing the incorporation of the next nucleotide. The unique template preparation process of PacBio involves the production of a circular double-stranded DNA molecule with a known adapter sequence complementary to the primers used to initiate the DNA synthesis. This allows the repetitive reading of a large template by the DNA polymerase (Rhoads and Au, 2015).

Nanopore-based DNA sequencing (developed by Oxford Nanopore Technologies) is another long-read sequencing platform based on the use of transmembrane proteins (alpha hemolysin or *Mycobacterium smegmatis* porin A, MspA) embedded in a lipid membrane to produce pores. These proteins are complemented with motor proteins, such as DNA polymerase (phi29), and DNA helicases, to avoid unwinding and boost the nucleotides through the nanopore. The DNA is sequenced by measuring changes in electric conductivity, characteristic of each nucleotide, while the DNA is crossing through the pores. Theoretically, 100 Kb of DNA can pass through each nanopore and be detected. (Slatko *et al.,* 2019)

The increase in the number of sequenced microbial genomes has been accompanied by the parallel development of new molecular biology techniques and bioinformatic tools that have revealed the huge biosynthetic potential of actinomycetes to produce many more natural products than those detected under laboratory conditions (Gómez-Escribano *et al.*, 2016). These tools have revolutionized the way that natural products are discovered by translating predicted secondary metabolites biosynthetic pathways into purified molecules with the activation and bioengineering of silent BGC and their direct expression in new heterologous chassis (Genilloud, 2017). These new bioinformatic analysis tools enable the search of certain classes of secondary metabolite biosynthetic pathways, as well as mining for genes encoding tailoring enzymes, resistance genes and global and pathway specific regulators (Ziemert *et al.*, 2016). Among these bioinformatic platforms, antiSMASH (Blin *et al.*, 2021a), PRISM (Skinnider *et al.*, 2020), BAGEL4

(van Heel *et al*., 2018) or RiPPMiner-Genome (Agrawal *et al.*, 2021) now facilitate the identification of new secondary metabolite.

antiSMASH (antibiotics & Secondary Metabolite Analysis SHell) is the most widely employed tool for the search of microbial secondary metabolites biosynthetic pathways in the genomes of bacteria, fungi and plants, using a rule-based approach to identify 71 different types of secondary metabolite BGCs such as non-ribosomal peptide synthetases (NRPS), type I and II polyketide synthases (PKS), lanthipeptides, thiopeptides, lasso peptides or sactipeptides, among others, and provides detailed predictions about the chemical novelty of the putative compounds produced by the cluster, supporting their dereplication (Blin *et al.*, 2021a).

PRISM (PRediction Informatics for Secondary Metabolomes) is another web Java application useful for the biosynthetic pathways and chemical structures predictions of genetically encoded PKs, NRPs, RiPPs (in the RiPP-PRISM version) and every class of bacterial natural antibiotic currently in clinical use, including aminoglycosides, nucleosides, β-lactams, alkaloids, and lincosamides among other classes of metabolites (Skinnider *et al.,* 2020).

BAGEL4 (Bayesian Analysis of Gene Essentiality) and RiPPMiner-Genome are other web servers that enable the specific identification of biosynthetic gene clusters involved in the production of RiPPs and bacterocins (van Heel *et al*, 2018; Agrawal *et al.*, 2021). The prediction of new RiPP BGCs by these tools is challenging, as these compounds represent a high diverse class and lack universal genetic markers. However, the increase on the discovery of new RiPPs is revealing several modifications shared between different classes, improving their prediction. RODEO, RiPPer, RRE-Finder or DeepRiPP are other examples of currently available tools specified in RiPP genome mining (Kloosterman *et al*., 2021). These prediction web tools are fed by different databases such as MiBiG (Kautsar *et al.*, 2020), antiSMASH-DB (Blin *et al*., 2021b) or CAMPR3 (Waghu *et al.*, 2016), which store many thousands of computationally predicted BGCs of a huge variety of natural products classes, providing the prediction tools with the required information to find out the function and novelty of the new predicted BGCs.

Currently, there are around 354 complete genome sequences available only from the genus *Streptomyces* in NCBI. The bioinformatic tools showed that a single *Streptomyces* genome usually encodes 25-50 BGCs. However, many of them are not actively expressed, or only lowly expressed, under standard laboratory growth conditions or their corresponding products cannot be detected with the normally used analytical methods (Mao *et al.*, 2018). In fact, transcriptome and proteomic analysis have suggested that the majority of secondary metabolism gene clusters in *Streptomyces* are not silent but expressed at a very low level under laboratory conditions (Rebets *et al.*, 2014).

These silent or cryptic BGC are much more numerous than the constitutively active BGCs by a factor of 5-10 (Rutledge and Challis, 2017). Genome mining analysis suggest that less than 10 % of the genetic potential of *Streptomyces* is currently being used, which implicates that this genus carries a huge genetic potential to be exploited for the discovery of new natural products.

In addition to this, metagenomic data indicate that today, almost 80 years after Selman Waksman introduced the genus of *Streptomyces* and with actinomycin discovered as the first antibiotic isolated from an actinomycete, these bacteria still represent an enormous reservoir for the identification of novel antibiotics (Genilloud, 2018). However, despite all the efforts employed, this biosynthetic potential is still elusive, as natural product research keeps relying on a classical culture-dependent approach. For the complete exploitation of the microbial biosynthetic potential, novel cultivation methods must be employed, new genetic manipulation tools developed, elaborated analytical and screening techniques implemented and knowledge on new secondary metabolite BGCs acquired in order to revolutionize the antibiotic drug discovery (Miethke *et al.*, 2021).

5. APPROACHES FOR THE ACTIVATION OF SILENT OR CRYPTIC BIOSYNTHETIC GENE CLUSTERS IN *STREPTOMYCES*

Unlocking the expression of silent or cryptic biosynthetic gene clusters to identify novel natural products has been a bottleneck in the exploitation of the whole biosynthetic potential in actinomycetes for many years. Different culture-based and bioengineering methods have been developed to activate silent or poorly expressed BGCs and increase production yields of some compounds.

5.1. Culture-based approaches to express silent BGCs

Cultured-based approaches have been employed for a long time. The traditional “One Strain Many Compounds” (OSMAC) approach has been a successful strategy for the isolation and identification of many new natural products from *Streptomyces* for two decades (Bode *et al.*, 2002) although the same strategy had been used before in industrial programs. The growth of strains in multiple and different nutritional and cultivation conditions following an OSMAC approach can lead to the activation of the biosynthesis of putative novel secondary metabolites. There is a large number of examples of the induction of the biosynthetis of new natural products following this approach, such as four new ansamycin-type polyketides from *Streptomyces* sp. C34, the cyclodepsipeptide NC-1 from *Streptomyces* sp. FXJ1.172, six new antifungal reedsmycins A-F and the two new cytotoxic alkaloids indotertine A and drimentine F from *Streptomyces* sp. CHQ-64.

The co-cultivation with other strains, which result in the physical interaction of the strains and cross-talks between their metabolic pathways, is another approach to enhance the production of secondary metabolites. The 12 different analogs of desferrioxamine resulted from the co-cultivation of *Streptomyces coelicolor* M145 with other actinomycete strains and the new luteoride and pseutorin derivatives and the new lasso peptide chaxapeptin from the co-cultivation between *Streptomyces leeuwenhoekii* and *Aspergillus fumigatus* are examples of new natural products generated by co-cultivation (Nguyen *et al.*, 2020).

Another widely used culture-dependent method for the activation of silent BGCs involves the use of elicitors or chemical stress factors. The elicitors are defined as exogenous small molecules which can stimulate secondary metabolism and enhance the production of cryptic secondary metabolites (Okada and Seyedsayamdost, 2017). A wide variety of molecules can act as elicitors in *Streptomyces*. For example, metals (Fe, Co, Cu, Sc, Ni, Zn or Mg) and rare earth elements (REEs) like lanthanum (La), vanadium (V) or scandium (Sc) can induce secondary metabolite production: Sc increased the production of actinorhodin in *S. coelicolor* A3(2) and streptomycin in *S. griseus* (Kawai *et al*., 2007) while La activated tacrolimus production in *S. tsukubaensis* (Wang *et al*., 2018b). Chemicals, like DMSO, can also induce the production of secondary metabolites in the low concentrations, by altering the metabolic flux toward primary metabolism (pentose phosphate pathway, glycolysis, and tricarboxylic acid cycle) to enhance the availability of secondary metabolite precursors. This effect has been observed in the biosynthesis of chloramphenicol in *S. venezuelae*, thiostrepton in *S. azureus* ATCC14921 or tacrolimus in *S. tsukubaensis* (Zong *et al*., 2021). Also, ethanol can enhance the biosynthesis of secondary metabolites by stimulating the pathway-specific transcriptional regulation and the oxidative stress signaling, as it has been reported in the biosynthesis of tetracenomycin C in *S. glaucescens* or chloramphenicol and jadomycin in *S. venezuelae* ATCC 10712 (ISP5230) (Wag *et al.,* 2018b). Interestingly, although nanoparticles have toxic effects on microorganisms, they have also been proved to induce the biosynthesis of secondary metabolites in small concentrations by oxidative stress. CuO, Al_2O_3 and carbon nanoparticles induced the production of actinorhodin in *S. coelicolor* M145 (Liu *et al*., 2019a; 2019b; 2019c). Also, enzymes and enzyme inhibitors have been used to enhance the antibiotic biosynthesis in *Streptomyces*. For example, the histone deacetylase (HDAC) suppresses the acetylation of histones and gene expression in eukaryotes. HDAC inhibitors, such as sodium butyrate or valproic acid, can adjust the structure of the prokaryotic nucleoid and induce the expression of secondary metabolite gene clusters, as the actinorhodin and undecylprodigiosin BGCs in *S. coelicolor* (Zong *et al.,* 2021). In addition, the phosphopantetheinyl transferase (PPTase) is involved in the biosynthesis of polyketides and non-ribosomal peptides by adding a 4'-phosphopantetheine group into the acyl carrier protein (ACP) domain. The overexpression of this PPTase can induce the production of oviedomycin in *S. antibioticus* NRRL 3238 or halichomycin in *S. ghanaensis* (Zong *et al*., 2021). Finally, antibiotics can be also used as elicitors, as they can induce the expression of BGCs in *Streptomyces* acting at the ribosome or RNA polymerase levels

5.2. Genetic tools to activate silent BGCs

There are two main genetic approaches to activate silent BGCs in *Streptomyces*. The first one involves *in situ* activation approaches including engineering of promoters, transcription factors and ribosomes. On the other hand, the heterologous expression includes different cloning strategies, biosynthetic pathways reconstruction methods and the use of different heterologous hosts.

5.2.1. In situ *activation of silent BGCs*

Several *in situ* refactoring approaches have been developed to activate silent clusters in the native strain, which needs to be genetically manipulable. Natural products BGCs are usually subjected to a very complex transcriptional and translational regulation. Switches in this regulation may activate gene expression and achieve the production of previously hidden metabolites.

Among all the genetic methods to activate silent or poorly expressed BGCs, the most widely used is the overexpression of positive transcription regulators under the control of constitutive or inducible promoters or the disruption of negative regulators. The most commonly used constitutive promoters include the promoter of the erythromycin resistance gene *ermEp*, *ermEp1* and its derivatives, the engineered *kasOP** promoter (strong promoter derived from the promoter of the SARP transcriptional activator *kasO*) and the phage I19 derived promoter SF14p. Among the inducible promoters, the most widely used are the thiostrepton-inducible promoter *PtipA*, the synthetic tetracycline-inducible promoter *tcp830* and the synthetic resorcinol-inducible and cumate-inducible promoters (Liu *et al*., 2021). These promoters carry out the overexpression of positive regulators such as SARP (*Streptomyces* antibiotic regulatory proteins) or LAL (Large ATP-binding regulators of the LuxR family) activators. One example of the use of this approach was the activation of the giant cryptic *stm* BGC in *S. ambofaciens* by overexpression of the LAL positive regulator SamR0469 under the control of the *ermEp* promoter to produce stambomycins A-D (Laureti *et al*., 2011). On the other hand, the disruption of negative regulators like TetR can also lead to the activation of BGCs. For example, the deletion of the genes encoding the TetR-family repressors RapY and RapS from the rapamycin BGC in *Streptomyces rapamycinicus* enhances the production of rapamycin (Yoo *et al*., 2015).

Transcriptional regulation engineering is also a widely used approach for the activation of silent or cryptic BGCs with certain RifR mutations by rifampicin in the central portion of the *rpoB* gene, encoding the β subunit of the RNA polymerase (amino acid substitutions at His437 is the most common mutation), which provides resistance to the antibiotic (Ochi *et al*., 2014). Thus, the introduction of mutations in the *rpoB* gene successfully enhanced the production of streptomycin in *S. griseus*, actinorhodin and undecylprodigiosyn in *S. coelicolor* and *S. lividans*, piperidamycin in *S. mauvicolor*, actinomycin in *S. antibioticus*, salinomycin in *S. albus*, sinefungin in *S. incarnatus* or formycin in *S. lavendulae* (Tanaka *et al.,* 2013). In addition to RpoB, mutations in the principal sigma factor of the RNA polymerase (HrdB) also showed to improve production of avermectin in *S. avermitilis* (Zhuo et al., 2010) and further secondary metabolites in other actinomycete strains like the enhancement of teicoplanin production in *Actinoplanes teichomycetus* (Wang *et al*., 2014b).

At translational level, ribosome engineering is also an approach used for the activation of the expression of silent BGCs, based on the generation of spontaneous mutations by in the *rpsL* gene, which encodes the ribosomal protein S12, providing resistance to streptomycin. The first increased production obtained using this approach was actinorhodin in a streptomycin-resistant

strain of *Streptomyces lividans*, having a Lys88Glu mutation in the *rpsL* gene (Shima *et al*., 1996). This technique has been successfully used to enhance the production of different natural products in several *Streptomyces* strains, such as actinomycin D in *S. antibioticus* and *S. parvulus*, avermectins in *S. avermitilis*, chloramphenicol in *S. coelicolor*, daptomycin in *S. roseosporus* or streptomycin in *S. griseus* (Zhu *et al*., 2019). Pan *et al*. (2013) found that disruption of *rimP*, encoding a 30S ribosome subunit assembly cofactor, enhanced the production of actinorhodin in *S. coelicolor* M145 and jadomycin in *S. venezuelae*. However, further investigations need to be done in a wider set of actinomycetes to include the disruption of *rimP* as a technique for ribosome engineering (Baltz, 2016).

5.2.2. Heterologous expression of BGCs

Heterologous expression is an efficient approach to activate the expression of silent BGCs identified by genome mining in those strains that cannot be genetically manipulated with the currently available genetic tools, which makes it more challenging to access their secondary metabolite potential.

To be heterologously expressed, the biosynthetic gene clusters must be first cloned into *Escherichia coli–Streptomyces* shuttle vectors capable to carry large DNA inserts, such as cosmids (derived from the *E. coli* λ bacteriophage) and fosmids (derived from the *E. coli* F^- plasmid), which can accommodate up to 40–50 Kb, or bacterial (BAC) and P1 phage (PAC) artificial chromosomes, which can harbor inserts from 100 Kb to 300 Kb (Xu and Wright, 2019). There are several techniques that allow the cloning of large DNA fragments containing BGCs into the corresponding vector (Figure A8).

The most traditional approach for the cloning of BGCs is the site-specific recombination system, involving the versatile *E. coli-Streptomyces* shuttle vector pSBAC (*Streptomyces* Bacterial Artificial Chromosome) (Liu *et al*., 2009). In this approach, a unique restriction site is introduced at both ends of the target BGC, along with the pSBAC-based backbone at one of the ends. Then, the bacterial chromosome is digested with the restriction enzyme and the product is self-ligated obtaining the pSBAC:BGC construction. With this method, Nah *et al.* (2015) cloned the 80 Kb tautomycin BGC after digestion with XbaI. The construct was introduced in *S. coelicolor* M145 and *S. lividans* TK21 for subsequent heterologous expression.

As an alternative of the use of restriction enzymes, the phage ΦBT1 *attP-int* machinery was used to develop an integrase-mediated site-specific recombination system (ISR) (Du *et al.*, 2015). First, an *attB* site is introduced in one of the ends of the BGC by a single crossover with the pUC119-based suicide plasmid. In the other end of the BGC, an *attP* site is also introduced using the pKC1139-based plasmid. The pIJ10500 plasmid containing the ΦBT1 integrase is also introduced to carry out the recombination between the *attP-attB* site. The result of this recombination is the pUC119-based plasmids containing the BGC and an *attR* site, and an *attL* site remaining in the genome. Using this method, Du *et al*. (2015) successfully cloned the

actinorhodin cluster from *S. coelicolor* M145 and the napsamycin and daptomycin BGCs from *S. roseosporus* NRRL 15998.

Another tool for the cloning of BGC includes the full-length RecET proteins from the Rac prophage to enhance homologous recombination between linear DNA molecules (LLHR) with 30-50 nucleotide homologous arms, promoting the direct cloning of linear BGCs to previously linearized plasmids (Fu *et al*., 2012). Because this LLHR approach resulted too laborious for large BGCs, Wang *et al*. (2018a) developed a new recombination RecET-based system (ExoCET) employing the in vitro T4 polymerase to enhance the annealing between the linear target BGC and the vector. The 106 Kb salinomycin BGC from *S. albus* was successfully cloned with this system.

The Transformation-Associated Recombination (TAR) system is also a cloning technique that uses the yeast *Saccharomyces cerevisiae* machinery to assembly a large size BGC into the *S. cerevisiae-E.coli-Streptomyces* linearized pCAP01 vector containing specific 1 Kb arms homologous to the ends of the target BGC (Yamanaka *et al*., 2014). For this TAR technique, spheroplast cells of *S. cerevisiae* are transformed with the BamHI-linearized pCAP01 and the genomic DNA digested with XbaI. Yamanaka *et al*. (2014) identified a novel daptomycin-like lipopeptide, taromycin A, cloning a 67 Kb cryptic BGC from *Saccharomonospora* sp. CNQ-490 and expressing it in *S. coelicolor* M1146. This TAR technique also allowed to heterologously express the first natural product BGC from the marine strain *Salinispora pacifica* to produce enterocin (Bonet *et al.*, 2015).

The DNA assembler method is another approach which allows the assembly of multiple short DNA fragments with short overlapping homologous arms, previously amplified by PCR, into a vector in yeast. Using DNA assembler, Shao *et al*., (2012 and 2013) cloned the aureothin and spectinabilin BGCs from *S. thioluteus* and *S. spectabilis*, respectively, and expressed them in *S. lividans*.

Gibson assembly is an *in vitro* technique similar to the latter involving the cloning of multiple DNA segments into a vector without the using of restriction enzymes or compatible restriction sites (Gibson *et al.*, 2009). A T5 exonuclease digests the 5' end of the DNA fragments, resulting in single strand overhangs that anneal with their homologous. One annealed, a polymerase and ligase fill the gaps and seal the nicks. Zhou *et al.* (2015) cloned successfully the 41 Kb conglobatin BGC from *S. conglobatus* using the Gibson assembly. The main challenge of the Gibson assembly technique is the cloning of many fragments of a BGC from high GC containing genomes, such as those of *Streptomyces*.

To overcome the limitations of the Gibson assembly, Jiang *et al*. (2015) developed an *in vitro* cloning technique called CATCH (Cas9-Assisted Targeting of CHromosome segments), which combines the RNA-guided Cas9 nuclease in-gel digestion of genomic DNA at the ends of the BGC, protecting it from shearing, and the Gibson assembly of the excised BGC into a plasmid containing homologous sequences with the ends of the target BGC. The authors cloned the 36 Kb jadomycin

BGC from *S. venezuelae*, the 32 Kb chlortetracycline BGC from *S. aureofaciens* and the bacillaene BGC from *Bacillus subtilis*.

Recently, a direct pathway cloning was developed by Greunke *et al*. (2018) named DiPac. This method is based on the high efficiency and long amplification ability of the Q5 high-fidelity polymerase and Gibson assembly. They amplified directly the 9.7 Kb phenazine BGC from *Serratia fonticola* DSM 4576 and assembled it by Gibson into the pET28b-PD1 plasmid. They also cloned the 55 Kb erythromycin BGC subdividing the cluster into five fragments of around 10 Kb, which were merged by HiFi DNA assembly. They also introduced a unique restriction site between the plasmid homology arms to linearize the plasmid after each round of assembly and enhance the insertion of all the fragments. However, this technique is still a challenge for large BGCs and requires the introduction of restriction sites for the cloning.

Once the BGC is cloned in the most appropriate plasmid, the construction must be transferred to a heterologous host. A suitable heterologous host must have a reduced secondary metabolism background, an easy genetic manipulation, a variety of natural products precursors (malonyl-CoA for PKS, amino acids for NRPS and RiPPs), a known regulatory network and the capacity to transcriptionally recognize exogenous DNA and express diverse types of BGCs (Xu and Wright, 2019). Based on this, some wild-type *Streptomyces* hosts, such as *Streptomyces albidoflavus* J1074, previously known as *S. albus* J1074 (Chater and Wilde, 1976; Labeda *et al*., 2014), and other strains with minimized genomes such as *Streptomyces coelicolor* M1152 and M1154 (Gómez-Escribano and Bibb, 2011), *Streptomyces avermitillis* SUKA5 and SUKA17 (Komatsu *et al*., 2013) or *Streptomyces lividans* SBT5 and SBT8 (Xu *et al*., 2016; Zhao *et al*., 2016), have been successfully used for heterologous expression. There are also other several actinomycete hosts such as *Amycolatopsis japonica* DSM 44213^{T} (Stegmann *et al.,* 2014) or *Saccharopolyspora erythraea* NRRL 23338 (Rodriguez *et al*., 2021) and other type of microorganisms, such as *E. coli, Pseudomonas putida, Saccharomyces cerevisiae* or *Bacillus subtilis*, that are suitable heterologous hosts.

However, heterologous expression has limitations, such as the need to clone and manipulate very large genome fragments (BGCs) that are usually laborious and time-consuming, the challenge of identifying a suitable host that avoids the foreign DNA degradation and provides all the necessary conditions for the biosynthesis of the corresponding natural products and the possible different codon usage of the foreign DNA and the heterologous host (Mitousis et al., 2020; Liu *et al.*, 2021).

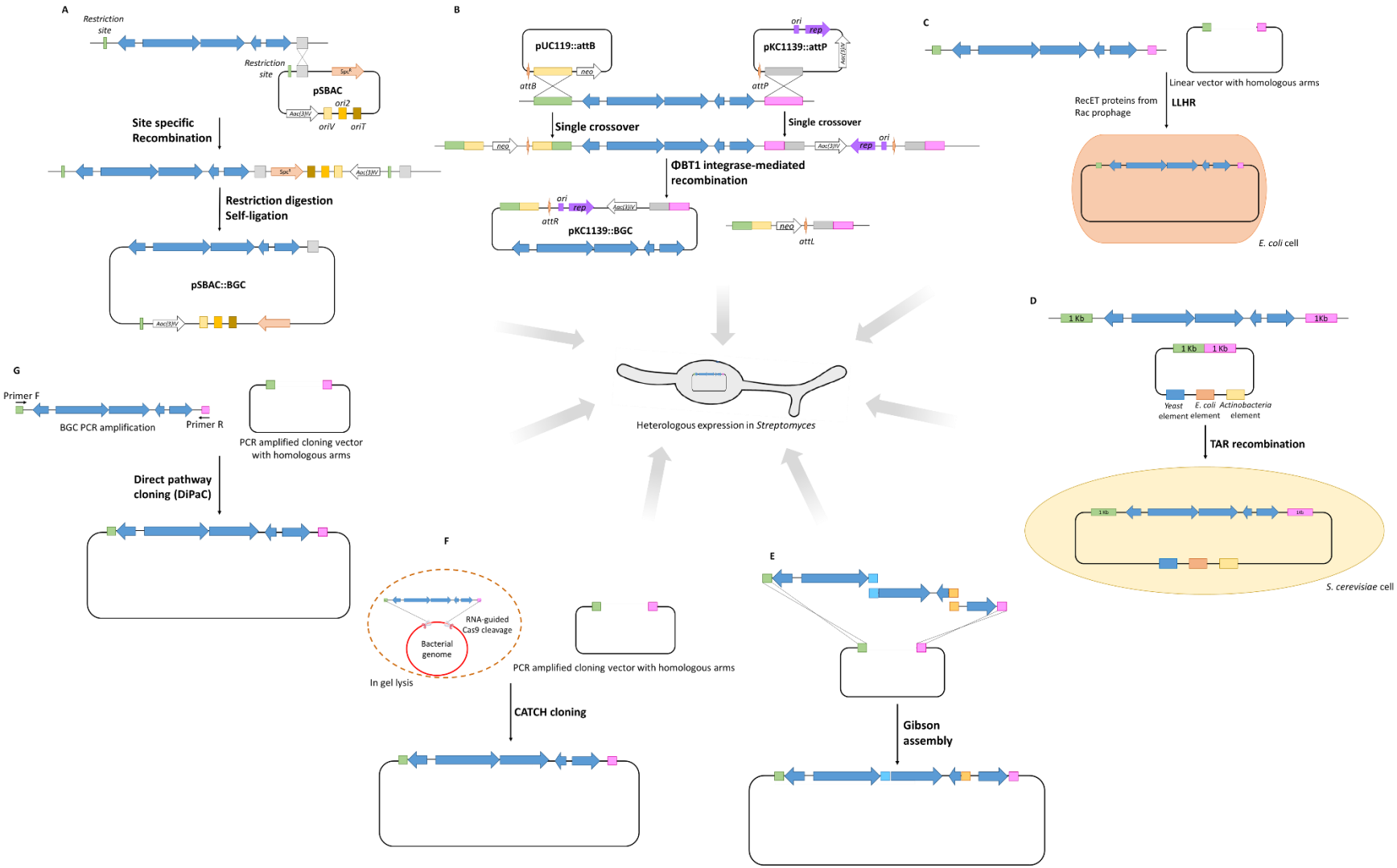


Figure A8. Different methods for the cloning of Biosynthetic Gene Clusters into shuttle vectors. A: Site-specific Recombination; B: Φ-BT1 phage integrase-mediated site-specific recombination (ISR); C: Linear-Linear Homologous Recombination (LLHR); D: Transformation-Associated Recombination (TAR); E: Gibson Assembly; F: Cas9-Assisted Targeting of CHromosome segments (CATCH) cloning); G: Direct Pathway Cloning (DiPaC).

6. ISOLATION AND IDENTIFICATION OF NEW NATURAL PRODUCTS

The identification of the putative natural products produced by a new biosynthetic gene cluster starts with the extraction of the secondary metabolites from the producer organisms. The choice of the extraction method will favor the presence of certain compound classes in the extracts (for example, the use of more polar solvents will result in a higher abundance of polar compounds in the extract). Then, these generated extracts can be screened against a panel of Gram-positive, Gram-negative and fungal pathogens to identify hits with potential antimicrobial bioactivity.

These hit extracts can contain a variety of natural products that require further separation and purification to isolate the putative bioactive natural products. This separation depends on the type of chemical and physical differences between the components and the main method employed is column chromatography. Different types of column chromatography can be used for the separation of natural products and the separation based on the different polar affinities of the natural products for the resin (adsorbent). The selection of the adsorbent or stationary phase and dissolvent or mobile phase is essential for the proper separation of the compounds. This type of separation can also be direct-phase or reversed-phase chromatography. In the direct-phase chromatography, the stationary phase is polar, like silica gel or alumina, retaining polar natural products longer in the columns than nonpolar ones. On the other hand, in the reversed-phase chromatography, the stationary phase is nonpolar, like C8 or C18, so nonpolar compounds are retained longer than polar (Zhang *et al*., 2018). Other separation methods are also widely used for the isolation of natural products, such as the separation based of the molecular size by membrane filtration or gel filtration chromatography, separation based on ionic strength (ion-exchange chromatography) where charged molecules can be retained and release changing the ionic strength of the mobile phase, preparative gas chromatography (Prep-GC) for the separation of volatile natural products or separation mediated by monoclonal antibodies.

Chemical dereplication is essential to carry out the rapid identification of known compounds and analogues in active samples (Chervin *et al*., 2017), which includes the determination of molecular mass and formula and the search in structural NP databases with taxonomic information. Such metadata are often compiled in proprietary databases, such as the Dictionary of Natural Products or Antibase, which comprises all NP structures reported with links to their biological sources (Pérez-Victoria *et al*., 2016), or the publicly available Natural Products Atlas, which includes a wide variety of information of 32.552 microbial compounds (van Santen *et al*., 2020). The most commonly used methods involve the use of liquid chromatography coupled to a photodiode array detector and a high-resolution mass spectrometer (LC-DAD-HRMS) (Chervin *et al*., 2017). The mass spectrometry provides high sensitivity and versatility, providing different types of data such as retention time, positive, negative and fragmentation, and has become the detection technique most frequently used as enables the detection of small amounts of compounds, even in trace-levels (Hubert *et al*., 2017). LC-coupled to high-resolution mass spectrometers such as Time-of-Flight (TOF) or Fourier Transform (FT) devices currently represent the gold standard high-throughput platforms for qualitative and quantitative metabolite profiling and enables routine acquisition of accurate molecular mass information,

which can provide unambiguous assignment of molecular formula for hundreds to thousands of metabolites (Fontana *et al*, 2020).

Another approach for the confirmation of the dereplication candidates obtained is the MS/MS analysis combined with *in silico* fragmentation tools. The use of MS/MS data can identify structurally related compound but requires MS/MS data in integrated compounds databases for their identification. However, a comprehensive experimental tandem mass spectrometry database of all NPs reported to date does not exist and the current MS/MS libraries only cover a limited number of natural products (Chervin *et al*., 2017). In this respect, the Global Natural Products Social (GNPS) molecular networking platform organizes thousands of sets of MS/MS data recorded from a given set of extracts and visualizes the relationship of the compounds as clusters of structurally related molecules (Wang *et al.*, 2016). This improves the efficiency of dereplication by enabling annotation of isomers and analogues of a given metabolite in a cluster.

Additionally, Nuclear Magnetic Resonance (NMR) spectroscopy is increasingly being used for unequivocal dereplication and determination of definitive molecular structures of natural products as it provides full structural information. Despite the lower sensitivity compared to MS-based analytical methods, the use of high field magnets, capillary and cryogenic probes allow the acquisition of proton spectra and 2D heteronuclear correlation experiments, which have progressively improved the sensitivity of the technique and facilitate the elucidation of the structures of minor compounds.

7. *STREPTOMYCES CACAOI* CA-170360 IS THE PRODUCER OF NEW BIOACTIVE NATURAL PRODUCTS

The global threat caused by the current antibiotic resistance crisis makes the discovery of novel antibacterial agents an urgent need. As a result of Fundación MEDINA's antibiotic discovery program, a family of eight new cyclic pentapeptides, known as pentaminomycins A-H, along with the already known BE-18257 A-C antibiotics, and a novel glycolanthipeptide called cacaoidin were isolated from MPG culture extracts of the strain *Streptomyces cacaoi* CA-170360, from MEDINA's microbial collection.

	Amino acid sequence
Pentaminomycin A	L-Val/D-Val/L-Trp/N5-OH-L-Arg/D-Leu
Pentaminomycin B	L-Leu/D-Val/L-Trp/N5-OH-L-Arg/D-Leu
Pentaminomycin C	L-Leu/D-Val/L-Trp/N5-OH-L-Arg/D-Phe
Pentaminomycin D	L-Val/D-Val/L-Trp/N5-OH-L-Arg/D-Phe
Pentaminomycin E	L-Phe/D-Val/L-Trp/N5-OH-L-Arg/D-Phe
Pentaminomycin F	L-Leu/D-Val/L-Trp/N5-OH-L-Arg/D-Pal
Pentaminomycin G	L-Phe/D-Val/L-Trp/N5-OH-L-Arg/D-Pal
Pentaminomycin H	L-Phe/D-Val/L-Trp/N5-OH-L-Arg/D-Leu

Table A3. Amino acid sequences of pentaminomycins A-H.

The new pentaminomycins A-H are cyclic pentapeptides containing the common core sequence Val-Trp-N5-OHArg including a modified amino acid residue, N5-hydroxyarginine. The structures were determined by a combination of mass spectrometry, NMR and Marfey's analyses (Carretero-Molina *et al.*, 2021). The amino acid sequences and structures of pentaminomycins A-H are shown in Table A3 and Figure A9 (Jang *et al.*, 2018; Hwang *et al.*, 2020; Carretero-Molina *et al.*, 2021).

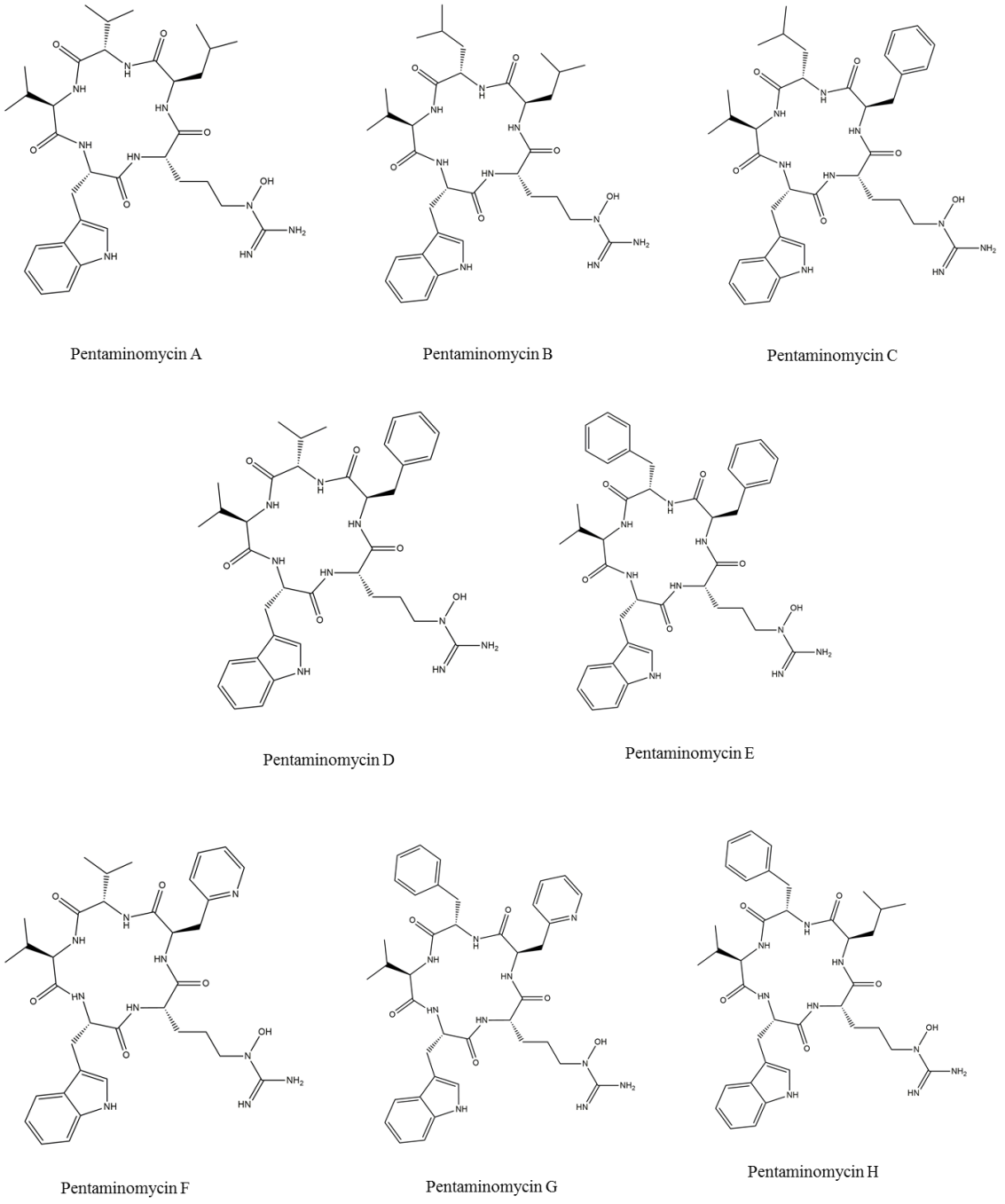


Figure A9. Structures of pentaminomycins A-H. Figure taken from Román-Hurtado et al*, 2021b.*

The bioactivities of the different members of pentaminomycins A-H were evaluated against a Gram-negative bacteria panel, including *Acinetobacter baumannii, Pseudomonas aeruginosa, Escherichia coli and Klebsiella pneumoniae*. Only pentaminomycins B and C displayed a moderate selective antibacterial activity against a clinical isolate of *A. baumannii* (Carretero-Molina *et al.*, 2021). Other authors have recently proposed distinct biological activities for some of the pentaminomycins. Pentaminomycin A has shown antimelanogenic activity against α-melanocyte stimulating hormone (α-MSH)-stimulated B16F10 melanoma cells (Jang *et al.*, 2018) and pentaminomycin C is active against Gram positive bacteria but no against Gram negative bacteria (Kaweewan *et al.*, 2020). Both pentaminomycins C and D act as autophagy inducers on HEK293 cells (Hwang *et al.*, 2020). Pentaminomycins F and G are characterized by containing the rare amino acid 3-(2-pyridyl)-alanine, resulting in the first non-ribosomal peptide reported with this residue (Carretero-Molina *et al.*, 2021).

The BE-18257 antibiotics are cyclic pentapeptides originally isolated from *Streptomyces misakiensis* BA18257 as endothelium-derived vasoconstrictor factor binding inhibitors (Nakajima *et al.*, 1991). The BE-18257 A-C pentapeptides contain three D-form amino acid residues, and were elucidated to be L-Leu, D-Trp, D-Glu, L-Ala, D-Val/D-*allo*-Ile/D-Leu, respectively (Figure A10).

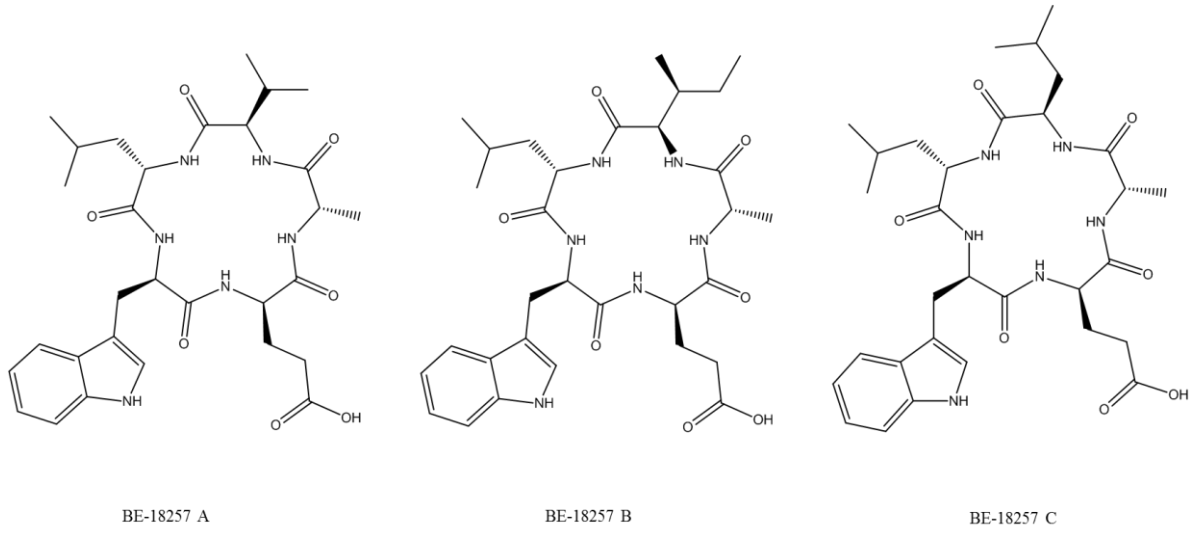


Figure A10. Structures of BE-18257 A-C antibiotics. Figure taken from Román-Hurtado et al*, 2021b.*

Interestingly, the *Streptomyces cacaoi* CA-170360 strain was also the producer of cacaoidin, a novel glycosylated lanthipeptide with remarkable unprecedented structural features.

Along with different proteinogenic amino acids, cacaoidin peptide presents posttranslational modified residues including a lanthionine ring, restricted thus far to lanthipeptides, and an N-terminal dimethylation, present only in linaridins, in an unprecedented *N,N*-dimethyl lanthionine system (NMe2Lan). The molecule also shows an unusually high number of D-amino acids, including D-2-aminobutyric acid (Abu), and a C-terminal S-[(Z)-2-aminovinyl-3-

methyl]-D-cysteine (AviMeCys) amino acid (Figure A11). The molecule also combines an O-glycosylation of a Tyr residue with the disaccharide β-6-deoxygulopyranosyl-(1→3)-α-rhamnopyranoside, which, up to date, has not been previously reported in any other natural product (Ortiz-López *et al.*, 2020).

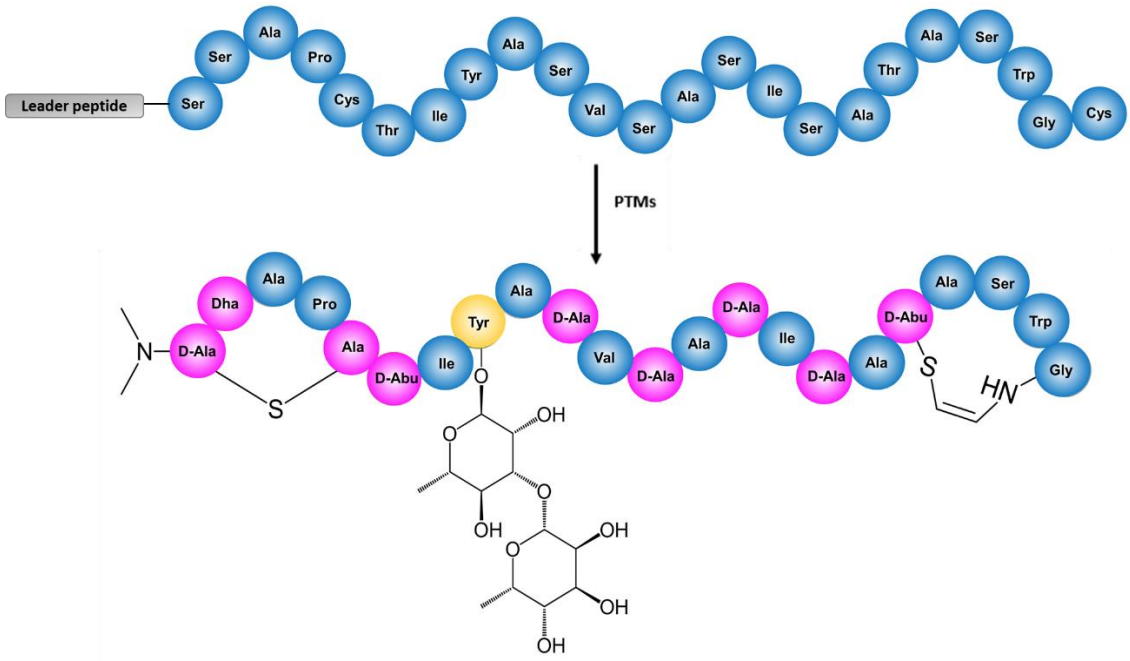


Figure A11. Structure of cacaoidin before (top) and after (bottom) the Post-Translational Modifications (PTMs). In the modified cacaoidin, absolute configuration for NMe2Ala(S)-1 is tentatively proposed as (2S) (equivalently, D-Ala). The amino acids modified by biosynthetic enzymes are colored in pink, the glycosylated tyrosine is colored in yellow, and the non-modified amino acids are colored in blue. Figure taken from Román-Hurtado et al*, 2021a.*

The antibacterial activity of cacaoidin was evaluated against different Gram-positive pathogens. Cacaoidin showed potent activity against methicillin-resistant *Staphylococcus aureus* (MRSA) (MIC 0,5 µg/mL) and a moderate activity against a set of *Clostridium difficile* strains (MIC 4-8 µg/mL). Furthermore, cacaoidin triggered the induction of the LiaRS (Lipid II cycle interfering antibiotic response regulator and sensor) bioreporter in *B. subtilis*, which suggested an interference with the Lipid II biosynthesis cycle and the cell wall synthesis as the specific target pathway (Ortiz-López *et al.*, 2020).

Cacaoidin is the first reported member of lanthidins, a new RiPP subfamily with structural elements found in lanthipeptides and linaridins (Montalbán-López *et al.*, 2021). Since linaridins are characterized by the presence of dehydrobutyrine (Dhb) residues, which are not present in cacaoidin, but lanthipeptides are characterized by lanthionine, which is present in cacaoidin, it has been proposed that lanthidins constitute the new class V lanthipeptides (Montalbán-López *et al.*, 2021).

B. Objectives

This PhD work was carried out in Fundación MEDINA, as part of the program focused on the discovery of new antibiotics. During a screening program to discover new antibacterial natural products from microorganisms, a family of eight new cyclic pentapeptides named pentaminomycins A-H, and the first member of class V lanthipeptides or lanthidins, cacaoidin, were isolated from the strain *Streptomyces cacaoi* CA-170360 and structurally characterized. Due to their novel and interesting structures and their bioactivities against Gram-negative and Gram-positive pathogens, respectively, it was important to study and analyze the genome of this strain and how these families of compounds were produced.

Thus, the general purpose of this work was the identification and study of the biosynthetic gene clusters of these families of new natural products produced by the strain *Streptomyces cacaoi* CA-170360.

As part of this main objective, the following specific aims were raised:

- Whole genome sequencing, assembly and annotation of the strain *Streptomyces cacaoi* CA-170360.

- Prediction of the secondary metabolite biosynthetic gene clusters present in the genome of the strain CA-170360 and identification of compound-linked gene clusters based on the secondary metabolites detected in culture.

- Identification, cloning and heterologous expression of the biosynthetic gene cluster responsible of the production of BE-18257 antibiotics and pentaminomycins A-H.

- Identification, cloning and heterologous expression of the biosynthetic gene cluster responsible of the production of cacaoidin.

- Generation of knockout strains by genetic replacement of several genes present in the biosynthetic gene clusters of cacaoidin and pentaminomycins A-H and BE-18257 antibiotics, in order to determine their function and generate derivatives of these compounds.

C. Materials and Methods

1. BACTERIAL STRAINS

The list of bacterial strains used in this work, along with their origin and some of the most relevant characteristics of their genotype, are shown in Table C1.

Streptomyces cacaoi CA-170360, from Fundación MEDINA's Culture Collection, was the first strain identified as the producer of cacaoidin, pentaminomycins A-H and cyclic pentapeptides BE-18257 A-C. It was isolated from the rhizosphere of *Brownanthus corallinus*, in the region of Namaqualand (South Africa) (Ortiz-López *et al.,* 2020)

Streptomyces albidoflavus J1074, previously identified as *S. albus* J1074 (Chater and Wilde, 1976; Labeda *et al*., 2014), kindly provided by José Antonio Salas, *Streptomyces coelicolor* M1152 and *Streptomyces coelicolor* M1154 (Gómez-Escribano and Bibb, 2010), generously provided by Mervyn Bibb, were employed as heterologous expression hosts.

Electrocompetent *Escherichia coli* NEB 10-β (New England BioLabs, Ipswich, MA, USA), *E. coli* ET12567 (MacNeil *et al.,* 1992) and *E. coli* ET12567/pUB307 (Flett *et al*., 1997) (also kindly provided by José Antonio Salas) were used for plasmid transformation and intergeneric biparental/triparental conjugation. One Shot TOP10 competent cells containing the pR9406 plasmid (Jones *et al*., 2013) were used as a helper strain in triparental conjugation to introduce large plasmids with a kanamycin resistance gene into *E. coli* ET12567 cells.

Methicillin resistant *Staphylococcus aureus* (MRSA COL MB5393) is a hospital-acquired penicillinase-negative strain (Dyke *et al.,* 1966) belonging to Fundación MEDINA's strain collection which was used in the screening of the antimicrobial activity.

Microorganism	Characteristics/Phenotype	Reference
Streptomyces cacaoi CA-170360	Producer strain of cacaoidin, pentaminomycins and BE-18257 antibiotics. Isolated from the rhizosphere of a specimen of *Brownanthus corallinus*, in the region of Namaqualand (South Africa).	Ortiz-López *et al*., 2020
Streptomyces albidoflavus J1074	*Ilv*-1, *sal*-2, *RM*-, Paulomycins producer	Chater and Wilde, 1976 Labeda *et al*., 2014
Streptomyces coelicolor M1152	Derivative of *S. coelicolor* M145 (SCP1$^-$SCP2$^-$), Δ*act* Δ*red* Δ*cpk* Δ*cda rpoB*(C1298T)	Gómez-Escribano *et al*., 2010
Streptomyces coelicolor M1154	Derivative of *S. coelicolor* M1152, Δ*act* Δ*red* Δ*cpk* Δ*cda rpoB*(C1298T) *rpsL*(A262G)	Gómez-Escribano *et al*., 2010
Streptomyces albidoflavus CA-295619	*Streptomyces albidoflavus* J1074/pCAP01, negative control harboring empty pCAP01 vector	Sánchez-Hidalgo *et al*., 2020
Streptomyces coelicolor CA-29560	*Streptomyces coelicolor* M1152/pCAP01, negative control harboring empty pCAP01 vector	Sánchez-Hidalgo *et al*., 2020
Streptomyces coelicolor CA-295621	*Streptomyces coelicolor* M1154/pCAP01, negative control harboring empty pCAP01 vector	Sánchez-Hidalgo *et al*., 2020
Streptomyces albidoflavus CA-300429	*S. albidoflavus* J1074/pCAO, strain containing the cacaoidin biosynthetic gene cluster	This work
Streptomyces albidoflavus CA-300381	*S. albidoflavus* J1074/pCCP1, strain containing the BE-18257 antibiotics biosynthetic gene cluster	This work
Streptomyces albidoflavus CA-300436	*S. albidoflavus* J1074/pCCP2, strain containing the BE-18257 antibiotics and pentaminomycins A-H biosynthetic gene cluster	This work
Streptomyces albidoflavus CA-300655	Derivative of *S. albidoflavus* CA-300429, Δ*cao4*::*aac(3)IV*	This work
Streptomyces albidoflavus CA-300647	Derivative of *S. albidoflavus* CA-300429, Δ*cao8*::*aac(3)IV*	This work
Streptomyces albidoflavus CA-300648	Derivative of *S. albidoflavus* CA-300429, Δ*cao16*::*aac(3)IV*	This work
Streptomyces albidoflavus CA-300649	Derivative of *S. albidoflavus* CA-300429, Δ*cao24*::*aac(3)IV*	This work
Streptomyces albidoflavus CA-301651	Derivative of *S. albidoflavus* CA-300436, Δ*cppA*::*aac(3)IV*	This work

Microorganism	Characteristics/Phenotype	Reference
Streptomyces albidoflavus CA-301341	Derivative of *S. albidoflavus* CA-300655, complemented with pEM4T/cao4	This work
Streptomyces albidoflavus CA-301580	Derivative of *S. albidoflavus* CA-300647, complemented with pEM4T/cao8	This work
Streptomyces albidoflavus CA-300647	Derivative of *S. albidoflavus* CA-301651, complemented with pEM4T/cppA	This work
Streptomyces coelicolor CA-301619	*Streptomyces coelicolor* M1152/pCAO, strain containing the cacaoidin biosynthetic gene cluster	This work
Streptomyces coelicolor CA-301632	*Streptomyces coelicolor* M1154/pCAO, strain containing the cacaoidin biosynthetic gene cluster	This work
Escherichia coli ET12567	Non-methylating donor strain in conjugations, *dam*-13::*Tn*9 *dcm*-6 *hsdM hsdS*	MacNeil *et al.*, 1992
Escherichia coli ET12567/pUB307	Helper/donor strain in conjugations, *E. coli* ET12567 containing self-transmissible plasmid pUB307, F-*dam*-13::*Tn*9 *dcm*-6 *hsdM hsdR zjj*202::*Tn*10 *recF*143 *galK*2 *galT*22 *ara*-14 *lacY*1 *xyl*-5 *leuB*6 *thi*-1 *tonA*31 *rpsL*136 *hisG*4 *tsx*-78 *mtl*-1 *glnV*44	Flett *et al*., 1997
Escherichia coli NEB 10-β	Strain for high-quality plasmid preparations, Δ*(ara-leu) 7697 araD139 fhuA* Δ*lacX74 galK16 galE15 e14-* φ*80*d*lacZ*Δ*M15 recA1 relA1 endA1 nupG rpsL* (Str^R) *rph spoT1* Δ*(mrr-hsdRMS-mcrBC)*	New England Biolabs
Escherichia coli One Shot TOP10/pR9406	Helper strain in conjugations containing pR9406, a Car^R derivative of pUB307, F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Jones *et al*., 2013
Staphylococcus aureus MRSA MB5393	Strain used as indicator in the antimicrobial assay	Dyke *et al.,* 1961

Table C1. Microbial strains used along this work.

2. PLASMIDS

For the development of this work, several plasmids have been used (Table C2).

pCR-Blunt (Invitrogen, Carlsbad, California) was used to clone the PCR fragments needed to create the knockouts and have them sequenced with the universal oligonucleotides M13 forward (-20) and M13 reverse (-24).

pCAP01 (Figure C1a) is a *S. cerevisiae-E. coli*-actinobacteria shuttle vector that can integrate cloned gene clusters into the genome of heterologous actinobacteria hosts due to its site-specific ϕC31 integrase (Yamanaka *et al*., 2014). It was used for the cloning of the cacaoidin, pentaminomycins A-H and cyclic pentapeptides BE-18257 BGCs.

pEFBAoriT (Figure C1b) is a pBluescript derivative containing the *oriT* region from pOJ260 and an apramycin resistance cassette (Horna, et al 2011; Fernández-Lozano, 2000) that has been used as backbone for the constructions needed to develop the knockouts. The PCR fragments are cloned on both sides of the apramycin cassette and after a double recombination event the target gene is replaced by this cassette.

pEM4T (Figure C1c) (Menéndez *et al.,* 2006) is a conjugative and multicopy vector (with *oriTRP4* as origin of transfer) that confers resistance to ampicillin in *E. coli* and to thiostrepton in *Streptomyces*. It was used to overexpress a previously disrupted gene under the control of the promoter *ermE*p*.

Plasmid	Characteristics	Reference
pCR-Blunt	*lac* promoter, *lacZα-cdcB* fusion gene, T7 promoter, M13 Forward and M13 Reverse priming sites, *kan*^{R}, *Sh ble* (Zeocin resistance gene), pUC origin (3.5 Kb)	Invitrogen
pCAP01	RP4 origin, *attP* site, TRP1 gene, CEN6_ARS4, URA3, ϕC31 integrase, *neo*^{R}/*kan*^{R} (9.022 Kb)	Yamanaka *et al*., 2014
pCAO	pCAP01 containing the *cao* cluster, responsible for the production of cacaoidin (51.5 Kb)	This work
pCPP1	pCAP01 containing the *cpp1* cluster, responsible for the production of BE-18257 antibiotics (37.7 Kb)	This work
pCPP2	pCAP01 containing the *cpp2* cluster, responsible for the production of BE-18257 antibiotics and pentaminomycins A-H (57 Kb)	This work
pEFBAoriT	*aac(3)IV*^{R}, pBluescript derivative containing an *oriT* (5.4 Kb), used as backbone for the constructions needed for the knockouts	Horna *et al*., 2011
pEFBA-caoA	*aac(3)IV*^{R}, *tsr*^{R}, pEFBAoriT derivative used for the replacement of *caoA*	This work
pEFBA-cao4	*aac(3)IV*^{R}, *tsr*^{R}, pEFBAoriT derivative used for the replacement of *cao4*	This work
pEFBA-cao8	*aac(3)IV*^{R}, *tsr*^{R}, pEFBAoriT derivative used for the replacement of *cao8*	This work
pEFBA-cao16	*aac(3)IV*^{R}, *tsr*^{R}, pEFBAoriT derivative used for the replacement of *cao16*	This work
pEFBA-cao24	*aac(3)IV*^{R}, *tsr*^{R}, pEFBAoriT derivative used for the replacement of *cao24*	This work
pEFBA-cppA	*aac(3)IV*^{R}, *tsr*^{R}, pEFBAoriT derivative used for the replacement of *cppA*	This work
pEM4T	*bla*^{R}, *tsr*^{R}, *oriTRP4, ermE**p. Used for the overexpression of genes	Menéndez *et al*., 2006
pEM4T-cao4	pEM4T derivative with *cao4* gene cloned into the EcoRI site	This work
pEM4T-cao8	pEM4T derivative with *cao8* gene cloned into the BamHI and EcoRI sites	This work
pEM4T-cao16	pEM4T derivative with *cao16* gene cloned into the BamHI and EcoRI sites	This work
pEM4T-cao24	pEM4T derivative with *cao24* gene cloned into the EcoRI site	This work
pEM4T-cppA	pEM4T derivative with *cppA* gene cloned into the BamHI and EcoRI sites	This work

Table C2. List of the plasmids employed in this work.

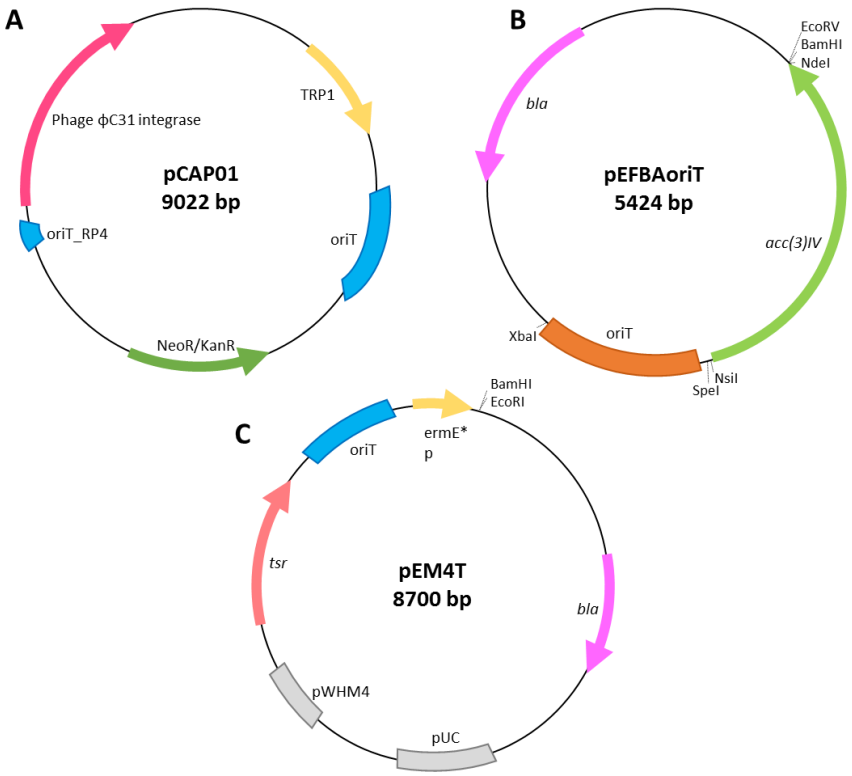


Figure C1. Map of the plasmids pCAP01 (A), pEFBAoriT (B) and pEM4T (C). These have been the most used plasmids in this work for the cloning of BGCs, generation of knockouts and gene complementation, respectively.

3. ANTIBIOTICS

Table C3 shows all the antibiotics used, in this work. These antibiotics were added to the cultivation media when needed.

Antibiotic	Solvent	Working concentration (µg/mL)	Reference
Apramycin	Distilled water	50	SIGMA A2024-5G
Ampicillin	Distilled water	50	SIGMA A1066-5G
Kanamycin	Distilled water	50	SIGMA K1377-5G
Chloramphenicol	Ethanol 96 %	25	SIGMA C0378-5G
Thiostrepton	DMSO	50	SIGMA T8902-1G
Nalidixic acid	0.5 N NaOH	25	SIGMA N8878-5G

Table C3. List of antibiotics used in this work in cultivation media.

4. CULTURE MEDIA

All the culture media compositions refer to the liquid formulation; for the agar-based formulation, 1.5-2 % agar was added. The composition is expressed in grams per liter of distilled water. The media were autoclaved at 121 °C for 20 minutes.

4.1. Culture media used for *E. coli*

- **Luria-Bertani** or **Lysogeny Broth Medium** (LB medium) (Miller, 1972) This medium was used for the cultivation of *E. coli* strains

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
pH adjusted to 7	

4.2. Culture media used for *Streptomyces*

- **Bennet** (Locci *et al*., 1969): Optimized medium for the sporulation of *Streptomyces*.

Glucose	10 g
NZ amine A	2 g
Yeast extract	1 g
Beef extract	1 g
pH adjusted to 7.2	

- **MA** (Fernández *et al*., 1998): This medium was employed for the growth and sporulation of *Streptomyces* after the biparental or triparental conjugation with *E. coli* strains.

MOPS	21 g
Glucose	5 g
Yeast extract	0.5 g
Beef extract	0.5 g
Casamino acids	1 g
pH adjusted to 7 with KOH	

- **ATCC-2** (Kieser *et al.,* 2000)**:** This liquid medium was used to obtain *Streptomyces* pre-inoculum to seed the different fermentation media.

Soluble starch	20 g
Glucose	10 g
NZ Amine type E	5 g
Meat extract	3 g
Peptone	5 g
Yeast extract	5 g
Calcium carbonate	1 g
pH adjusted to 7	

The following fermentation media were used for the production and isolation of secondary metabolites:

- **YEME**

Yeast extract	3 g
Peptone	5 g
Oxoid malt extract	3 g
Glucose	10 g
Sucrose	340 g
$MgCl_2$-$6H_2O$	2 mL

- **R2YE**

Yeast extract	5 g
Sucrose	103 g
K_2SO_4	0.25 g
$MgCl_2$-$6H_2O$	10.12 g
Glucose	10 g
Casamino acids	0.1 g
KH_2PO_4 0.5 %	10 mL
$CaCl_2$-$2H_2O$ 3.68 %	80 mL
L-proline 20 %	15 mL
TES buffer 5.73 % adjusted to pH 7.2	100 mL
Trace element solution	2 mL
NaOH 1N	5 mL

Trace element solution (g/100 mL):	
$ZnCl_2$	0.004 g
$FeCl_3 \cdot 6H_2O$	0.02 g
$CuCl_2 \cdot 2H_2O$	0.001 g
$MnCl_2 \cdot 4H_2O$	0.001 g

$Na_2B_4O_7 \cdot 10H_2O$	0.001 g
$(NH_4)_6Mo_7O_{24} \cdot 4H_2O$	0.001 g

- **KH4**

Glucose	4 g
Yeast extract	4 g
Malt extract	10 g
$CaCO_3$	2 g

- **MPG**

Glucose	10 g
Millet	20 g
Cottonseed flour	20 g
MOPS	20 g
pH adjusted to 7	

- **FR23**

Glucose	5 g
Soluble starch from potato	30 g
Cottonseed flour	20 g
Cane molasses	20 g
pH adjusted to 7	

- **DEF-15**

Sucrose	40 g
$ClNH_4$	2 g
Na_2SO_4	2 g
K_2HPO_4	1 g
$Cl_2Mg \cdot 6H_2O$	1 g
Trace element solution	1 mL
$CaCO_3$	2 g
pH adjusted to 7	

Trace element solution (g/100 mL):

$MnCl_2 \cdot 4H_2O$	0.1 g
$ZnCl_2$	0.1 g
$FeCl_2 \cdot 4H_2O$	0.1 g
NaI	0.05 g

5. CULTIVATION CONDITIONS

5.1. *E. coli* cultures

E. coli strains were cultivated overnight in agar LB plates supplemented with the appropriate antibiotics at 37 °C and single colonies were picked to inoculate liquid LB cultures, flasks or tubes, supplemented with the appropriate antibiotics that were shaken at 250 rpm at 37 °C.

5.2. *Streptomyces* cultures and fermentations

Streptomyces strains were seeded in tubes containing in 10 mL of medium ATCC-2 from glycerol stocks (5 %) and incubated at 28 °C for 2-3 days, 220 rpm, 70 % humidity. These cultures were used to seed Bennet plates, which were incubated at 28 °C for 3-4 days to obtain spores, and also to seed different fermentation media (in 500 mL flasks containing 125 mL of medium or 40 mL EPA vials containing 10 mL of medium) that were incubated at 28 °C, 220 rpm, 70 % humidity for the days needed to produce secondary metabolites.

5.3. Strains preservation

E. coli glycerol stocks were prepared from overnight *E. coli* cultures and diluted with sterile 50 % (w/v) glycerol to a final concentration of 25 %. The glycerol stocks were stored at -80 °C.

Streptomyces strains were preserved from liquid cultures or from spores. ATCC-2 cultures were diluted with sterile 40 % (v/v) glycerol to a final concentration of 20 % and stored at -80 °C. On the other hand, a confluent lawn of spores from a Bennet plate was directly transferred to a sterile 20 % (v/v) glycerol tube. These spore glycerol stocks were stored at -80 °C.

6. DNA EXTRACTION

6.1. *Streptomyces* genomic DNA extraction

Streptomyces genomic DNA was extracted and purified from axenic cultures grown on ATCC-2 liquid medium on an orbital shaker at 28 °C, 220 rpm and 70 % relative humidity, following the protocol described by Kieser *et al*., 2000. A total of 1.5 mL of mycelium suspension was transferred to Eppendorf tubes and centrifuged at 13.000 rpm (Biofuge-fresco, Heraeus) for 5 minutes at 4 °C. The supernatant was removed, and the pellet was incubated with 150 µL TE buffer containing 3 mg/ml lysozyme at 37 °C for at least 1 h. Then, 800 µL of extraction buffer (50 mM EDTA pH 8.5, 0.2 % SDS) containing 1 mg/ml proteinase K were added and the sample was incubated for 30 min at 70 °C. After this time, it was centrifuged at 13.000 rpm for 20 minutes at 4 °C, and 800 µL from the supernatant were transferred to a new tube, where 80 µL 3 M KAc, pH 5.2 were added. The sample was incubated at -20 °C for 20 minutes and then centrifuged for 10 minutes at 13.000 rpm, 4 °C. 700 µL from the supernatant were transferred to a new tube and 800 µL isopropanol were added, incubating it at -20 °C for 10 minutes. After another

centrifugation for 10 minutes at 13.000 rpm, 4 °C, the supernatant was discarded, and the pellet was washed with 500 µL 70 % ethanol. The sample was air-dried at 37 °C and finally the DNA was resuspended in 100 µL Milli Q water.

6.2. Plasmid DNA extraction (minipreparations)

The QIAprep Spin Miniprep Kit from QIAGEN was employed for the extraction and purification of plasmid DNA from *E. coli*, following the manufacturer's instructions. This QIAprep Miniprep procedure is based on the modified alkaline lysis method of Birnboim and Doly (1979) of bacterial cells followed by adsorption of plasmid DNA onto silica in the presence of high salt concentrations.

The pellet of overnight *E. coli* cultures grown at 37 °C, was lysed under alkaline conditions and the lysate was subsequently neutralized and adjusted to high-salt-binding conditions. After lysate clearing, the samples were purified on the QIAprep silica membranes, which allowed the selective adsorption of plasmid DNA in high-salt buffers and the elution in low-salt buffers. RNA, cellular proteins, endonucleases or other metabolites are not retained on the membrane but are removed in the flow through. Finally, the plasmid DNA was eluted with 50 µL of Milli-Q water.

7. DNA MANIPULATION TECHNIQUES

7.1. Polymerase Chain Reaction (PCR)

This technique was used to amplify targeted DNA fragment sequences by cyclic repetition of three steps: template DNA denaturation, hybridization of oligonucleotides and polymerase extension (Saiki *et al.,* 1985; Mullis *et al.,* 1986).

- **DNA denaturation.** The denaturation of the two strands of the DNA template is carried out by its incubation at high temperature (95 °C).
- **Annealing.** This consists in the hybridization of the oligonucleotides with the complementary sequences of the template DNA, The Tm for each primer was calculated according to the next formula (Sambrook *et al.,* 2001):

$$\mathbf{Tm = 4 \times (\% G+C) + 2 \times (\% A+T)}$$

- **Extension.** DNA polymerization was performed at 72 °C and the extension time depending on the size of the DNA fragment to be amplified (normally 1 minute per kilobase).

7.1.1. *Oligonucleotides*

Complementary oligonucleotides were designed to provide the 3'-OH end to begin the synthesis of the new DNA strand using the single DNA template.

The whole genome sequence of *Streptomyces cacaoi* CA-170360 was used to design the oligonucleotides employing the software Geneious® 9.1.8 and considering certain determining factors such as the length, the melting temperature, the GC content, the hairpin and self-dimer temperatures, the introduction of certain restriction sites at the beginning of the sequence to promote the later cloning of the PCR products into the plasmids.

All the oligonucleotides designed for this work are shown in Tables C4-C9:

Oligonucleotide	Sequence (5'-3')	Application
Glyco1-sgRNA	TAATACGACTCACTATAGGACGACTCACGTGTCAAAGAGTTTTAGAGCTAGAAATAGCAA	pCAO synthesis
Glyco2-sgRNA	TAATACGACTCACTATAGGGGGCGAGATGCCATTCCAAGGTTTTAGAGCTAGAAATAGCAA	
sgRNA-F	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	
sgRNA-R	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACT	
pCAP01-Glyco-F	CGTGCGGTGGACCGCGCCGTGACCCCCTTG**TCGAGACTTGAGGTACCTGT**	
pCAP01-Glyco-R	GGGCCGGGTTCCAGCCGGTGATGCCGTCTT**TCGAGGTTACTAGTCGATCT**	
Glyco1-check-F	CGTGTGGTGAGAGAACTGA	Checking of pCAO
Glyco1-check-R	TCTGCTCTGATGCCGCATAG	
Glyco2-check-F	CCGCTGAGTGGTATGAGCTT	
Glyco2-check-R	TCGCTCCGATGTTCTGAAA	

Table C4. Primers used for the cloning and checking of the cao *biosynthetic gene cluster. The underlined sequences are the targets where Cas9 must cut and the bold sequences belong to the pCAP01 plasmid.*

Oligonucleotide	Sequence (5'-3')	Application
Penta1-sgRNA	TAATACGACTCACTATAGATGATCCAGAATCCGTGCTTGTTTTAGAGCTAGAAATAGCAA	pCCP1 and pCCP2 synthesis
Penta2-sgRNA	TAATACGACTCACTATAGACCCCAGACTTCAGCGTTTGGTTTTAGAGCTAGAAATAGCAA	
Penta3-sgRNA	TAATACGACTCACTATAGGAACTGAAGGCACAACCAAAGTTTTAGAGCTAGAAATAGCAA	
sgRNA-F	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTC	
sgRNA-R	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACT	
pCAP01-Penta1-F	AGGCTAGTCAGGGGTACCGGGCCCCTCAAA**TCGAGACTTGAGGTACCTGT**	
pCAP01-Penta1-R	TCGGAAAGCGGCTGAAGGTCTCTCCCAAGC**TCGAGGTTACTAGTCGATCT**	
pCAP01-Penta2-F	AAGGCTAGTCAGGGGTACCGGGCCCCTCAA**TCGAGACTTGAGGTACCTGT**	
pCAP01-Penta2-R	GGCCAACTGGCCTGCTACCTGCGCCATTTG**TCGAGGTTACTAGTCGATCT**	
Penta1.1-check-F	TCTTAGGACGGATCGCTTGC	Checking of pCPP1 and pCCP2
Penta1.2-check-F	CTGAGCTTCCACGTCCCC	
Penta1-check-R	GGTGTGGCTGAGCAGCTG	
Penta2-check-F	CCGCTGAGTGGTATGAGCTT	
Penta2-check-R	AGCACATCCGTCTGTTCGAG	

Table C5. Primers used for the cloning and checking of the cpp1 *and* cpp2 *biosynthetic gene clusters. The underlined sequences are the targets where Cas9 must cut and the bold sequences belong to the pCAP01 plasmid.*

Final construction	Oligonucleotide	Sequence (5'-3')	Application
pEFBA-caoA	Frag1-F	TAAGCA**ACTAGT**TGCTGCTGGAGATGCTGG	Generation of the mutant Δ*caoA*
	Frag1-R	TAAGCA**ATGCAT**ATGCTGTTCACCCCCTTTG	
	Frag2-F	TAAGCA**GGATCC**ATCGAGATCTGATCCAGCATC	
	Frag2-R	TAAGCA**GATATC**CGGATGGAGTCGCGCAG	
pEFBA-cao4	Frag1-F	TAAGCA**ACTAGT**AACGGGGAGGCAGCACA	Generation of the mutant Δ*cao4*
	Frag1-R	TAAGCA**ATGCAT**TATCTCGGTTGCCATACGTGT	
	Frag2-F	TAAGCA**CATATG**TCGACGGGGTGAAGGAAC	
	Frag2-R	TAAGCA**GATATC**TGTACCTGGGCGACAACTTC	
pEFBA-cao8	Frag1-F	TAAGCA**ACTAGT**CCGGTCAGCTCGCTCAC	Generation of the mutant Δ*cao8*
	Frag1-R	TAAGCA**ATGCAT**TGCCGTCCTCACTCGCC	
	Frag2-F	TAAGCA**CATATG**GGTTCCTGAGCAGCGGAG	
	Frag2-R	TAAGCA**GATATC**CCGGGAACAGCAGGGAC	
pEFBA-cao16	Frag1-F	TAAGCA**ACTAGT**ACGAGGTGCTGGACATGG	Generation of the mutant Δ*cao16*
	Frag1-R	TAAGCA**ATGCAT**GCAGATGATGTCGATGCG	
	Frag2-F	TAAGCA**GGATCC**TAGACCTGAGACGAGGGG	
	Frag2-R	TAAGCA**GATATC**GAAGGGGATCATCAAAGAGC	
pEFBA-cao24	Frag1-F	TAAGCA**ACTAGT**TCGGTGGAGTCGGTCTTC	Generation of the mutant Δ*cao24*
	Frag1-R	TAAGCA**ATGCAT**GTACGCACTGAAGATCAAGA	
	Frag2-F	TAAGCA**CATATG**GGAAGGAGCACGGTCATGAT	
	Frag2-R	TAAGCA**GATATC**ACAGCACAGTACGCAATCCT	
pEFBA-cppA	Frag1-F	TAAGCA**ATGCAT**CGGCGCTGGTGGATGAC	Generation of the mutant Δ*cppA*
	Frag1-R	TAAGCA**ACTAGT**CGGATTCGGACATGCAAAGC	
	Frag2-F	TAAGCA**GGATCC**GGAAGAAGTGGTCGCGGG	
	Frag2-R	TAAGCA**GATATC**ATCTCTACAGCTTCGTCCGC	

Table C6. Primers used for the generation of knockouts through genetic replacements of the different genes caoA*,* cao4*,* cao8*,* cao16*,* cao24 *and* cppA *for an apramycin resistance cassette. The underlined and bold sequences are the restriction sites for the cloning of the PCR products into the plasmids (ACTAGT: SpeI; ATGCAT: NsiI; GGATCC: BamHI; CATATG: NdeI; GATATC: EcoRV).*

Oligonucleotide	Sequence (5'-3')	Application
Apramycin-F	GGTGGAGTGCAATGTCGT	Amplification of the apramycin resistance cassette to check the generation of knockouts
Apramycin-R	TCGCATTCTTCGCATCCC	
Thiostrepton-F	TAAGCA**TCTAGA**GCACGTACTAAGGCGCG	Amplification of the thiostrepton resistance cassette to clone into the pEFBA constructions
Thiostrepton-R	TAAGCA**TCTAGA**CTCACTGCCCGCTTTCC	

Table C7. Primers for the amplification of apramycin and thiostrepton resistance cassettes, for the constructions of the pEFBA plasmids and the checking of the generation of knockouts. The underlined and bold sequences are the restriction sites for the cloning of the PCR products into the plasmids (TCTAGA: XbaI; ATGCAT: NsiI; CATATG: NdeI).

Oligonucleotide	Sequence (5'-3')	Application
Cao4-F	TAAGCA**GAATTC**CGCGCACTGACCGTCTG	Amplification of the *cao4* gene for its complementation in the knockout CA-300655
Cao4-R	TAAGCA**GAATTC**CTGCTGCCGGACTGACC	
Cao8-F	TAAGCA**GGATCC**CCAACATCGACCCGGACC	Amplification of the *cao8* gene for its complementation in the knockout CA-300647
Cao8-R	TAAGCA**GAATTC**GTCACGCCGCCACACTC	
Cao16-F	TAAGCA**GGATCC**GCTCGACGGCTGTTCGG	Amplification of the *cao16* gene for its complementation in the knockout CA-300648
Cao16-R	TAAGCA**GAATTC**ACGTCCATGGCCCACAG	
Cao24-F	TAAGCA**GAATTC**GACATCTCGCCTCATCGCTG	Amplification of the *cao24* gene for its complementation in the knockout CA-300649
Cao24-R	TAAGCA**GAATTC**TGTCGGCCGTGCAGTCG	
CppA-F	TAAGCA**GGATCC**CCAACACACCTCCACGCA	Amplification of the *cppA* gene for its complementation in the knockout CA-301651
CppA-R	TAAGCA**GAATTC**CCCCAGACTTCAGCGTTTGA	

Table C8. Primers used to amplify the genes caoA, cao4, cao8, cao16, cao24, cppA *for their complementation in the corresponding knockout. The underlined and bold sequences are the restriction sites for the cloning of the PCR products into the plasmids (GGATCC: BamHI; GAATTC: EcoRI).*

Oligonucleotide	Sequence (5'-3')	Application
Cao4-check-F	GCGGATCGTCAAGGAGTA	Internal oligonucleotides to check of *cao4* knockout gene in the knockout CA-300655
Cao4-check-R	GTTCCTTCACCCCGTCGA	
Cao8-check-F	CATCTGCCCCACCCACAAC	Internal oligonucleotides to check of *cao8* knockout gene in the knockout CA-300647
Cao8-check-R	GATCTCCGGGTGCCACAC	
Cao16-check-F	CCGGATCTACAGGCGGTCTA	Internal oligonucleotides to check of *cao16* knockout gene in the knockout CA-300648
Cao16-check-R	TGTGAACACATGGGTCCGC	
Cao24-check-F	GACTCAATGTGCACGTCG	Internal oligonucleotides to check of *cao24* knockout gene in the knockout CA-300649
Cao24-check-R	GTGCGGGCAGTCGGTG	
CppA-check-F	CTGCTCAGCCACACCG	Internal oligonucleotides to check of *cppA* knockout gene in the knockout CA-301651
CppA-check-R	CCAACTCCTCGAACAGCT	

Table C9. Primers used to check the absence of cao4, cao8, cao16, cao24 *and* cppA *genes in the respective knockout strains.*

7.1.2. *PCR amplification conditions*

Amplifications were carried out in a thermocycler C1000 Touch thermal Cycler (BIO-RAD, Hercules, Carlifornia, USA) using the Q5 polymerase (New England Biolabs, Ipswich, MA, USA) in a final 50 µL volume reactions.

The general composition of the amplification mix was:

5x Q5 Reaction Buffer	10 µL
5x Q5 High GC Enhancer	10 µL
10 mM dNTPs	2 µL
10 µM Forward Primer	2.5 µL
10 µM Reverse Primer	2.5 µL
Template DNA	< 1000 ng
Q5 High-Fidelity DNA Polymerase	0.5 µL
Nuclease-Free Water	to50 µL

In terms of the conditions applied in each PCR, the variations were defined by the optimum annealing temperature of the two oligonucleotides used in each reaction, the size of the DNA fragment to amplify and the potential formation of secondary structures. In these cases, specific changes were done in the hybridization temperature and the extension time. The standard conditions used in the amplification reactions using the Q5 DNA polymerase are shown in the following table:

Step	Temperature	Time
Initial Denaturation	98 °C	30 seconds
35 Cycles	98 °C	10 seconds
	Tm	30 seconds
	72 °C	30 seconds/Kb
Final Extension	72 °C	10 minutes
Hold	4 °C	∞

Table C10. Standard conditions used in the PCR amplification reactions.

Once the PCR reaction was over, the final result was checked in an agarose gel, to confirm the size and amount of the amplified fragment.

7.1.3. *PCR products purification*

The resulting amplification products were purified using the GFX PCR DNA and Gel Band Purification Kit (Sigma-Aldrich, St Louis, Missouri, USA) following the guidelines from the trading house.

7.2. Agarose gel electrophoresis

DNA electrophoresis in agarose gels were carried out following the recommendations of Sambrook and Russell (2001).

For the preparation of the gels electrophoresis-grade agarose (Invitrogen) was dissolved by heating in 1x TAE buffer (diluted from 50x TAE buffer: 20mM Trizma Base, 10 mM Glacial acetic acid, 0.5 M pH 8 EDTA). Agarose concentrations between 0.7-2 % were used, according to the size of the DNA fragments to separate. 10.000x SYBR Safe DNA gel stain (Invitrogen) was added to the gels to facilitate the visualization of the fragments after the electrophoresis. Once solidified, the gel was submerged in an electrophoresis tray filled with 1x TAE buffer. Before their application on the gel, the DNA samples were mixed with 1/10 volume of 10x loading buffer (50 % glycerol, 49.75 % TE buffer, 0.25 % bromophenol blue) which provides density to the samples and plays as an indicator of the migration front in the gel. The HindIII digest of lambda DNA (λ-HindIII) (ThermoFisher Scientific, Waltham, Massachusetts, USA) and the TrackIt 1 Kb/100 bp DNA (Invitrogen, Waltham, Massachusetts, USA) ladders were used for sizing and approximate quantification of the samples.

The electrophoresis was subjected to a voltage of 70-100 V and the time of running was different depending on the size of the gels and the DNA fragments. Once the electrophoresis was over, the bands were visualized in a UV-transilluminator CN-08 Infinity.

7.3. Purification of DNA fragments from agarose gels

After the electrophoretic separation of the DNA fragments, the agarose piece corresponding to the DNA fragment of interest was cut with a scalpel. This piece was plunged in a solution from the GFX PCR DNA and Gel Band Purification Kit (Sigma-Aldrich, St Louis, Missouri, USA) that dissolves the agarose, denaturalizes proteins and promotes the join of the DNA to the column's matrix. Once the DNA was retained, the purification was carried out as described in the kit's protocol.

7.4. DNA concentration and purity determination

The quantification of the DNA fragments was estimated in agarose gels by comparison with a DNA ladder with an already known concentration, the λ-HindIII and the TrackIt 1 Kb/100 bp DNA ladders.

A NanoDrop ONE C spectrophotometer (Thermofisher Scientific, Waltham, Massachusetts, USA) was employed to determine the DNA concentration and purity. This equipment calculates the concentration of the sample in ng/µL according to its absorbance at a wavelength of 260 nm (A_{260}).

For the purity determination, two parameters are used: the ratios A_{260}/A_{280} and A_{260}/A_{230}. The two coefficients should have values between 1.8-2.0 (A_{260}/A_{280}) and between 2.0-2.2 (A_{260}/A_{230}) to indicate that the DNA sample is pure. The presence of proteins may decrease the coefficient A_{260}/A_{280} (< 1.8) as many amino acids absorb at the wavelength of 280 nm (Sambrook and Russel, 2001). An A_{260}/A_{280} > 2 indicates a possible contamination with RNA. On the other hand, a low A_{260}/A_{230} ratio may be the result of carbohydrate carryover, residual phenol from nucleic acid extraction, residual guanidine or glycogen used for precipitation. A high A_{260}/A_{230} ratio may be the result of using an inappropriate solution for the blank measurement.

7.5. DNA digestion with restriction endonucleases

DNA digestions with restriction endonucleases were performed with the buffer and optimal incubation conditions recommended by the trading house (New England BioLabs, Ipswich, MA, USA). High fidelity restriction enzymes were employed, which have the same specificity as the native enzymes, but have been engineered to reduce star activity and performance all in a single 10x CutSmart Buffer. For the digestion mix, 1 µg DNA, 1x CutSmart Buffer and 1 µL restriction enzyme were mixed and incubated at 37 °C for 1 hour.

7.6. Techniques for the DNA ligation

7.6.1. *Vector dephosphorylation*

The aim of the dephosphorylation is to remove the 5'-phosphate ends in the digested vectors to prevent their recircularization and increase the efficiency of the cloning. This process was carried out by adding 1 µL of the calf intestinal alkaline phosphatase (CIAP, New England BioLabs, Ipswich, MA, USA) to the digested vector and incubating it at 37 °C for 30 minutes.

7.6.2. *DNA Ligation*

The ligation of DNA fragments obtained by restriction enzyme digestion into the compatible cloning sites of the vector, was performed using the Blunt/TA Ligase Master Mix protocol from New England BioLabs (Ipswich, MA, USA). The linearized and dephosphorylated vector (20-100 ng) was mixed with a 3-fold molar excess of the digested and purified DNA

fragment and the Blunt/TA Master Mix. This mix was incubated at 25 °C for 15 minutes. The product of this reaction was used for the transformation of competent *E. coli* cells.

Blunt-ended PCR products were cloned into pCR®-Blunt plasmid using a 10:1 molar ratio of insert:vector and the ExpressLink™ T4 DNA Ligase (5 U/ µL), according to manufacturer´s protocol (Thermo-Fisher). The ligation product was used to transform *E. coli* NEB 10-β cells, which were spreaded in LB plates supplemented with kanamycin, IPTG and X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside). If the DNA insert was properly ligated, the *lacZ* gene would be disrupted, impairing the hydrolyzation of X-Gal to galactose and 5-bromo-4-chloro-3-hydroxyindole (which dimerizes and is oxidized to the insoluble dark-blue compound 5,5'-dibromo-4,4'-dichloro-indigo) generating white colonies. If the DNA fragment was not appropriately cloned, LacZ would hydrolyze X-Gal and would generate *E. coli* blue colonies. Only white cells were picked, and their plasmids were sequenced with the universal oligonucleotides M13 forward (-20) and M13 reverse (-24).

7.7. Cas9-Assisted Targeting of Chromosome segments (CATCH) cloning of BGCs

The cloning of biosynthetic gene clusters was performed following the CATCH technique (Jiang *et al.,* 2015), where the target genome sequences were excised from bacterial chromosomes *in vitro* by the RNA-guided Cas9 nuclease at two previously designated loci and ligated to the cloning vector pCAP01 by Gibson assembly. The CATCH cloning was performed as Jiang and Zhu described (2016).

The first step in CATCH cloning was to design effective single guide RNA (sgRNA) pairs to cleave the gene clusters from the genome of *Streptomyces cacaoi*. The CRISPy-web online tool (http://crispy.secondarymetabolites.org/) was employed to design these 20 nt target RNA sequences which must be upstream a PAM (Protospacer-Adjacent Motif) sequence 'NGG' (Tong *et al.,* 2018) that is the target where the Cas9 endonuclease cuts. The templates for the *in vitro* transcription of sgRNAs were obtained by the overlapping PCR of three oligonucleotides, including a target-specific oligo containing a T7 promoter sequence (TAATACGACTCACTATA) and a target sequence with a G as the starting nucleotide that is required for T7 RNA polymerase *in vitro* transcription and two universal oligos (sgRNA-F and sgRNA-R, Tables C4 and C5) of an sgRNA constant sequence contained in the crRNA-tracrRNA chimera. The overlapping PCR was performed using a melting temperature of 55 °C and an extension time of 15 s. The product was purified by phenol/chloroform/isoamyl alcohol (25:24:1, pH>7.8) extraction and isopropanol precipitation, with a consequent resuspension in RNase-free water. This overlapping PCR product was used as template for the *in vitro* transcription using the HiScribe T7 Quick High Yield RNA synthesis kit. The resulting sgRNA was purified by phenol/chloroform isoamyl alcohol (25:24:1, pH>5.0) extraction and isopropanol precipitation, which ensured the recovery yield of the RNA molecules and effectively avoided RNase contamination. The sgRNA was resuspended in RNase-free water and its concentration and purity was checked spectrophotometrically (optimal parameters were around 300-1000 ng/µL and A_{260}/A_{280} >2.0). The sample was stored at -80 °C for up to 3 months.

CATCH cloning involved in-gel digestion of genomic DNA, which protects the genomes from extensive shearing and allows the isolation and cloning of long and intact DNA sequences. To prepare the agarose plugs for the subsequent in-gel Cas9 digestion, the strain *Streptomyces cacaoi* was cultured in ATCC-2 for 2 days on an orbital shaker at 28 °C, 220 rpm and 70 % relative humidity and the bacterial cells were embedded in low-melting-temperature 2 % CleanCut agarose plugs (CHEF Bacterial Genomic DNA Plug Kit, BioRad). The plugs were treated with lysozyme and proteinase K and successively washed four times with 1x wash buffer (CHEF Bacterial Genomic DNA Plug Kit, BioRad). In the third wash, PMSF was added to inactivate the remaining proteinase K and 0.1x wash buffer was used during the last wash for diluting the EDTA (Keijser *et al.,* 2000).

For the in-gel RNA-guided Cas9 digestion of the genomic DNA, the agarose plugs were first equilibrated in RNase-free cleavage buffer (100 mM HEPES pH 7.5, 750 mM KCl, 0.5 mM EDTA pH 8, 50 mM MgCl2, DEPC-treated water) and then transferred into a new tube containing cleavage buffer and the Cas9 endonuclease from *S. pyogenes* (New England BioLabs, Ipswich, MA, USA) preassembled with the sgRNA pairs and incubated at 37 °C for 2 h. After the digestion, the plugs were washed, melted and subjected to digestion by GELase agarase (Epicentre Technologies, Madison, WI), followed by ethanol precipitation of the digested DNA and gentle resuspension in DNase-free water.

The Cas9-digested DNA was cloned into the previously prepared pCAP01 vector (Figure C1a) by Gibson assembly. XhoI-digested pCAP01 was amplified with primers containing a 20 nt sequence that annealed to the vector template ends and a 30 nt overhang overlapping with the terminal sequences of the Cas9-digested target cluster. The PCR reaction was performed using a melting temperature of 65 °C and 5 min of extension time. Because contamination by the template plasmids could affect the cloning positive rates, the PCR products were treated with DpnI. For the Gibson Assembly, a 2x Gibson Assembly Master Mix (New England BioLabs, Ipswich, MA, USA) was used, containing an exonuclease that creates single-stranded 3’ overhangs that facilitates the annealing of the fragments, a polymerase that filled in gaps within each annealed fragment and a DNA ligase that sealed nicks in the assembled DNA. 30-50 ng/µL of pCAP01 vector were mixed with the Cas9 digested genomic DNA and the 2x Gibson assembly mix, incubating at 50 °C for 1 h. After the ligation, the Gibson products were used for the electrotransformation of NEB 10-β electrocompetent *E. coli* cells.

The colonies grown on kanamycin selective LB agar plates were validated by PCR and double digestions with restriction enzymes. Different pairs of oligonucleotides were designed, one PCR primer on the pCAP01 vector and the other on the insert at both ends. Around 20 colonies were picked for each cloning and the plasmids were extracted as described in section 6.2. The plasmids were examined by agarose electrophoresis and those with the correct length were validated by PCR and restriction endonucleases digestion.

7.8. Generation of knockouts by genetic replacement

The genetic replacement allows the specific inactivation of one or several genes of the genome of a strain. This technique relies on the substitution of genes for another DNA fragment that confers the strain a different phenotype, as resistance to a determined antibiotic.

The first step was the PCR amplification of two DNA fragments around 2 Kb contiguous to the gene to be replaced (flanked by proper restriction sites, oligo 1-SpeI, oligo 2-NsiI, oligo 3-BamHI/NdeI and oligo 4-EcoRV). The PCR products were cloned into pCRBlunt and sequenced by Secugen S. L. (Madrid) to check that the DNA fragment sequences were correct.

Once checked, the fragments were double-digested with the appropriate pair of restriction enzymes (SpeI-NsiI or BamHI/NdeI-EcoRV) and they were cloned sequentially in the conjugative plasmid pEFBAoriT (Horna *et al.,* 2011) (Figure C1b), a suicide, non-replicative vector in *Streptomyces*, at both sides of the apramycin resistance gene *acc(3)IV* (Am^R), checking the cloning of each fragment by digestions with the appropriate pair of restriction enzymes.

The pEFBAoriT includes, besides the apramycin resistance gene, an ampicillin resistance gene, but it only selects in *E. coli*, so an additional resistance gene must be cloned in the vector to verify that the double crossover takes place successfully in *Streptomyces*. To that end, once the two fragments were cloned, the vector was digested with XbaI and a thiostrepton resistance gene (Tsr^R), obtained from pEM4T by PCR amplification with XbaI restriction sites at both ends, was inserted into the plasmid. The general sequence of the generation of the construction is shown in the Figure C2. After this process, the new construction was ready to be transferred by conjugation to the heterologous hosts.

The resulting knockout had the apramycin resistance gene in the same transcriptional direction of the replaced gene. This apramycin resistance gene did not have a transcriptional terminator to avoid polar effects on the expression of downstream genes.

The selection of positive transconjugants relied in their resistance to apramycin and sensitivity to thiostrepton (Am^R / Tsr^S), which corroborate that the double crossover happened, and the gene of interest was successfully substituted by the apramycin resistance gene. If the knockout strain had shown resistance to both antibiotics, just a single crossover would have occurred, which implied that the whole plasmid would have been incorporated into the genome. Moreover, the double crossovers were checked by PCR amplification using internal primers of the deleted gene, internal primers of the apramycin resistance gene and combinations with external primers annealing with the fragments flanking the target gene.

The positive transconjugants were fermented in 10 mL R2YE vials and incubated at 28 °C, 220 rpm, 70 % humidity for 13 days. After the extraction of the secondary metabolites produced, the changes in the structure and production of the compounds were studied by LC-HRESIMS(+)-TOF.

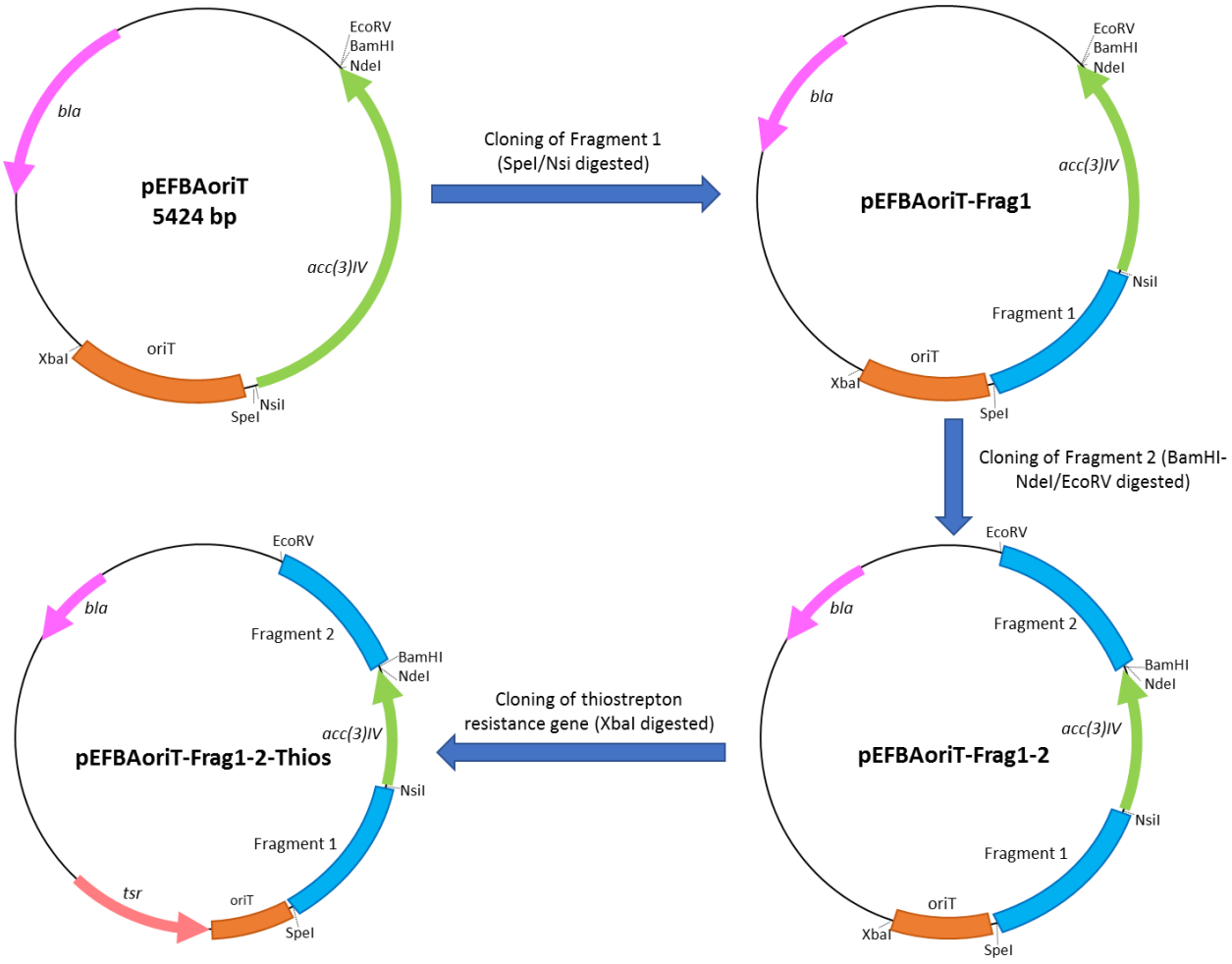


Figure C2. Generation of pEFBAoriT-derived gene replacement constructions.

7.9. Knockout complementation

A genetic replacement leads to the loss of function of one or several genes. The complementation of the replaced gene is needed in the knockout strains to restore and verify the function of the gene. The multicopy pEM4T vector was used for the overexpression of the previously replaced genes under the control of the strong constitutive promoter *ermE**p. pEM4T vector (Figure C1c) is a conjugative plasmid with the origin of transfer *oriTRP4* that conferred resistance to ampicillin in *E. coli* and to thiostrepton in *Streptomyces*.

The genes to overexpress were amplified by PCR using oligonucleotides with tails containing BamHI (forward primer) and EcoRI (reverse primer), or both EcoRI (forward and reverse primers) restriction sites (Table C8), in order to insert the gene in the correct direction downstream the promoter. After the amplification, the DNA fragments were cloned into the pCR-Blunt vector and sequenced. Once the sequence was confirmed, the vector was digested with BamHI and EcoRI, or just EcoRI and the gene was cloned into the previously digested pEM4T. The final construction was transferred to NEB 10-β *E. coli* cells and after the plasmid isolation, it was checked by digestion to ensure that the direction of the gene was appropriate.

Then, an electroporation of ET12567/pUB307 *E. coli* cells was performed followed by the subsequent conjugation with the spores of the knockout strain. The transconjugants were streaked into MA plates supplemented with thiostrepton and nalidixic acid and the complemented strains were fermented in 10 mL R2YE vials and incubated at 28 °C, 220 rpm, 70 % humidity for 13 days. After the extraction of the secondary metabolites, the restoration of the function of the knocked out gene was studied by LC-HRESIMS(+)-TOF.

8. DNA SEQUENCING

8.1. Genomic DNA sequencing

The extracted genomic DNA was fully sequenced *de novo*, assembled and annotated by Macrogen (Seoul, Korea; http://www.macrogen.com/), using a combined strategy of Illumina HiSeq 2500 and PacBio RSII platforms. The PacBio long-reads were assembled with Canu (v1.7) (Koren *et al.,* 2017). After assembly, Illumina reads were applied for accurate genome sequence using Pilon (v1.21) (Walker *et al.,* 2014). To validate the accuracy of the assembly, Illumina reads were mapped to the assembly result. After mapping, the consensus sequence was generated.

8.2. Sanger sequencing of plasmid constructions and PCR products

The sequencing reactions were performed in a BigDye® Terminator v3.1 sequencer by Secugen S.L. (Madrid) using a system based on the Sanger technique (Sanger, 1977) The reaction required at least 1 µg purified plasmid DNA along with 1.5 µL of 5 µM primer in a total of 15 µL reaction. PCR products were usually cloned into pCRBlunt to assure better reads of sequencing. This service normally provides reads up to 900 pb.

The analysis of the results obtained by sequencing were carried out using the bioinformatic software Geneious® 9.1.8 (www.geneious.com).

9. BIOINFORMATIC ANALYSIS

The whole genome sequence of *Streptomyces cacaoi* CA-170360 was analyzed with antiSMASH v6.0.1 (Blin *et al.,* 2021), BAGEL4 (van Heel *et al.,* 2018), PRISM (Skinnider *et al.,* 2020) and RiPPMiner (Agrawal *et al.,* 2017) to determine the number of predicted specialized metabolite gene clusters. All these web tools depend on the information from literature and databases such as MiBiG (Kautsar *et al.,* 2020), Bactibase (Hammami *et al.,* 2010) or CAMPR3 (Waghu *et al.,* 2016). AntiSMASH is the most widely employed tool for the search of microbial secondary metabolites biosynthetic pathways in the genomes of bacteria, fungi and plants, using a rule-based approach to identify a wide variety of BGCs such as non-ribosomal peptide synthetases (NRPS), type I and II polyketide synthases (PKS), lanthipeptides, thiopeptides, lasso peptides or sactipeptides, among others, and provides detailed predictions on the putative compound produced (Blin *et al.,* 2021a). PRISM (PRediction Informatics for Secondary Metabolomes) is another web Java application useful for the chemical structure prediction of

genetically encoded type I and II PKs and NRPs. BAGEL4 and RiPPMiner are other web servers that enables the specific identification of biosynthetic gene clusters in prokaryotic DNA involved in the production of Ribosomally synthesized and Post-translationally modified Peptides (RiPPs) and bacterocins (van Heel *et al.,* 2018; Agrawal *et al.,* 2017).

BLAST (Basic Local Alignment Search Tool) (Johnson *et al.,* 2008) and HHpred based on profile hidden Markov model (HMM) comparisons (Zimmermann *et al.,* 2018) were also employed to predict the function of each gene involved in the biosynthesis of the compounds.

A phylogenetic tree based on nearly complete 16S rRNA sequences was generated with the MEGA X software (http://www.megasoftware.net) (Kumar et al., 2018), using the Neighbor-Joining method corrected with the Tamura 3-parameter algorithm (Tamura, 1992) and a bootstrap of 1,000 replicates. *Micromonospora auratinigra* DSM 44815^T 16S rRNA gene was used as outgroup. Genomic distances were calculated with the Kimura-2 parameter model also included in the MEGA X software.

10. GENETIC TRANSFORMATION

Genetic transformation was used for the incorporation of plasmid constructions into *E. coli* cells through heat shock (chemical transformation) or electrical shock (electroporation).

10.1. Transformation

For the chemical transformation, commercial chemically competent NEB 10-β *E. coli* cells (New England BioLabs, Ipswich, MA, USA) were used. In this method, the plasmid DNA entry was favored by the modification of the cellular membrane caused by a heat shock at 42 °C. For that, 50 µL *E. coli* cells were mixed with 1-5 µL containing 1 pg-100 ng of plasmid DNA. After the heat shock, 950 µL fresh LB was added and the mix as incubated at 37 °C for an hour at 250 rpm, and different dilutions were spread onto selection plates with the proper antibiotics. These plates were incubated overnight at 37 °C.

10.2. Electroporation

This type of transformation was performed in a MicroPulser electroporator (BioRad), using a cuvette with 2 mm separation between the electrodes.

Depending on the purpose of the work, different electrocompetent cells were used. For the obtention of high amounts of a large plasmid construction the commercial NEB 10-β electrocompetent *E. coli* (New England BioLabs, Ipswich, MA, USA) was used. ET12567 and ET12567/pUB307 *E. coli* cells were used to electroporate plasmid DNA for a conjugation process.

50 µL electrocompetent cells were mixed with 1 µL DNA solution in an electroporation cuvette and a pulse of 2.5 kV was applied for 5 ms to get the plasmid inserted in the cells. Then, the cells were diluted as quickly as possible in 950 µL fresh LB and were incubated at 37 °C, 250

rpm for an hour to allow the development of resistance. Finally, different dilutions were spreaded onto selective plates containing the plasmid-selective antibiotic. In the case of ET12567 and ET12567/pUB307 *E. coli* cells, which have intrinsic antibiotic resistance to chloramphenicol and to chloramphenicol and kanamycin, respectively, these antibiotics were also added. These plates were incubated overnight at 37 °C.

10.3 Preparation of electrocompetent *E. coli* cells

Electrocompetent *E. coli* ET12567 and ET12567/pUB307 cells were obtained from exponentially growing cultures, inducing the electrocompetence with 10 % glycerol at 4 °C (MacNeil *et al.,* 1992).

First, a frozen glycerol stock of bacterial cells was streaked out onto a fresh LB plate, which was incubated overnight at 37 °C. The second day, the starter culture of cells was prepared from a single colony of *E. coli*, inoculating 20 mL of fresh LB. This starter culture was grown at 37 °C, 250 rpm overnight.

Then, 1 L LB medium was inoculated with 10 mL from starter culture and was let grown at 37 °C, 250 rpm until the OD_{600} reached 0.35-0.4. The culture was immediately put on ice and the rest of the procedure was carried out in the cold room and using ice. The 1 L culture was split into four parts by pouring about 250 mL into ice-cold centrifuge bottles. The cells were harvested by centrifugation at 1000 g for 20 min at 4 °C. The supernatant was discarded, and each pellet was resuspended in 200 mL of ice-cold ddH_20 (shaking as gentle as possible). The cells were harvested again by centrifugation at 1000 g for 20 min at 4 °C and resuspended in 50 mL of ice-cold 10 % glycerol. The resuspensions were combined into 2 centrifuge bottles and harvested the cells. Each pellet was resuspended in 40 mL of ice-cold 10 % glycerol and transferred to a 50 mL Falcon tube. After the last centrifugation, the cells were resuspended in 1 mL of ice-cold 10 % glycerol. The resuspension was divided into 40 μL aliquots in sterile 1.5 mL microfuge tubes and snapped freeze with liquid nitrogen. The frozen cells were stored at -80 °C.

11. INTERGENERIC CONJUGATION OF *STREPTOMYCES*

The transference of genetic material from *E. coli* hosts into *Streptomyces* strains was carried out by intergeneric conjugation following the protocol described by Mazodier *et al* (1989), which requires bifunctional *E. coli-Streptomyces* vectors. The heterologous hosts *Streptomyces albidoflavus* J1074 and *Streptomyces coelicolor* M1152 and M1154 as receptor strains while *E. coli* ET12567 and ET12567/pUB307 (Flett *et al.,* 1997) cells were used as the donor strains in triparental and biparental conjugations, respectively.

11.1. Biparental conjugation

In biparental conjugations, ET12567/pUB307 *E. coli* cells were electroporated with the target plasmid as described in section 10.2 and selective agar plates, supplemented with kanamycin, chloramphenicol and the plasmid antibiotic resistance, were incubated overnight at 37 °C. Then, one colony was picked into selective LB liquid tubes and grew overnight at 37 °C and 250 rpm. The following day the pellet obtained from 1.5 mL LB culture was washed three times with 1 mL fresh LB and then resuspended in 100 µL fresh LB. 50 µL host strain spores were preactivated at 50 °C for 10 minutes. Finally, the ET12567/pUB307/plasmid culture and the preactivated spores were mixed and spread onto two MA plates (50 and 100 µL onto each plate) that were incubated overnight at 28 °C. On the next day, the MA plates were overlayed with 1.5 mL sterilized Milli-Q water containing nalidixic acid and the appropriate selective antibiotic. The plates were incubated at 28 °C for 3-5 days to let the transconjugants grow.

11.2. Triparental conjugation

Triparental conjugation was useful for conjugating large plasmids and those containing a kanamycin resistance gene as a selectable marker into *Streptomyces.* Since pUB307 contains a kanamycin resistance gene, it is impossible to select for a vector with a kanamycin resistance gene if *E. coli* ET12567/pUB307 is used as the helper strain. Depending on the heterologous host, the approximation was different.

When the heterologous host was *Streptomyces albidoflavus* J1074, as this strain is defective in the SalI (SalGI) restriction-modification system for methylated DNA (Chater and Wilde, 1976), NEB 10-β *E. coli* cells (previously electroporated with a kanamycin resistance-containing vector) could be used as donor strain. An overnight liquid culture of these NEB 10-β *E. coli* cells and an overnight culture of ET12567/pUB307 *E. coli* cells (kanamycin and chloramphenicol-resistant) were grown. The pellets obtained from 1.5 mL of the two cultures were washed with 1 mL of fresh LB (x3) and then resuspended in 100 µL fresh LB, which were mixed with 50 µL preactivated spores of *S. albidoflavus*. Three MA plates were seeded (50, 100 and 100 µL onto each plate) and on the following day they were overlayed with 1.5 mL sterilized Milli-Q water containing nalidixic acid and kanamycin and incubated at 28 °C for 3-5 days to let the transconjugants grow.

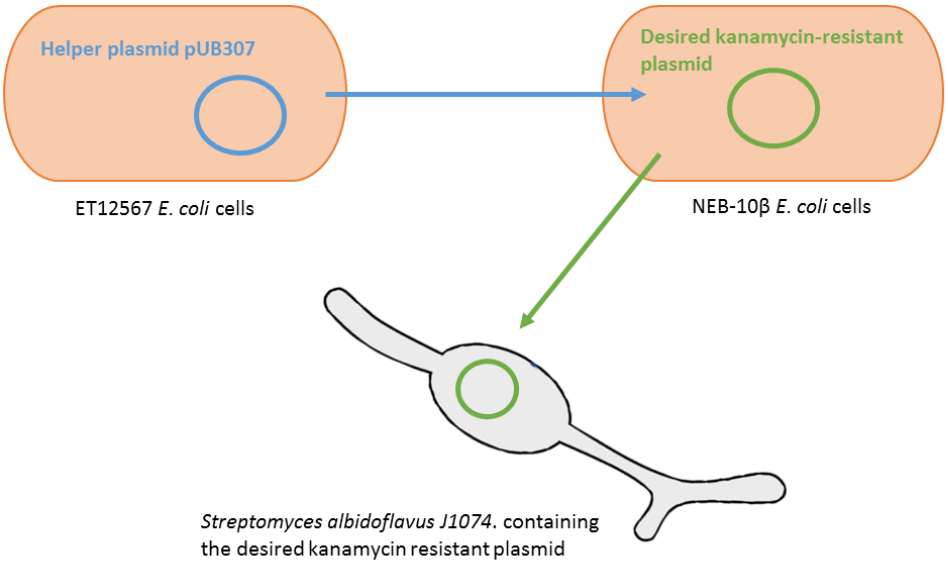


Figure C3. Scheme of triparental conjugation in Streptomyces albidoflavus*. The pUB307 helper plasmid of ET12567 cells is transferred into E. coli NEB 10-β containing the desired kanamycin-resistant plasmid. The pUB307 plasmid mobilizes the desired plasmid into* S. albidoflavus.

However, *Streptomyces coelicolor* strains M1152 and M1154 have restriction systems for methylated DNA (González-Cerón et al., 2009), so NEB 10-β E. coli cells could not be used as direct donor strain. In this case, a triparental *E. coli* conjugation was first performed to obtain a *E. coli* ET12567 donor strain. Therefore, an overnight liquid culture of NEB 10-β *E. coli* cells (previously electroporated with the kanamycin resistance-containing vector), an overnight culture of TOP10 *E. coli* cells containing the self-transmissible helper plasmid pR9406 (ampicillin-resistant) and an overnight culture of ET12567 *E. coli* cells (chloramphenicol-resistant), were grown. The pellets obtained from 1.5 mL of the three cultures were washed with 1 mL of fresh LB (x3) and then resuspended in 100 µL fresh LB. 15 µL of each culture were spotted at the same location on an LB agar plate without antibiotics so that the three strains were mixed. This plate was incubated overnight at 37 °C. The next day, the mix of strains were streaked onto LB plates containing kanamycin, ampicillin and chloramphenicol to select those ET12567 *E. coli* derivatives containing both the helper plasmid pR9406 and the desired kanamycin-resistant vector (Figure C3). Single colonies of these *E. coli* ET12567/pR9406/plasmid cells were treated as described in section 11.1 and mixed with 50 µL preactivated spores of the host strain. Two MA plates were seeded (50 and 100 µL onto each plate) and on the following day they were overlayed with 1.5 mL sterilized Milli-Q water containing nalidixic acid and kanamycin and incubated at 28 °C for 3-5 days to let the transconjugants grow.

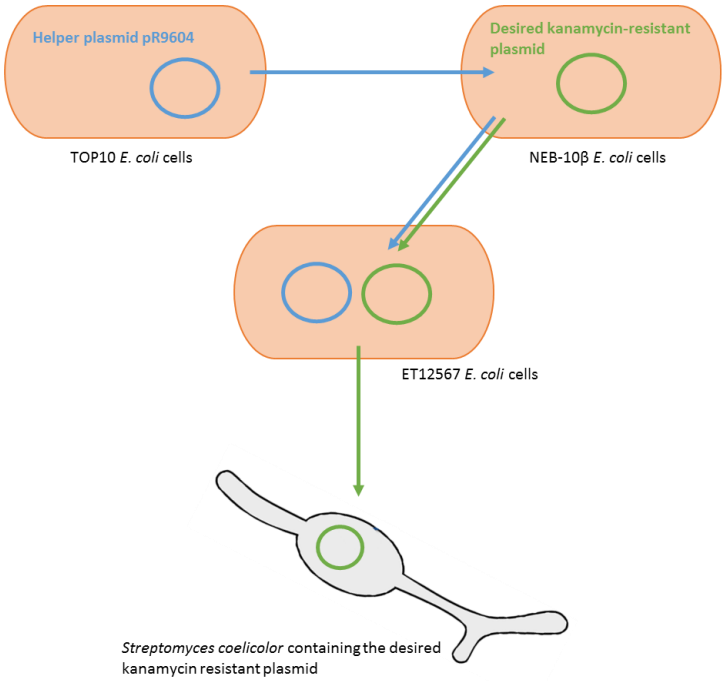


Figure C4. Scheme of triparental conjugation in Streptomyces *sp. the pR9406 helper plasmid is transferred into* E. coli *NEB 10-β containing the desired kanamycin-resistant plasmid. The pR9406 plasmid mobilizes the desired plasmid into* E. coli *ET12567 resulting in a chloramphenicol, ampicillin and kanamycin resistant strain. Standard* E. coli/Streptomyces *conjugation is carried out to transfer the desired plasmid into* Streptomyces coelicolor.

12. SECONDARY METABOLITES ANALYSIS

12.1. Cultivation and preparation of extracts

Depending on the purpose of the fermentation, the extraction of the secondary metabolites was carried out differently.

12.1.1. *Preparation of extracts from small-scale fermentations*

The small-scale fermentations were performed to confirm the presence or absence of target compounds. They were carried out in 10 mL of medium in 40 mL fermentation vials.

The extraction of the fermentation broths was carried out with acetone (1:1, v/v) and stirring at 220 rpm for 2 hours. The solvent with the secondary metabolites extracted was separated from the biomass by centrifugation in a SpeedVac Plus centrifuge at 2,300 rpm for 15 minutes. Then, the acetone was removed by a stream of nitrogen overnight in a GeneVac 4T-24 and the next day the extracts were resuspended to a final ratio of 20 % DMSO/water. The resulting microbial extracts were filtered, and the presence of the compounds was analyzed by LC-HRMS and MS/MS.

12.1.2. *Preparation of extracts from large-scale fermentations*

The large-scale fermentations were performed to purify a high amount of compound in order to elucidate its structure or evaluate its potential microbial activity. These fermentations were carried out in 500 mL flasks containing 125 mL of fermentation medium.

In this case, the extraction of secondary metabolites in the large-scale was different in order to improve the extraction of cacaoidin and its derivates (Ortiz-López *et al.,* 2020). The first steps of the large-scale extraction were the same as the small-scale fermentation: extraction with acetone (1:1, v/v), separation of the organic solvent from the biomass and removal of the acetone. Once the organic solvent is removed under a N_2 stream, the pH of the resulting aqueous phase was adjusted with NaOH to 10.0, and three liquid-liquid extractions with MEK (2:1, v/v) were performed. The organic solvent, containing the extracted compounds, was rotary evaporated to dryness. This residue was used to purify the compound by reverse phase HPLC (High-Performance Liquid Chromatography).

12.2. Dereplication based on mass spectrometry and molecular formula by LC-HRMS and MS/MS

For the dereplication of compounds, the extracts were analyzed by LC-HRMS using the AMDIS (Automated Mass Spectral Deconvolution and Identification System) tool developed by the National Institute of Standards and Technology (NIST) for the extraction of pure component MS spectra from complex chromatograms (Stein, 1999). Peaks from UV/Vis LC trace at 210 nm were detected by integration in the acquisition instrument software Agilent ChemStation. The HRMS spectra was acquired by a Bruker maXis QTOF mass spectrometer, where MS/MS could also be obtained (Martín *et al.,* 2014). The acquisition was performed in the ESI+ mode, switching polarity for cases where no ionization was achieved. Extraction of pure component HRMS spectra from the raw data was carried out by the instrument software Bruker DataAnalysis and the accurate mass for the extracted components was interpreted internally.

The UV/Vis spectra, the retention time, and positive and negative mass spectra of the samples were compared to the corresponding LC-UV-MS data of known microbial metabolites stored in the proprietary database. This is the Fundación MEDINA's reference library which contains annotated secondary metabolite data obtained under identical conditions to those for the samples under analysis; this library includes 405 fungal metabolites and 478 metabolites from bacteria and actinomycetes. Such library is dynamic and is continuously populated with NPs identified in Fundación MEDINA from different drug discovery programs. This library contains identification information and characteristic analytical data such as the retention time, accurate mass and UV profile of all pure components extracted from all samples and components analyzed. When the structure of a pure component is identified, the database records are marked as "fingerprint", with its trivial name and molecular formula annotated and automatically propagated to all the samples containing that specific component. The extracted component data (retention time, UV and HRMS spectra) of every sample analyzed by LC-HRMS are searched against the library.

For those components not already flagged with their name in the MEDINA-HRMS library, the interpreted accurate mass was searched for in the Chapman and Hall/CRC Press Dictionary of Natural Products (DNP), which is considered the gold standard of natural products databases containing the calculated accurate mass based on the molecular formula of the compounds (Buckingham, 2015). The database contains also information on biological source and bioactivity. It can be searched by trivial name, accurate mass, molecular weight, molecular formula and UV absorption maxima. This DNP calculated accurate mass retrieves the first hit list of candidates and, at the same time, the likely molecular formula of the compound. If the molecules in the hit list contain characteristic λ_{max}, the experimental UV spectrum is checked for its compatibility with the reported absorption maxima. Additionally, the taxonomic information listed in the DNP is also considered for the hit list refinement.

Moreover, a complementary and more robust strategy based on the acquisition of HRMS/MS spectra for the target unknown component was followed. This tandem MS spectra were obtained with the same LC-HRMS system, just selecting the target parent ion to be fragmented by collision-induced dissociation (CID). The experimental MS/MS spectrum obtained was first used for the molecular formula confirmation using the SmartFormula 3D application (Pérez-Victoria *et al.,* 2016). The obtained peaks (fragments) list was then used as query input for searching ChemSpider and PubChem using the MetFrag Web tool (Wolf *et al.,* 2010) based on *in silico* fragmentation for the computer-assisted identification of metabolites (FingerID).

12.3. Isolation and purification of cacaoidin and its variants

The protocol described by Ortiz-López *et al.* (2020) was followed to isolate and purify the compounds. After the MEK extraction and the evaporation to dryness, the resulted crude extract was divided into successive portions of 150 mg dissolved in dimethyl sulfoxide (DMSO) and directly pre-purified by preparative RP-HPLC using an Agilent Zorbax SB-C_8, (21.2 x 250 mm, 7 µm) and applying a linear H_2O-CH_3CN gradient (both solvents containing 0.1 % trifluoroacetic acid, TFA) with an initial 30 % and a final 45 % CH_3CN for 35 min, followed by 100 % CH_3CN for 10 min collecting 80 fractions in vials of 20 mL. The UV detection was achieved at 210 and 280 nm. The fractions corresponding to the peaks observed in the chromatograms were submitted to LC-HRESIMS(+)-TOF and those fractions from all the portions of the extract where the compounds were detected were pooled; then the organic solvent (CH_3CN) was carefully evaporated under a N_2 stream and the resulting aqueous sample was freeze-dried. This final step was very important as if the TFA was highly concentrated, it could affect the structure of the compounds.

Two additional rounds of purification were performed to complete the isolation of cacaoidin. First, the powder obtained from the first preparative RP-HPLC was dissolved in methanol and submitted to a semipreparative RP-HPLC (Agilent Zorbax SB-C8, 21.2 x 250 mm, 7 µm), applying a linear H_2O-CH_3CN gradient (both solvents containing 0.1 % TFA) with an initial 34 % and a final 38 % CH_3CN for 35 min, followed by 100 % CH_3CN for 10 min, collecting 80 fractions in a 96-deepwell plate. The fractions in which the compounds were detected by LC-HRESIMS(+)-TOF were pooled and carefully freeze-dried as described before. Finally, a semipreparative RP-HPLC was performed using a Phenyl (10 x 150 mm, 5 µm) column, to detach aromatic compounds

from cacaoidin, with a linear H_2O-CH_3CN gradient (both solvents containing 0.1 % trifluoroacetic acid, TFA) with an initial 25 % and a final 45 % CH_3CN for 35 min, and ending with 10 min at 100 % CH_3CN, collecting 80 fractions in a 96-deepwell-plate. The fractions corresponding to the resulted peak were collected, pooled and then freeze-dried. A sample dissolved in methanol was analyzed by LC-HRESIMS(+)-TOF.

13. DETERMINATION OF ANTIMICROBIAL ACTIVITIES

The purified compounds were assayed against the Gram-positive strain Methicillin Resistant *Staphylococcus aureus* (MRSA) MB5393. First, the compounds were two-fold serially diluted in DMSO, providing 10 concentrations starting at 32 µg/mL for the antimicrobial assay. From these 10 concentrations, that which inhibited 90 % of the growth of the microorganism after overnight incubation was determined as the MIC.

Antibacterial assay against MRSA was carried out as described in Martín *et al.,* (2013) and Carretero-Molina *et al.,* (2020). The first day, a MRSA thawed stock inoculum suspension from a cryovial was streaked onto a Brain Heart Infusion (BHI) agar plate and incubated overnight at 37 °C. Isolated colonies were inoculated into 25 mL of BHI broth medium in 250 mL Erlenmeyer flasks and incubated overnight at 37 °C with shaking at 220 rpm. For the assay, this overnight culture was diluted in order to obtain an assay inoculum of 1.1 x 10^6 CFU/mL. 90 µL/well of the diluted inoculum were mixed with 1.6 µL/well of compound dissolved in DMSO and 8.4 µL/well of BHI medium. Vancomycin and ramoplanin were used as positive controls. Absorbance at OD_{612} was measured with an EnVision Microplate Reader (PerkinElmer) at *zero time* (T_0) and then the plates were immediately incubated at 37 °C for 20 h. After this period, the assay plates were shaken using the DPC Micromix-5 and once more the absorbance at OD_{612} was measured at *final time* (T_f). The processing and analyzing of the data were performed by a Genedata Screener software (Genedata, Inc., Basel, Switzerland). This software calculated the percentage of growth inhibition using the following normalization:

$$\% \text{ Inhibition} = 100 \times \{1 - [(T_f \text{ Sample} - T_0 \text{ Sample}) - (T_f \text{ Blank} - T_0 \text{ Blank})]/[(\text{Tf Growth} - T_0 \text{ Growth}) - (T_f \text{ Blank} - T_0 \text{ Blank})]\}$$

The software also calculated an RZ' factor, which predicted the robustness of the assay (Zhang *et al.,* 1999). An RZ' factor around 0.9 meant a high quality of the assay.

D. Results

CHAPTER I. CHARACTERIZATION OF THE STRAIN *STREPTOMYCES CACAOI* CA-170360

1.1. Whole Genome Sequencing of *Streptomyces cacaoi* CA-170360

Streptomyces cacaoi CA-170360 was cultured on ATCC-2 liquid medium for the subsequent extraction and purification of the genomic DNA, which were carried out as described in the Materials and Methods section.

The genome of CA-170360 was fully sequenced by Macrogen (Seoul, Korea; http://www.macrogen.com/) using a combination of *de novo* PacBio RSII and Illumina HiSeq 2500 platforms. *De novo* PacBio sequencing yielded two linear, non-overlapping contigs of 5,971,501 bp and 2,704,079 bp, and a total genome size of 8,671,580 bp, with a GC content of 73.17 % (Table D1). No extrachromosomal elements were identified.

Contig Name	Length (bp)	GC %	Depth	Circular
Contig 1	5,971,501	73.2	105	No
Contig 2	2,704,079	73.0	103	No
Total	8,671,580	73.17	104	

Table D1. PacBio Assembly.Results Length(bp): The number of bases in each contig. GC %: GC content. Depth: The number of reads that overlap each contig. Circular: 5' end and 3' ends are connected.

The Illumina sequencing was performed to correct PacBio frameshifts caused by the high GC content of the genome. A total of 163 contigs were obtained, with a N50 of 169,024 bp.

Number of contigs	Contig sum (bp)	N50	GC %	Longest contig (bp)	Shortest contig (bp)	Average contig length (bp)
163	8,567,810	169,024	73.25	493,879	1,010	52,563

Table D2. Illumina Assembly Summary. Contigs sum: The total number of bases in the Contigs, N50: half of all bases reside in contigs of this size or longer. GC %: GC content.

The 163 contigs obtained with Illumina were mapped against the two contigs obtained with PacBio, and after Illumina correction, two contigs of 5,971,081 bp and 2,704,105 bp were generated, yielding a genome size of 8,675,186 bp and a GC content of 73.2 %.

1.2. Genome mining analysis

The CA-170360 genome sequence was analyzed with antiSMASH v6.0.1., PRISM, BAGEL4 and RiPPMiner and 33 putative BGCs were predicted, including terpenes, nucleosides, non-ribosomal peptide synthetases (NRPS), polyketide synthases (PKS) or ribosomally synthesized and post-translationally modified peptides (RiPPs) (Table D3). Only seven of these putative BGCs had a similarity above 70 % with other already known biosynthetic clusters, suggesting that this *S. cacaoi* strain had the potential to synthetize, among others, already known natural products such as actinonine, bezastatin, naseseazine C, puromycin, desferrioxamine E, candicidin, ectoine or any derivatives. The whole results predicted by antiSMASH are shown in Table D3.

1.2.1. Identification of pentaminomycins and BE-18257 antibiotics biosynthetic gene cluster

During the screening program in Fundación MEDINA, the cyclic pentapeptides Pentaminomycins and BE-182577 antibiotics were isolated and structurally elucidated from *S. cacaoi* CA-170360. The presence of the two families of cyclic pentapeptides suggested that they were expected to be produced by different non-ribosomal peptide synthetases (NRPS) pathways (Román-Hurtado *et al.,* 2021). antiSMASH analysis predicted eleven BGCs containing NRPS genes in the genome of CA-170360, eight of them were present in contig 1 while three were contained in contig 2. From all these NRPS BGCs, only one cluster from contig 1 (Region 1.2, Table D3) could correlate, according to the number of modules and domains of two neighboring NRPS genes, with the structure of the two types of cyclic pentapeptides. This cluster (*cpp* cluster) was identified as the putative pathway for the synthesis of both BE-18257 A-C and pentaminomycins A–H (GenBank number MW038823).

The *cpp* gene cluster (around 48 Kb) contains 15 ORFs, including two NRPS genes, each of them containing five adenylation (A) domains. A BLAST analysis allowed us to propose the function for each protein (Table D3 and Figure D1).

Figure D1. cpp *biosynthetic gene cluster. The BGC contains two NRPS genes,* cppB *and* cppM *(blue), one PBP-type TE gene* cppA *(green), and the genes present between* cppB *and* cppM *(cppC-cppL): two cytochromes P450 (yellow), two genes* potentially involved in regulation *(red) and other genes with unknown functions (grey). Figure from Román-Hurtado et al., 2021.*

Region	Type	Contig	From	To	Most similar known cluster	Similarity
Region 1.1	terpene	1	98,485	119,906	daptomycin	3 %
Region 1.2	NRPS	1	605,914	689,333	CDA1b-CDA4b	17 %
Region 1.3	NRPS-like	1	829,449	871,904	bottromycin A2	45 %
Region 1.4	lanthipeptide-class-III	1	2,247,981	2,270,755	chrysomycin	5 %
Region 1.5	nucleoside	1	2,271,456	2,294,026	puromycin	87 %
Region 1.6	T2PKS	1	2,926,240	2,998,839	murayaquinone	12 %
Region 1.7	ectoine	1	3,008,333	3,018,707	kosinostatin	13 %
Region 1.8	NRPS-like	1	3,143,724	3,184,224	CDA1b-CDA4b	12 %
Region 1.9	**NRPS-like**	**1**	**3,691,051**	**3,733,403**	**bezastatin derivates**	**90 %**
Region 1.10	siderophore	1	4,054,391	4,069,211	ficellomycin	3 %
Region 1.11	NRPS	1	4,218,079	4,273,066	herboxidiene	2 %
Region 1.12	terpene	1	4,399,617	4,420,069	xiamycin A	13 %
Region 1.13	RiPP-like	1	4,465,078	4,476,010		
Region 1.14	NRPS, T1PKS	1	4,655,386	4,753,157	caerulomycin A	20 %
Region 1.15	terpene	1	4,832,016	4,851,953	abyssomicin M X	9 %
Region 1.16	NRPS	1	4,918,244	5,064,565	surugamide A / surugamide D	57 %
Region 1.17	betalactone	1	5,134,671	5,164,815		
Region 1.18	terpene	1	5,234,725	5,261,358	hopene	69 %
Region 1.19	**NRPS**	**1**	**5,390,971**	**5,437,581**	**actinonin**	**100 %**
Region 1.20	CDPS	1	5,605,890	5,626,711	naseseazine C / C3-aryl pyrroloindolines	66 %
Region 1.21	lanthipeptide-class-V, T1PKS, NRPS	1	5,659,705	5,785,380	coelibactin	45 %
Region 1.22	NRPS,transAT-PKS,NRPS-like	1	5,863,192	5,956,851	phthoxazolin	18 %
Region 2.1	betalactone, NRPS, NRPS-like, T3PKS, T1PKS	2	1	133,759	totopotensamide A B	64 %
Region 2.2	lassopeptide	2	209,846	232,318	incednine	2 %
Region 2.3	**T1PKS, NRPS-like, CDPS**	**2**	**474,094**	**628,505**	**candicidin**	**76 %**
Region 2.4	NRPS, NRPS-like	2	675,827	727,759	vazabitide A	8 %
Region 2.5	NRPS	2	801,794	862,517	surugamide A / surugamide D	9 %
Region 2.6	terpene	2	1,237,857	1,267,491	carotenoid	54 %
Region 2.7	**CDPS**	**2**	**1,340,288**	**1,361,013**	**naseseazine C / C3-aryl pyrroloindolines**	**100 %**
Region 2.8	NRPS	2	1,773,968	1,838,639	cyclomarin D	8 %
Region 2.9	**siderophore**	**2**	**2,039,297**	**2,051,102**	**desferrioxamine E**	**100 %**
Region 2.10	**ectoine**	**2**	**2,110,793**	**2,121,197**	**ectoine**	**100 %**
Region 2.11	lanthipeptide-class-I	2	2,264,164	2,289,552		

Table D3. antiSMASH v6.0.1 analysis from Streptomyces cacaoi *CA-170360 whole genome sequence. antiSMASH detected a total of 33 biosynthetic gene clusters (22 in the first contig and 11 in the second contig). The most similar BGCs (above 70 % similarity) are highlighted in bold.*

ORF	Length (aa)	Closest BLAST match [Organism] GenBank Reference	Putative Function in the biosynthesis of cyclic pentapeptides	Identity(%)/ Similarity(%)
cppA	502	Hypothetical protein DEH18_05445 [*Streptomyces* sp. NHF165] QHF93414.1	Cyclization/ release	99/99
cppB	6187	Non-ribosomal peptide synthase/polyketide synthase [*Streptomyces cacaoi*] WP_158102276.1	Biosynthesis of BE.18257 antibiotics	99/99
cppC	172	MULTISPECIES: DUF2975 domain-containing protein [*Streptomyces*] WP_030891799.1	Unknown	100/100
cppD	102	Helix-turn-helix domain-containing protein [*Streptomyces cacaoi*] WP_149564434.1	Regulation	99/99
cppE	420	MULTISPECIES: sensor histidine kinase [*Streptomyces*] WP_051857187.1	Regulation	99/100
cppF	280	Hypothetical protein SCA03_67000 [*Streptomyces cacaoi* subsp. *cacaoi*] GEB54149.1	Unknown	100/100
cppG	131	MULTISPECIES: DUF742 domain-containing protein [*Streptomyces*] WP_030891786.1	Unknown	100/100
cppH	186	MULTISPECIES: ATP/GTP-binding protein [*Streptomyces*] WP_030891784.1	Unknown	100/100
cppI	451	MULTISPECIES: cytochrome P450 [*Streptomyces*] WP_030891781.1	Biosynthesis of 5-OH-Arginine	100/100
cppJ	431	Cytochrome P450 [*Streptomyces cacaoi*] WP_086815207.1	Biosynthesis of 5-OH-Arginine	99/99
cppK	271	3-hydroxybutyryl-CoA dehydratase [*Streptomyces cacaoi* subsp. *cacaoi*] GEB54144.1	Unknown	99/99
cppL	141	MULTISPECIES: hypothetical protein [*Streptomyces*] WP_141275837.1	Unknown	100/100
cppM	5901	Non-ribosomal peptide synthase/polyketide synthase [*Streptomyces* sp. NHF165] WP_159784853.1	Biosynthesis of Pentaminomycins A-H	99/99
cppN	69	MULTISPECIES: hypothetical protein [*Streptomyces*] WP_030890086.1	Unknown	100/100
cppO	175	MULTISPECIES: hypothetical protein [unclassified *Streptomyces*] WP_030890089.1	Unknown	99/100

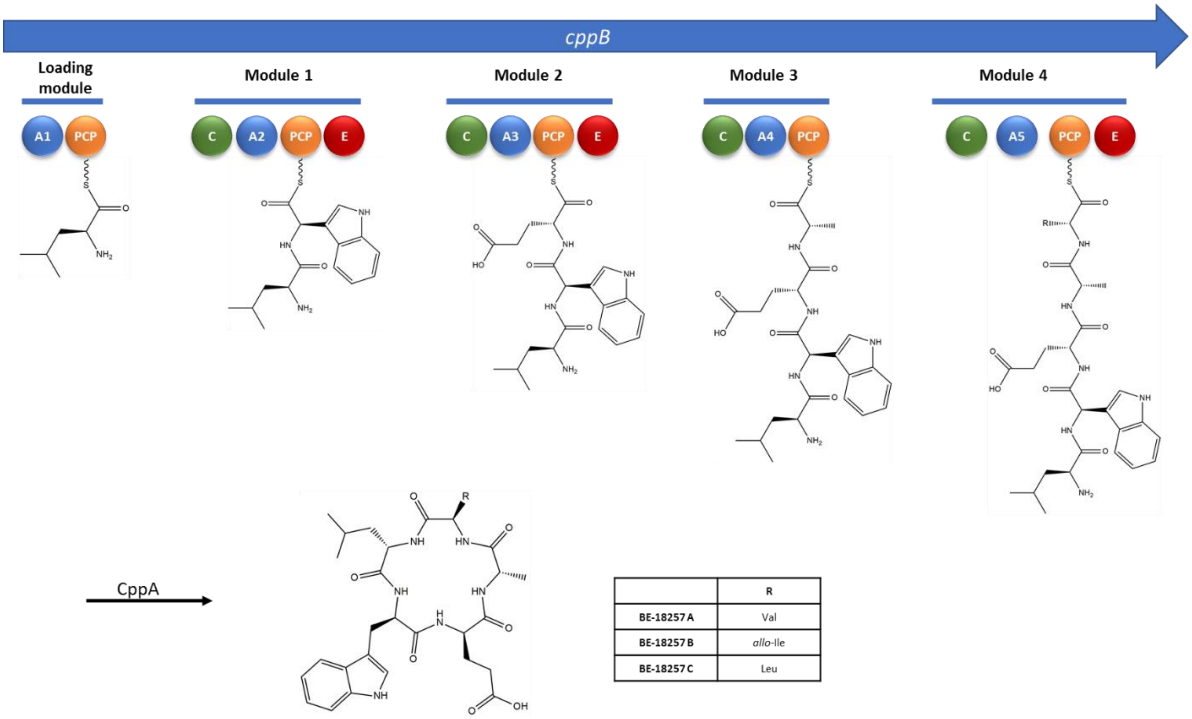
Table D4. Closest BLAST homolog for each ORF in cpp *BGC.*

The first NRPS gene *cppB* contains five adenylation domains and their amino acid specificity prediction (antiSMASH) leads to the peptide sequence Leu (A1), Trp (A2), Leu/Ser (A3), Ala (A4) and Val/Leu (A5), assuming fulfillment of the collinearity rule (Figure D2). Three epimerization (E) domains are located in the second, third and fifth modules, and they would be involved in the isomerization of an L- to D- amino acid, resulting in the final sequence L-Leu, D-Trp, D-Leu/Ser, L-Ala, D-Val/Leu, which is in accordance with the amino acid sequence of BE-18257 A-C (L-Leu, D-Trp, D-Glu, L-Ala, D-Val/D-*allo*-Ile/D-Leu). Therefore, these results suggest that the first NRPS gene *cppB* may be involved in the biosynthesis of peptidic skeleton of BE-18257 A–C antibiotics that would be completed by a final cyclization step.

The amino acid specificity prediction of the second NRPS gene *cppM,* which contains two E domains and five adenylation domains leads to the amino acid sequence Val/Leu/Phe (A1), Val (A2), Trp (A3), Arg (A4) and Leu/Phe (A5), assuming fulfillment of the collinearity rule. The two E domains are located in the second and fifth modules, so the final amino acid sequence would be L-l/Leu/Phe, D-Val, L-Trp, L-Arg, D-Leu/Phe, in agreement with the amino acid sequence of pentaminomycins A-H (L-Val/L-Leu/L-Phe, D-Val, L-Trp, L-N5-OH-Arg, D-Leu/D-Phe/D-2-Pal) (Figure D3). Subsequent modifications such as hydroxylation and cyclization would complete the biosynthesis of the pentaminomycins A-H.

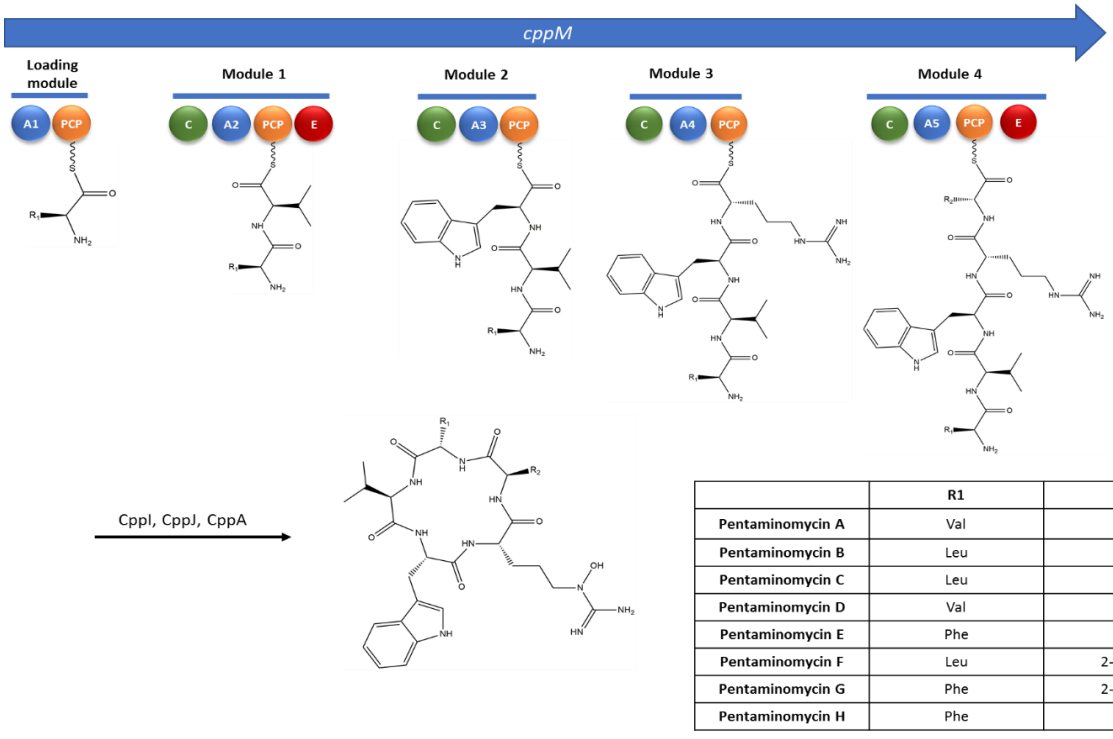
However, none of the two NRPS genes of the *cpp* cluster contains a TE domain responsible for the release and cyclization of the pentapeptides. Nevertheless, a PBP-type stand-alone protein with a beta-lactamase conserved domain (*cpp*A) was found upstream the first NRPS gene *cppB* and could act as a releasing enzyme.

The *cpp* cluster includes, between the two NRPS *cppB* and *cppM* genes, eight additional ORFs, including two cytochrome P450 enzymes (*cppI* and *cppJ*), regulatory genes and other genes of unknown function. Interestingly, downstream the second NRPS gene (*cppM*), some genes related to tryptophan biosynthesis are present.



	R
BE-18257A	Val
BE-18257B	*allo*-Ile
BE-18257C	Leu

Figure D2. Proposed biosynthetic pathway for the BE-18257 A–C antibiotics with the non-ribosomal peptide synthetase CppB modular organization. A1-A5, adenylation domains; PCP, peptidyl carrier protein; C, condensation domain; E, epimerase domain; CppA, PBP-type TE. Figure taken from Román-Hurtado et al., 2021b.



	R1	R2
Pentaminomycin A	Val	Leu
Pentaminomycin B	Leu	Leu
Pentaminomycin C	Leu	Phe
Pentaminomycin D	Val	Phe
Pentaminomycin E	Phe	Phe
Pentaminomycin F	Leu	2-Pyr-Ala
Pentaminomycin G	Phe	2-Pyr-Ala
Pentaminomycin H	Phe	Leu

Figure D3. Proposed biosynthetic pathway for the pentaminomycins A–H with the non-ribosomal peptide synthetase CppM modular organization. A1-A5, adenylation domains; PCP, peptidyl carrier protein; C, condensation domain; E, epimerase domain; CppI and CppJ, cytochromes P450; CppA, PBP-type TE. Figure taken from Román-Hurtado et al., 2021b.

1.2.1.1. Genome mining of cpp-like BGCs in Streptomyces *genomes*

A tblastn search of the CppB and CppM protein sequences against both nucleotide and Whole Genome Sequence (WGS) databases from NCBI showed that the *cpp* cluster is also present in the genomes of different strains of *S. cacaoi* described in Figure D4.

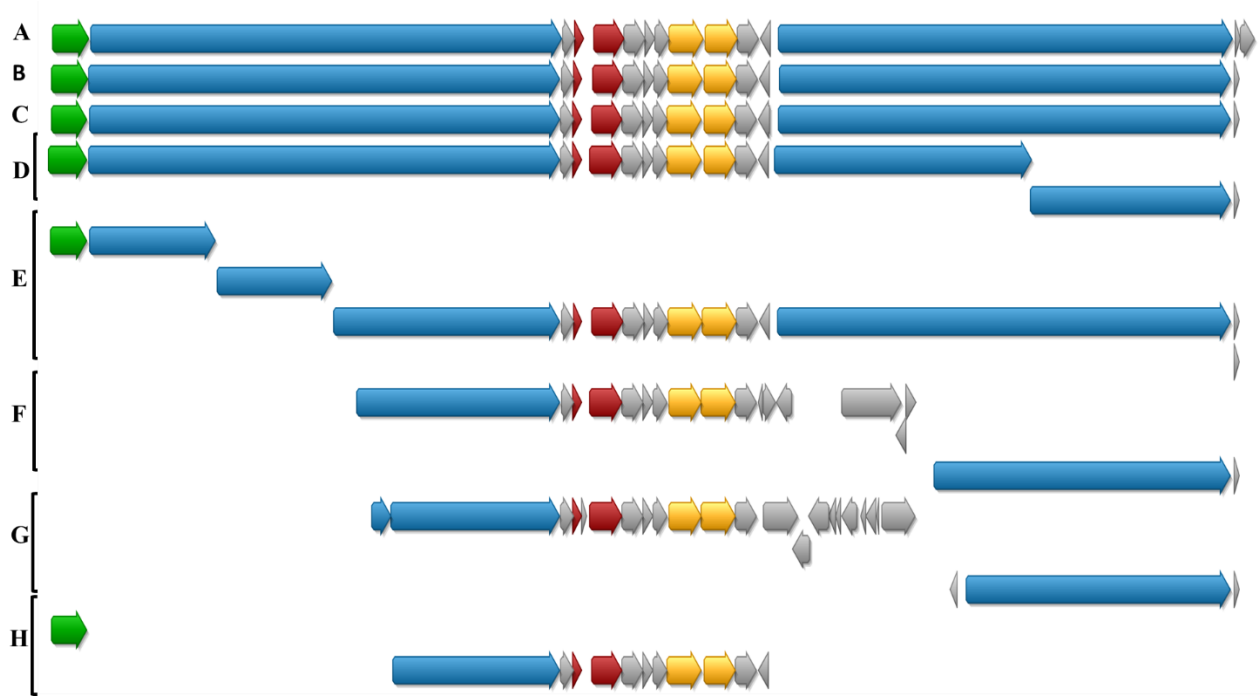


Figure D4. Schematic representation of the alignment of cpp BGC from Streptomyces cacaoi *CA-170360 and the homologous genome sequences found in NCBI. A,* S. cacaoi *CA-170360; B,* S. cacaoi *NHF165; C,* S. cacaoi *DSM 40057; D,* S. cacaoi *NRRL B-1220; E,* S. cacaoi *OABC16; F,* Streptomyces *sp. NRRL S-1868; G,* Streptomyces *sp. NRRL F-5053 and H,* S. cacaoi *NBRC 12748. Due to the high fragmentation in some of these S. cacaoi genomes, the corresponding BGC was found in different contigs and was not complete. Figure taken from Román-Hurtado et al., 2021b.*

Interestingly, all of them belong to *Streptomyces cacaoi* species: *S. cacaoi* NHF165 (CP029241.1), *S. cacaoi* DSM 40057 (JABELW000000000.1), *S. cacaoi* subsp. *cacaoi* NRRL-1220 (MUBL01000486), *S. cacaoi* OABC16 (VSKT010000024), *Streptomyces* sp. NRRL S-1868 (JOGD01000003), *Streptomyces* sp. NRRL F-5053 (JOHT01000009) and *S. cacaoi* NBRC 12748 (BJMM01000002.1). The *pen* cluster described by Hwang *et al.* (2020) in the strain *Streptomyces* sp. GG23, which has been also identified as a strain of *Streptomyces cacaoi,* has not been included in this analysis because the sequence is not yet available. Nevertheless, the comparison of the homologies described in the *pen* and in the *cpp* clusters clearly shows that they are highly similar.

1.2.2. Identification of cacaoidin biosynthetic gene cluster

The genome sequence of *S. cacaoi* CA-170360 was analyzed with several BGC prediction tools such as antiSMASH, BAGEL4, PRISM or RiPPMiner, but neither of them could predict the BGC responsible for cacaoidin biosynthesis. However, the current version of antiSMASH can predict a class-V lanthipeptide BGC in the genome of corresponding to cacaoidin (Region 1.21, Table D3).

Thus, the C-terminal sequence of cacaoidin (Thr-Ala-Ser-Trp-Gly-Cys) was used to query in a tBLASTn using the whole genome sequence of the producing strain. A 162 bp Open Reading Frame (ORF) was found to encode this sequence and helped to elucidate the final primary structure of the peptide. Cacaoidin structural gene *caoA* encodes a 23-amino acid C-terminal core peptide (SSAPCTIYASVSASISATASWGC) following a predicted 30-amino acid N-terminal leader peptide (MGEVVEMVAGFDTYADVEELNQIAVGEAPE) (Figure D5). Neither the leader nor the core peptide sequences showed high sequence similarity with any other lanthipeptide or linaridin.

Considering the structural characteristics of cacaoidin and the functional genes predicted to be necessary for its biosynthesis (Ortiz-López *et al.,* 2020), a putative 30 Kb BGC (*cao* cluster) containing 27 ORFs could be associated to the biosynthesis after BLAST analysis and HHpred secondary structure prediction (Figure D5, Table D5) (GenBank number MT210103).

In the *cao* cluster, no homologous genes of known dehydratases or cyclases commonly found in the four classes described for lanthipeptides nor in the class of linaridins could be identified. However, some Cao proteins from the *cao* BGC presented a certain degree of homology with the three-component lanthionine synthetase from the lexapeptide BGC (Xu *et al*., 2020): Cao9 shows a 49.6 % similarity with LxmK, Cao14 has a 40.5 % similarity with LxmX and Cao7 shows a 44.9 % similarity with LxmY. Interestingly, both LxmY and Cao7 contain a HopA1 conserved domain (PFAM17914) that has been described in the HopA1 effector protein from *Pseudomonas syringae* (Park *et al.,* 2015).

The *cao* BGC also encodes a putative cypemycin decarboxylase CypD homologue (CaoD) containing a conserved phosphopantothenoylcysteine (PPC) synthetase/decarboxylase domain (Ding *et al.,* 2018a; Sit *et al.,* 2011), CaoD shares a 54.8 % similarity with LxmD from the lexapeptide BGC.

The *cao* BGC lacks homologues of the methyltransferases CypM or LxmM, but it encodes the putative O-methyltransferase Cao4, containing the conserved SAM-dependent Methyltransf_2 domain, which might be involved in the N-terminal dimethylation.

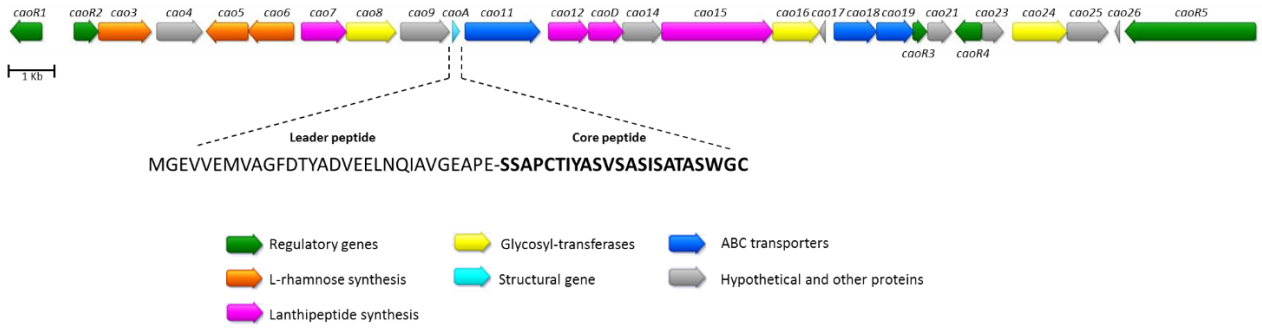


Figure D5. Schematic representation of the BGC of cacaoidin, where caoA *codes for the precursor peptide. The sequences of the leader and core peptides of cacaoidin are shown. Figure taken from Román-Hurtado.*

Another characteristic structural feature of cacaoidin is its unusually high number of D-amino acids, but no dehydrogenases were found in the *cao* cluster. However, Cao12 showed a 58.6 % similarity with the recently described $F_{420}H_2$-dependent reductase LxmJ, involved in the introduction of D-amino acids in the lexapeptide (Xu *et al.,* 2020). Both proteins also contained a $F_{420}H_2$-dependent oxidoreductase (MSMEG_4879 family) conserved domain.

Processing of leader peptide is another key step in the post-translational modification of RiPPs (Repka *et al.,* 2017). In the cacaoidin cluster, Cao15 encodes a putative Zn-dependent peptidase belonging to the M16 peptidase family that may be involved in the leader peptide processing.

The disaccharide moiety β-6-deoxygulopyranosyl-(1→3)-*α*-rhamnopyranoside of cacaoidin has not been previously reported (Ortiz-López *et al.,* 2020). Three of the four proteins required for the synthesis of *α*-l-rhamnose (*rmlA, rmlB* and *rmlD*) are encoded in the *cao* cluster, and a BLAST search of RmlC against CA-170360 whole genome sequence also shows the presence of a *rmlC* gene and additional *rmlA, rmlB* and *rmlD* homologous genes outside the cacaoidin cluster.

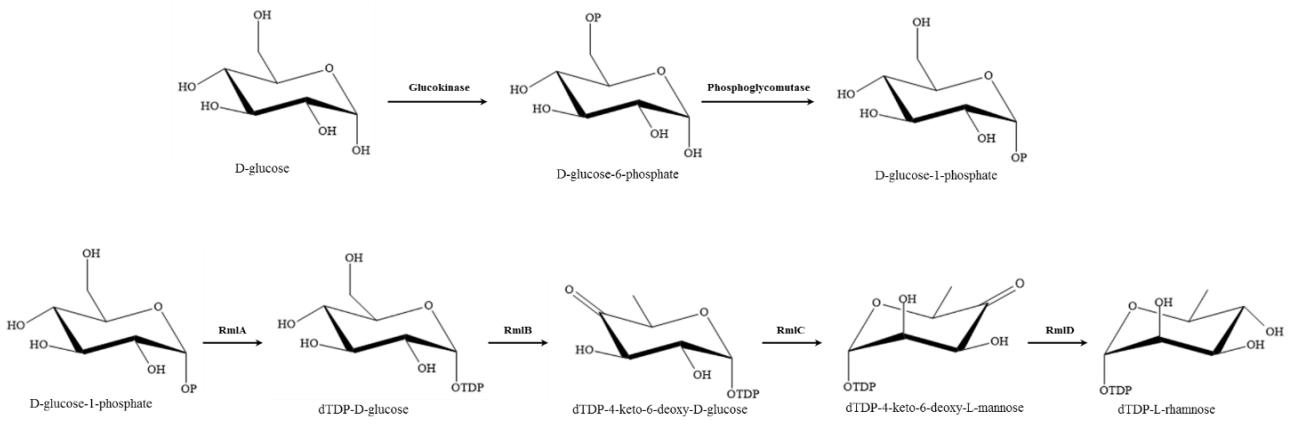


Figure D6. Schematic representation of the biosynthesis of dTDP-L-rhamnose from D-glucose. Figure taken from Román-Hurtado et al., 2021a.

The biosynthetic genes involved in the biosynthesis of NDP-L-gulose or NDP-6-deoxy-L-gulose in the antitumor antibiotics bleomycin, tallysomycin and zorbamycin (Chen *et al.,* 2020; Galm *et al.,* 2011) (Figure D7) were used as query to search for genes involved in the biosynthesis of L-gulose. However, no homologous genes were found in the *cao* cluster. A BLAST search in the whole genome sequence of CA-170360 identified some protein homologues, including a D-glycero-beta-D-manno-heptose 1-phosphate adenylyltransferase homologous to BlmC/TlmC/ZbmC (48 % similarity); a NAD-dependent epimerase/dehydratase homologous to BlmE/TlmE (38.7 % similarity); a GDP-mannose 4,6-dehydratase with 62 % similarity to ZbmL; and a dTDP-glucose 4,6-dehydratase with 34 % similarity with ZbmG. However, none of these proteins were found associated within the same cluster in the CA-170360 genome.

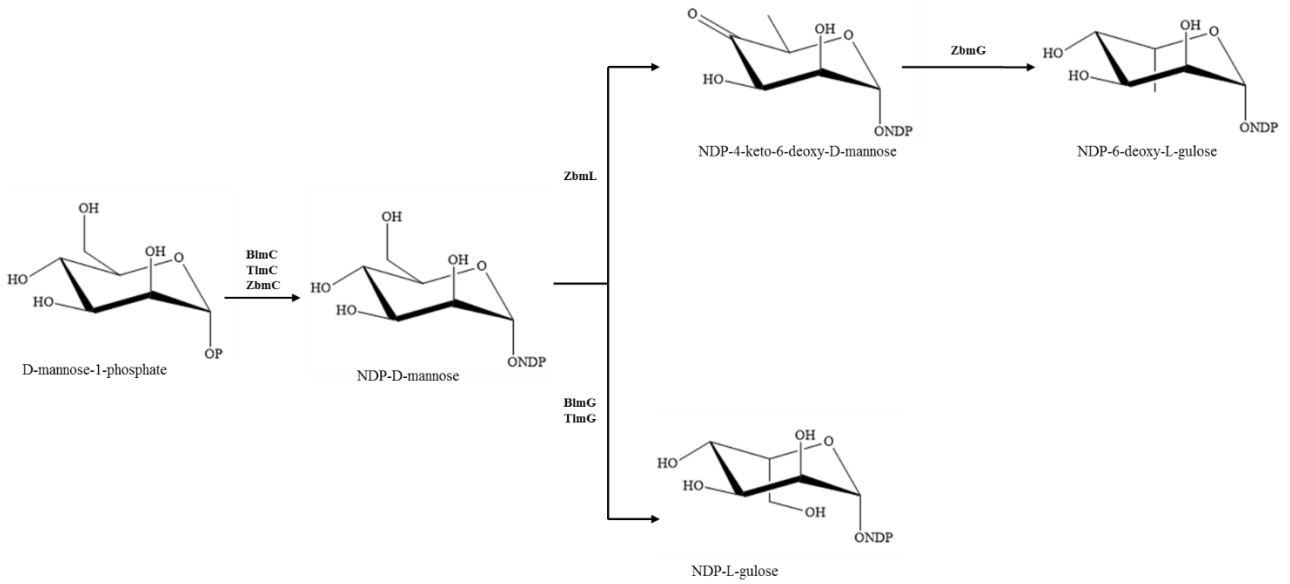


Figure D7. Proposed pathway for the β-L-6-deoxy-gulose sugar biosynthesis for the BLM, TLM and ZMB compounds. Figure taken from Román-Hurtado et al., 2021a.

As one of the unusual structural features of cacaoidin, the disaccharide is O-linked to the aromatic ring of the tyrosine residue (Ortiz-López *et al.,* 2020). The *cao* cluster contains three glycosyltransferases (GTs) probably involved in this glycosylation: Cao8, Cao16 and Cao24. Cao8 and Cao16 belong to the family GT-2 and show 42 % identity (54 % similarity) and 43 % identity (52 % similarity), respectively, with an UDP-Glc:alpha-D-GlcNAc-diphosphoundecaprenol beta-1,3-glucosyltransferase WfgD, which catalyzes the addition of Glc, the second sugar moiety of the O152-antigen repeating unit, to GlcNAc-pyrophosphate-undecaprenol (Brockhausen *et al.,* 2008). In contrast, Cao24 belongs to the family GT-4 that has a GT4_GtfA-like domain and a conserved RfaB domain, involved in the cell wall and membrane biosynthesis (Pradel *et al.,* 1992).

ORF	Closest BLAST homolog	%Identity/ Similarity	Conserved Domains	Putative Function
caoR1	DNA-binding response regulator [*Streptomyces*] WP_086815549.1	100/100	CitB (NarL/FixJ family, contains REC and HTH domains); HTH_LUXR	Positive regulator
caoR2	Helix-turn-helix transcriptional regulator [*Streptomyces* sp. NRRL F-5053] WP_159401550.1	98.44/99	HTH_XRE superfamily	Negative regulator
cao3	Hypothetical protein SCA03_05120 [*Streptomyces cacaoi* subsp. *cacaoi*] GEB47961.1	97.81/100	RmlD_sub_bind	L-rhamnose synthesis
cao4	Methyltransferase [*Streptomyces cacaoi*] WP_149563574.1	100/100	Methyltrans_2 (O-methyltransferase)	N-terminal dimethylation
cao5	dTDP-glucose 4,6-dehydratase [*Streptomyces cacaoi*] WP_149563573.1	98.16/100	dTDP_GD_SDR-e (dDTP-D-glucose 4,6-dehydratase)	L-rhamnose biosynthesis
cao6	MULTISPECIES: glucose-1-phosphate thymidylyltransferase [*Streptomyces*] WP_030874943.1	100/100	RmlA_long (glucose-1-phosphate thymidylyltransferase)	L-rhamnose biosynthesis
cao7	T3SS effector HopA1 family protein [*Streptomyces cacaoi*] WP_149563571.1	100/100	HopA1 superfamily	Unknown
cao8	Glycosyltransferase family 2 protein [*Streptomyces* sp. NRRL S-1868] WP_051855540.1	99.48/100	Glycos_transf_2	Glycosylation
cao9	Phosphotransferase [*Streptomyces* sp. NRRL S-1868] WP_051855537.1	100/100	PKc_like superfamily (protein kinase catalytic domain)	Unknown
caoA	Hypothetical protein SCA03_05190 [*Streptomyces cacaoi* subsp. *cacaoi*] WP_169729690.1	100/100	No putative conserved domains detected	Structural gene
cao11	ABC transporter ATP-binding protein [*Streptomyces* sp. NRRL F-5053] WP_030890022.1	99.66/100	MdlB (ABC-type multidrug transport system, ATPase and permease component)	Cacaoidin biosynthesis
cao12	MULTISPECIES: LLM class flavin-dependent oxidoreductase [*Streptomyces*] WP_030890020.1	100/100	SsuD (Flavin-dependent oxidoreductase)	Cacaoidin biosynthesis

ORF	Closest BLAST homolog	%Identity/ Similarity	Conserved Domains	Putative Function
caoD	Hypothetical protein SCA03_05220 [*Streptomyces cacaoi* subsp. *cacaoi*] GEB47971.1	99.63/100	PRK05579 (bifunctional phosphopantothenoylcysteine decarboxylase/phosphopantothenate synthase)	AviMeCys biosynthesis
cao14	MULTISPECIES: hypothetical protein [unclassified *Streptomyces*] WP_030890015.1	99.34/100	No putative conserved domains detected	Unknown
cao15	Insulinase family protein [*Streptomyces* sp. NHF165] WP_159787532.1	99.42/100	PqqL (predicted Zn-dependent peptidase)	Leader peptide cleavage
cao16	Glycosyltransferase family 2 protein [*Streptomyces cacaoi*] WP_086816368.1	100/100	Glycos_transf_2	Glycosylation
cao17	Hypothetical protein [*Streptomyces cacaoi*] WP_158102314.1	91.7/100	No putative conserved domains detected	Unknown
cao18	ABC transporter ATP-binding protein [*Streptomyces* sp. NRRL F-5053] WP_030890006.1	100/100	CcmA (ABC-type multidrug transport system)	Cacaoidin biosynthesis
cao19	MULTISPECIES: ABC transporter permease [*Streptomyces*] WP_030874928.1	99.65/100	ABC2_membrane_3 (ABC-2 family transporter protein)	Cacaoidin biosynthesis
caoR3	Hypothetical protein SCA03_05280 [*Streptomyces cacaoi* subsp. *cacaoi*] QKI29080.1	100/100	HTH_XRE superfamily	Negative regulator
cao21	Hypothetical protein [*Streptomyces* sp. NRRL S-1868] WP_158933426.1	100/100	No putative conserved domains detected	Unknown
caoR4	TetR/AcrR family transcriptional regulator [*Streptomyces cacaoi*] WP_086816362.1	100/100	AcrR (DNA-binding transcriptional regulator)	Negative regulator
cao23	Hypothetical protein SCA03_05310 [*Streptomyces cacaoi* subsp. *cacaoi*] GEB47980.1	99.39/100	SRPBCC superfamily	Unknown

ORF	Closest BLAST homolog	%Identity/ Similarity	Conserved Domains	Putative Function
cao24	MULTISPECIES: glycosyltransferase family 4 protein [*Streptomyces*] WP_030874924.1	100/100	GT4_AmsD-like	Glycosylation
cao25	Hypothetical protein [*Streptomyces cacaoi*] WP_086816358.1	99.06/100	No putative conserved domains detected	Unknown
cao26	No homologues found	-	-	Unknown
caoR5	Tetratricopeptide repeat protein [*Streptomyces* sp. NRRL S-1868] WP_030874922.1	99.41/100	BTAD (Bacterial Transcriptional Activation)	Positive regulator

Table D5. Closest BLAST homolog for each Open Reading Frame (ORF) in cacaoidin Biosythetic Gene Cluster (BGC.)

Furthermore, the *cao* BGC also encodes three ABC transporters (*cao11*, c*ao18* and *cao19*) and five different transcriptional regulators: one LuxR (CaoR1), two HTH-type XRE (CaoR2 and CaoR3), one TetR (CaoR4) and one SARP (CaoR5). XRE and TetR have been described as transcriptional repressors (Wood *et al.,* 1990; Cuthbertson and Nodwell*.,* 2013) while LuxR and SARP have been designated as transcriptional activators (Chen *et al.,* 2011; Li *et al.,* 2019). The remaining five genes identified in the *cao* cluster (*cao17, cao21, cao23, cao25* and *cao26*) do not have any defined functions. Cao23 belongs to the SRPBCC (START/RHO_alpha_C/PITP/Bet_v1/CoxG/CalC) superfamily of proteins (Radauer *et al.,* 2008; Iyer *et al.,* 2001). This superfamily contains aromatase/cyclase (ARO/CYC) domains, such as those described for tetracenomycin (Ames *et al.,* 2008) and Smu.440 (Nan *et al.,* 2009). However, the function of Cao23 cannot be proposed yet.

The HHpred analysis of each ORF was also used for the detection of RiPP precursor peptide Recognition Elements (RREs) (Burkhart *et al.,* 2015). However, no RREs were found in the Cao proteins, suggesting the possibility of alternative leader peptide recognition domains that are unrelated to the already known RREs.

1.2.2.1. Additional lanthidin clusters in public databases

To date, only the class V lanthipeptides cacaoidin (Román-Hurtado *et al.,* 2021a), pristinin A (Kloosterman *et al.,* 2020) and lexapeptide (Xu *et al.,* 2020) BGCs have been described. To study if more lanthidin-encoding clusters could be found within actinomycetes, a BLAST search against the NCBI whole genome shotgun sequences database was performed, and sequences with a high degree of homology to cacaoidin BGC were found in *Streptomyces cacaoi* species: *S. cacaoi* NHF165 (CP029241.1), *S. cacaoi* DSM 40057 (JABELW000000000.1), *S. cacaoi* subsp. *cacaoi* NRRL-1220 (MUBL01000486), *S. cacaoi* OABC16 (VSKT010000024), *Streptomyces* sp. NRRL S-1868 (JOGD01000003), *Streptomyces* sp. NRRL F-5053 (JOHT01000009) and *S. cacaoi* NBRC 12748 (BJMM01000002.1) (Figure D8, Table D6). Interestingly, these were the same strains in whose genomes the *cpp* cluster was found. An alignment of the precursor peptide of the cacaoidin in all homologous clusters showed that no variations in the protein sequence were found. No other cacaoidin-derived peptides or pathways were found in the databases, indicating that the cacaoidin BGC is very conserved.

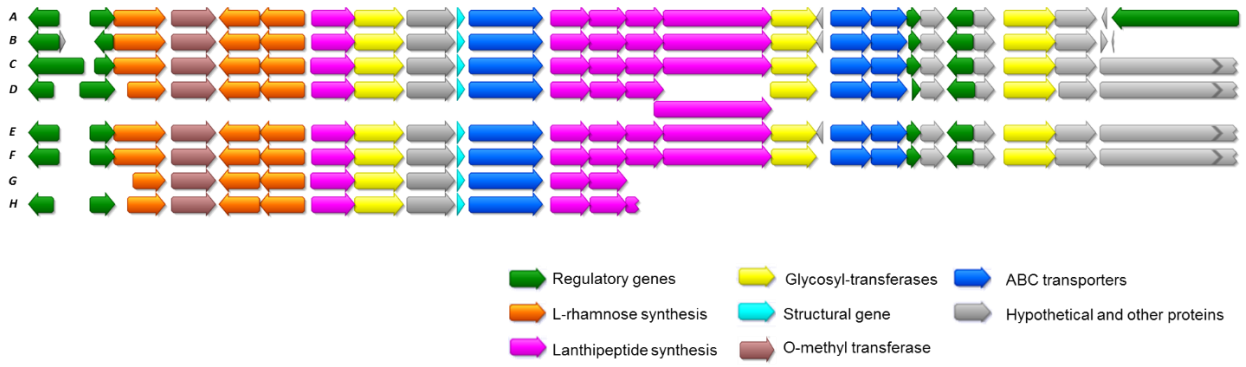


Figure D8. Schematic representation of the alignment of cacaoidin BGC from Streptomyces cacaoi *CA-170360 and the highly homologous clusters found in NCBI. All of them belong as well to strains of* S. cacaoi. *A:* S. cacaoi *CA-170360; B:* Streptomyces *sp. NRRL S-1868; C:* Streptomyces *sp. NRRL F-5053; D:* Streptomyces cacaoi *NHF 165; E*: S. cacaoi *subsp.* cacaoi *OABC16; F:* S. cacaoi *subsp.* cacaoi *NBRC 12748; G:* S. cacaoi *subsp.* cacaoi *NRRL B-1220;H:* Streptomyces cacaoi *DSM 40057. Figure taken from Román-Hurtado et al., 2021a.*

A phylogenetic tree generated using the Neighbor-Joining method and corrected with the Tamura 3-parameter algorithm (Tamura, 1992; Saitou and Nei, 1987) showed the close relatedness of strain *Streptomyces cacaoi* CA-170360 with the other species of *S. cacaoi* subsp. *cacaoi* that contained the *cao* cluster, which was highly supported by the bootstrap values (Figure D12). Moreover, when the 16S rDNA sequences of the unclassified strains *Streptomyces* sp. NRRL F-5053 and *Streptomyces* sp. NRRL S-1868 were analyzed in EzBiocloud, a 100 % sequence similarity was confirmed with *Streptomyces cacaoi* subsp. *cacaoi* NRRL B-1220, indicating that the cacaoidin BGC is so far limited to this specific clade of highly related strains associated to the subspecies *S. cacaoi* subsp. *cacaoi*, with no identifiable orthologs in other *Streptomyces* species. Other members of the subspecies *S.cacaoi* subsp. *asoensis* were included in the study, but they were clustered in a very distant phylogenetic branch from *S.cacaoi* subsp. *cacaoi* (Figure D9). In fact, none of the three whole genome sequences available in NCBI from strains of *S. cacaoi* subsp. *asoensis* contained any region homologous to the *cao* cluster, supporting that the *cao* BGC is a characteristic trait of the subspecies *S. cacaoi* subsp *cacaoi*.

Nevertheless, the analysis of below-threshold scores of CaoA BLAST results, together with the search of HopA1 domain-containing proteins similar to Cao7, allowed us to find some additional pathways that could encode new lanthidins (Figure D10). The alignment of the hypothetical precursor peptides shows the presence of some conserved residues that possibly could be involved in the leader peptide recognition by biosynthetic enzymes (Figure D11). In addition, the analysis of the ORFs present in all these clusters shows that all of them share a HopA1 domain-containing protein, a $F_{420}H_2$-dependent oxidoreductase, a CypD-related protein, a Zn-dependent or S9 peptidase and a putative phosphotransferase (Figure D10). Furthermore, a comparative analysis between Cao4 and other O-methyltransferases found in the putative clusters of class V lanthipeptides or lanthidins showed that they share around 50 % of homology with Cao4 (Figure D10). All of them have also the AdoMet_MTases superfamily domain, characteristic of the SAM-dependent methyltransferases and a Methyltransf_2 domain, which seems to be characteristic of the O-methyltransferases present in the putative class V lanthipeptides BGCs.

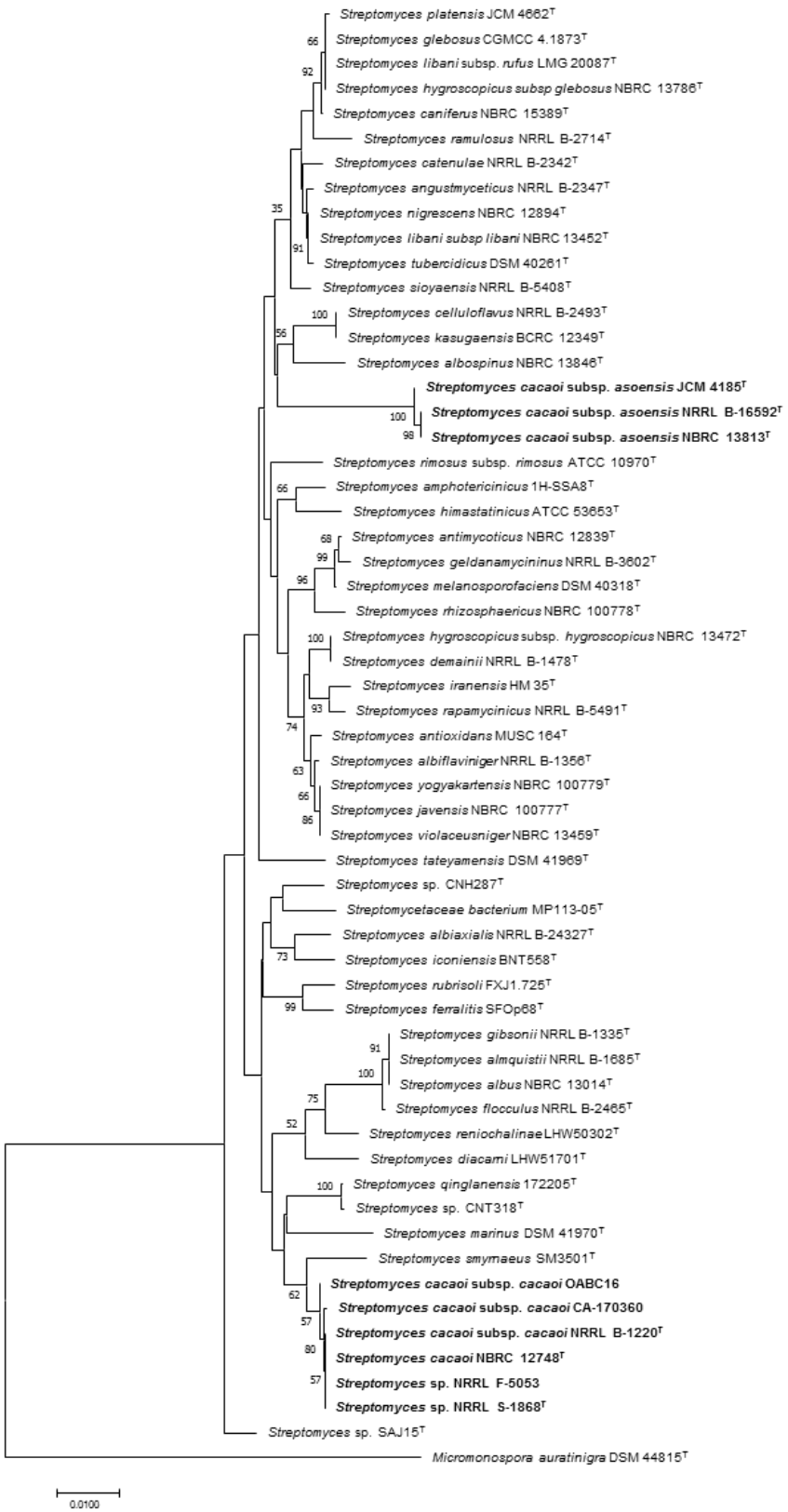


Figure D9. Neighbor-joining tree built with Mega X based on nearly complete 16S rRNA gene sequences of CA-170360, the 50 closest type strains of the genus Streptomyces and three strains of Streptomyces cacaoi subsp. asoensis. Micromonospora auratinigra *DSM 44815 (T) was used as an outgroup. The numbers at the nodes indicate the bootstrap value (%) based on NJ analysis of 1000 replicates; only values higher than 50 % are shown. Figure taken from Román-Hurtado et al., 2021a*

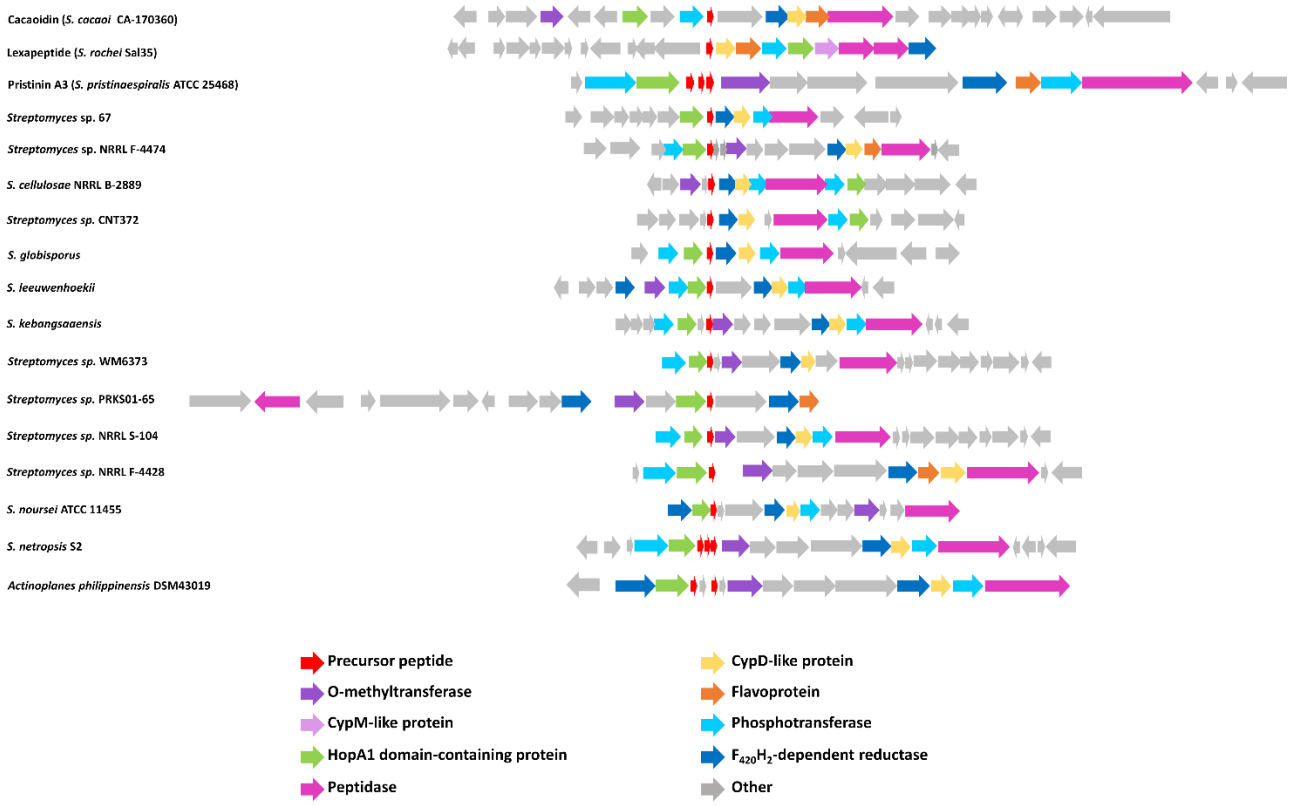


Figure D10. Schematic representation of cacaoidin BGC and the BGC of lexapeptide and other putative lanthidins (class V lanthipeptides). Figure taken from Román-Hurtado et al., 2021a.

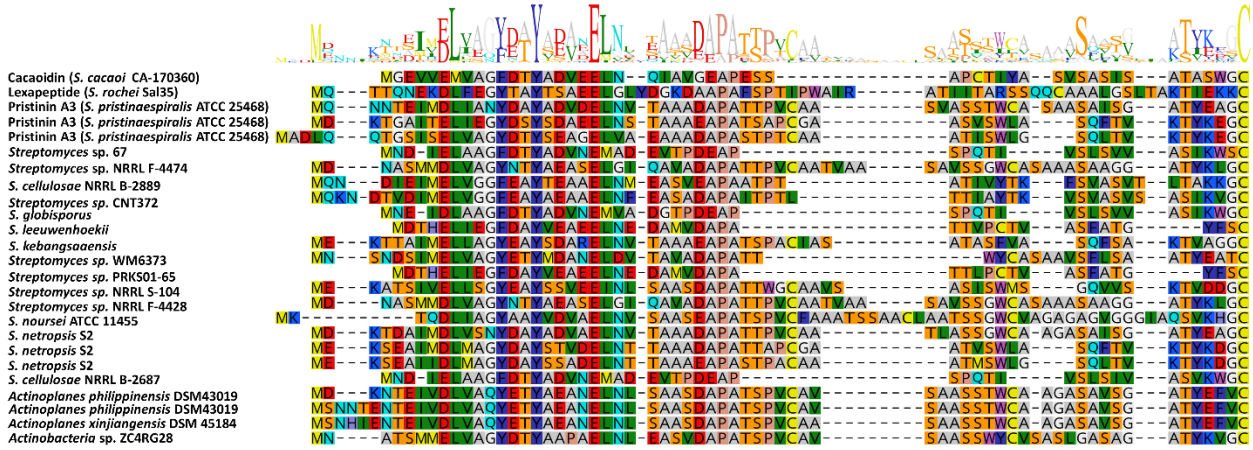


Figure D11. Alignment of the protein sequence of cacaoidin and putative lanthidins (class V lanthipeptides) precursor peptides. Sequence logo shows conserved residues. Figure modified from Román-Hurtado et al., 2021a.

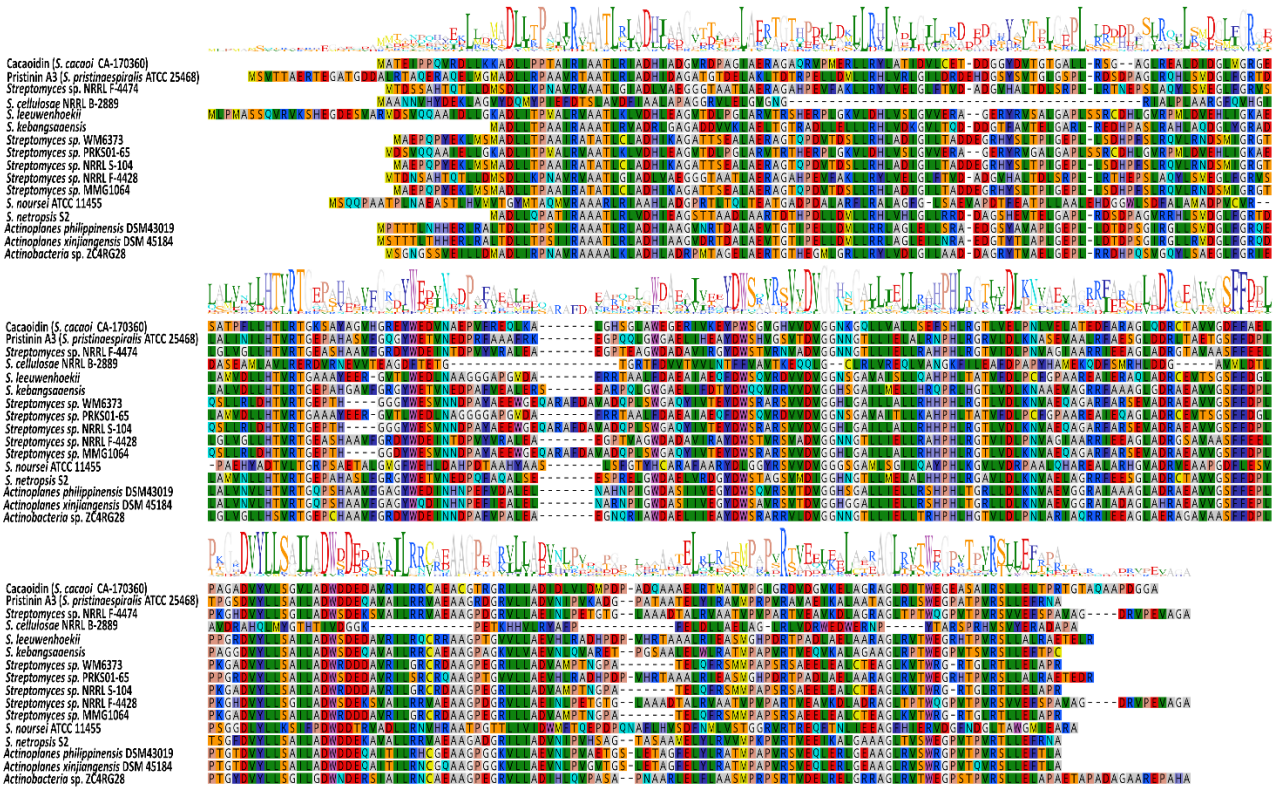


Figure D12. Alignment of the protein sequences of the O-methyltransferase present in cao *BGC and other O-methyltransferases found in the putative lanthidins BGCs (class V lanthipeptides). Sequence logo shows conserved residues.*

.

	Streptomyces cacaoi subsp. *cacaoi* NBRC 12748		*Streptomyces* sp. NRRL F-5053		*Streptomyces* sp. NRRL S-1868		*Streptomyces cacaoi* subsp. *cacaoi* NRRL B-1220		*Streptomyces cacaoi* subsp. *cacaoi* OABC16		*Streptomyces cacaoi* subsp. *cacaoi* NHF 165		*Streptomyces cacaoi* subsp. *cacaoi* DSM 40057	
	% Identity	% Similarity	% Identity	% Similarity	% Identity	% Similarity	% Identity	% Similarity	% Identity	% Similarity	% Identity	% Similarity	% Identity	% Similarity
caoR1	100	100	99.2	99.6	100	100	-	-	99.6	100	100	100	100	100
caoR2	97.9	99	98.1	98.1	34.4	40.6	-	-	97.9	99	94.92	100	97.92	100
cao3	97.8	98.1	99	99	96.6	97.1	99.6	100	98.1	98.3	98.69	100	99.67	74
cao4	99.1	99.7	99.7	100	99.1	99.7	99.1	99.7	100	100	99.71	100	99.14	100
cao5	97.9	98.8	97.5	98.2	98.2	98.8	97.9	98.8	98.2	98.8	98.16	100	97.85	100
cao6	100	100	99.7	100	100	100	100	100	99.2	99.7	99.72	100	100	100
cao7	99.7	99.7	98.9	99.4	99.1	99.4	99.4	99.7	99.7	100	99.42	100	99.71	100
cao8	98.7	99.5	99.5	99.5	99.5	99.7	98.7	99.5	99.2	99.5	98.96	100	98.69	100
cao9	99.7	99.7	99.7	100	100	100	99.7	99.7	99.5	100	99.74	100	99.74	100
caoA	100	100	100	100	100	100	100	100	100	100	100	100	100	100
cao11	99.1	100	99.7	100	99.1	100	99.1	100	99.5	100	99.14	100	99.14	100
cao12	99.7	100	100	100	99.7	100	99.7	100	99.7	100	100	100	99.68	100
caoD	99.6	100	92.1	92.1	93.3	93.3	99.6	100	94.6	95.7	98.46	97	99.23	97
cao14	99.3	99.3	99.3	99.3	98.7	99	-	-	99	99	99.13	100	-	-
cao15	98.6	99	98.6	98.8	99	99.3	-	-	98.4	98.8	99.31	92	-	-
cao16	100	100	99.7	99.7	99.4	99.4	-	-	100	100	99.72	100	-	-
cao17	-	-	-	-	100	100	-	-	100	100	100	100	-	-
cao18	99.1	99.1	100	100	99.7	99.7	-	-	99.4	99.4	99.08	100	-	-
cao19	99.7	99.7	99.7	99.7	99.7	99.7	-	-	99.3	99.7	99.65	100	-	-
caoR3	100	100	100	100	99	100	-	-	99.1	100	100	100	-	-
cao21	99.5	99.5	99.5	99.5	100	100	-	-	98.9	99.5	98.9	100	-	-
caoR4	100	100	100	100	100	100	-	-	100	100	100	92	-	-
cao23	99.4	99.4	100	100	98.8	98.8	-	-	99.4	100	100	93	-	-
cao24	100	100	100	100	100	100	-	-	100	100	100	100	-	-
cao25	99.1	99.1	96.9	97.8	98.8	99.1	-	-	98.4	98.8	97.6	91	-	-
cao26	-	-	-	-	-	-	-	-	-	-	-	-	-	-
caoR5	-	-	-	-	-	-		-		-	-	-	-	-

Table D6. BLAST homology between the ORFs in cacaoidin BGC of CA-170360 and NCBI found clusters.

1.3. OSMAC study, growth curves and production of secondary metabolites

The strain *Streptomyces cacaoi* CA-170360 was previously described to produce the novel lanthidin cacaoidin (Ortiz-López *et al.,* 2020), the family of cyclic pentapeptides BE-18257 A–C antibiotics (Román-Hurtado *et al.,* 2021b) and, to a much lesser extent, the pentaminomycins A–H (Carretero-Molina *et al.,* 2021).

To identify the best production conditions of each family of compounds, an OSMAC approach (Bode *et al.,* 2002) was performed. The strain was grown in a total of six production media (YEME, R2YE, KM4, MPG, FR23 and DEF-15) and different fermentation times in 40 mL EPA vials containing 10 mL of each medium. The MPG medium was included in this OSMAC study as it was the original medium where cacaoidin and pentaminomycins A-H were first detected in Fundación MEDINA. On the other hand, the YEME, R2YE and KM4 media were selected as they were successfully employed to produce new class III lantibiotics, such as erythreapeptin, avermipeptin and griseopeptin from *Saccharopolyspora erythraea*, *Streptomyces avermitilis* and *Streptomyces griseus*, respectively (Völler *et al.,* 2012). These media were also successfully used for the heterologous expression of the labyrinthopeptins BGC, a class III lantibiotic from *Actinomadura namibiensis* in different *Streptomyces* host strains (Krawczyk *et al.,* 2013). Finally, the FR23 and DEF-15 are complex and defined media, respectively, that are usually employed in Fundación MEDINA for the production of secondary metabolites (Trabelsi *et al.,* 2016; Sánchez-Hidalgo *et al.,* 2012). The generated extracts were analyzed by LC-HRESIMS(+)-TOF and the production titers were estimated based on the Extracted Ion Chromatograms (EIC) peak areas for each of the *m/z*ions.

Cacaoidin production was studied at 28 °C at five different times (7, 10, 14, 17 and 21 days) and the highest production rate was detected in media FR23 at 17 days with 10-folder increase regarding the production level in the original medium where cacaoidin was first detected (MPG), followed KM4 by and R2YE at the same time. Interestingly, the original medium MPG showed relative lower production rates in comparison to the other media (Figure D13).

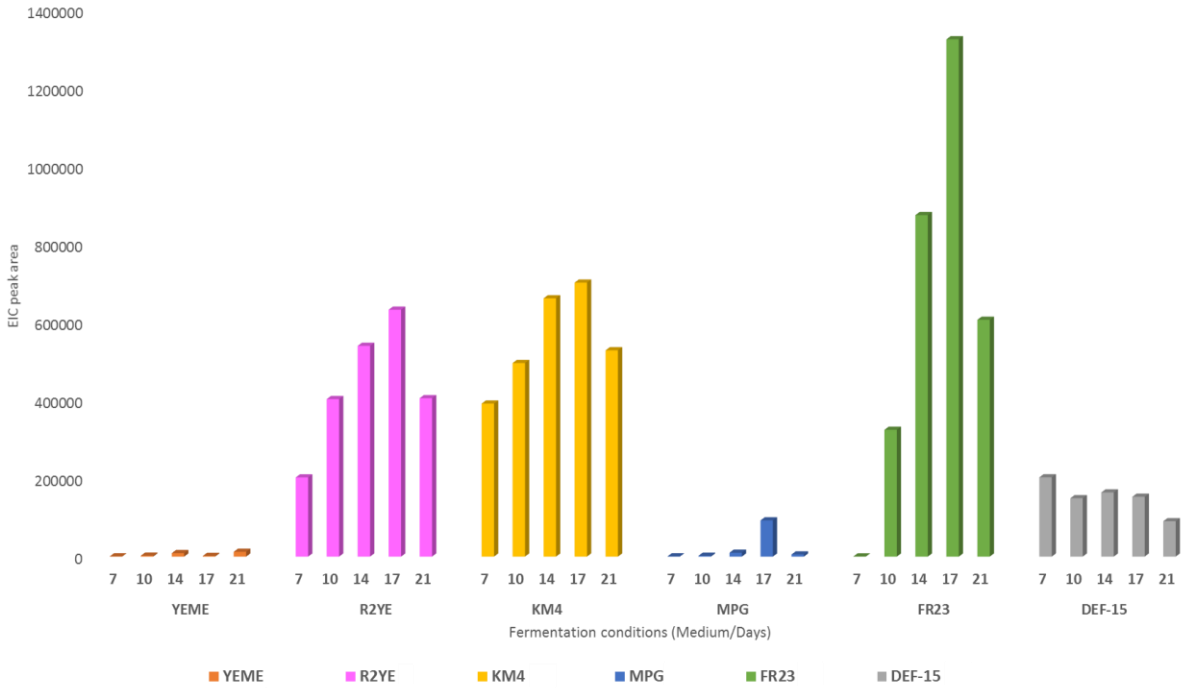


Figure D13. Production of cacaoidin by S. cacaoi *CA-170360 in six different media, YEME (orange), R2YE (pink), KM4 (yellow), MPG (blue), FR23 (green) and DEF-15 (grey) at five different times (7, 10, 14, 17 and 21 days. The average extracted ion chromatogram (EIC) peak area for the* m/z *793.057±0.005 ion from triplicate culture extracts of the strain CA-170360 is represented.*

Cyclic pentapeptides productions were studied in the same six media than cacaoidin but at three different times (7, 14, and 21 days) selected based in previous fermentation studies. Production of BE-18257 A-C was the highest at the three times in KM4 and FR23 media, and at 21 days in MPG medium, whereas the detection levels of pentaminomycins were very low. Pentaminomycins A-H were mostly produced in YEME and R2YE and required longer incubations of 14 and 21 days, although the production was still very low (Figure D14). Pentaminomycin H coeluted with pentaminomycin C in the LC-HRMS analytical conditions employed. Therefore, the peak might contain only one of the isomers or a mixture of both.

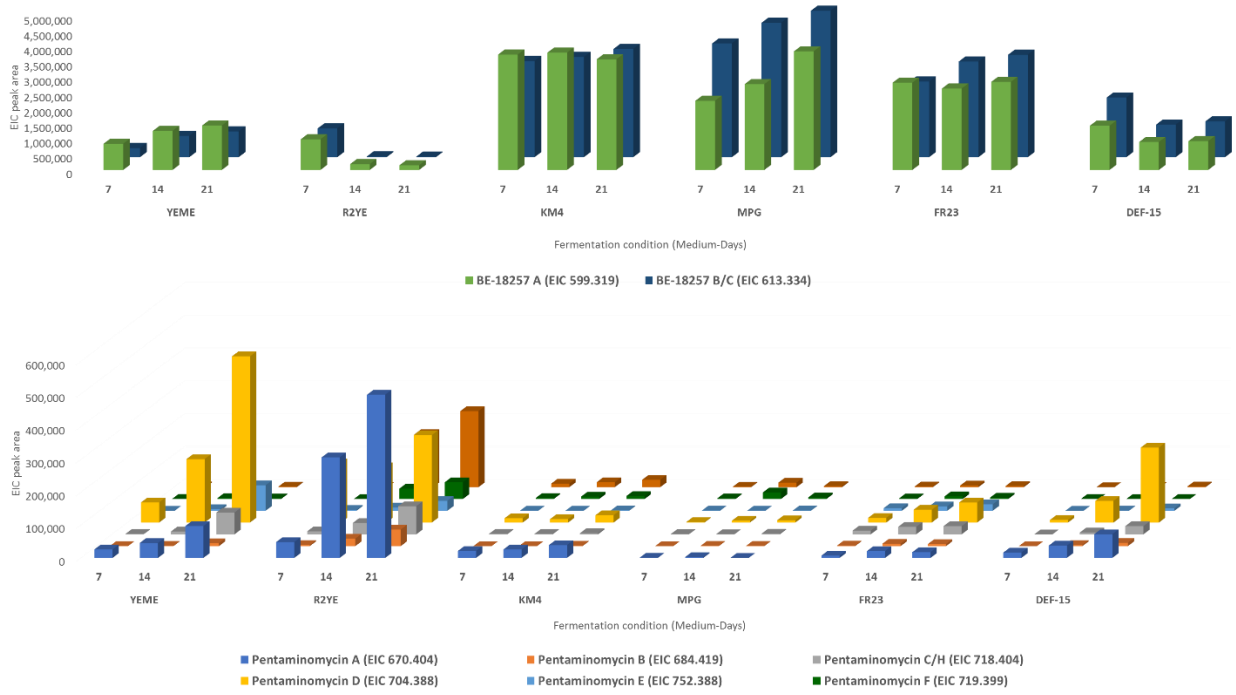


Figure D14. Production of BE-18257 antibiotics (top) and pentaminomycins A-H (bottom) by strain S. cacaoi *CA-170360 in six different media at three different times (7, 14 and 21 days). The average extracted ion chromatogram (EIC) peak area from triplicate culture extracts of the strain CA-170360 is represented.*

The growth rate of *S. cacaoi* CA-170360 was also studied during the 21 days of the OSMAC experiment in the six different media. For this purpose, a 10 mL fermentation tube was used each day to measure the wet weight of the strain in each medium. At T_0 the wet weight was obtained through a fermentation tube with no inoculum. MPG and FR23 media are complex media with a high concentration of non-soluble components that showed a high wet weight at T_0. As shown in Figure D15, CA-170360 starts to grow exponentially until it reaches a maximum peak at a time that depends on the medium used. After this maximum of growth, the biomass usually begins to decrease due to the nutrient consumption or the lysis of bacterial cells, except in the R2YE medium where after the growth maximum at 12 days the strain enters into a stationary phase until the 21st day of incubation (Figure D15).

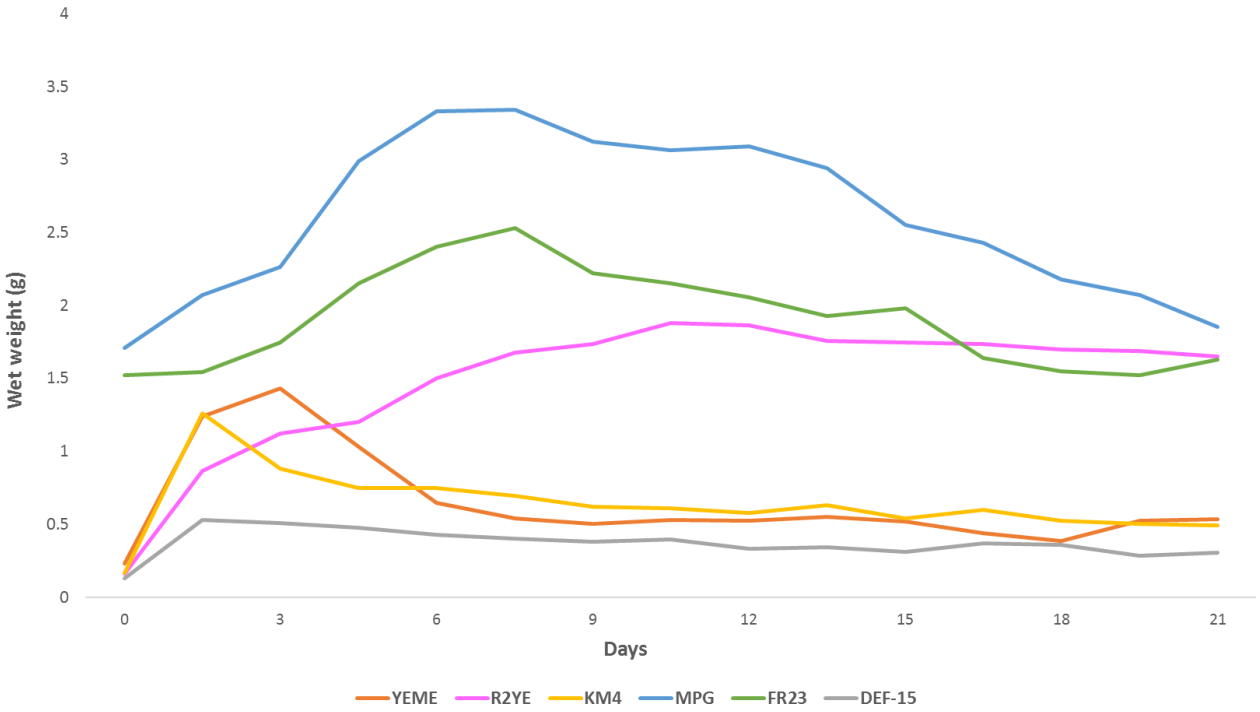


Figure D15. Growth curves of S. cacaoi *CA-170360 for 21 days in the six different media from OSMAC study (YEME in orange, R2YE in pink, KM4 in yellow, MPG in blue, FR23 in green and DEF15 in grey).*

1.4. Metabolomic analysis

Besides the BGCs of cacaoidin, BE-18257 antibiotics and pentaminomycins A-H, the antiSMASH results revealed the presence of other putative BGCs with ≥70 % similarity with known BGCs in the genome of *Streptomyces cacaoi* CA-170360 (Table D3). This meant that the strain had the potential to synthesize a long list of compounds, such as puromycin, bezastatin derivates, actinonin, candicidin, naseseazine C, desferrioxamine E and ectoine (Table D3).

The extracts obtained in the OSMAC study described in section 1.3 were used to perform a metabolomic analysis and check which compounds from the prediction list were being produced by CA-170360. The LC-HRESIMS(+)-TOF analysis detected the presence of puromycin (*m/z* 472.230±0.005) mainly in the R2YE and KM4 media, proferrioxamine D2 (*m/z* 587.340±0.005) in YEME and R2YE media, nocardamine or desferrioxamine E (*m/z* 601.355±0.005) abundantly in YEME, R2YE and, in a lesser extent, in MPG media, terragine E or dehydronocardamine (*m/z* 585.361±0.005) in R2YE medium, actinonin (*m/z* 386.265±0.005) abundantly in MPG and FR23 and, in a lesser extent, in R2YE and KM4 media, ferrioxamine E (*m/z* 654.267±0.005) poorly in DEF-15, and naseseazines A (*m/z* 539.240±0.005) and B (*m/z* 565.256±0.005), which had the same molecular mass than naseseazine C, in all media but in very low amounts (Figure D16).

However, the production of candicidin, bezastatin derivatives and ectoine was not detected, in any of the fermentation conditions tested.

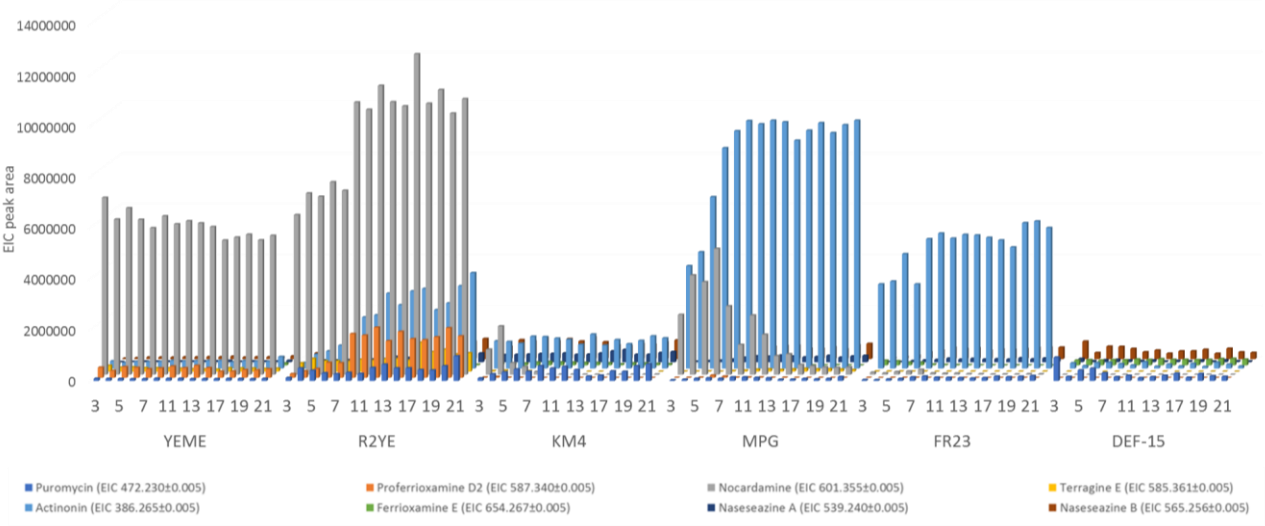


Figure D16. Metabolites produced by S. cacaoi *CA-170360 in the OSMAC study with six different media (YEME, R2YE, KM4, MPG, FR23 and DEF-15) other than cacaoidin, BE-18257 antibiotics and pentaminomycins A-H.*

CHAPTER II. CLONING AND HETEROLOGOUS EXPRESSION OF THE CLUSTER RESPONSIBLE OF THE BIOSYNTHESIS OF PENTAMINOMYCINS A-H AND BE-18257 ANTIBIOTICS

2.1. Cloning and heterologous expression of the *cpp* cluster

The strain *S. cacaoi* CA-170360 is reluctant to genetic manipulation, limiting the obtention of knockout mutants to confirm the involvement of the different gene clusters in the biosynthesis of cacaoidin and the pentaminomycins A-H and BE-18257 antibiotics. For this reason, the cloning and heterologous expression of the *cao* and *cpp* BGCs were crucial to confirm that the genes included in both clusters were sufficient for the biosynthesis of both families of antibiotics and analyze the function of several of these genes by the generation of knockout mutants.

As mentioned before, the *cpp* BGC contained 15 ORFS, with two NRPS genes, each containing five adenylation (A) domains, the first (*cppB*) supposedly involved in the biosynthesis of BE-18257 A–C antibiotics, and the second (*cppM*) in the production of pentaminomycins A-H. The *cpp* cluster also lacks a TE domain to release and cyclize the pentapeptides but contains a PBP-type stand-alone protein (*cpp*A) that may be involved in the release and cyclization of both family of compounds.

To demonstrate that this identified *cpp* cluster is involved in the biosynthesis of both BE-18257 A-C and pentaminomycins A-H, two different fragments of the BGC were separately cloned by Cas9-assisted targeting of chromosome segments (CATCH) cloning (Jiang *et al.,* 2015) into the vector pCAP01 (Yamanaka *et al.,* 2014). For the *in vitro* transcription of the sgRNAs needed for the Cas9-directed cleavage, two universal oligos (sgRNA-F and sgRNA-R) were used along with a specific oligopeptide containing the T7 promoter sequence and the target sequence with a G as the starting nucleotide (Penta1-sgRNA, Penta2-sgRNA and Penta3-sgRNA) (Tables C4 and C5), obtaining a crRNA-tracrRNA chimera. These products were subjected to *in vitro* transcription and were purified as described in Materials and Methods, section 7.7, and the resulting sgRNAs were used to guide the Cas9 endonuclease to cleave the target sequences. Two DNA sequences were obtained with the three specific oligos. The first genome sequence cloned, using the Penta1-sgRNA and Penta2-sgRNA, was a 28.7 Kb fragment containing the PBP-type gene (*cppA*), the first NRPS gene (*cppB*) and the genes present between *cppB* and *cppM* (*cppC-L*) to obtain pCPP1; the second one, obtained with Penta1-sgRNA and Penta3-sgRNA, was a 48 Kb fragment including all the genes supposed to be required for the biosynthesis of both antibiotics; this fragment included the previously described 28.7 Kb fragment and the *cppM* gene, together with two genes encoding hypothetical proteins downstream *cppM* (*cppN-O*), to obtain pCCP2 (Figure D18). Both genome sequences lacked the genes downstream *cppO*, which Hwang *et al* (2020) proposed to be involved in tryptophan biosynthesis. These two resulting DNA fragments were cloned into the pCAP01 plasmid by Gibson assembly. The oligonucleotides pairs pCAP01-Penta1-F and pCAP01-Penta1-R and pCAP01-Penta2-F and pCAP01-Penta2-R were used for the preparation of pCAP01 to clone the, BE-18257 and pentaminomycins A-H BGCs.

The final Gibson assembly products of the CATCH cloning of pCPP1 and pCPP2 were transformed into *E. coli* NEB 10-β electrocompetent cells, which were streaked into kanamycin-

supplemented LB plates. Twenty clones were selected and validated by PCR with the primers Penta1.1-check-F, Penta1-check-R, Penta1.2-check-F, Penta2-check-F and Penta2-check-R, and by restriction digestion with HindIII and NdeI from New England BioLabs (Ipswich, MA, USA). Although the efficiency was not very high, a clone containing the proper construction was obtained for each case (pCCP1 and pCCP2).

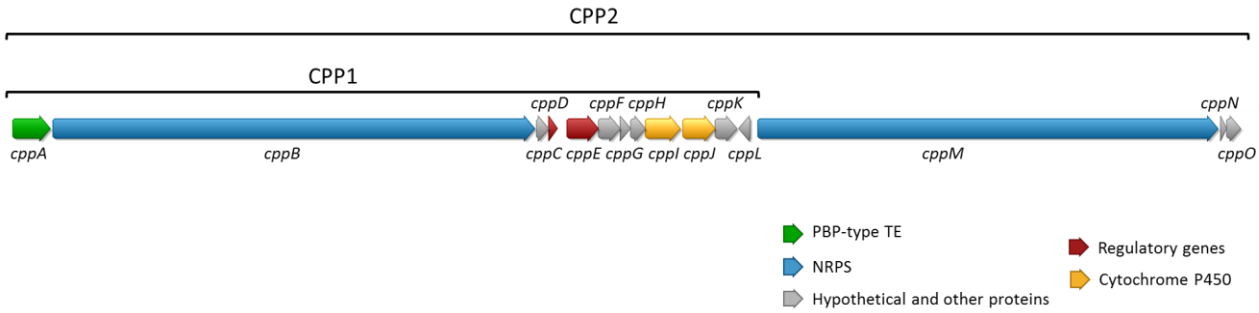


Figure D17. cpp *biosynthetic gene cluster. The two fragments cloned by CATCH into vector pCAP01 are indicated: the 28.7 Kb CPP1 fragment contains the PBP-type TE gene (*cppA*), the first NRPS gene (*cppB*) and the genes present between* cppB *and* cppM *(*cppC-cppL*); the 48 Kb CPP2 fragment includes the above described 28.7 Kb fragment and* cppM, cppN *and* cppO *genes.*

Then, the pCPP1 and pCPP2 plasmids were transformed into *E. coli* NEB 10-β competent cells, and one of the clones harboring pCPP1 and another one harboring pCPP2 were selected to perform intergeneric conjugations. Since pCPP1 and pCPP2 contain the kanamycin-resistant marker, the non-methylating Cm^R Km^R *E. coli* ET12567/pUB307 could not be directly electroporated with the constructions. Thus, two triparental intergeneric conjugations were performed, using *E. coli* NEB 10-β/pCPP1 and ET12567/pUB307 or *E. coli* NEB 10-β/pCPP2 and ET12567/pUB307 as donor strains, and spores of *S. albidoflavus* J1074 as recipient strain. Transconjugants were verified by their resistance to kanamycin and checked by PCR with primers from Table C5 to confirm the integration of the cloned BGCs into the chromosome of *S. albidoflavus* J1074.

Five positive transconjugants from each conjugation, together with the negative control (*S. albidoflavus* J1074/pCAP01 CA-295619) and the wild-type strain *S. cacaoi* CA-170360, were grown in liquid MPG and R2YE media (to favor the detection of BE-18257 antibiotics and pentaminomycins, respectively) during 14 days at 28 °C, and acetone extracts from the cultures whole broths were prepared. After removing the solvent, the residue was resuspended in 20 % DMSO/water and analyzed by LC-HRESIMS-TOF.

The analysis of extracts from pCPP1 and pCPP2 transconjugants confirmed the presence of peaks at 3.46 min and 3.77 min, coincident with the retention time of elution of the three authentic standards BE-18257 A–C isolated from the CA-170360 strain (Figures D18 and D19). The detection levels of the BE-18257 A–C molecules in the pCPP1 transconjugants (which lacked

the pentaminomycins NRPS gene) were much higher than in the pCPP2 transconjugants (which carried in addition the pentaminomycins NRPS gene).

The analysis of the pCPP2 transconjugants also confirmed the presence of peaks coincident with the retention time of elution of the pentaminomycins A, C/H, D and E, isolated from CA-170360, which were absent in the pCPP1 transconjugants (Figures D20-D26). Neither of these peaks were found in the fermentations of the negative control *S. albidoflavus* J1074/pCAP01 CA-295619.

The correlation between the retention time, UV spectrum, exact mass and isotopic distribution between the BE-18257 and pentaminomycins from *S. cacaoi* CA-170360 and the components isolated from the transconjugants *S. albidoflavus*/pCPP1 and *S. albidoflavus*/pCPP2 (Figure D20-D26) unequivocally confirmed that they corresponded to BE-18257 A–C in the case of *S. albidoflavus*/pCPP1 and to BE-18257 A–C and pentaminomycins C/H, D and E in the case of *S. albidoflavus*/pCPP2. In the pCPP2 transconjugants, ions suggesting the presence of pentaminomycins B, F and G were detected, but the low production levels of these compounds impaired the obtention of proper mass spectra (Figure D21, D25 and D26). The detection levels of all the cyclopentapeptides in the heterologous hosts was lower than in the *S. cacaoi* strain, in which the pentaminomycins were already poorly produced. Consequently, the productions of pentaminomycins in the heterologous host *S. albidoflavus*/pCPP2 was still at the detection limit from most of the compounds. These results clearly demonstrate that the first NRPS gene (*cppB*) is the responsible of the biosynthesis of BE-18257 antibiotics, and that the second NRPS gene (*cppM*) synthesizes pentaminomycins. The results also suggest that the cluster-located PBP-type protein is involved in the cyclization of both compounds, and that the *cpp* BGC can be considered an atypical case in which two types of independent compounds are processed by the same enzyme.

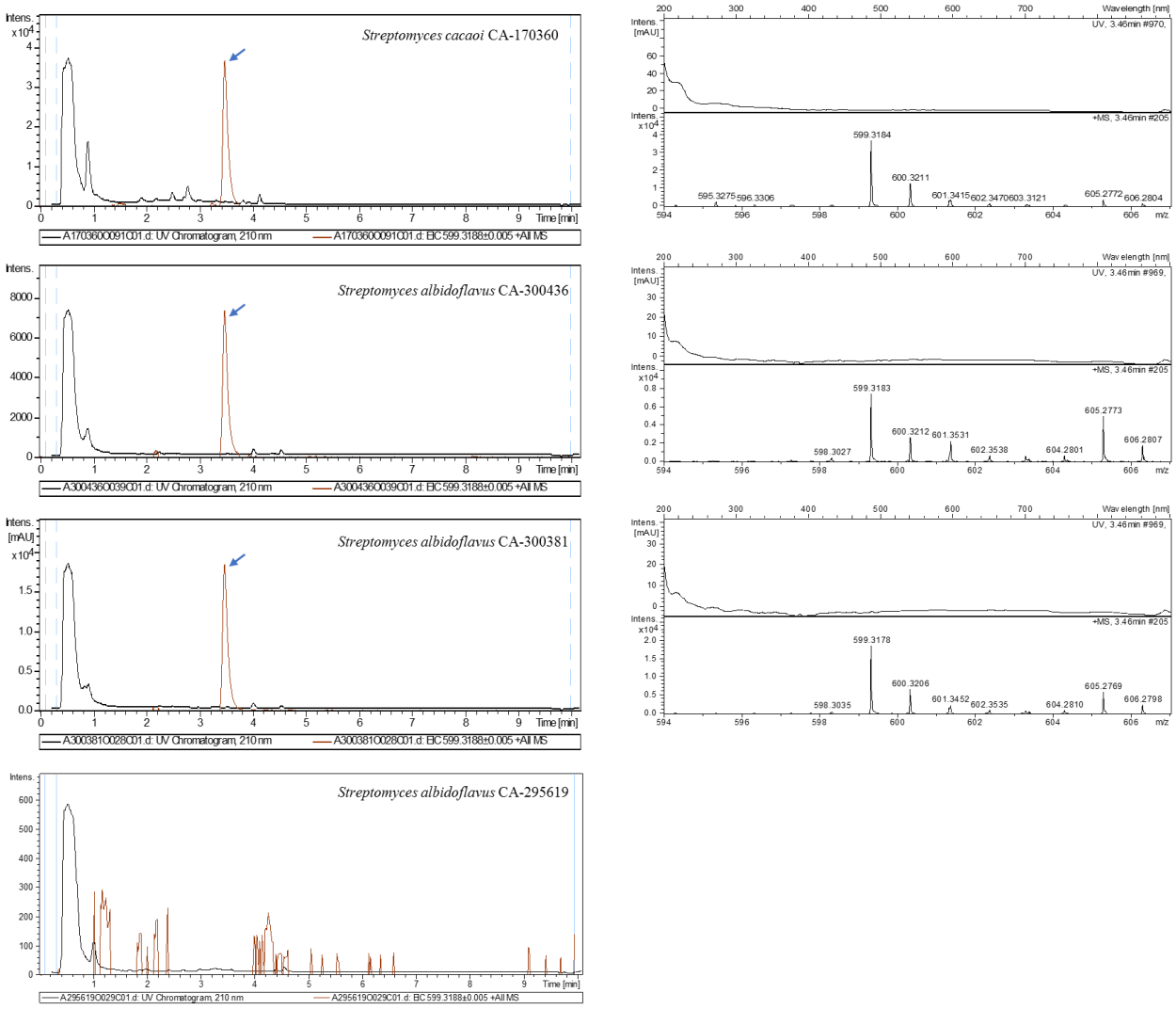


Figure D18. ***BE-18257 A production. Left.*** *Chromatograms of UV absorbance at 210 nm and extracted ion chromatogram at* m/z *= 599.3188 ± 0.005, $C_{30}H_{43}N_6O_7^+$ of BE-18257 A (blue arrows) from original producing strain* Streptomyces cacaoi *CA-170360, the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP2 CA-300436 and* Streptomyces albidoflavus *J1074/pCPP1 CA-300381 and the negative control* Streptomyces albidoflavus *J1074/pCAP01 CA-295619.* ***Right.*** *Experimental UV and positive mass spectra from $C_{30}H_{43}N_6O_7^+$ (calculated value: 599.3188) adduct from original producing strain* Streptomyces cacaoi *CA-170360 (top) and the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP2 (middle) and* Streptomyces albidoflavus *J1074/pCPP1 (bottom). No UV or mass spectra was obtained in the negative control.*

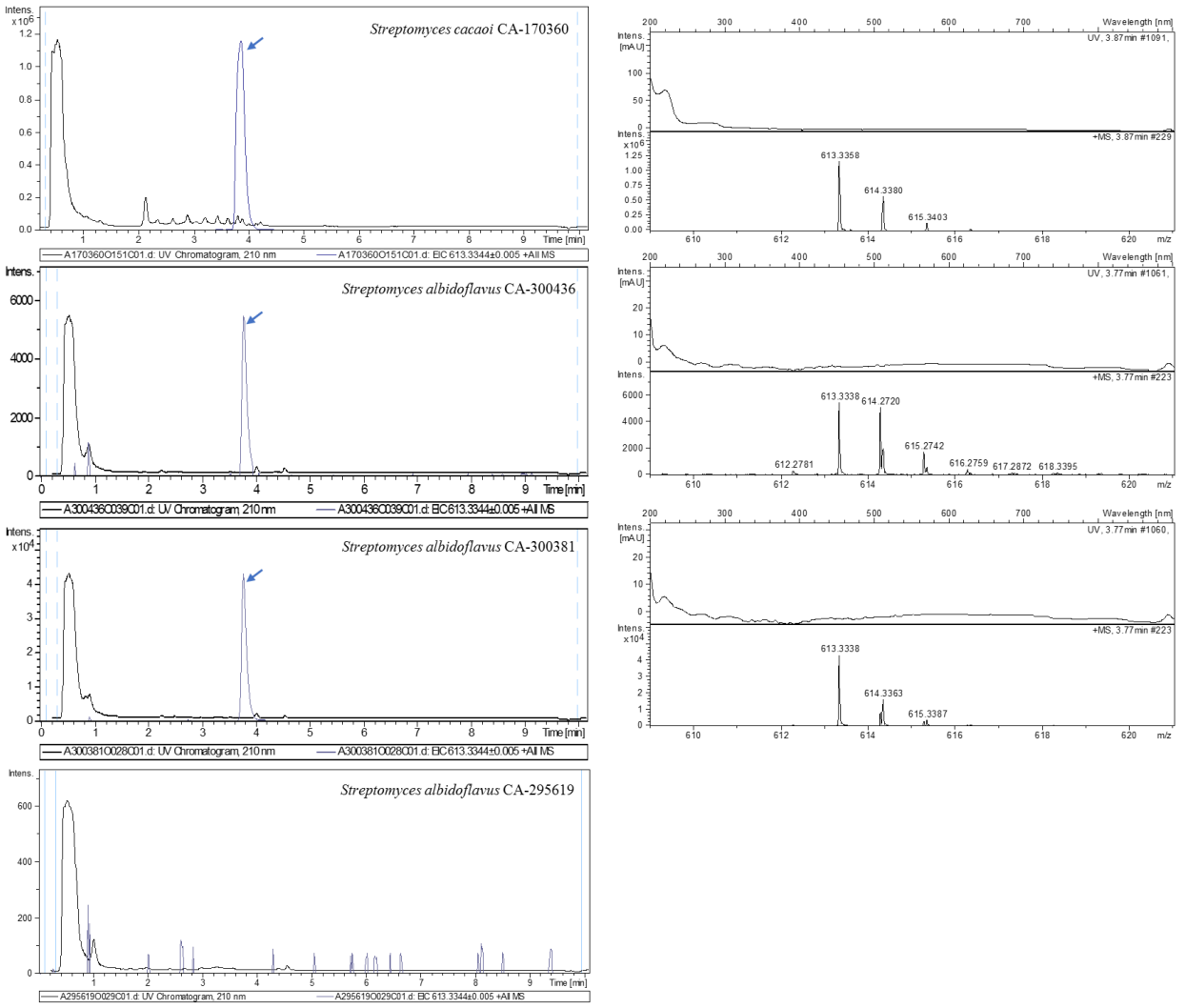


Figure D19. ***BE-18257 B/C production. Left.*** *Chromatograms of UV absorbance at 210 nm and extracted ion chromatogram at* m/z *= 613.3344 ± 0.005, $C_{31}H_{45}N_6O_7^+$ of BE-18257 B/C (blue arrows) from original producing strain* Streptomyces cacaoi *CA-170360, the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP2 CA-300436 and* Streptomyces albidoflavus *J1074/pCPP1 CA-300381 and the negative control* Streptomyces albidoflavus *J1074/pCAP01 CA-295619.* ***Right.*** *Experimental UV and positive mass spectra from $C_{31}H_{45}N_6O_7^+$ (calculated value: 613.3344) adduct from original producing strain* Streptomyces cacaoi *CA-170360 (top) and the heterologous producing strains* Streptomyces *albidoflavus J1074/pCPP2 (middle) and* Streptomyces *albidoflavus J1074/pCPP1 (bottom). No UV or mass spectra was obtained in the negative control.*

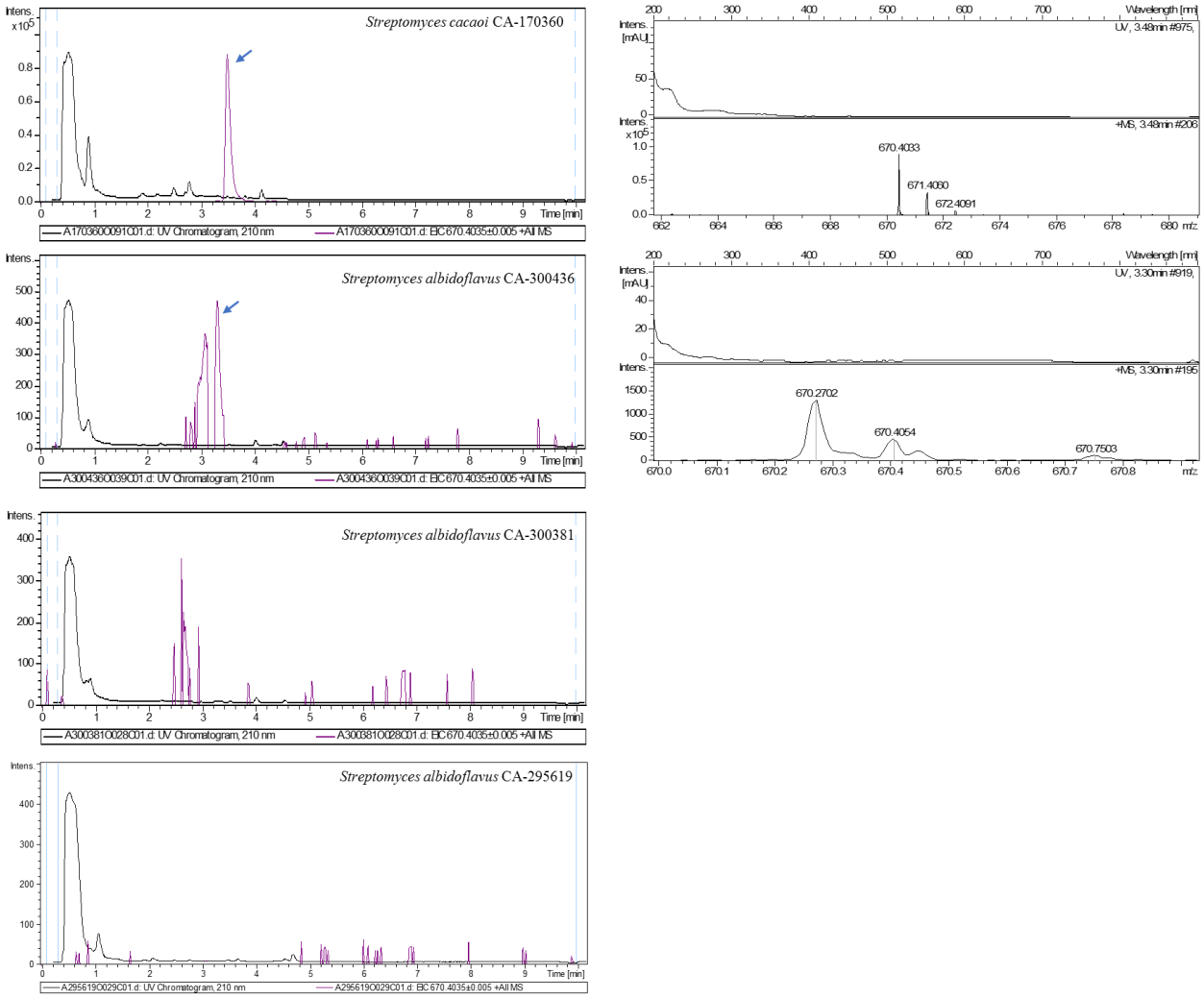


Figure D20. ***Pentaminomycin A production. Left.*** *Chromatograms of UV absorbance at 210 nm and extracted ion* m/z *= 670.4035 ± 0.005,* $C_{33}H_{52}N_9O_6^+$ *of pentaminomycin A (blue arrows) from original producing strain* Streptomyces cacaoi *CA-170360, the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP2 CA-300436 and* Streptomyces albidoflavus *J1074/pCPP1 CA-300381 and the negative control* Streptomyces albidoflavus *J1074/pCAP01 CA-295619.* ***Right.*** *Experimental UV and positive mass spectra from* $C_{33}H_{52}N_9O_6^+$ *(calculated value: 670.4035) adduct from original producing strain* Streptomyces cacaoi *CA-170360 (top) and the heterologous producing strain* Streptomyces albidoflavus *J1074/pCPP2 (middle). No UV or mass spectra was obtained in the heterologous producing strain* Streptomyces albidoflavus *J1074/pCPP1, as it did not carry the NRPS gene required for the production of pentaminomycins, nor in the negative control.*

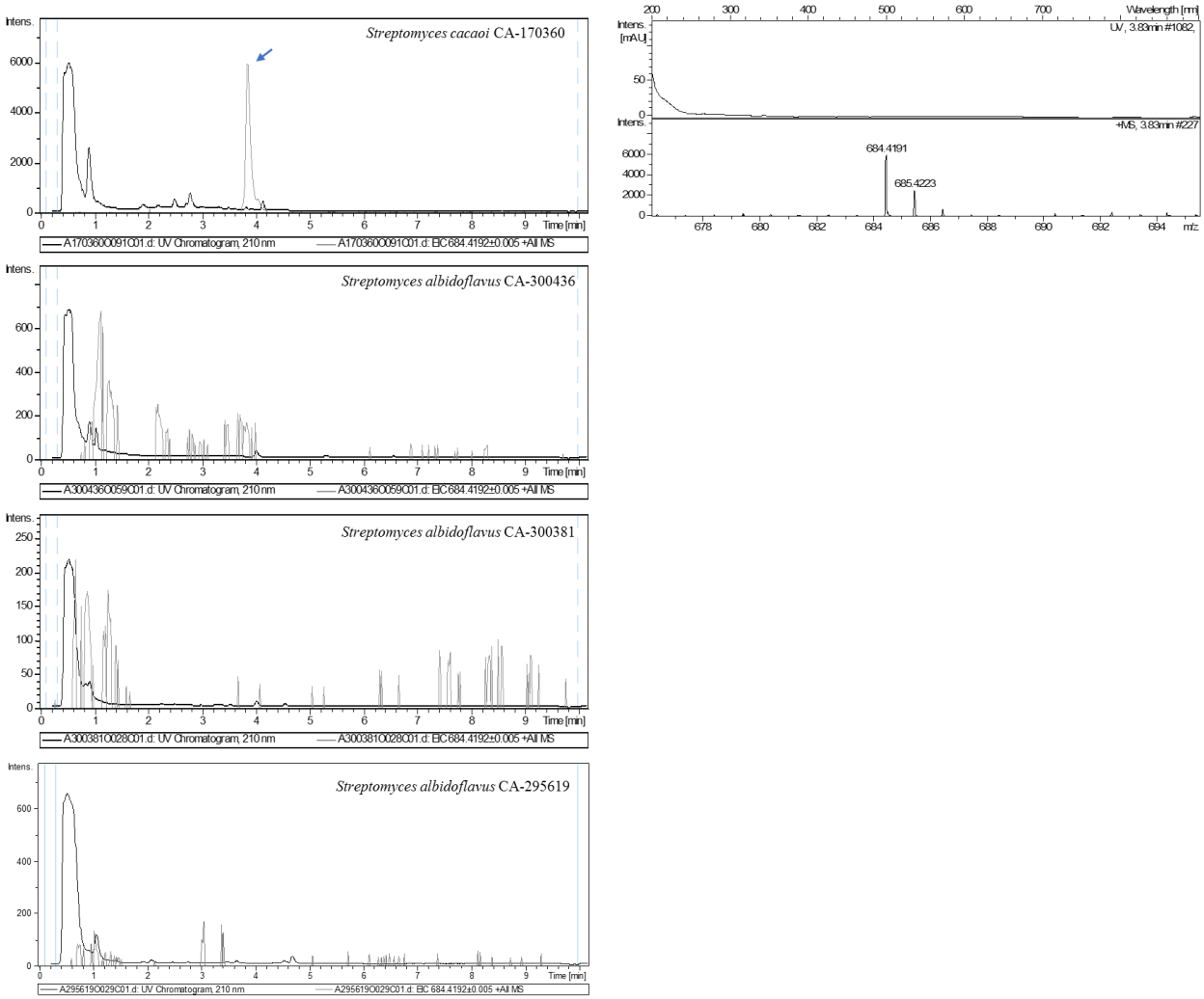


Figure D21. ***Pentaminomycin B production. Left.*** *Chromatograms of UV absorbance at 210 nm and extracted ion chromatogram at* m/z *= 684.4192 ± 0.005, $C_{34}H_{54}N_9O_6^+$ of pentaminomycin B (blue arrows) from original producing strain* Streptomyces cacaoi *CA-170360, the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP2 CA-300436 and* Streptomyces albidoflavus *J1074/pCPP1 CA-300381 and the negative control* Streptomyces albidoflavus *J1074/pCAP01 CA-295619.* ***Right.*** *Experimental UV and positive mass spectra from $C_{34}H_{54}N_9O_6^+$ (calculated value: 684.4192) adduct from original producing strain* Streptomyces cacaoi *CA-170360. No UV or mass spectra was obtained in the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP1, as it did not carry the NRPS gene required for the production of pentaminomycins, nor in the negative control. However, no UV or mass spectra was obtained in the heterologous producing strain* Streptomyces albidoflavus *J1074/pCPP2 either.*

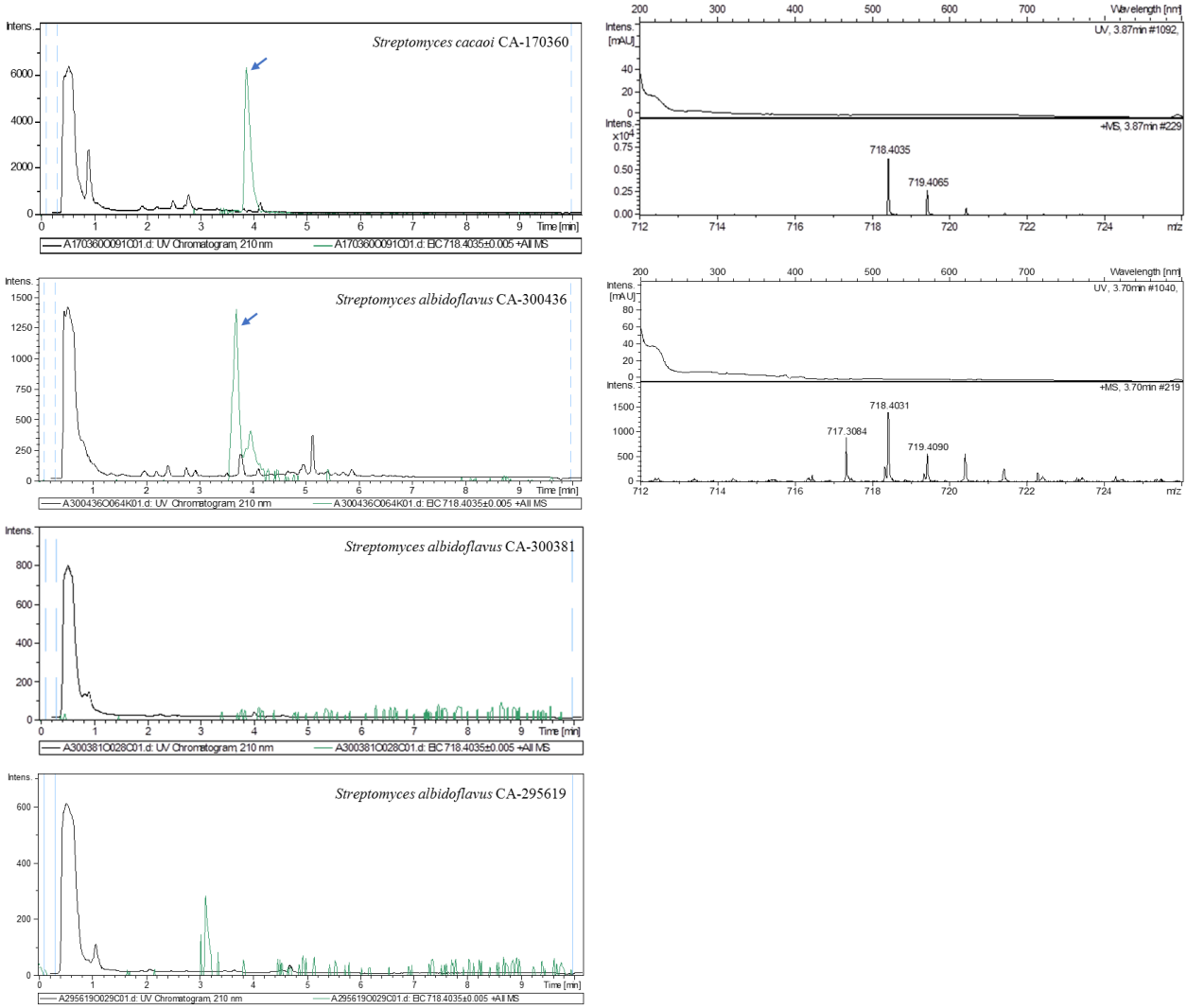


Figure D22. ***Pentaminomycin C/H production. Left.*** *Chromatograms of UV absorbance at 210 nm and extracted ion chromatogram at* m/z *= 718.4035 ± 0.005, $C_{34}H_{54}N_9O_6^+$ of pentaminomycin C/H (blue arrows) from original producing strain* Streptomyces cacaoi *CA-170360, the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP2 CA-300436 and* Streptomyces albidoflavus *J1074/pCPP1 CA-300381 and the negative control* Streptomyces albidoflavus *J1074/pCAP01 CA-295619.* ***Right.*** *Experimental UV and positive mass spectra from $C_{34}H_{54}N_9O_6^+$ (calculated value: 718.4035) adduct from original producing strain* Streptomyces cacaoi *CA-170360 (top) and the heterologous producing strain* Streptomyces albidoflavus *J1074/pCPP2 (middle). No UV or mass spectra was obtained in the heterologous producing strain* Streptomyces albidoflavus *J1074/pCPP1, as it did not carry the NRPS gene required for the production of pentaminomycins, nor in the negative control.*

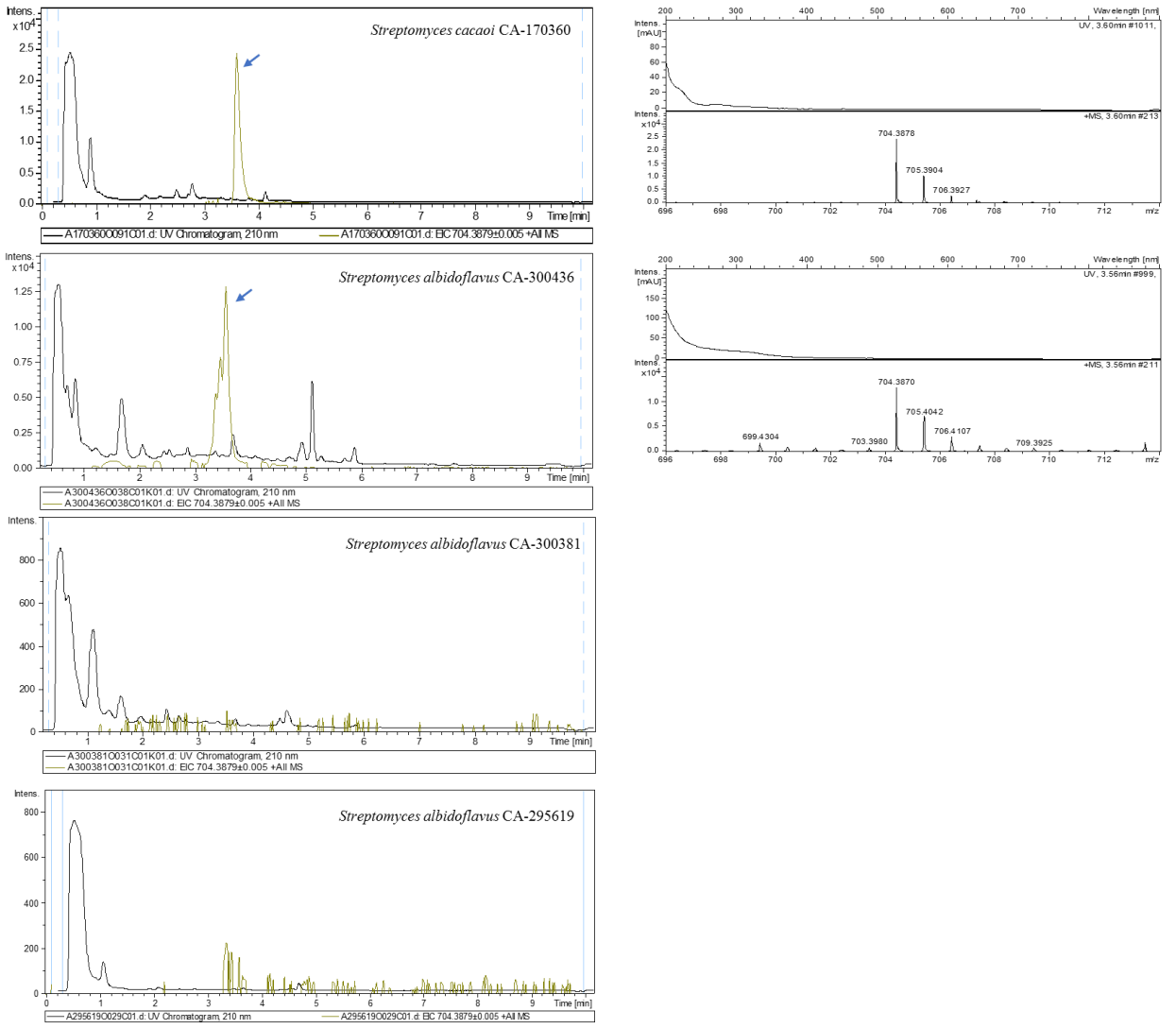


Figure D23. ***Pentaminomycin D production. Left.*** *Chromatograms of UV absorbance at 210 nm and extracted ion chromatogram at* m/*z = 704.3879 ± 0.005, $C_{36}H_{50}N_9O_6^+$ of pentaminomycin D (blue arrow) from original producing strain* Streptomyces cacaoi *CA-170360, the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP2 CA-300436 and* Streptomyces albidoflavus *J1074/pCPP1 CA-300381 and the negative control* Streptomyces albidoflavus *J1074/pCAP01 CA-295619.* ***Right.*** *Experimental UV and positive mass spectra from $C_{36}H_{50}N_9O_6^+$ (calculated value: 704.3879) adduct from original producing strain* Streptomyces cacaoi *CA-170360 and the heterologous producing strain* Streptomyces albidoflavus *J1074/pCPP2. No UV or mass spectra was obtained in the heterologous producing strain* Streptomyces albidoflavus *J1074/pCPP1, as it did not carry the NRPS gene required for the production of pentaminomycins, nor in the negative control.*

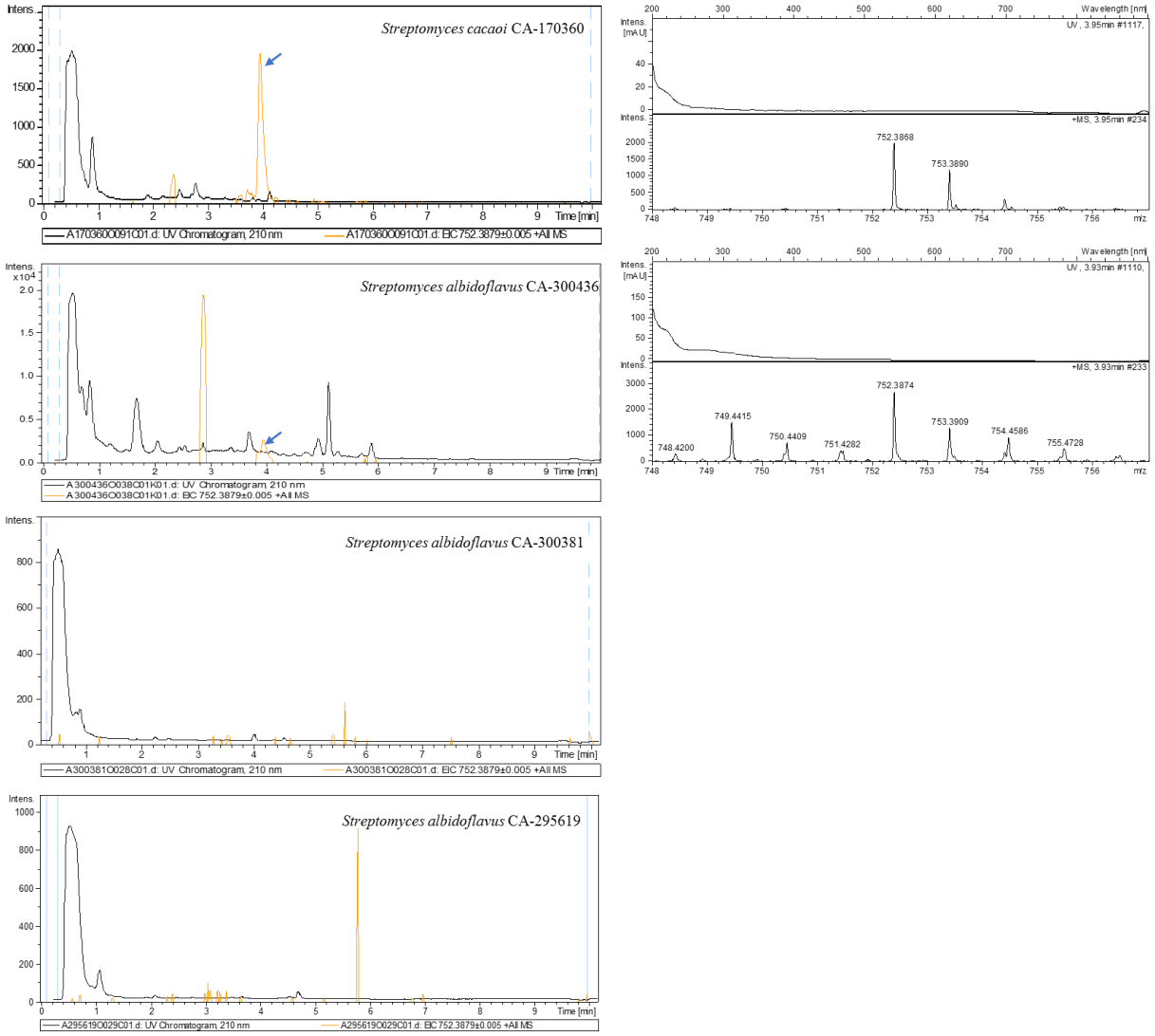


Figure D24. ***Pentaminomycin E production. Left.*** *Chromatograms of UV absorbance at 210 nm and extracted ion chromatogram at* m/z *= 752.3879 ± 0.005, $C_{40}H_{50}N_9O_6^+$ of pentaminomycin E (blue arrows) from original producing strain* Streptomyces cacaoi *CA-170360, the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP2 CA-300436 and* Streptomyces albidoflavus *J1074/pCPP1 CA-300381 and the negative control* Streptomyces albidoflavus *J1074/pCAP01 CA-295619.* ***Right.*** *Experimental UV and positive mass spectra from $C_{40}H_{50}N_9O_6^+$ (calculated value: 752.3879) adduct from original producing strain* Streptomyces cacaoi *CA-170360 (top) and the heterologous producing strain* Streptomyces albidoflavus *J1074/pCPP2 (middle). No UV or mass spectra was obtained in the heterologous producing strain* Streptomyces albidoflavus *J1074/pCPP1, as it did not carry the NRPS gene required for the production of pentaminomycins, nor in the negative control.*

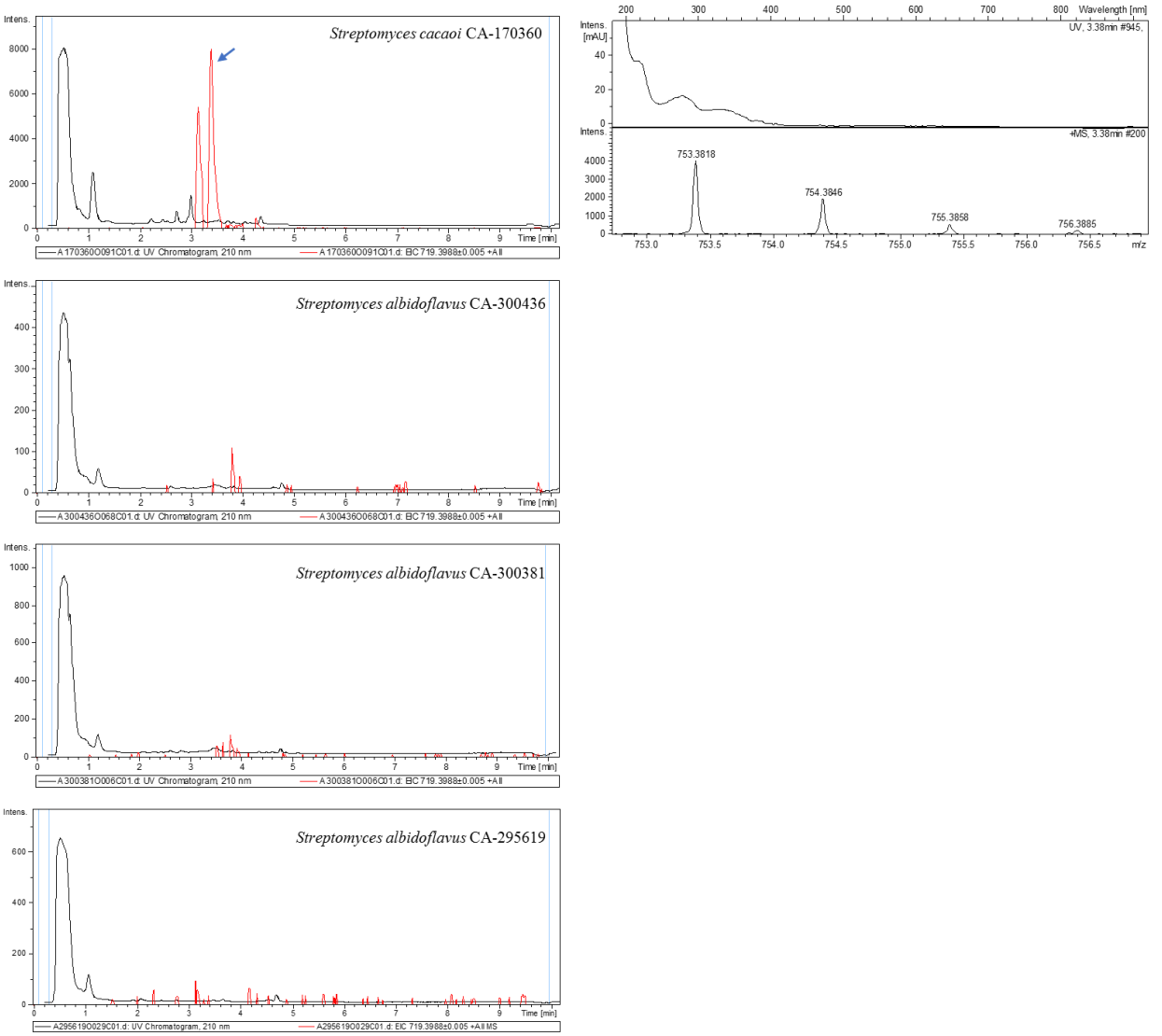


Figure D25. ***Pentaminomycin F production. Left.*** *Chromatograms of UV absorbance at 210 nm and extracted ion chromatogram at* m/z *= 719.3988 ± 0.005, $C_{36}H_{50}N_{10}O_6^+$ of pentaminomycin G (blue arrow) from original producing strain* Streptomyces cacaoi *CA-170360, the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP2 CA-300436 and* Streptomyces albidoflavus *J1074/pCPP1 CA-300381 and the negative control* Streptomyces albidoflavus *J1074/pCAP01 CA-295619.* ***Right.*** *Experimental UV and positive mass spectra from $C_{36}H_{50}N_{10}O_6^+$ (calculated value: 719.3988) adduct from original producing strain* Streptomyces cacaoi *CA-170360. No UV or mass spectra was obtained in the heterologous producing strain* Streptomyces albidoflavus *J1074/pCPP1, as it did not carry the NRPS gene required for the production of pentaminomycins, nor in the negative control. However, no UV or mass spectra was obtained with the heterologous producing strain* Streptomyces albidoflavus *J1074/pCPP2 either.*

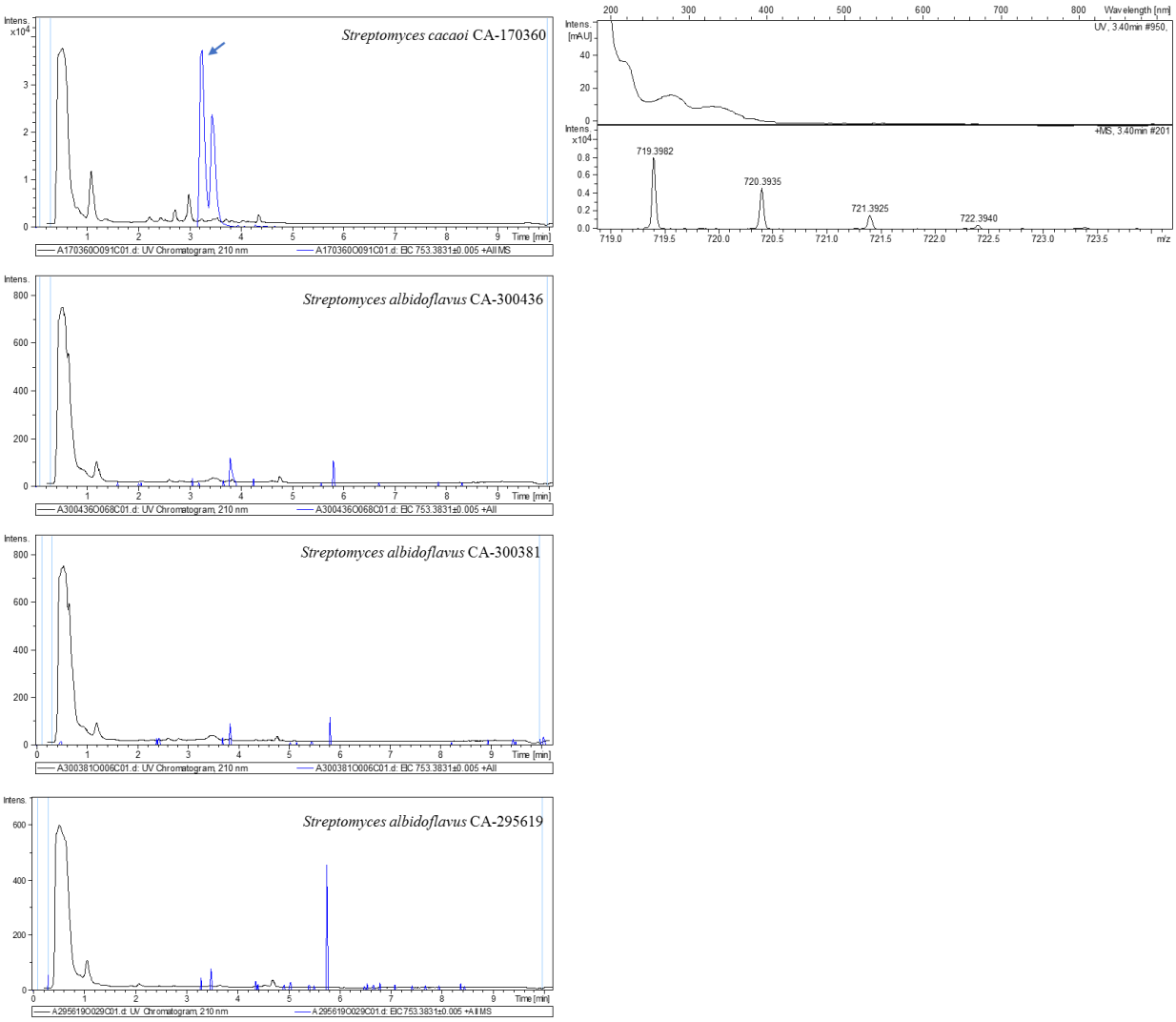


Figure D26. ***Pentaminomycin G production. Left.*** *Chromatograms of UV absorbance at 210 nm and extracted ion chromatogram at* m/z *= 753.3831 ± 0.005,* $C_{39}H_{48}N_{10}O_6^+$ *of pentaminomycin G (blue arrow) from original producing strain* Streptomyces cacaoi *CA-170360, the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP2 CA-300436 and* Streptomyces albidoflavus *J1074/pCPP1 CA-300381 and the negative control* Streptomyces albidoflavus *J1074/pCAP01 CA-295619.* ***Right.*** *Experimental UV and positive mass spectra from* $C_{39}H_{48}N_{10}O_6^+$ *(calculated value: 753.3831) adduct from original producing strain* Streptomyces cacaoi *CA-170360. No UV or mass spectra was obtained in the heterologous producing strain* Streptomyces albidoflavus *J1074/pCPP1, as it did not carry the NRPS gene required for the production of pentaminomycins, nor in the negative control. However, no UV or mass spectra was obtained with the heterologous producing strain* Streptomyces albidoflavus *J1074/pCPP2 either.*

2.2. OSMAC study in the pCPP2 heterologous host

An OSMAC approach was carried out in order to identify the best production conditions of the family of pentaminomycins A-H and BE-18257 antibiotics in the new heterologous host *S. albidoflavus* J1074/pCPP2 (CA-300436). The analysis included a total of four production media (MPG, KM4, R2YE and FR23) for 14 days. This OSMAC study brought to light the low production levels of these compounds compared to the levels in the *S. cacaoi* strain, especially in the case of pentaminomycins. In the heterologous host 300436, while the highest production of BE-18257 antibiotics took place in the FR23 and MPG media, pentaminomycins were found to be produced the most in R2YE medium, especially pentaminomycins D and E (Figure D27). Although the production levels cannot be comparable, the production levels of BE-18257 antibiotics in the *S. cacaoi* CA-170360 were also higher in MPG and FR23 media, although the production is very high in KM4 medium as well, and lower in R2YE. However, as happens in the heterologous host, all the pentaminomycins A-H were better produced in R2YE medium (Figure D27).

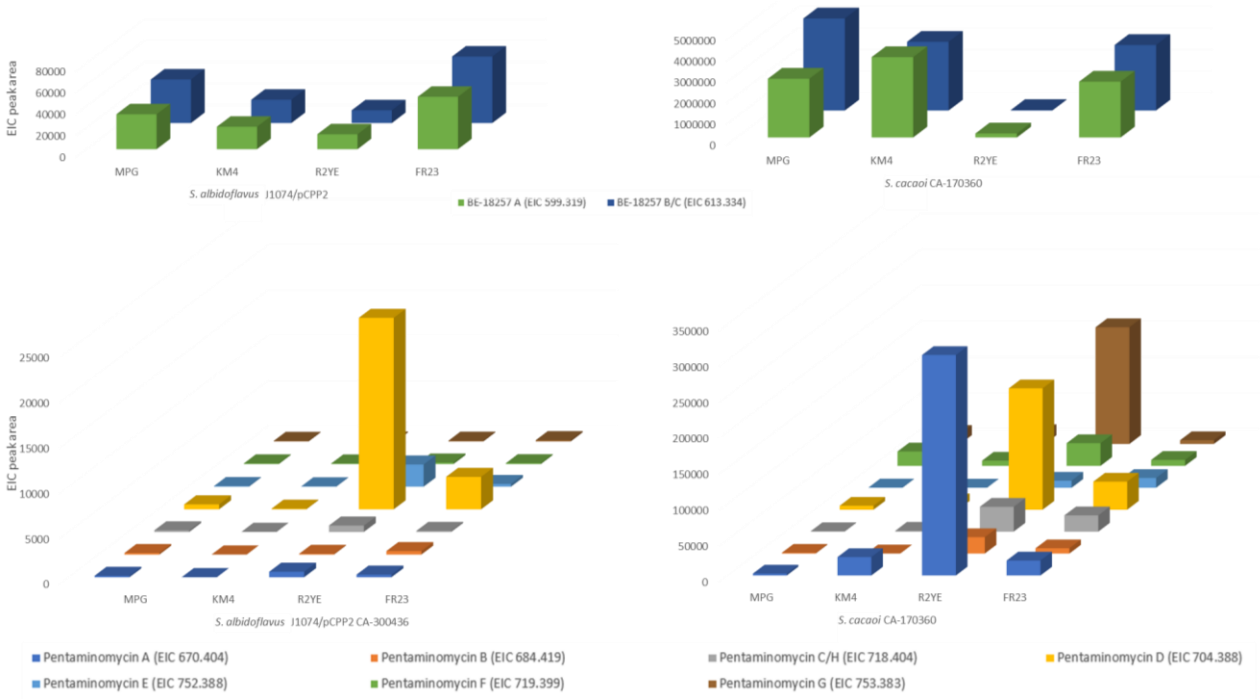


Figure D27. Comparison of the production of BE-18257 antibiotics (top) and pentaminomycins A-H (bottom) by the strains S. albidoflavus *J1074/pCPP2 CA-300436 (left) and* S. cacaoi *CA-170360 (right) in four different media at 14 days. The average extracted ion chromatogram (EIC) peak area from triplicate culture extracts is represented.*

A time course of the strain CA-300436 was performed in R2YE and MPG media for 14 days. The production of BE-18257 antibiotics remained fairly constant the whole time, being higher in MPG medium, while for the production of pentaminomycins, the highest production was achieved at 13 days in R2YE (Figure D28), especially for pentaminomycins D, C/H and E. The production of pentaminomycins in MPG medium was much lower.

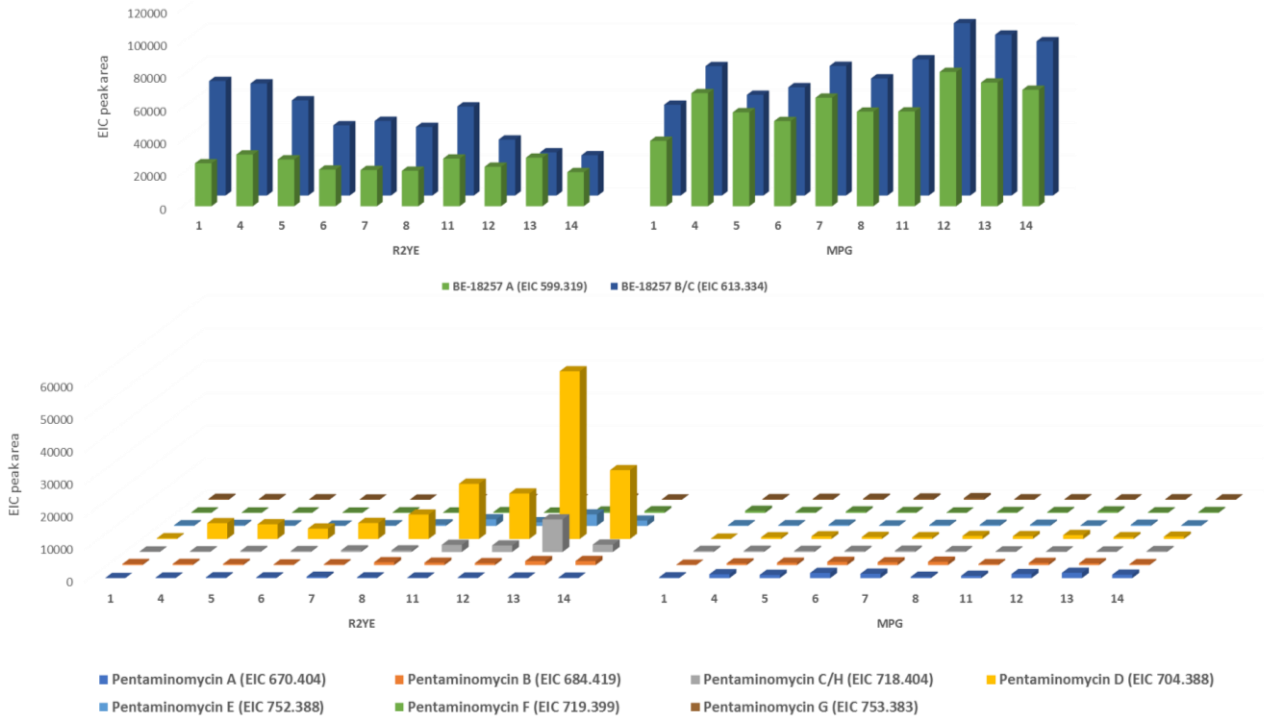


Figure D28. Time course of BE-18257 antibiotics (top) and pentaminomycins A-H (bottom) production by the strain S. albidoflavus *J1074/pCPP2 CA-300436 in R2YE and MPG media for 14 days.*

2.3. Generation of *cppA* knockout by genetic replacement

As described above, the two NRPS genes of the *cpp* BGC, *cppB* and *cppM*, have been demonstrated to be responsible of the biosynthesis of BE-18257 antibiotics and pentaminomycins A-H, respectively. However, this cluster lacks a TE domain to release and cyclize the pentapeptides and instead contains a PBP-type stand-alone protein (*cpp*A) that might be involved in the release and cyclization of the peptide chains of both type of compounds.

In order to prove the involvement of *cppA* in the release and cyclization of the cyclic pentapeptides, a genetic replacement approach was designed as described before using the primers in Table C6 to finally generate the pEFBA-cppA construction (Figure D29). This vector was transferred to *S. albidoflavus* J1074/pCPP2 CA-300436 via conjugation with ET12567/pUB307 *E. coli* cells. Ten transconjugants were grown in liquid ATCC-2 medium and submitted to resistance analysis, streaking in MA plates containing apramycin and MA plates containing thiostrepton. Those transconjugant which were resistant to apramycin and sensitive to thiostrepton ($Am^R/Thios^S$) were finally checked by PCR using the internal pair of primers CppA-check-F and CppA-check-R amplifying a 356 bp DNA fragment (Table C9) of the deleted gene and the external and internal primers pEFBA-cppA-Frag2-F and Apramycin-F amplifying a 3.4 Kb fragment (Table C6 and C7, respectively) (Figure D29). As positive PCR control, the pEFBA-cppA construction was used. A positive transconjugant, CA-301651 (*S. albidoflavus* CA-300436, *ΔcppA::aac(3)IV*), together with the negative control (*S. albidoflavus* J1074/pCAP01 CA-295619) and the CA-300436 strain, were grown in 10 mL liquid R2YE medium tubes for 13 days.

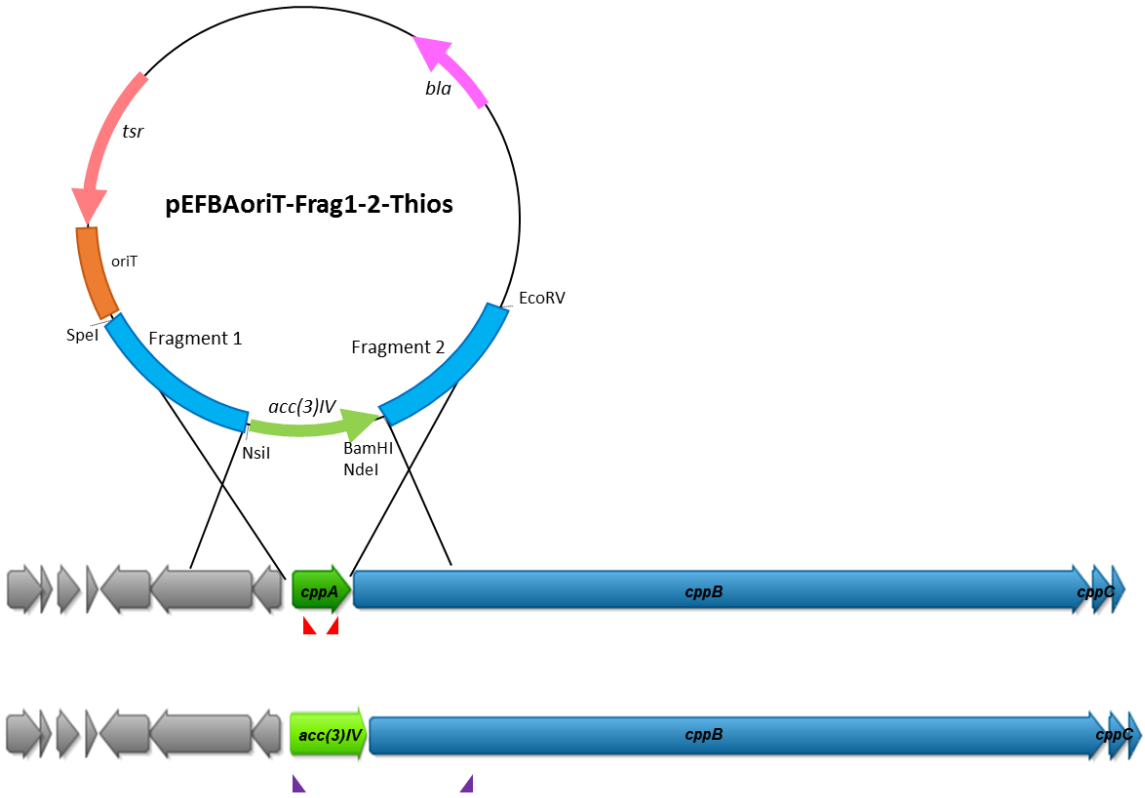


Figure D29. Scheme of the genetic replacement of cppA *with the apramycin resistance cassette in the* cpp *cluster. In the figure, the internal pair of primers CppA-check-F and CppA-check-R for the checking of the deletion of the gene are shown in red and the external and internal primers pEFBA-cppA-Frag2-F and Apramycin-F for the checking of the insertion of the apramycin cassette are shown in purple.*

Then, the extraction of the compounds was performed as described in Materials and Methods and they were submitted to LC-HRESIMS(+)-TOF to study the production of the cyclic pentapeptides. No production of BE-18257 antibiotics or pentaminomycins A-H, neither their cyclic nor their lineal structures, were detected in the *cppA* knockout strain (Figures D30-D35). These results confirmed that, as it was proposed before (Kaweewan *et al.,* 2020; Hwang *et al.,* 2020), CppA is a PBP-type TE involved in the release and cyclization of the cyclic pentapeptides.

Finally, in order to verify that CppA was the responsible of the release of both pentaminomycins A-H and BE-18257 antibiotics, an overexpression of the previously replaced gene *cppA* was performed to study if the production of the cyclic compounds could be restored. First, *cppA* was amplified using the CppA-F and CppA-R pair of primers (Table C8), purified and cloned into the pEM4T conjugative vector. The final pEM4T-cppA construction was transferred to the *cppA* knockout *S. albidoflavus* CA-301651 via conjugation with ET12567/pUB307 *E. coli* cells. Five thiostrepton-resistant transconjugants were grown in 10 mL liquid R2YE medium for 13 days and the extracts generated were analyzed by LC-HRESIMS(+)-TOF. The only compounds whose production was restored were pentaminomycins D and C/H (Figures D33 and D34). No production of the rest of pentaminomycins nor the BE-18257 antibiotics could be restored (Figure D30, D31, D32 and D35).

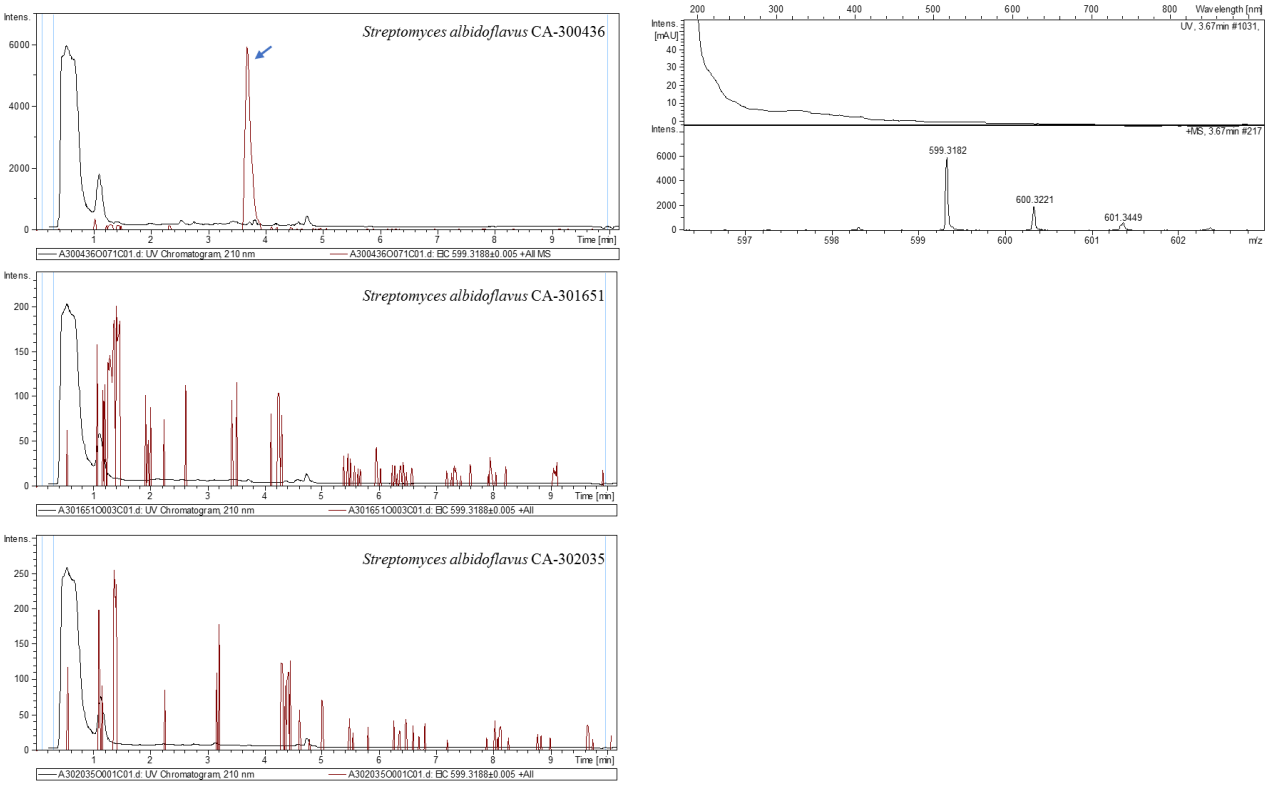


Figure D30. ***BE-18257 A production in the knockout and complemented strains. Left.*** *Chromatograms of UV absorbance at 210 nm and extracted ion chromatogram at* m/z *= 599.3188 ± 0.005,* $C_{30}H_{43}N_6O_7^+$ *of BE18257 A (blue arrow) from the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP2 CA-300436 (top), the* cppA*-knockout* Streptomyces albidoflavus *CA-301651 (middle) and the complemented pEM4T-cppA* Streptomyces albidoflavus *CA-302035 (bottom).* ***Right.*** *Experimental UV and positive mass spectra from* $C_{30}H_{43}N_6O_7^+$ *(calculated value: 599.3188) adduct from the heterologous producing strains* Streptomyces albidoflavus *CA-300436. No UV or mass spectra was obtained with the* cppA*-knockout* S. albidoflavus *CA-301651 as it was expected. However, no UV or mass spectra was obtained with the complemented pEM4T-cppA CA-302035 strain either, thus the production of BE18257 A could not be restored.*

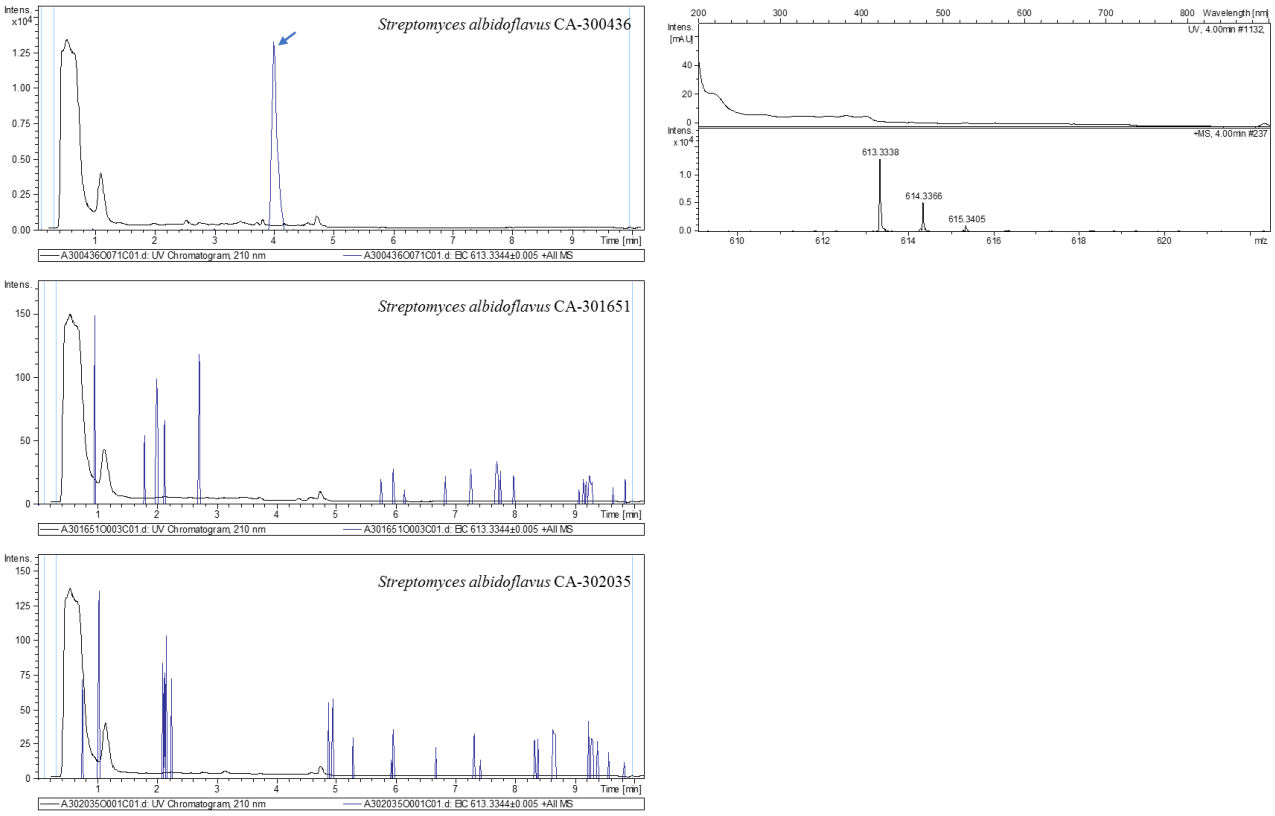


Figure D31. ***BE-18257 B/C production in the knockout and complemented strains. Left.*** *Chromatograms of UV absorbance at 210 nm and extracted ion chromatogram at* m/z *= 613.3344 ± 0.005, $C_{31}H_{45}N_6O_7^+$ of BE18257 B/C (blue arrow) from the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP2 CA-300436 (top), the* cppA*-knockout* Streptomyces albidoflavus *CA-301651 (middle) and the complemented pEM4T-cppA* Streptomyces albidoflavus *CA-302035 (bottom).* ***Right.*** *Experimental UV and positive mass spectra from $C_{31}H_{45}N_6O_7^+$ (calculated value: 613.3344) adduct from the heterologous producing strains* Streptomyces albidoflavus *CA-300436. No UV or mass spectra was obtained with the* cppA*-knockout* S. albidoflavus *CA-301651 as it was expected. However, no UV or mass spectra was obtained with the complemented pEM4T-cppA CA-302035 strain either, thus the production of BE18257 B/C could not be restored.*

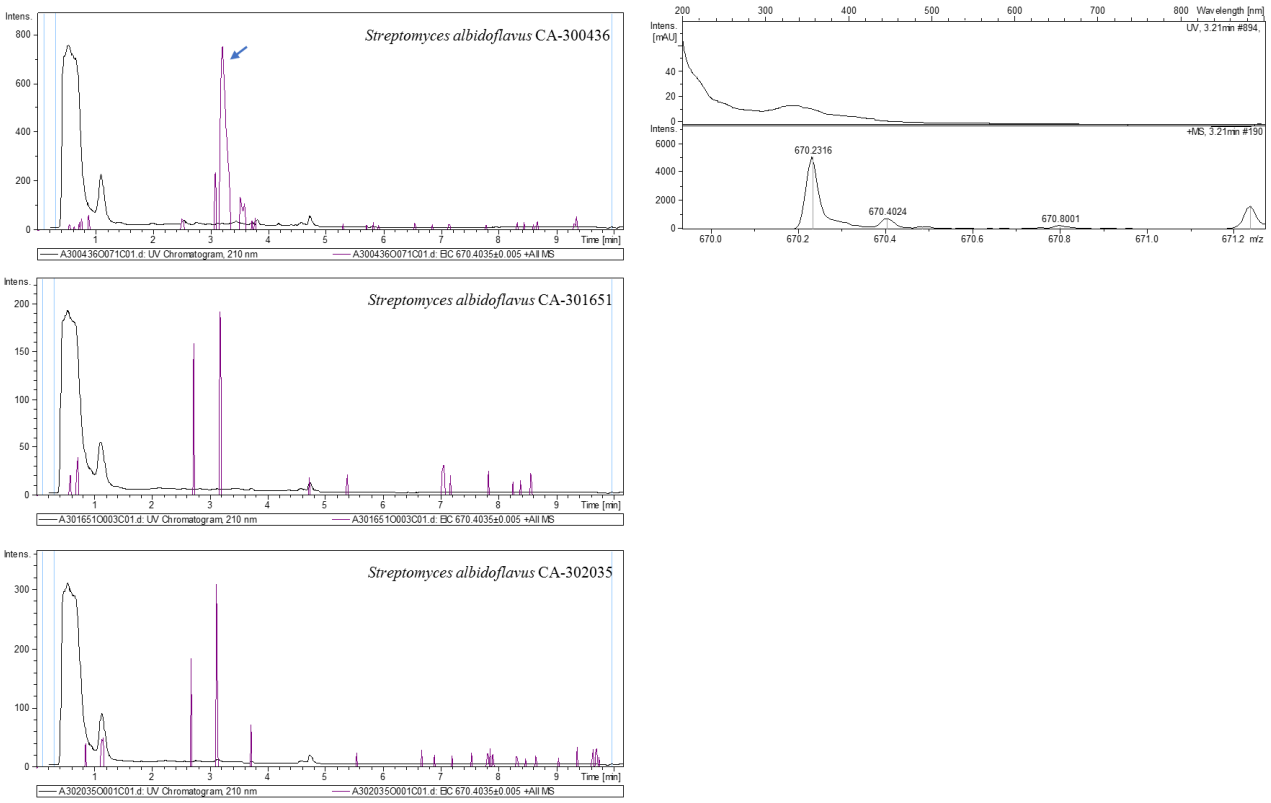


Figure D32. ***Pentaminomycin A production in the knockout and complemented strains. Left.*** *Chromatograms of UV absorbance at 210 nm and extracted ion chromatogram at* m/z *= 670.4035 ± 0.005, $C_{33}H_{52}N_9O_6^+$ of pentaminomycin A (blue arrow) from the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP2 CA-300436 (top), the* cppA*-knockout* Streptomyces albidoflavus *CA-301651 (middle) and the complemented pEM4T-cppA* Streptomyces albidoflavus *CA-302035 (bottom).* ***Right.*** *Experimental UV and positive mass spectra from $C_{33}H_{52}N_9O_6^+$ (calculated value: 670.4035) adduct from the heterologous producing strains* Streptomyces albidoflavus *CA-300436. No UV or mass spectra was obtained with the* cppA*-knockout* S. albidoflavus *CA-301651 as it was expected. However, no UV or mass spectra was obtained with the complemented pEM4T-cppA CA-302035 strain either, thus the production of pentaminomycin A could not be restored.*

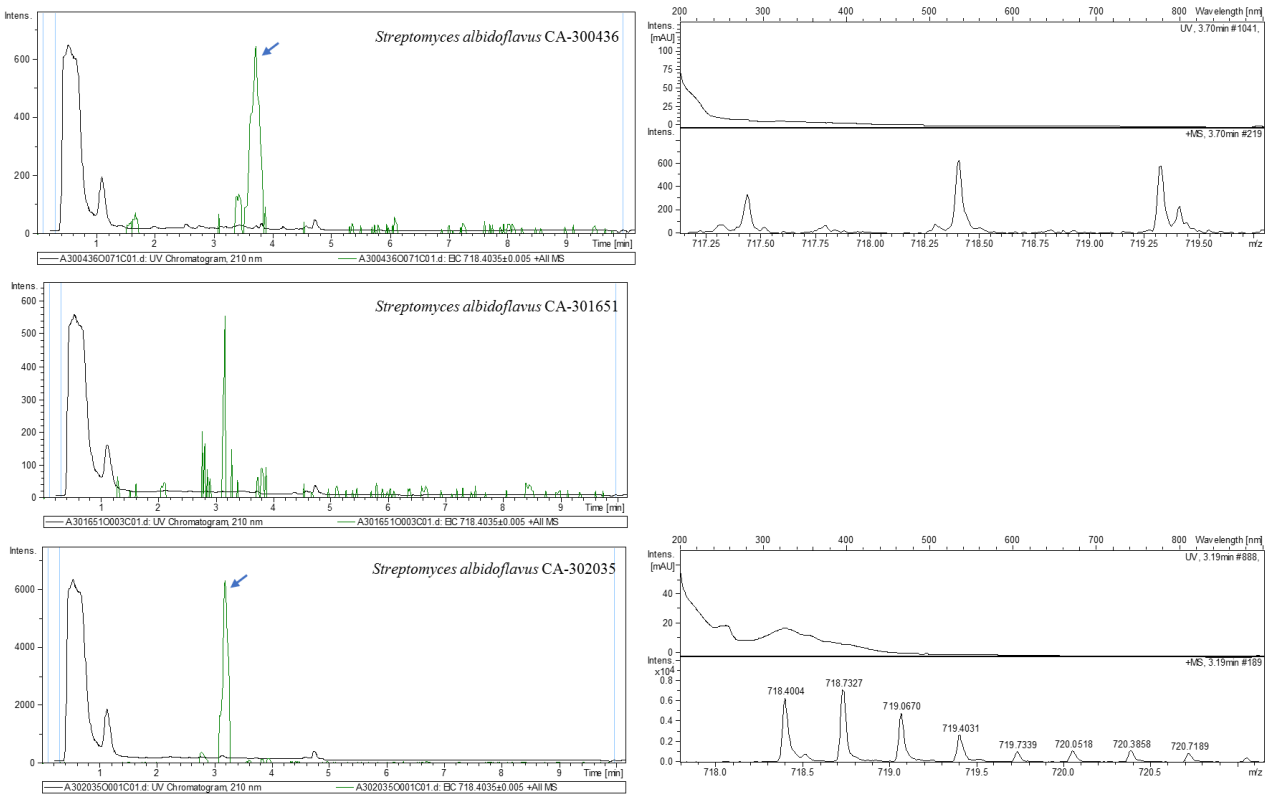


Figure D33. ***Pentaminomycin C/H production in the knockout and complemented strains. Left.*** *Chromatograms of UV absorbance at 210 nm and extracted ion chromatogram at* m/z *= 684.4192 ± 0.005, $C_{34}H_{54}N_9O_6^+$ of pentaminomycin C/H (blue arrow) from the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP2 CA-300436 (top), the* cppA*-knockout* Streptomyces albidoflavus *CA-301651 (middle) and the complemented pEM4T-cppA* Streptomyces albidoflavus *CA-302035 (bottom).* ***Right.*** *Experimental UV and positive mass spectra from $C_{34}H_{54}N_9O_6^+$ (calculated value: 684.4192) adduct from the heterologous producing strains* Streptomyces albidoflavus *CA-300436. No UV or mass spectra was obtained with the* cppA*-knockout* S. albidoflavus *CA-301651 as it was expected., Experimental UV and positive mass spectra from $C_{34}H_{54}N_9O_6^+$ (calculated value: 684.4192) adduct from the complemented pEM4T-cppA CA-302035 strain was obtained, thus the production of pentaminomycin C/H could be restored.*

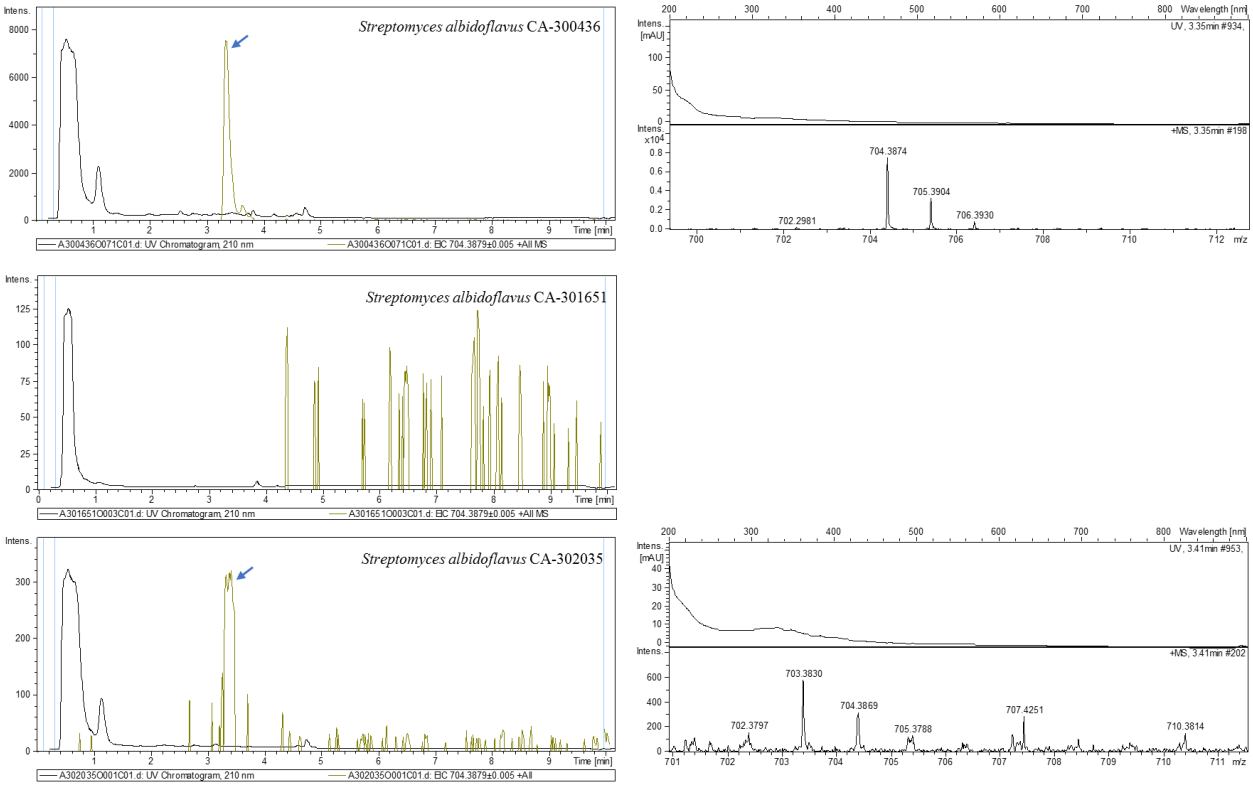


Figure D34. ***Pentaminomycin D production in the knockout and complemented strains. Left.*** *Chromatograms of UV absorbance at 210 nm and extracted ion chromatogram at* m/z *= 704.3879 ± 0.005, $C_{36}H_{50}N_9O_6^+$ of pentaminomycin D (blue arrow) from the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP2 CA-300436 (top), the* cppA*-knockout* Streptomyces albidoflavus *CA-301651 (middle) and the complemented pEM4T-cppA* Streptomyces albidoflavus *CA-302035 (bottom).* ***Right.*** *Experimental UV and positive mass spectra from $C_{36}H_{50}N_9O_6^+$ (calculated value: 704.3879) adduct from the heterologous producing strains* Streptomyces albidoflavus *CA-300436. No UV or mass spectra was obtained with the* cppA*-knockout* S. albidoflavus *CA-301651 as it was expected., Experimental UV and positive mass spectra from $C_{36}H_{50}N_9O_6^+$ (calculated value: 704.3879) adduct from the complemented pEM4T-cppA CA-302035 strain was obtained, thus the production of pentaminomycin D could be restored.*

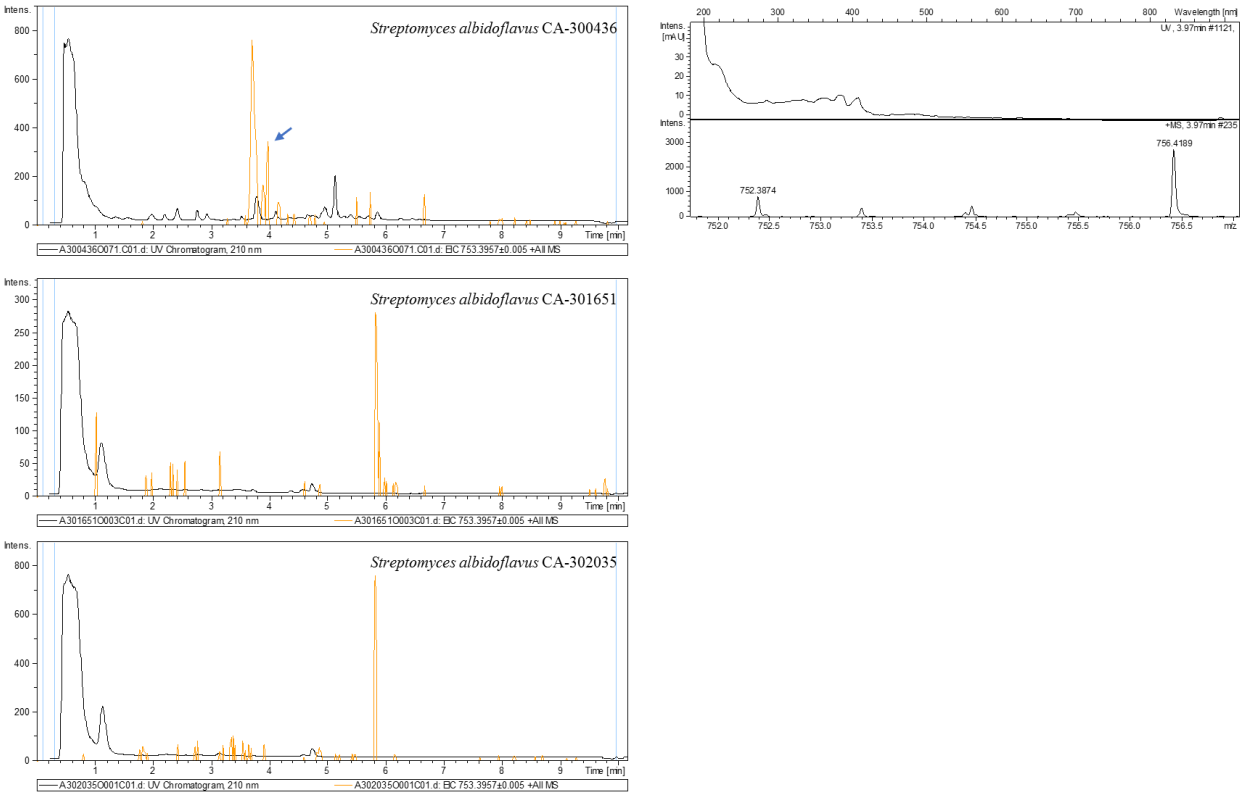


Figure D35. ***Pentaminomycin E production in the knockout and complemented strains. Left.*** *Chromatograms of UV absorbance at 210 nm and extracted ion chromatogram at* m/z *= 752.3879 ± 0.005, $C_{40}H_{50}N_9O_6^+$ of pentaminomycin E (blue arrow) from the heterologous producing strains* Streptomyces albidoflavus *J1074/pCPP2 CA-300436 (top), the* cppA*-knockout* Streptomyces albidoflavus *CA-301651 (middle) and the complemented pEM4T-cppA* Streptomyces albidoflavus *CA-302035 (bottom).* ***Right.*** *Experimental UV and positive mass spectra from $C_{40}H_{50}N_9O_6^+$ (calculated value: 752.3879) adduct from the heterologous producing strains* Streptomyces albidoflavus *CA-300436. No UV or mass spectra was obtained with the* cppA*-knockout* S. albidoflavus *CA-301651 as it was expected. However, no UV or mass spectra was obtained with the complemented pEM4T-cppA CA-302035 strain either, thus the production of pentaminomycin E could not be restored.*

CHAPTER III. CLONING AND HETEROLOGOUS EXPRESSION OF THE *CAO* CLUSTER RESPONSIBLE OF THE BIOSYNTHESIS OF CACAOIDIN

3.1. Cloning and heterologous expression of *cao* cluster

To confirm that the genes included in the *cao* cluster were sufficient for the biosynthesis of the antibiotic, the cacaoidin BGC was cloned and heterologously expressed in the genetically amenable hosts *Streptomyces albidoflavus* J1074, *Streptomyces coelicolor* M1152 and *Streptomyces coelicolor* M1154. As described before, the CATCH method was followed using the specific oligopeptides Glyco1-sgRNA and Glyco2-sgRNA, to clone a 40 Kb region containing the *cao* BGC into the pCAP01 vector. The oligonucleotides pair pCAP01-Glyco-F and pCAP01-Glyco-R was used for the preparation of pCAP01 to clone the 40 Kb *cao* cluster, yielding pCAO. This pCAO construction was transformed into *E. coli* NEB 10-β electrocompetent cells, which were streaked into kanamycin supplemented LB plates. Twenty clones were selected and validated with the oligonucleotides Glyco1-check-F and Glyco1-check-R, and Glyco2-check-F and Glyco2-check-R were used, and two independent double digestions with HindIII/NdeI and XbaI/EcoRV restriction endonucleases were performed, which gave two different fragmentation patterns that allowed us to identify the correct pCAO construction. A clone containing the proper construction was obtained.

The strain *E. coli* NEB 10- β/pCAO was used as donor in triparental conjugations, although, as described in Materials and Methods, the process was differently performed depending on the host.

When *Streptomyces albidoflavus* J1074 was the heterologous host, as it does not have restriction systems for methylated DNA, the NEB 10-β cells containing the kanamycin-resistant pCAO construction were mixed with ET12567/pUB307 *E. coli* cells and spores of *S. albidoflavus* J1074. The pUB307 helper plasmid of ET12567 cells was transferred into the NEB 10-β cells containing pCAO and this pUB307 plasmid mobilized pCAO into *S. albidoflavus*.

When *Streptomyces coelicolor* M1152 and M1154 were the heterologous hosts, NEB 10-β *E. coli* cells could not be used as direct donor strain as *S. coelicolor* strains contain restriction systems for methylated DNA. Thus, the *E. coli* NEB 10-β/pCAO cells were conjugated with TOP10 *E. coli* cells containing the self-transmissible helper plasmid pR9406 and ET12567 *E. coli* cells. The pR9406 helper plasmid mobilized the pCAO from the NEB 10-β to the ET12567 cells. These ET12567/pR9406/pCAO cells were mixed with spores of *S. coelicolor* to perform a biparental conjugation.

The transconjugants were verified phenotypically, relying in their resistance to kanamycin, and genetically via PCR amplification with the checking primers of Table C4. Ten positive transconjugants of each conjugation, together with the negative controls, *S. albidoflavus*/pCAP01 CA-295619, *S. coelicolor* M1152/pCAP01 CA-295620 and *S. coelicolor* M1154/pCAP01 CA-295621, and the wild-type strain *S. cacaoi* CA-170360, were grown in liquid MPG, KM4, FR23 and R2YE media for 14 days at 28 °C to confirm the production of the targeted antibiotic. After acetone extraction of the cultures, organic solvent was evaporated, and the aqueous extracts containing 20 % DMSO were analyzed via LC-HRESIMS(+)-TOF and MS/MS.

The analysis of the extracts from *S. albidoflavus* J1074/pCAO transconjugants (strain CA-300429) confirmed the presence of peaks at 3.35 min, coincident with the retention time of elution of cacaoidin in the wild-type strain and purified cacaoidin standard, only in R2YE medium (Figure D36). No cacaoidin was detected in MPG, KM4 or FR23 media, in contrast of what happened in the wild-type strain, where cacaoidin was detected in the rest of media and the highest production was achieved in FR23 medium (Figure D36). Also compared with CA-170360, the production in the heterologous host *S. albidoflavus* was much more reduced. On the other hand, no cacaoidin was detected in the *S. coelicolor*/pCAO extracts.

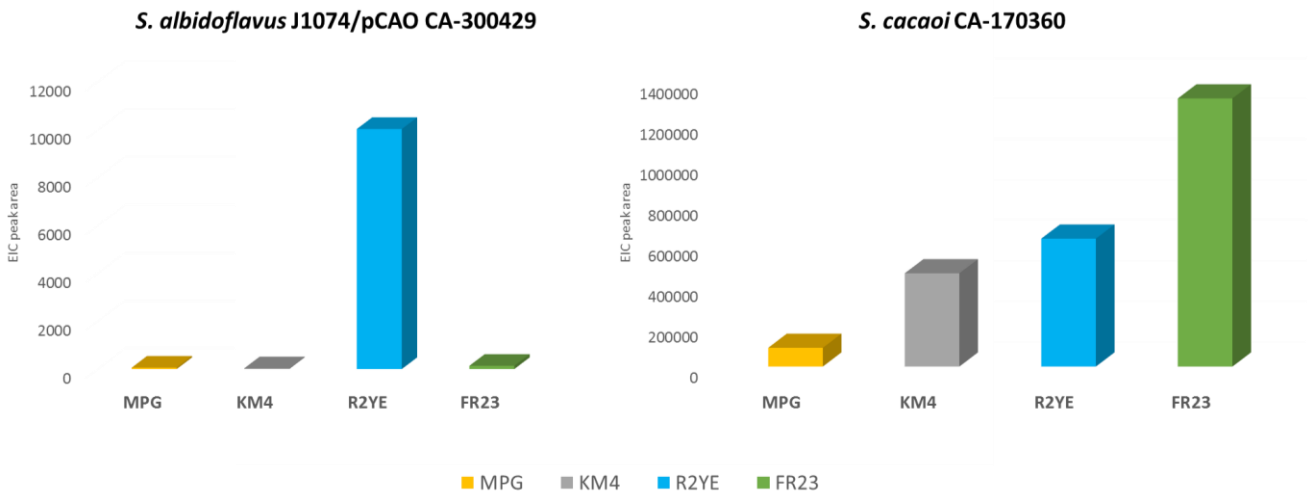


Figure D36. Comparison of the production of cacaoidin by the strains S. albidoflavus *J1074/pCAO (CA-300429) (left) and* S. cacaoi *CA-170360 in four different media (MPG, KM4, R2YE and FR23) at 14 days. The average extracted ion chromatogram (EIC) peak area from triplicate culture extracts is represented.*

The detection of peaks at 3.35 min in the *S. albidoflavus* J1074/pCAO extract suggested the presence of cacaoidin in the heterologous extracts, as it coincided with the retention time of cacaoidin in the original *S. cacaoi* CA-170360 strain. In the negative control *S. albidoflavus* J1074/pCAP01 CA-295619, this peak was not present. Furthermore, the perfect correlation between the UV spectrum, exact mass and the identical isotopic distribution of the ions M + NH_4^+ + $2H^+$ (*m/z* 793.0571) and M + $2NH_4^+$ (*m/z* 1197.5952) supported the production of cacaoidin in the heterologous *S. albidoflavus*/pCAO host (Figure D37). Finally, the coincident MS/MS fragmentation patterns of cacaoidin standards and the components isolated from *S. albidoflavus* J1074/pCAO, undoubtedly demonstrated that they corresponded to cacaoidin (Figures D38).

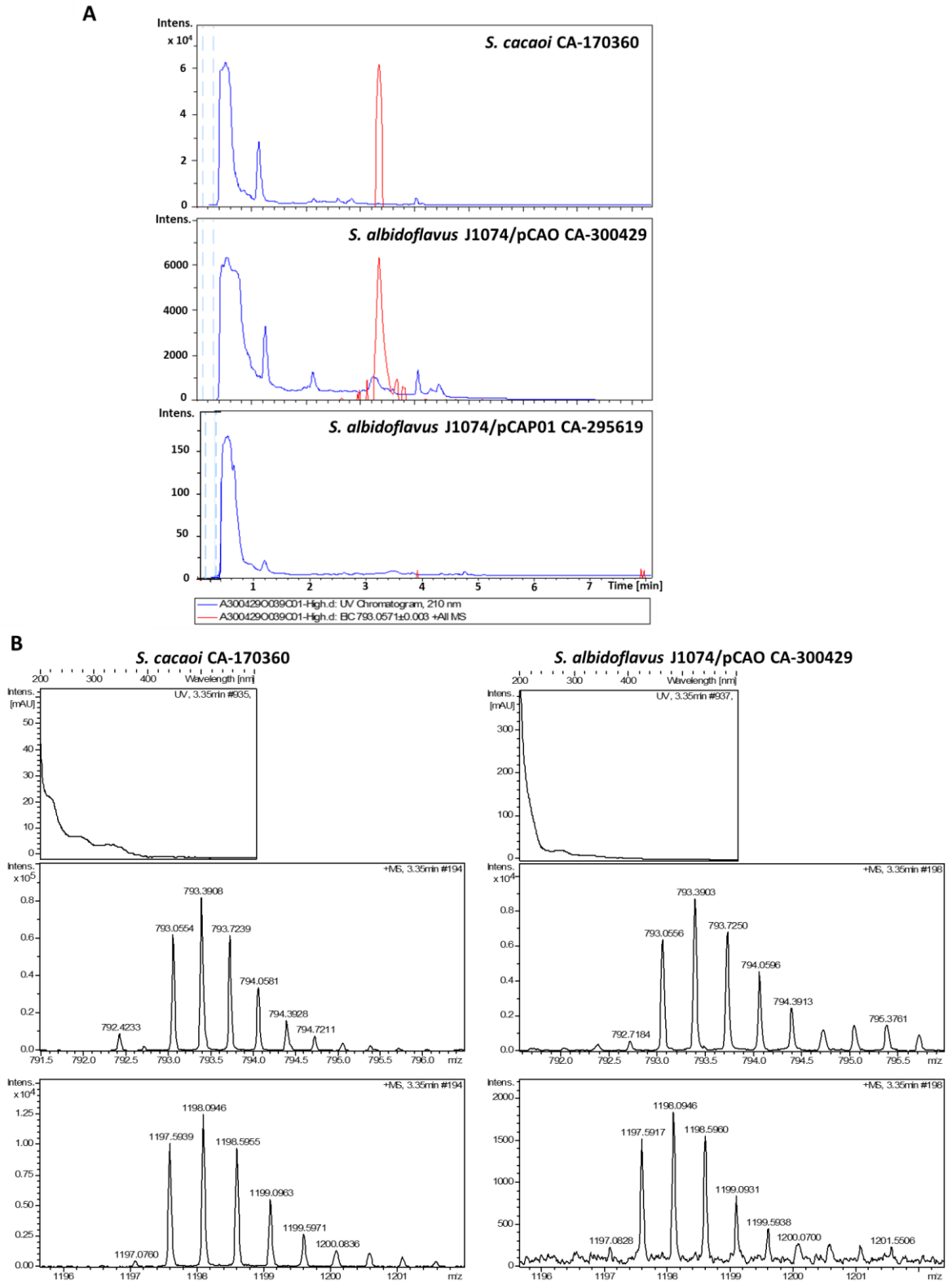


Figure D37. ***(a)*** *Chromatograms of UV absorbance at 210 nm (blue trace) and extracted ion chromatogram at* m/z *= 793.0571 ± 0.003, $C_{107}H_{162}N_{24}O_{32}S_2 + NH_4^+ + 2H^+$ of cacaoidin (MDN-0207) (red trace) from original producing strain Streptomyces cacaoi CA-170360, the heterologous producing strain* Streptomyces albidoflavus *J1074/pCAO and the strain* Streptomyces albidoflavus *J1074/pCAP01 as negative control.* ***(b)*** *Experimental UV and positive mass spectra from $C_{107}H_{162}N_{24}O_{32}S_2 + NH_4^+ + 2H^+$ (calculated value: 793.0571) (top) and $C_{107}H_{162}N_{24}O_{32}S_2 + 2NH_4^+$ (calculated value: 1197.5952) (bottom) adducts from heterologous producing strain (right) and from original producing strain CA-170360 (left).*

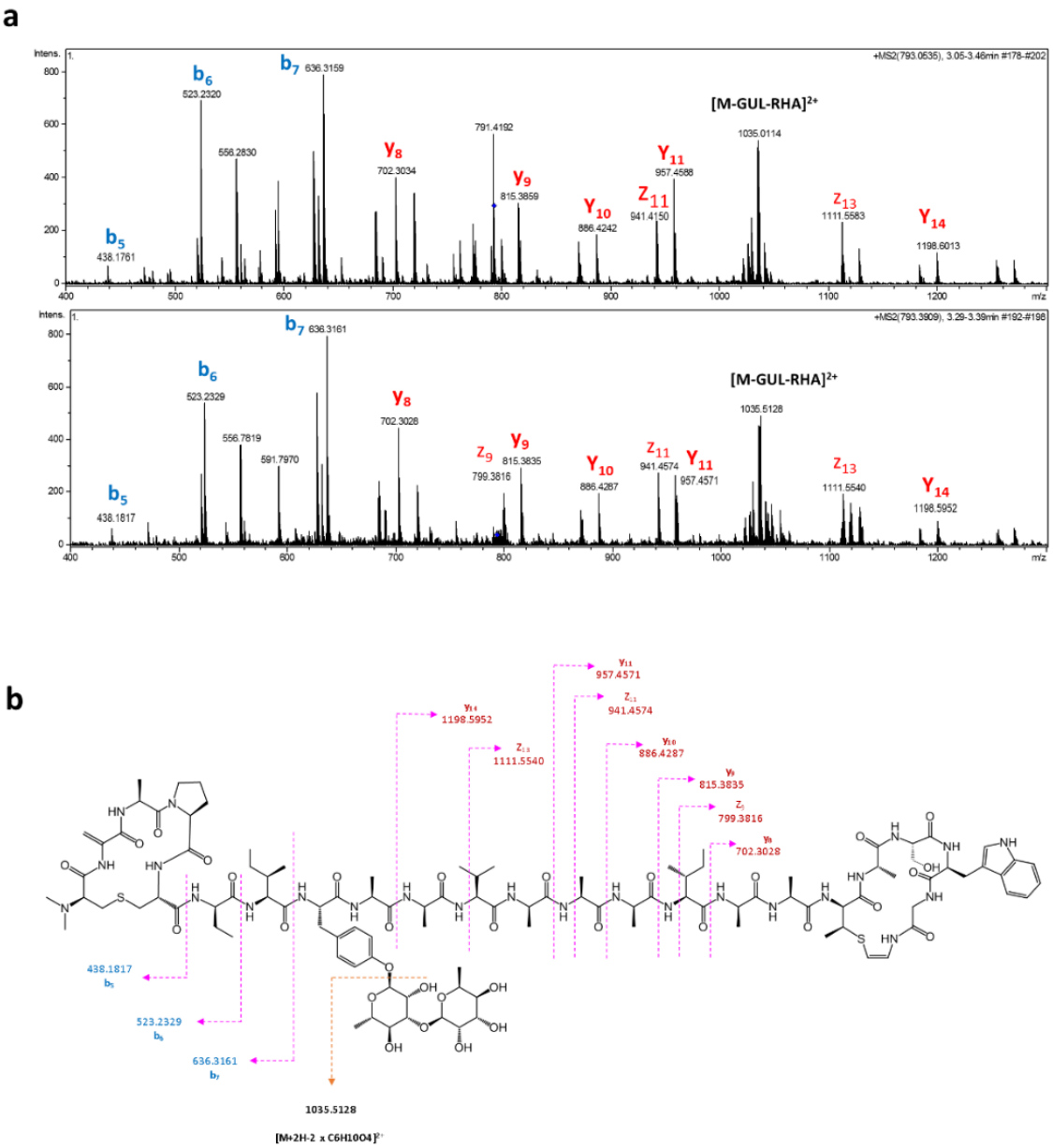


Figure D38. ***(a)*** *Tandem MS/MS fragmentation pattern of* $C_{107}H_{162}N_{24}O_{32}S_2 + NH_4^+ + 2H^+$ *from strain CA-170360 (top) and* S. albidoflavus *J1074/pCAO (bottom). Annotation of b-, y- and z- ions is shown.* ***(b)*** *Key MS/MS fragmentation of* $[M + 2H + NH_4]^{3+}$ *ion.*

The production rate was much lower in the heterologous host CA-300429 than in the wild type strain, as it can be observed in Figure D37a. The yield of cacaoidin production was quantified in both the heterologous and the wild-type strains. The cacaoidin production yield of 0.14 mg/L in the heterologous strain grown in R2YE was reduced approximately 4-fold with respect to the 0.57 mg/L obtained in the wild-type strain CA-170360 in the same medium and fermentation format (10 mL). These results clearly confirm that the *cao* BGC cloned in pCAO contains all the genes required to ensure the biosynthesis of cacaoidin, although the production yield is dependent on the strain used.

3.2. Cacaoidin production in the heterologous host CA-300429

Once proved that the heterologous cacaoidin production was restricted to R2YE medium, a time course was performed for 14 days to see when the cacaoidin production was the highest. The production increased gradually until reached its maximum at 13 days, followed by a decrease in the following days (Figure D39).

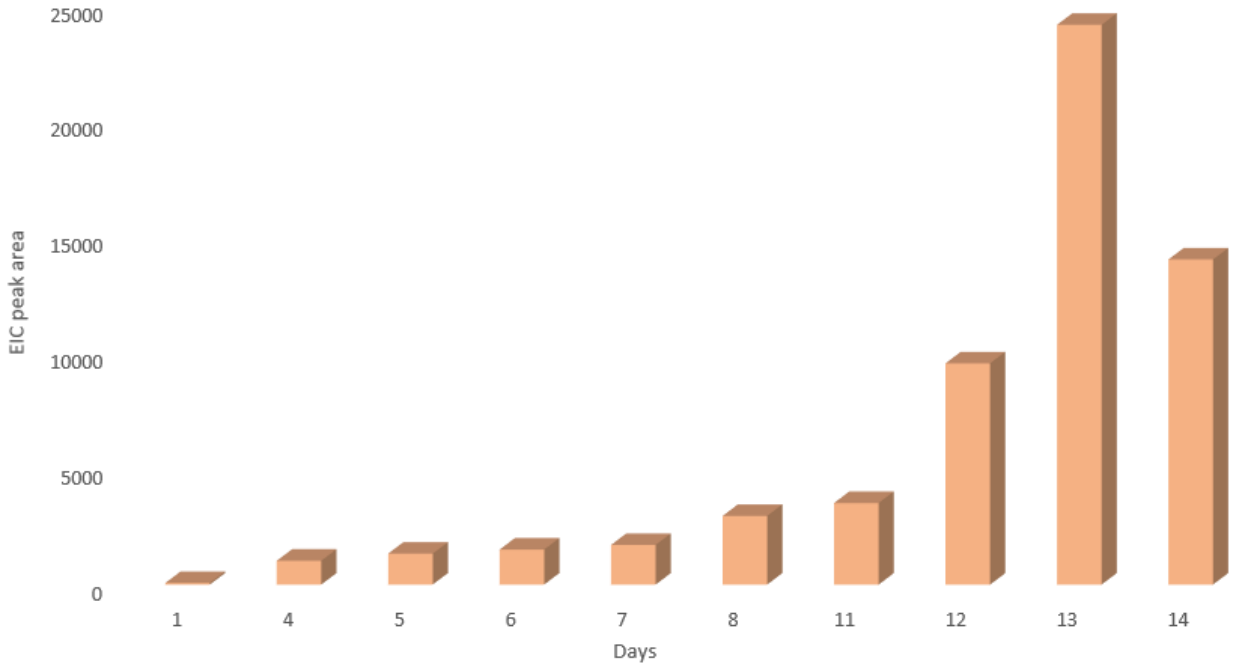


Figure D39. Time course of cacaoidin production by the strain S. albidoflavus *J1074/pCAO CA-300429 in R2YE medium for 14 days. The average extracted ion chromatogram (EIC) peak area from triplicate culture extracts is represented.*

3.2.1. Identification of a new variant of cacaoidin

In a more exhaustive study of the extracts from the strain CA-300429 in R2YE medium, a more abundant cacaoidin-related component was detected by LC-MS. It was a new variant of cacaoidin with a $[M+H]^+$ ion at *m/z* 2328.1432±0.01, indicative of the molecular formula $C_{107}H_{162}N_{24}O_{30}S_2$, corresponding to cacaoidin with two oxygen atoms less, from now on named cacaoidin-2O. This was further supported by the presence of $M+NH_4^+ +H^+$ and $M+NH_4^+ +2H^+$ ions detected at *m/z* 1173.589±0.01 and 782.728±0.01, respectively. These ions were not present in the original *Streptomyces cacaoi* strain (Figures D40 and D41).

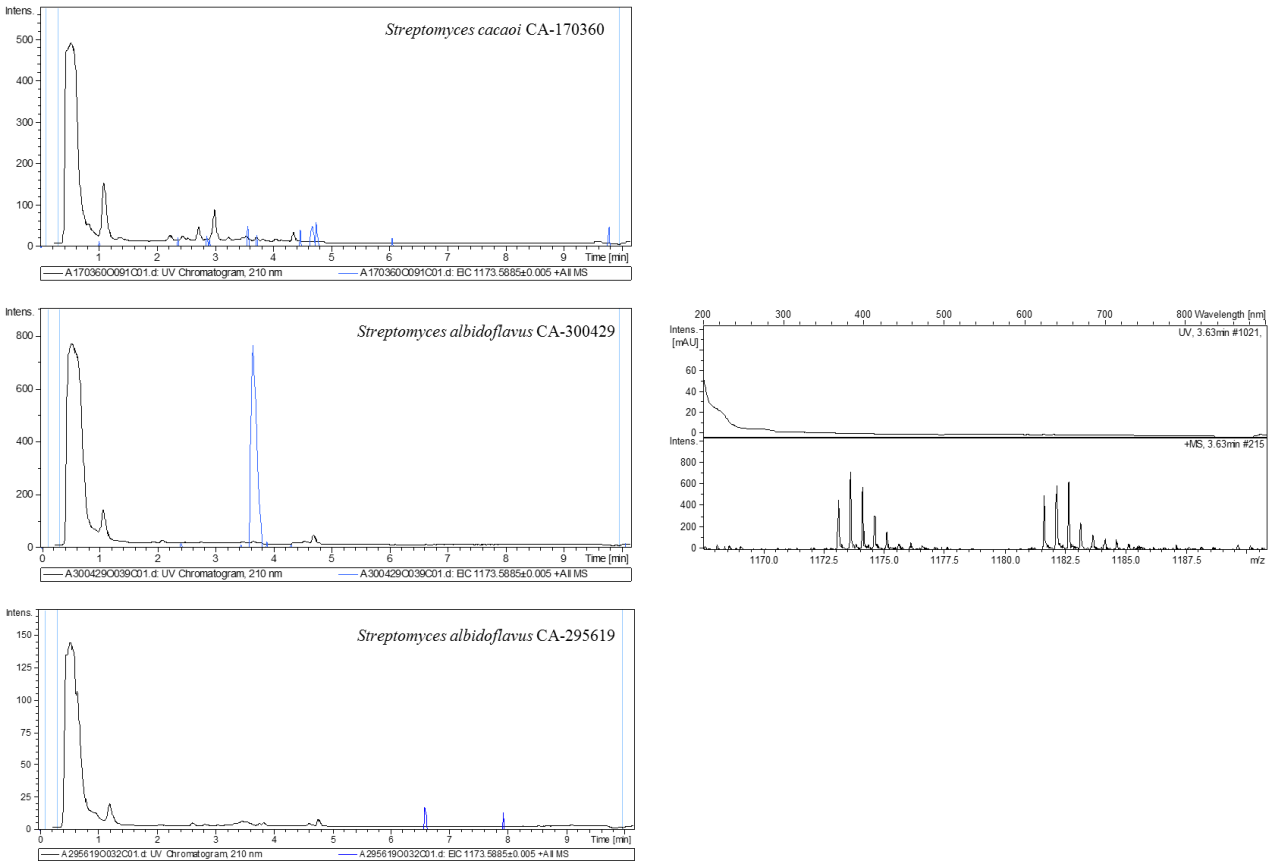


Figure D40. ***Left.*** *Chromatograms of UV absorbance at 210 nm (black trace) and extracted ion chromatogram at* m/z *= 1173.5885 ± 0.005,* $C_{107}H_{162}N_{24}O_{30}S_2 + NH_4^+ + H^+$ *(blue trace) of cacaoidin-2O from original cacaoidin producing strain* Streptomyces cacaoi *CA-170360 (top), the heterologous producing strain* Streptomyces albidoflavus *CA-300429 (middle) and the heterologous strain* Streptomyces albidoflavus/*pCAP01 CA-295619 as negative control (bottom).* ***Right.*** *Experimental UV and positive mass spectra from* $C_{107}H_{162}N_{24}O_{30}S_2 + NH_4^+ + H^+$ *(calculated value: 1173.5885) adducts from heterologous producing strain. No experimental UV and positive mass spectra for the cacaoidin-2O could be obtained from the original cacaoidin producing strain* Streptomyces cacaoi *CA-170360 or the negative control.*

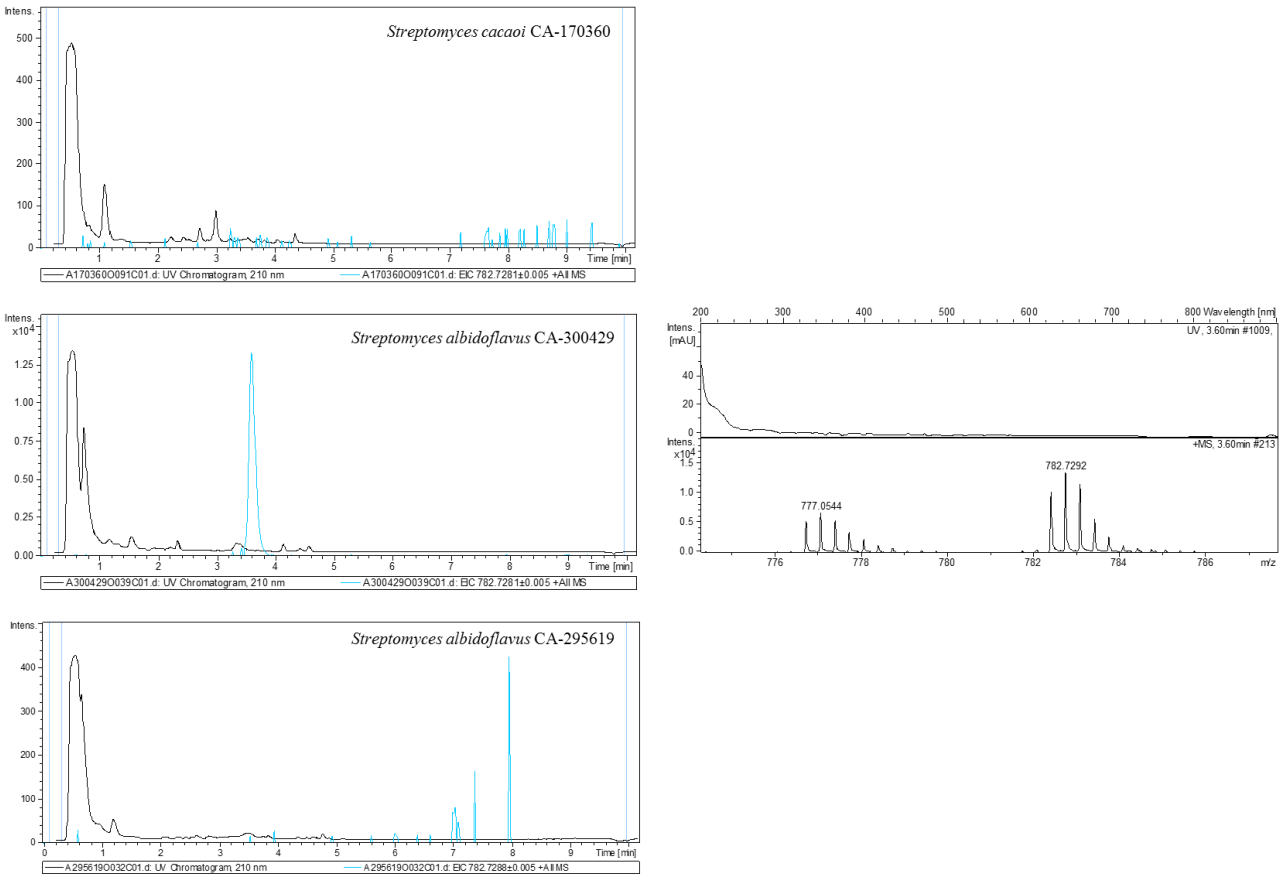


Figure D41. ***Left.*** *Chromatograms of UV absorbance at 210 nm (black trace) and extracted ion chromatogram at* m/z *= 782.7288 ± 0.005, $C_{107}H_{162}N_{24}O_{30}S_2 + NH_4^+ + 2H^+$ (clear blue trace) of cacaoidin-2O from original cacaoidin producing strain* Streptomyces cacaoi *CA-170360 (top), the heterologous producing strain* Streptomyces albidoflavus *CA-300429 (middle) and the heterologous strain* Streptomyces albidoflavus/*pCAP01 CA-295619 as negative control (bottom).* ***Right.*** *Experimental UV and positive mass spectra from $C_{107}H_{162}N_{24}O_{30}S_2 + NH_4^+ + 2H^+$ (calculated value: 782.7288) adducts from heterologous producing strain. No experimental UV and positive mass spectra for the cacaoidin-2O could be obtained from the original cacaoidin producing strain* Streptomyces cacaoi *CA-170360 or the negative control.*

The LC-HRESIMS(+)-TOF analysis and MS/MS fragmentation proved that a loss of an oxygen atom happened in each sugar in the cacaoidin-2O. The same aglycon (*m/z* 1035.0122 ± 0.005, M + $2H^+$) and fragments of cacaoidin generated in the MS/MS fragmentation were detected in the heterologous host (Figure D42).

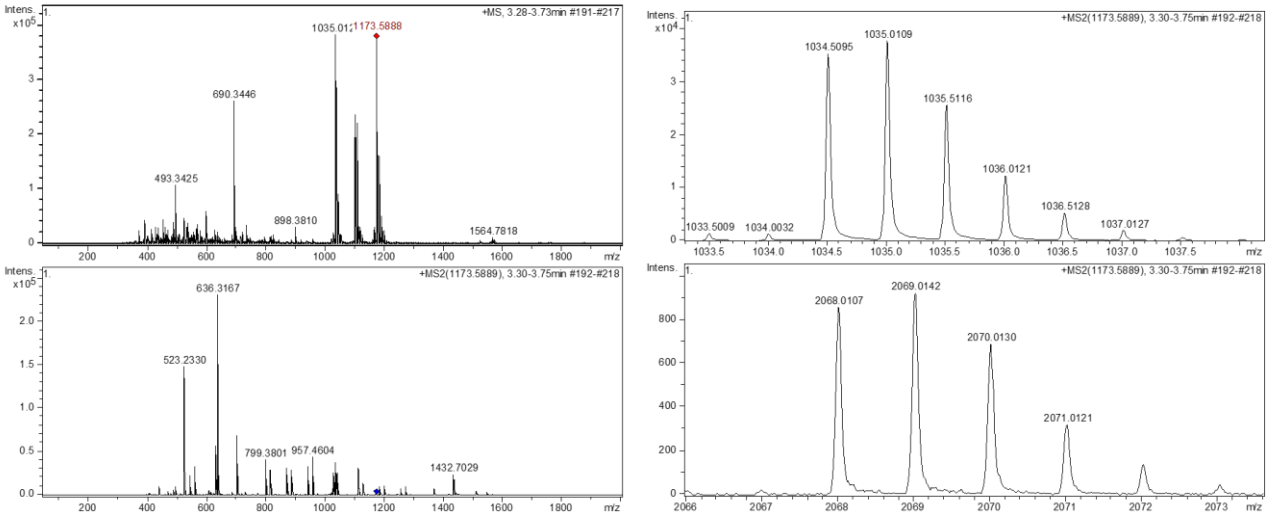


Figure D42. ***Left.*** *Positive mass spectra with the adducts originated from cacaoidin-2O (top). The ion* m/z *1035.0122 ± 0.005, corresponding to the aglycon of cacaoidin, was detected. MS/MS fragmentation pattern of the* $C_{107}H_{162}N_{24}O_{30}S_2$ *+*$NH4^+$ *+* $2H^+$ *ion* m/z *1173.5888 ± 0.005 (red spot from the positive mass spectra) from cacaoidin-2O (bottom). The same fragments as in the original cacaoidin were obtained with the MS/MS fragmentation.* ***Right.*** *Isotopic pattern of the ion* $C_{95}H_{142}N_{24}O_{24}S_2$ *+* $2H^+$ (m/z *1035.0122 ± 0.005) (top) and the* $C_{95}H_{142}N_{24}O_{24}S_2$ *+* H^+ (m/z *2069.0172 ± 0.005) (botton), corresponding to the aglycon of the cacaoidin-2O.*

Furthermore, in the MS/MS fragmentation, the loss of one sugar and one oxygen atom were detected in the MS/MS fragmentation (Figure D43). The presence of two fragment ions at *m/z* 2069.0172 and 2199.0802 in the MS/MS spectrum indicated the sequential loss of two deoxyhexose units, each of them lacking one oxygen atom compared to the corresponding monosaccharide units present in cacaoidin (Figure D43).

When the production of this new variant was compared with that of the wild-type cacaoidin in the heterologous host CA-300429, it was highlighted that the new cacaoidin-2O was the major compound (EIC peak area 92689), as the production of the wild-type cacaoidin was much lower (EIC peak area 9796).

As the new cacaoidin-2O was produced in such higher levels than the wild type cacaoidin, the generated knock outs were characterized relying on this new variant.

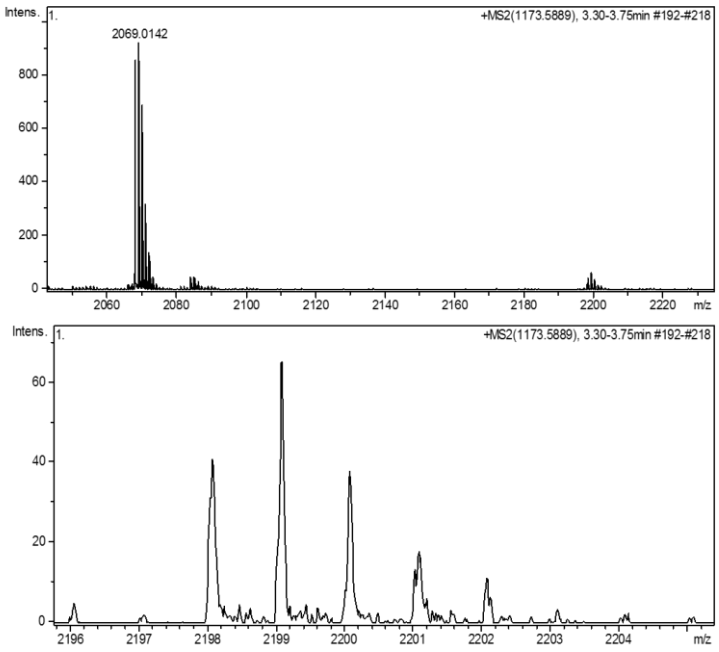


Figure D43. ***Top.*** *MS/MS fragmentation pattern obtaining the ion $C_{95}H_{142}N_{24}O_{24}S_2 + H^+$ (*m/z *2069.0172 ± 0.005), corresponding to the aglycon, and the ion $C_{101}H_{151}N_{24}O_{27}S_2 + H^+$ (*m/z *2199.0802 ± 0.005), corresponding to the monoglycosylated cacaoidin with the loss of one oxygen atom.*

3.2.2. Heterologous host CA-300429 large scale-up fermentation to produce cacaoidin-2O

In order to purify enough amount of cacaoidin-2O and be able to characterize the new disaccharide of the molecule and evaluate its microbial activity, a total of 10 L fermentation of CA-300429 in R2YE medium were prepared in 125 mL flasks, reaching a total of 80 flasks. To seed all these flasks, 0.5 L of inoculum were needed (6.25 mL x 80 flasks), and thus 10 liquid ATCC-2 50 mL-flasks were seeded from a 10 mL ATCC-2 tube and incubated for 2-3 days at 28 °C for 13 days, 220 rpm, 70 % humidity. The 80 flasks were seeded from these ATCC-2 50 mL flasks and incubated at 28 °C for 13 days, 220 rpm, 70 % humidity.

After the 13 days of fermentation, the extraction of the new cacaoidin-2O was performed with acetone (1:1, vol/vol) shaking at 220 rpm for 2 hours. Then, the organic solvent was separated from the biomass by centrifugation and vacuum filtration, and the acetone was removed O/N under N_2 stream. The resulting aqueous broth (aprox. 8 L) was adjusted to pH 10 with 10 N NaOH and then MEK was added (3 x 4 L MEK) to improve the extraction of the new cacaoidin derivate. The organic solvent was rotary evaporated to dryness. From this 10 L fermentation, 4.3547 g of powder extract were obtained.

3.2.3. Isolation and purification of the cacaoidin-2O

In previous studies with the original cacaoidin, it was stated that the purification of the compound was daunting as different attempts to pre-purify the wild type cacaoidin extract by MPLC resulted in the loss of the molecule (Ortiz-López *et al.,* 2020). This, along with the limiting measures of the columns, made necessary the partition of the cacaoidin-2O extract in successive portions of 150 mg which were directly purified by preparative RP-HPLC as described in Materials and Methods, collecting 80 fractions in vials of 20 mL. The peaks observed in the chromatogram (Figure D44) were submitted to LC-HRESIMS(+)-TOF analysis and the fractions 27-28, eluting at 15-16 min, were selected as they corresponded to the cacaoidin-2O. After all the preparative RP-HPLCs, the 27-28 fractions of each HPLC, were collected. The organic solvent of these fractions was carefully evaporated under N_2 stream and the resulted aqueous sample was freeze-dried.

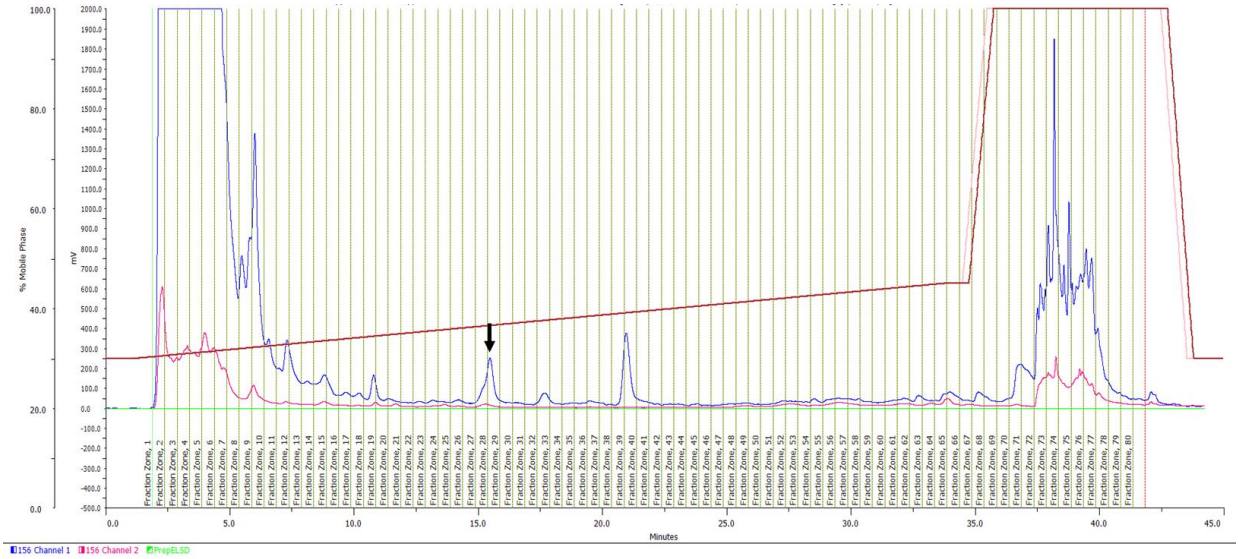


Figure D44. HRMS-guided isolation of cacaoidin-2O. Preparative RP-HPLC purification of the cacaoidin-2O extract. LC-UV preparative chromatogram recorded at both 210 (blue trace) and 280 nm (pink trace).

This pool of fractions 27-28 of all the preparative RP-HPLCs was subjected to two additional rounds of purification in order to obtain the cacaoidin-2O completely isolated. These two rounds of purification were carried out following the protocol described in Materials and Methods, resulting in the total isolation and purification of the cacaoidin-2O (Figure D45).

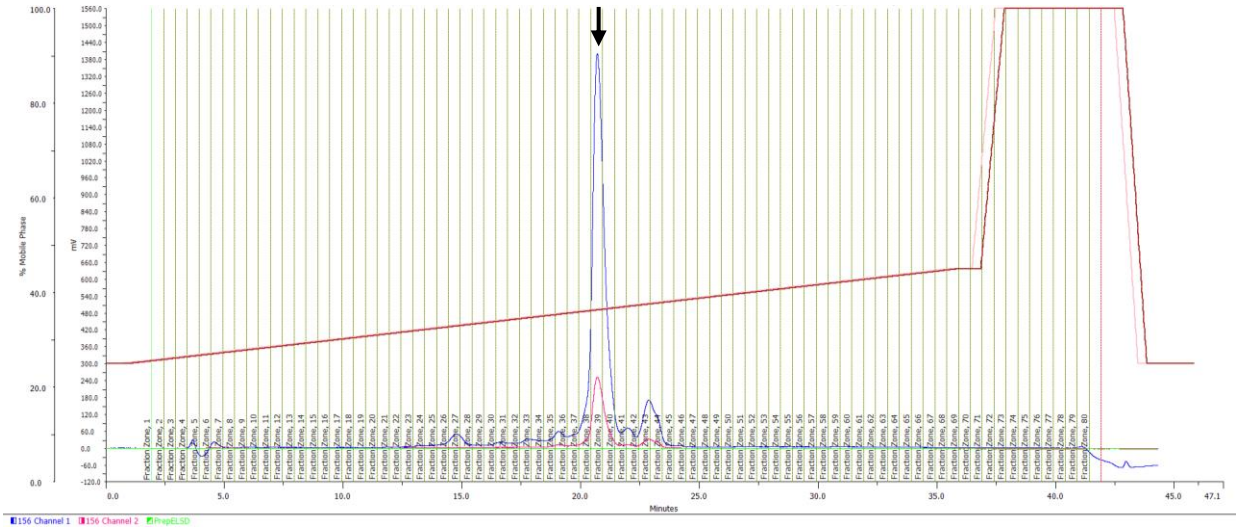


Figure D45. HRMS-guided isolation of cacaoidin-2O. Semipreparative RP-HPLC purification of the fractions 24 to 26 from the semipreparative pre-purification. LC-UV preparative chromatogram recorded at both 210 (blue trace) and 280 nm (pink trace).

The fractions 37 to 39 of the last RP-HPLC performed with the Phenyl column were collected and pooled, as they corresponded with the pure cacaoidin-2O. The organic solvent of this pool of fractions was carefully evaporated and the resting aqueous component was freeze-dried, obtaining a final 1.5 mg of amorphous white powder of cacaoidin-2O.

This pure cacaoidin-2O is currently under structural determination by NMR in order to stablish the identity of the new disaccharide bonding with the tyrosine residue. Also, part of the pure molecule was used to generate aglycon (see section D3.4) and in antimicrobial assays against MRSA. In this assays, cacaoidin-2O showed the same inhibition percentage at the same concentration than cacaoidin, suggesting that the difference in the disaccharide does not have an impact in the bioactivity. However, due to the low amounts obtained, the MIC value could not be determined.

3.2.4. Culture media enrichment with L-rhamnose

As described before, the difference between cacaoidin and cacaoidin-2O lies in the disaccharide, which has an oxygen atom less in each sugar moiety. In order to favour the production of cacaoidin in the heterologous host, the fermentation tubes were enriched with L-rhamnose, which is one of the components of the disaccharide in cacaoidin (Ortiz-López *et al.,* 2020). The purpose of this study was to test if, in the presence of an additional source of L-rhamnose, the heterologous host was able to incorporate L-rhamnose to the tyrosine residue of the core peptide instead of the new sugar moieties. Based on the composition of the R2YE medium that includes 10 g/L glucose and 103 g/L sucrose, different variants of the R2YE medium containing different concentrations of L-rhamnose were used to grow the heterologous host *S. albidoflavus* CA-300429 and the original producer *S. cacaoi* CA-170360 strains:

R2YE medium	Glucose (g/L)	Sucrose (g/L)	Rhamnose (g/L)
S+G+ (control)	10	103	-
S-G+	10	-	-
S+G-R+	-	103	10
S-G-R+	-	-	10
S+G+R+	10	103	10
S-G+R+	10	-	10
S+G5+R5+	5	103	5
S-G5+R5+	5	-	5

Table D7. Composition of the different R2YE media variants. S+G+ is the R2YE medium with no modifications.

The two strains were fermented in 40 mL vials containing 10 mL of all these variants of R2YE medium for 13 days. After extraction of the secondary metabolites following the process described in Section C12.1.1, the extracts were analyzed by HRESIMS(+)-TOF to study if the production of cacaoidin-2O had been derived to the original cacaoidin in the heterologous strain *S. albidoflavus* CA-300429. In the case of the original *S. cacaoi* CA-170360 strain, cacaoidin was detected in all the fermentation media, while the cacaoidin-2O was never detected (Figure D48). On the other hand, in the heterologous *S. albidoflavus* CA-300429 both cacaoidins were detected only when the strain was grown in the R2YE media variants supplemented with sucrose. In these cases, the heterologous strain CA-300429 produced cacaoidin-2O in a higher extent, while the production of cacaoidin was much lower (Figure D46). In the heterologous host, the production of both cacaoidins were comparable when glucose or L-rhamnose were used as carbon sources, but in combination with sucrose, which seemed essential for the production. No other variants of cacaoidin were detected in either strain in the different fermentation conditions.

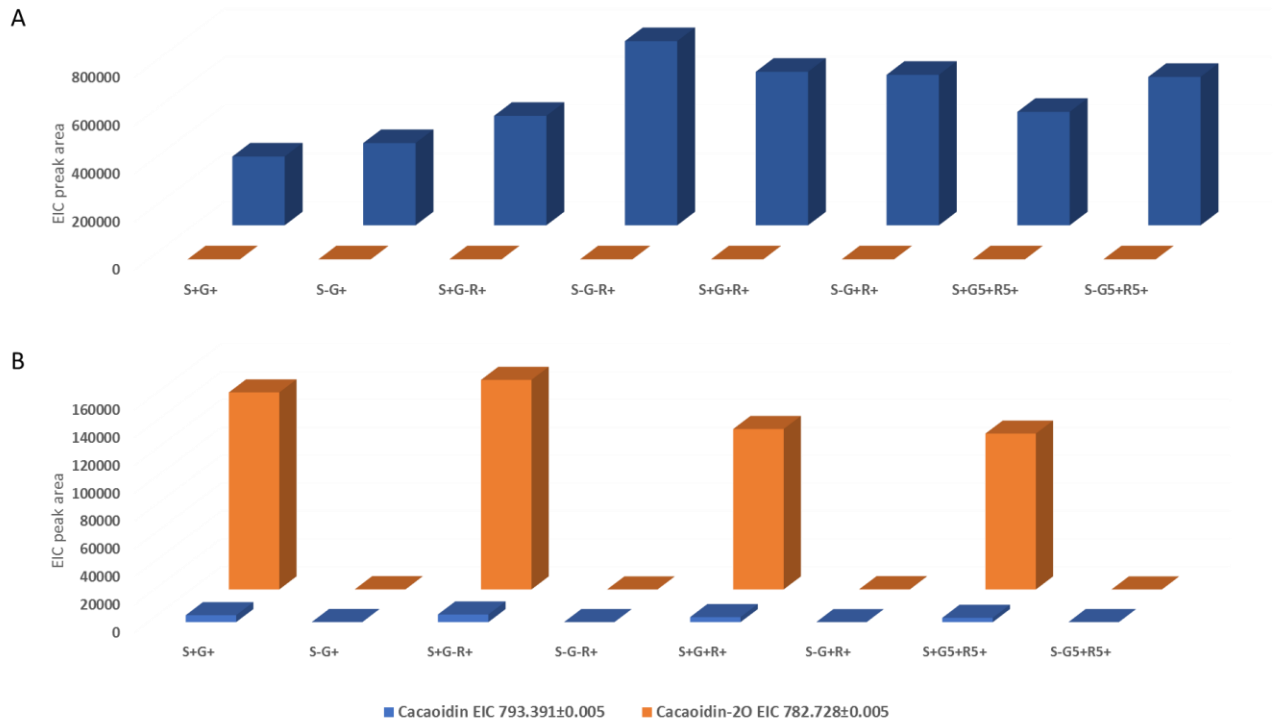


Figure D46. Production of original cacaoidin (blue) and cacaoidin-2O (orange) in the original Streptomyces cacaoi *CA-170360 strain (A) and the heterologous host* Streptomyces albidoflavus *CA-300429. The average extracted ion chromatogram (EIC) peak area from triplicate culture extracts of the strain CA-170360 is represented.*

3.3. Generation of knockouts in the heterologous host CA-300429

To study the role of some genes in the biosynthesis of cacaoidin-2O, and with the aim of obtaining aglycon, monoglycosylated and non-methylated derivatives, different knockouts were performed.

3.3.1. Generation of the knockout of the O-methyltransferase gene cao4

Cao4 was determined by BLAST to be a putative O-methyltransferase containing the conserved Methyltransf_2 domain. Despite the low similarity (26.9 %) found between Cao4 and LxmM, the methyltransferase that catalyzes the lexapeptide N-terminal dimethylation (Xu *et al.*, 2020), both proteins share the AdoMet_MTases superfamily domain, characteristic of the SAM-dependent methyltransferases. In order to verify the implication of Cao4 in the dimethylation of the N-terminal alanine, in both cacaoidin and cacaoidin-2O, a knockout of *cao4* was carried out in the heterologous host. Primer pairs pEFBA-cao4-Frag1-F and pEFBA-cao4-Frag1-R, and pEFBA-cao4-Frag2-F and pEFBA-cao4-Frag2-R from Table C6 were used to amplify two DNA fragments of 1.965 Kb and 1.965 Kb up- and downstream *cao4,* respectively. These DNA fragments were digested with the required restriction enzymes (SpeI/NsiI and NdeI/EcoRV, respectively) and were sequentially cloned into the previously digested pEFBA-oriT plasmid. Finally, the

thiostrepton resistance gene was cloned in the XbaI restriction site of the pEFBA-oriT plasmid. The constructions were checked by digestion after each step of the cloning.

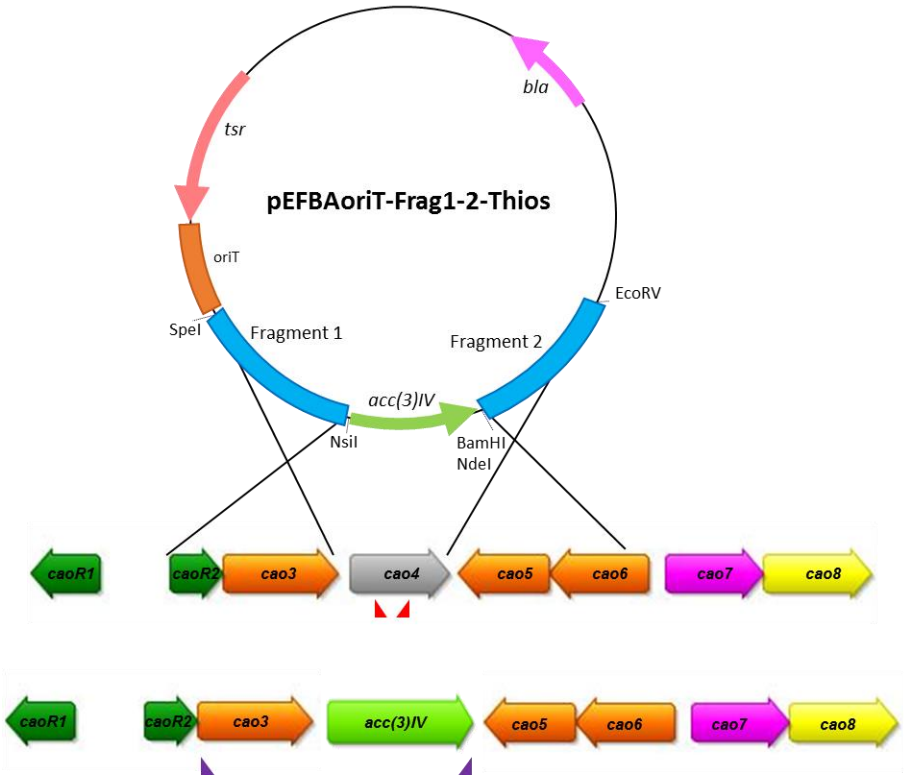


Figure D47. Scheme of the genetic replacement of cao4 *with the apramycin resistance cassette in the* cao *cluster. In the figure, the internal pair of primers Cao4-check-F and Cao4-check-R for the checking of the deletion of the gene are shown in red and the external and internal primers pEFBA-Cao4-Frag1-F and Apramycin-R for the checking of the insertion of the apramycin cassette are shown in purple.*

This final pEFBA-cao4 construction was transferred to the heterologous host *S. albidoflavus* CA-300429 via intergeneric conjugation as described in the Section C11.1 and the obtained transconjugants were checked by the growing phenotype (resistance to apramycin and sensitivity to thiostrepton) and by PCR amplification with Cao4-check-F and Cao4-check-R primers (Table C9), which amplify a 455 bp fragment inside *cao4,* to check the deletion of the gene, and pEFBA-cao4-Frag1-F and Apramycin-R primers (Tables C6 and C7, respectively), which amplify a 3.5 Kb fragment, to check the insertion of the apramycin resistance cassette (Figure D47). Those transconjugants which were resistant to apramycin and sensible to thiostrepton and whose PCR amplifications showed the presence of the apramycin cassette instead of *cao4*, were fermented, along with the heterologous host CA-300429 as control, in 40 mL vials containing 10 mL R2YE and incubated at 28 °C, 220 rpm, 70 % humidity for 13 days. Then, the extracts generated in these fermentations were processed as described in Section C 12.1.1 and analyzed by HRESIMS(+)-TOF scavenging for the presence of a peak with a $M+2H^+ + NH_4^+$ ion at *m/z* 773.3843, indicative of the molecular formula $C_{105}H_{158}N_{24}O_{30}S_2$, corresponding with the *N,N*-demethylated cacaoidin-2O. This analysis confirmed the presence of this ion at 3.46 min in the *cao4*-knockout strain CA-

300655 (Figure D48), while the cacaoidin-2O was not detected (Figure D49). In the control heterologous host CA-300429 the ion corresponding to the *N,N*-demethylated cacaoidin was also detected but in a much lesser extent (Figure D48).

To restore the function of Cao4 and check if the heterologous *N,N*-methylated cacaoidin-2O was produced again, the *cao4*-knockout strain CA-300655 was complemented with a construction overexpressing the *cao4* gene.

First, *cao4* was amplified using the primers Cao4-F and Cao4-R (Table C8). The PCR was digested with EcoRI and cloned into the previously EcoRI-digested pEM4T. The ligation was transferred to NEB 10-β *E. coli* cells and then several colonies were checked by digestion with ClaI to ensure that the direction of the gene was the appropriate to be transcribed under the control of the *ermE*p* promoter. Once checked, the pEM4T-cao4 construction was transferred to spores of the Δ*cao4* knockout strain CA-300655 by intergeneric conjugation. The transconjugants were streaked into MA plates supplemented with thiostrepton and nalidixic acid and then checked by PCR amplification with the Cao4-check-F and Cao4-check-R (Table C9).

The positive transconjugants were fermented in 40 mL vials containing 10 mL R2YE, along with the heterologous host CA-300429 and the *cao4*-knockout CA-300655, and incubated at 28 °C, 220 rpm, 70 % humidity for 13 days. The extracts generated were analyzed by HRESIMS(+)-TOF and the *N*-dimethylated cacaoidin-2O was detected in the complemented strain *Streptomyces albidoflavus* CA-301341. (Figure D49).

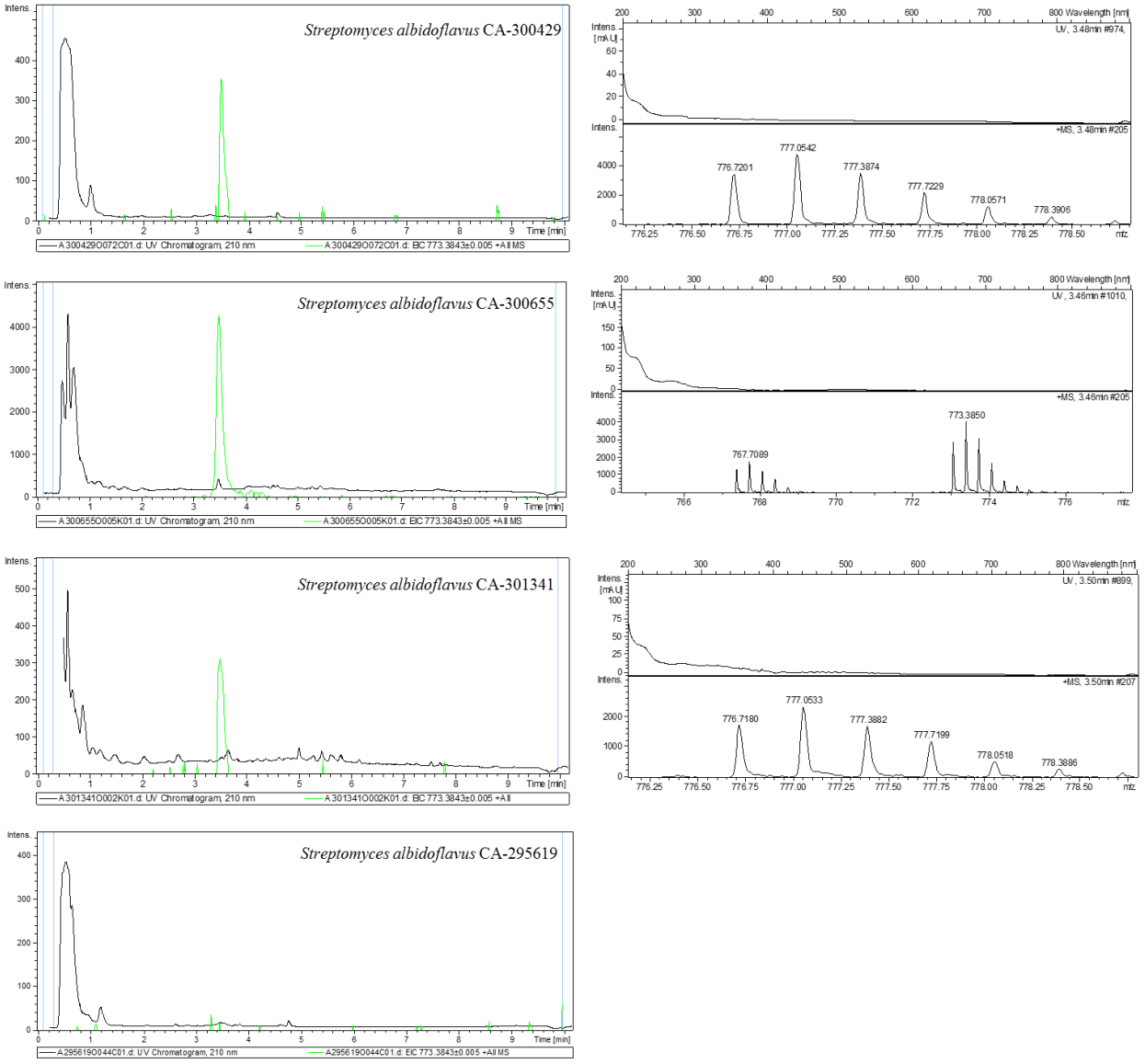


Figure D48. ***Left.*** *Chromatograms of UV absorbance at 210 nm (black trace) and extracted ion chromatogram at* m/z *= 773.3843 ± 0.005, $C_{105}H_{158}N_{24}O_{30}S_2 + NH_4^+ + 2H^+$ (green trace) of demethylated cacaoidin-2O from heterologous producing strain* Streptomyces albidoflavus *CA-300429, the cao4-knockout strain* Streptomyces albidoflavus *CA-300655, the cao4-complemented strain* Streptomyces albidoflavus *CA-301341 and the negative control* Streptomyces albidoflavus*/pCAP01 CA-295619.* ***Right.*** *Experimental UV and positive mass spectra from $C_{105}H_{158}N_{24}O_{30}S_2 + NH_4^+ + 2H^+$ (calculated value: 773.3843) adducts from heterologous producing strain (top), the cao4-knockout strain* Streptomyces albidoflavus *CA-300655 (middle) and from cao8-complemented strain CA-301580 (bottom).*

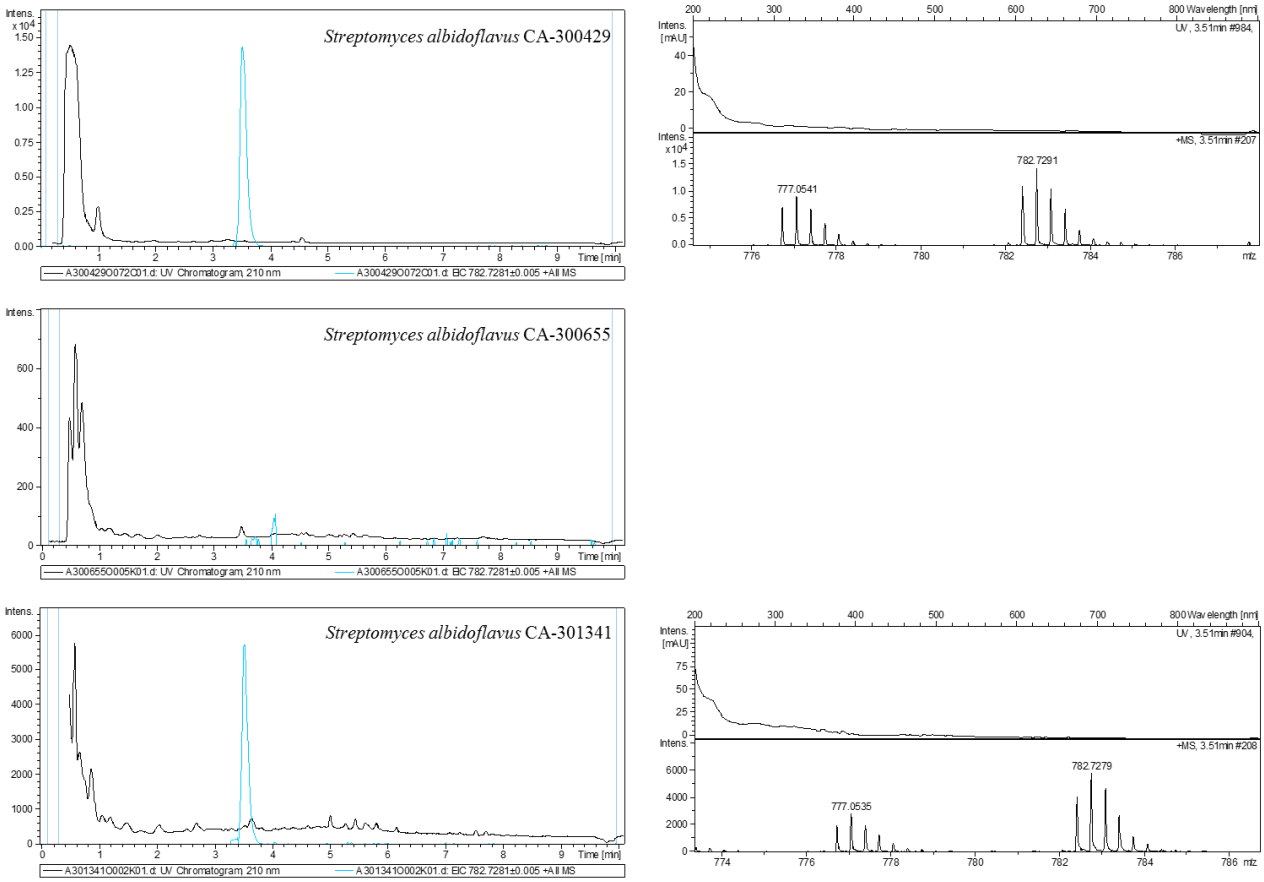


Figure D49. ***Left.*** *Chromatograms of UV absorbance at 210 nm (black trace) and extracted ion chromatogram at* m/z *= 782.7288 ± 0.005, $C_{107}H_{162}N_{24}O_{30}S_2 + NH_4^+ + 2H^+$ (clear blue trace) of cacaoidin-2O from heterologous producing strain* Streptomyces albidoflavus *CA-300429 (top), the cao4-knockout strain* Streptomyces albidoflavus *CA-300655 (middle) and the cao4-complemented strain* Streptomyces albidoflavus *CA-301341.* ***Right.*** *Experimental UV and positive mass spectra from $C_{107}H_{162}N_{24}O_{30}S_2 + NH_4^+ + 2H^+$ (calculated value: 782.7288) adducts from heterologous producing strain (top) and from cao4-complemented strain CA-301341 (bottom). No experimental UV and positive mass spectra for this cacaoidin-2O could be obtained from the cao4-knockout strain CA-301655.*

3.3.2. Generation of the knockout of the glycosyltransferase gene cao8

As mentioned before, three glycosyltransferase genes were identified in the *cao* cluster: *cao8*, *cao16* and *cao24*. However, only two sugar moieties in the form of disaccharide were detected in the structure. Thus, the three GTs might work in a cooperative manner to achieve effective glycosylation. In order to elucidate the role played by each GT in the formation of the disaccharide and its joining to the tyrosine residue, knockouts of each GT gene were generated by genetic replacement as described in Section C7.8.

Cao8 is a glycosyltransferase belonging to the family GT-2 and shows 42 % identity (54 % similarity) with an UDP-Glc:alpha-D-GlcNAc-diphosphoundecaprenol beta-1,3-glucosyltransferase WfgD, which catalyzes the addition of Glc, the second sugar moiety of the O152-antigen repeating unit, to GlcNAc-pyrophosphate-undecaprenol (Brockhausen *et al.,* 2008). For the generation of the knockout of *cao8*, fragments of 1.99 and 2.35 Kb, located up- and downstream *cao8,* respectively, were amplified using the pair of primers pEFBA-cao8-Frag1-F and pEFBA-cao8-Frag1-R, and pEFBA-cao8-Frag2-F and pEFBA-cao8-Frag2-R (Table C6), and they were digested with the required restriction enzymes (SpeI/NsiI and BamHI/EcoRV) and sequentially cloned into the previously digested pEFBA-oriT plasmid, where a thiostrepton resistance gene was also cloned to confirm the double crossover. The constructions were checked by digestion after each step of the cloning.

The final pEFBA-cao8 construction was transferred to the heterologous host *S. albidoflavus* CA-300429 via intergenic conjugation and the transconjugants obtained were checked based on the growing phenotype (resistance to apramycin and sensitivity to thiostrepton) and the lack of PCR amplification of a 398 bp fragment with Cao8-check-F and Cao8-check-R primers (Table C9), to check the deletion of the gene, and the PCR amplification of a 3.5 Kb fragment with primers pEFBA-cao8-Frag1-F and Apramycin-R (Tables C6 and C7, respectively), to confirm the gene replacement (Figure D50). Once checked, the positive transconjugants were fermented in 40 mL vials containing 10 mL R2YE, along with the heterologous host CA-300429, and incubated at 28 °C, 220 rpm, 70 % humidity for 13 days. The extraction was carried out as described in Section C 12.1.1 and the HRESIMS(+)-TOF analysis detected neither the presence of cacaoidin-2O nor any of its putative monoglycosylated or non-glycosylated intermediates (Figure D50).

The strain CA-301537 is one of the *cao8* knockout transconjugants which did not produce cacaoidin-2O nor any variant of it. Complementation with pEM4T-cao8 was needed in this case to restore the function of *cao8* and check if the production of cacaoidin-2O was detected.

First, *cao8* was amplified using the primers Cao8-F and Cao8-R (Table C8) and digested with the restriction enzymes BamHI and EcoRI. This digested product, after purification, was cloned into the previously BamHI/EcoRI-digested pEM4T. The new construction was transferred to NEB 10-β *E. coli* cells and then checked by digestion with restriction enzymes. Once checked, it was transferred to spores of the *cao8* knockout strain CA-301537 by conjugation. The transconjugants were streaked into MA plates supplemented with thiostrepton and nalidixic acid and then checked by PCR amplification with the Cao8-check-F and Cao8-check-R (Table C9).

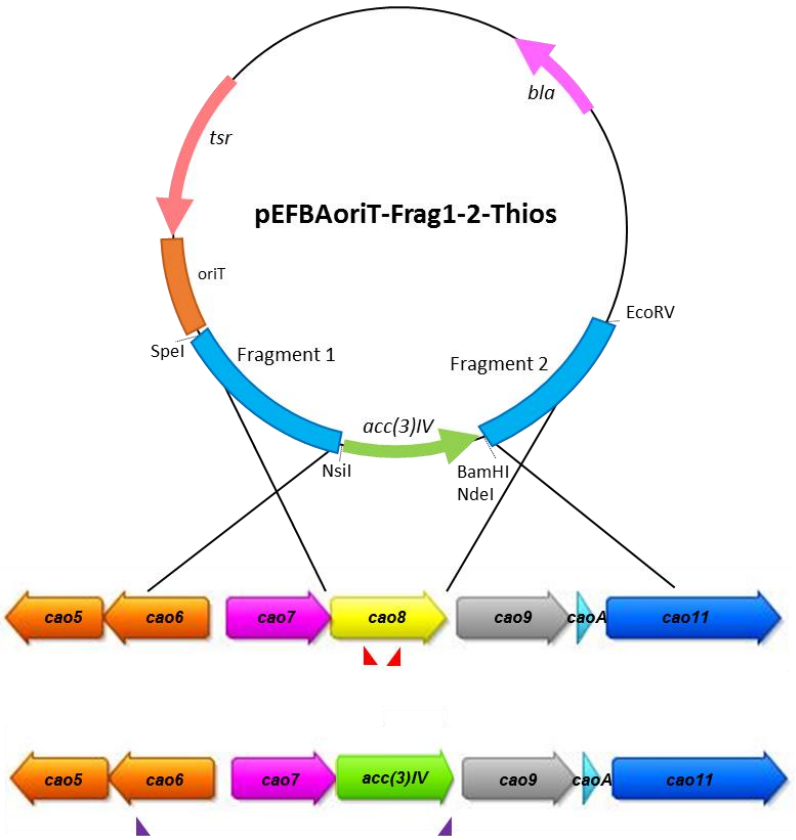


Figure D50. Scheme of the genetic replacement of cao8 *with the apramycin resistance cassette in the* cao *cluster. In the figure, the internal pair of primers Cao8-check-F and Cao8-check-R for the checking of the deletion of the gene are shown in red and the external and internal primers pEFBA-Cao8-Frag1-F and Apramycin-R for the checking of the insertion of the apramycin cassette are shown in purple.*

The positive transconjugants were fermented in 10 mL R2YE tubes, along with the heterologous host CA-300429 and the *cao8*-knockout CA-301537, and incubated at 28 °C, 220 rpm, 70 % humidity for 13 days. The extracts generated were analyzed by HRESIMS(+)-TOF and cacaoidin-2O was detected, indicating that the function of Cao8 was restored (Figure D51).

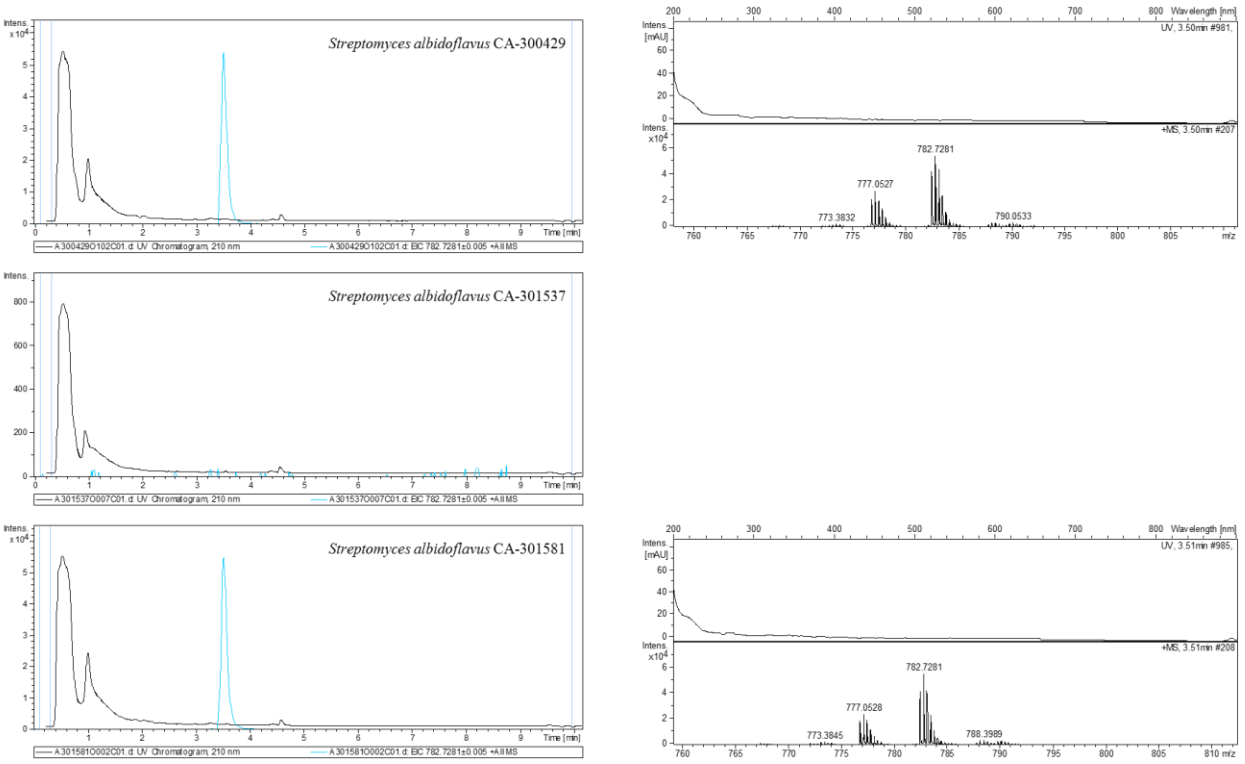


Figure D51. ***Left.*** *Chromatograms of UV absorbance at 210 nm (black trace) and extracted ion chromatogram at* m/z *= 782.7281 ± 0.005, $C_{107}H_{162}N_{24}O_{30}S_2 + NH_4^+ + 2H^+$ (clear blue trace) of cacaoidin-2O from heterologous producing strain* Streptomyces albidoflavus *CA-300429 (top), the* cao8*-knockout strain* Streptomyces albidoflavus *CA-301537 (middle) and the* cao8*-complemented strain* Streptomyces albidoflavus *CA-301581.* ***Right.*** *Experimental UV and positive mass spectra from $C_{107}H_{162}N_{24}O_{30}S_2 + NH_4^+ + 2H^+$ (calculated value: 782.7281) adducts from heterologous producing strain (top) and from cao8-complemented strain CA-301581 (bottom).*

3.3.3. *Generation of the knockout of the glycosyltransferase gene* cao16

For the generation of the knockout of *cao16*, encoding a GT-2 glycosyltransferase, fragments of 2.02 and 1.56 Kb, located up- and downstream *cao16,* respectively, were amplified using the pair of primers pEFBA-cao16-Frag1-F and pEFBA-cao16-Frag1-R, and pEFBA-cao16-Frag2-F and pEFBA-cao16-Frag2-R (Table C6), and they were digested with the pair of restriction enzymes SpeI/NsiI and BamHI/EcoRV, respectively, and sequentially cloned into the previously digested pEFBA-oriT plasmid. Finally, a thiostrepton resistance gene was also cloned to confirm the double crossover. The constructions were checked by digestion after each step of the cloning.

The final pEFBA-cao16 construction was transferred to the heterologous host *S. albidoflavus* CA-300429 via intergenic conjugation and the transconjugants obtained were checked based on the growing phenotype (resistance to apramycin and sensitivity to thiostrepton) and PCR amplification with Cao16-check-F and Cao16-check-R primers (Table C9), which amplify a 461 bp fragment, to check the deletion of the gene, and with pEFBA-cao16-Frag1-F and Apramycin-R primers (Tables C6 and C7, respectively), which amplify a 3.5 Kb fragment, to check the insertion of the apramycin cassette (Figure D52). Once checked, the positive transconjugants were fermented in 10 mL R2YE tubes, along with the heterologous host CA-300429, and incubated at 28 °C, 220 rpm, 70 % humidity for 13 days. The extraction was carried out as described in Section C 12.1.1 and the HRESIMS(+)-TOF analysis did not detect the presence of cacaoidin-2O or any of its putative monoglycosylated not aglycosylated intermediates (Figure D53).

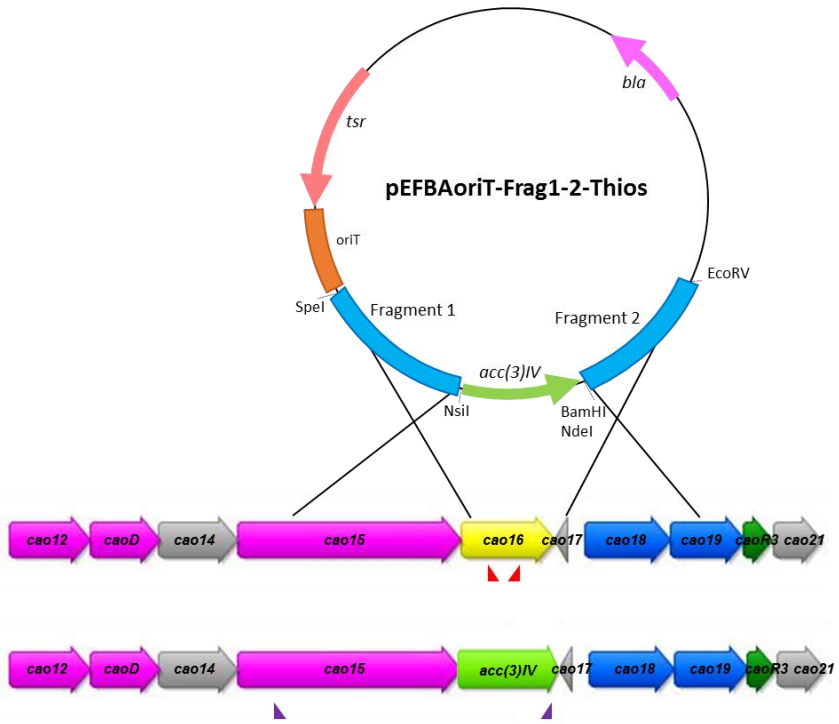


Figure D52. Scheme of the genetic replacement of cao16 *with the apramycin resistance cassette in the* cao *cluster. In the figure, the internal pair of primers Cao16-check-F and Cao16-check-R for the checking of the deletion of the gene are shown in red and the external and internal primers pEFBA-Cao16-Frag1-F and Apramycin-R for the checking of the insertion of the apramycin cassette are shown in purple.*

The resulted strain CA-303085 is one of the *cao16* knockout transconjugants which did not produce cacaoidin nor any variant of it. Complementation with pEM4T-cao16 was needed to study if *cao16* is involved in the production of cacaoidin-2O once its function is restored.

First, *cao16* was amplified using the primers Cao16-F and Cao16-R (Table C8) and digested with the restriction enzymes BamHI and EcoRI. This digested product, after purification, was cloned into the previously BamHI/EcoRI-digested pEM4T. The new construction was transferred to NEB 10-β *E. coli* cells and then checked by digestion with restriction enzymes. Once checked, it was transferred to spores of the *cao16* knockout strain CA-303085 by conjugation. The transconjugants were streaked into MA plates supplemented with thiostrepton and nalidixic acid and then checked by PCR amplification with the Cao16-check-F and Cao16-check-R (Table C9).

The positive transconjugants were fermented in 10 mL R2YE tubes, along with the heterologous host CA-300429 and the *cao16* knockout CA-303085 and incubated at 28 °C, 220 rpm, 70 % humidity for 13 days. The extracts generated were analyzed by HRESIMS(+)-TOF and cacaoidin-2O production was restored (Figure D193).

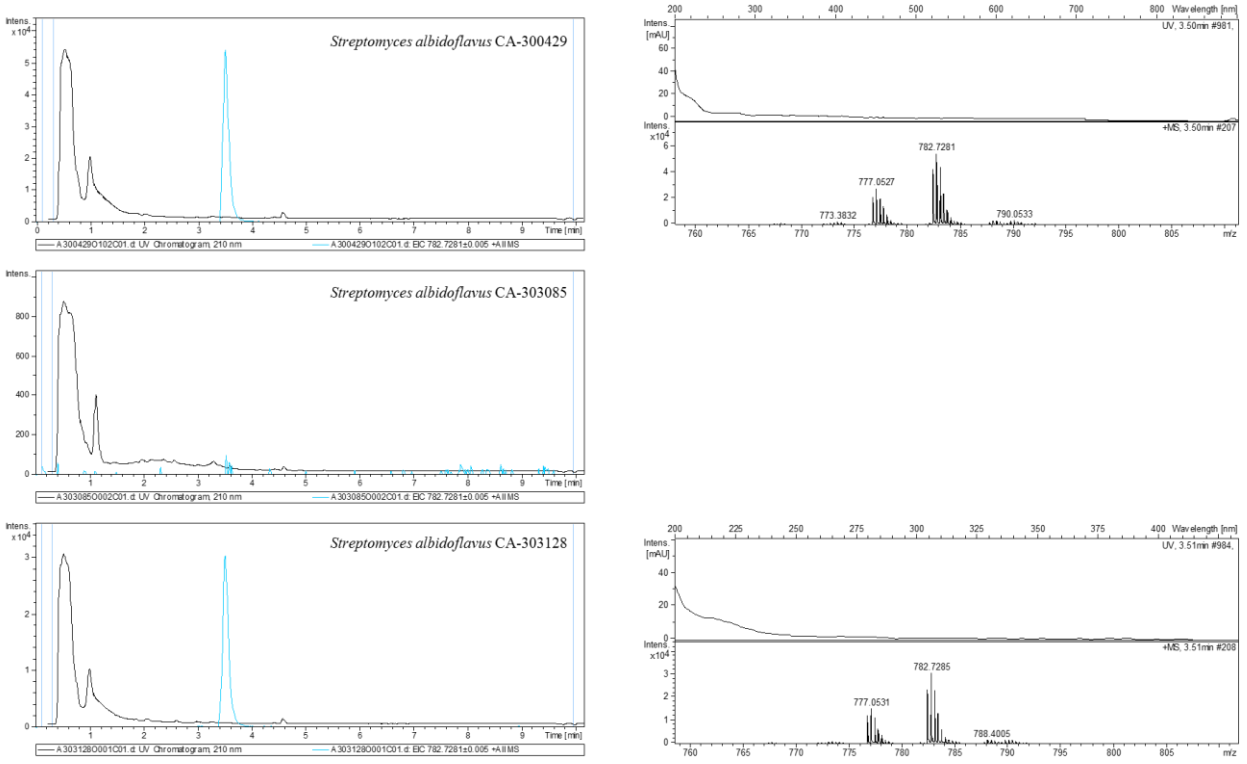


Figure D53. ***Left.*** *Chromatograms of UV absorbance at 210 nm (black trace) and extracted ion chromatogram at* m/z *= 782.7281 ± 0.005, $C_{107}H_{162}N_{24}O_{30}S_2 + NH_4^+ + 2H^+$ (clear blue trace) of cacaoidin-2O from heterologous producing strain* Streptomyces albidoflavus *CA-300429 (top), the cao16-knockout strain* Streptomyces albidoflavus *CA-303085 (middle) and the cao16-complemented strain* Streptomyces albidoflavus *CA-303128.* ***Right.*** *Experimental UV and positive mass spectra from $C_{107}H_{162}N_{24}O_{30}S_2 + NH_4^+ + 2H^+$ (calculated value: 782.7281) adducts from heterologous producing strain (top) and from cao16-complemented strain CA-303128 (bottom).*

3.3.4. Generation of the knockout of the glycosyltransferase gene cao24

To disrupt the GT-4 glycosyltransferase Cao24 encoding gene, the pair of primers pEFBA-cao24-Frag1-F and pEFBA-cao24-Frag1-R, and pEFBA-cao24-Frag2-F and pEFBA-cao24-Frag2-R (Table C6) were used to amplify two fragments located up and downstream of 2.086 and 1.809 Kb. These two DNA fragments were cloned into the previously digested pEFBA-oriT plasmid, along with the thiostrepton resistance gene. The constructions were checked by digestion after each step of cloning.

The final pEFBA-cao24 was transferred to spores of the heterologous host *S. albidoflavus* CA-300429 via intergenic conjugation and the transconjugants obtained were selected based on their sensitivity to antibiotics (resistance to apramycin and sensitivity to thiostrepton) and on the PCR amplification with Cao24-check-F and Co24-check-R primers (Table C9), which amplify a 579 bp fragment, to check that the gene was deleted, and with pEFBA-cao24-Frag1-F and Apramycin-R primers (Tables C6 and C7, respectively), which amplify a 3.5 Kb fragment, to check that the apramycin cassette had replaced the gene (Figure D54).

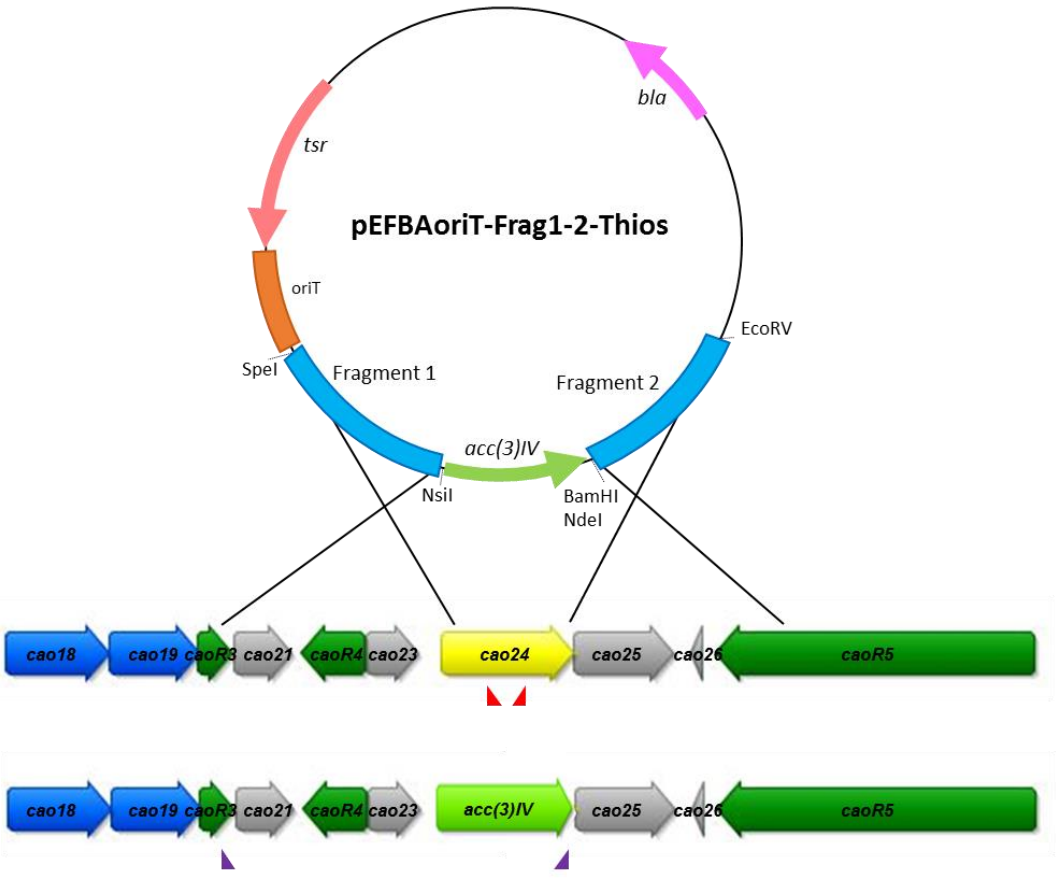


Figure D54. Scheme of the genetic replacement of cao24 *with the apramycin resistance cassette in the* cao *cluster. In the figure, the internal pair of primers Cao24-check-F and Cao24-check-R for the checking of the deletion of the gene are shown in red and the external and internal primers pEFBA-Cao24-Frag1-F and Apramycin-R for the checking of the insertion of the apramycin cassette are shown in purple.*

Once checked, the right transconjugants were fermented in 10 mL R2YE tubes and incubated at 28 °C, 220 rpm, 70 % humidity for 13 days. After the generation of extracts as described in Section C 12.1.1, the HRESIMS(+)-TOF analysis was carried out, where the cacaoidin-2O was still detected at the same levels than the heterologous strain CA-300429 (Figure D55). This fact suggest that Cao24 is not involved in the glycosylation of cacaoidin.

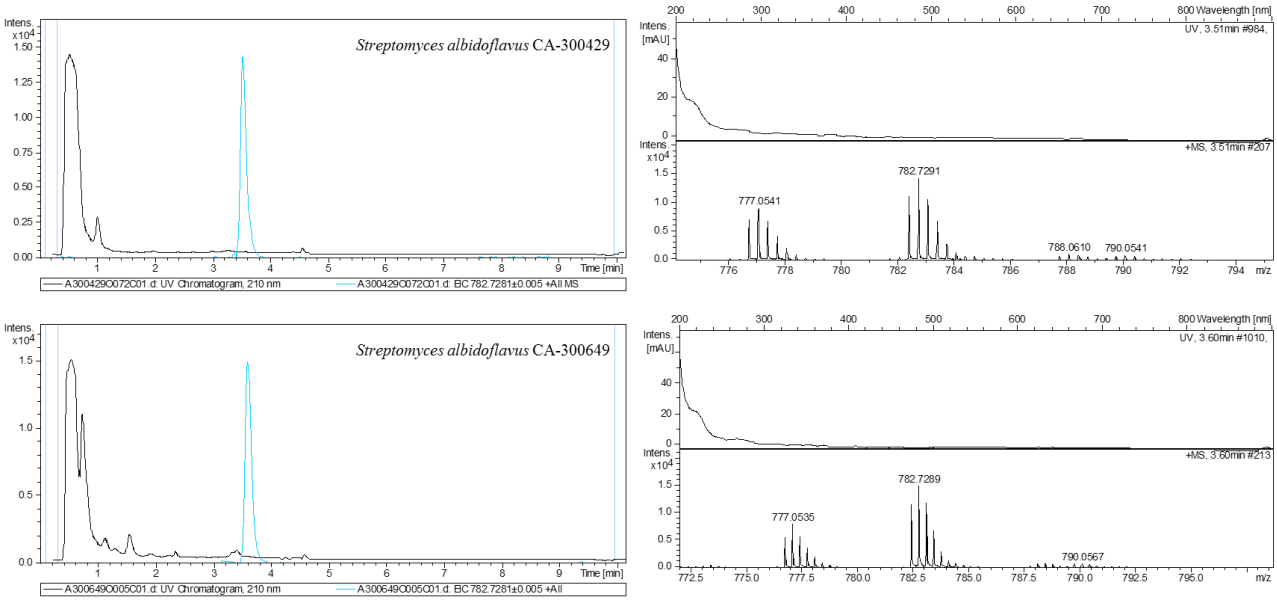


Figure D55. ***Left.*** *Chromatograms of UV absorbance at 210 nm (black trace) and extracted ion chromatogram at* m/z *= 782.7281 ± 0.005, $C_{107}H_{162}N_{24}O_{30}S_2 + NH_4^+ + 2H^+$ (clear blue trace) of cacaoidin-2O from heterologous producing strain* Streptomyces albidoflavus *CA-300429 (top), the cao24-knockout strain* Streptomyces albidoflavus *CA-300649 (bottom).* ***Right.*** *Experimental UV and positive mass spectra from $C_{107}H_{162}N_{24}O_{30}S_2 + NH_4^+ + 2H^+$ (calculated value: 782.7281) adducts from heterologous producing strain (top) and from cao24-knockout strain CA-300649 (bottom).*

3.4. Generation of cacaoidin aglycon from cacaoidin-2O

As none of the knockouts performed with the glycosyltransferases genes led to the production of cacaoidin aglycon, other methods were tested to generate it. As cacaoidin is susceptible to be hydrolyzed at acid conditions (Ortiz-López *et al.,* 2020), different approaches were used to hydrolyze the molecule, but only the incubation of cacaoidin-2O with 35.6 % ACN/H_2O containing 0.1 % TFA at 4 °C for ten days successfully generated aglycon. This transition from cacaoidin-2O to aglycon was monitored by LC-HRESIMS(+)-TOF every few days until the whole cacaoidin-2O was converted into aglycon. Surprisingly, the incubation of the original cacaoidin under the same conditions did not generate aglycon.

Once the conversion was completed, the aglycon was purified to remove the possible waste products generated during the reaction. A semipreparative RP-HPLC was performed using an Agilent Zorbax SB-C_8 (9.4 x 250 mm, 5 µm) column and a linear H_2O-CH_3CN gradient (both solvents containing 0.01 % trifluoroacetic acid, TFA) with an initial 34 % and a final 42 % CH_3CN

for 35 min, and ending with 10 min at 100 % CH_3CN, collecting 80 fractions in a 96-deepwell-plate (Figure D56).

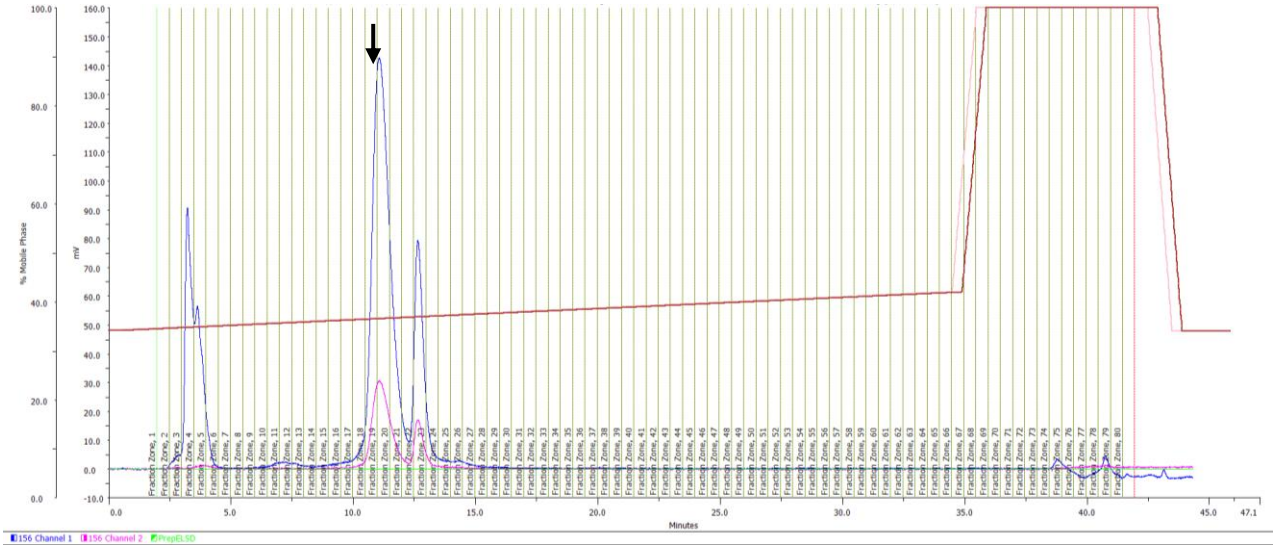


Figure D56. HRMS-guided isolation of the cacaoidin aglycon. Semipreparative RP-HPLC purification of cacaoidin -2O led to the purification of the aglycon. LC-UV preparative chromatogram recorded at both 210 nm (blue trace) and 280 nm (pink trace).

The fractions 17 to 21 were collected and pooled, as they corresponded with the aglycon. The organic solvent of this pool of fractions were evaporated obtaining a final 0.3 mg of aglycon.

E. Discussion

As a result of the continued natural products discovery programs developed at Fundación MEDINA, the strain *Streptomyces cacaoi* CA-170360 was known to produce cacaoidin, the first member of class V lanthipeptides (Ortiz-López *et al.,* 2020) and pentaminomycins A-H and BE-18257 antibiotics (Carretero-Molina *et al.,* 2021) in a liquid fermentation of the strain in MPG medium after 13 days of incubation, but no further studies had been performed to date on the production conditions of all these bioactive molecules.

Both cacaoidin and pentaminomycins are novel and interesting structures and show bioactivity against Gram-positive pathogens (methicillin-resistant *Staphylococcus aureus* MRSA and *Clostridium difficile*) (Ortiz-López *et al.,* 2020) and Gram-negative pathogens (*Acinetobacter baumannii*) (Carretero-Molina *et al.,* 2021) respectively. Thus, the whole genome sequence of the producer strain *Streptomyces cacaoi* CA-170360 and its biosynthetic analysis by antiSMASH v6.0.1. (Blin *et al.,* 2021a), identifying as many as 33 predicted BGCs, provided a much deeper insight of the potential of this strain to produce additional secondary metabolites.

OSMAC approach in *Streptomyces cacaoi* CA-170360

The One Strain Many Compounds (OSMAC) strategy has been shown as a simple and powerful first approach to activate silent gene clusters and increase the low production titers of molecules of interest in microorganisms (Bode *et al.,* 2002). Production of microbial metabolites is dependent on culture conditions, such as media composition, temperature, pH, salinity, culture status, availability of biosynthetic precursors, addition of signaling molecules, as well as the interaction with other microbial strains such as in co-culturing experiments. Modifications of these factors may result in an improvement of the production and in a change in the diversity of secondary metabolites. Traditional culture methods using just a few limited number of media only ensure the biosynthesis of a small number of secondary metabolites from all those encoded in the genome of a microbial strain. An OSMAC strategy considering all these parameters can provide a simple, quick and effective approach for enhancing the number of wide chemical diversity of secondary metabolites by activation of silent gene clusters, and in consequence to improve the chemical diversity produced by microbial strains in culture conditions (Pan *et al.,* 2019).

This OSMAC approach was therefore applied to identify the best production conditions of each family of compounds by testing six production media (YEME, R2YE, KM4, MPG, FR23 and DEF-15) and different fermentation times up to 28 days of incubation. These experiments performed in small volume fermentation vials permitted to identify improved production conditions for cacaoidin and pentaminomycins A-H respect to the original detection medium.

The highest cacaoidin production rate was detected in FR23, R2YE and KM4 media after 17 days of incubation, while the production was much lower in the original MPG medium where cacaoidin was first detected. On the other hand, the OSMAC study showed that, interestingly, the production of the BE-18257 antibiotics were higher in the KM4, FR23 and MPG media while pentaminomycins A-H are poorly produced in these media and, on the other hand, in the media where the BE-18257 antibiotics were detected in lower quantities (YEME and R2YE), the

production of pentaminomycins A-H were the highest, suggesting that these media ensure the biosynthesis of pentaminomycins to the detriment of the BE-18257 molecules.

To our knowledge, CA-170360 is the first strain reported to produce all pentaminomycins A-H described to date and the three BE-18257 antibiotics (Carretero-Molina *et al.,* 2021), although BE-18257 B and C, and pentaminomycins C and H coelute in the LC-HRMS analytical conditions employed and the peak might contain either isomer or a mixture of both. Kaweewan *et al.* (2020) only detected the antibiotics BE-18257 A and B and pentaminomycin C in *S. cacaoi* NBRC 12748^{T}, while Hwang *et al.* (2020) reported the detection of the antibiotics BE-18257 A and B and pentaminomycins C-E in *S. cacaoi* GG23. Pentaminomycins A and B were first described by Jang *et al.* (2018) from cultures of *Streptomyces* sp. RK88-1441, but no BE-18257 antibiotics were reported. These strains may have the capacity to synthesize all the pentaminomycins and BE-18257 antibiotic variants detected in strain CA-170360, but most probably the culture conditions used did not ensure the production of all the analogs.

The antiSMASH analysis of the genome of *S. cacaoi* CA-170360 predicted 33 putative BGCs in addition to the *cpp* and *cao* gene clusters, whereof seven had a similarity above 70 % with clusters already known to be involved in the biosynthesis of the natural products actinonine, bezastatin, naseseazine C, puromycin, desferrioxamine E, candicidin, ectoine or any potential derivates, suggesting the high biosynthetic potential of this *S. cacaoi* strain.

The OSMAC study has also allowed us to confirm from the LC-HRESIMS(+)-TOF analysis the presence of puromycin, proferrioxamine D2, nocardamine or desferrioxamine E, terragine E or dehydronocardamine, actinonin, ferrioxamine E, and naseseazines A and B. However, no traces of candicidin, bezastatin derivates or ectoine were detected in any of the fermentation conditions tested, suggesting that these pathways might be cryptic or silent in the strain or that the compounds are only produced at trace level in the culture conditions used below the minimal detection levels.

These results evidence the importance of combine the genome mining analysis with an OSMAC-based approach in order to exploit the whole secondary metabolites potential of the strains, establish the optimal conditions for the highest production of the compounds and better understand the genetic and regulatory mechanism controlling their biosynthesis (Almeida *et al*., 2019).

Characterization and heterologous expression of the cluster responsible of pentaminomycins A-F and antibiotics BE-18257

From the 33 BGCs predicted, the 48-Kb *ccp* cluster was proposed to be putatively responsible of the production of both pentaminomycins A-H and BE-18257 antibiotics as it contained two NRPS genes *cpp*B and *cpp*M, each of them containing five adenylation (A) domains, with substrate specificities matching those required for the biosynthesis of both types of cyclic pentapeptides. Additional ORFs between both NRPS genes are also found.

The gene organization of the *cpp* cluster and the amino acid incorporation by each domain of both NRPS genes was highly similar to those also recently proposed, during the course of this

study, for these BGCs in the pentaminomycin C producer *Streptomyces cacaoi* subsp. *cacaoi* strains NBRC 12748^T and the pentaminomycins C-E producer *Streptomyces* sp. GG23 (Kaweewan *et al.,* 2020; Hwang *et al.,* 2020). The first NRPS gene (*cppB*) was proposed to determine the synthesis of a peptide with the amino acid sequence L-Leu, D-Trp, D-Leu/Ser, L-Ala, D-Val/Leu, assuming fulfillment of the co-linearity rule, which corresponded well with the amino acid sequence of BE-18257 A-C (L-Leu, D-Trp, D-Glu, L-Ala, D-Val/D-*allo*-Ile/D-Leu). Therefore, these results suggested that the first NRPS gene (*cpp*B*)* is involved in the biosynthesis of BE-18257 A–C antibiotics and that further cyclization would be required to complete the biosynthesis of these molecules. On the other hand, the second NRPS gene (*cppM)* would generate the following amino acid sequence L-l/Leu/Phe, D-Val, L-Trp, L-Arg, D-Leu/Phe, which agrees with the amino acid sequence of pentaminomycins A-H (L-Val/L-Leu/L-Phe, D-Val, L-Trp, L-N5-OH-Arg, D-Leu/D-Phe/D-2-Pal). Subsequent modifications such as hydroxylation and cyclization would be needed to complete the biosynthesis of the pentaminomycins A-H.

The heterologous expression of the *cpp* cluster was achieved in this work, cloning separately two different fragments of the BGC, containing both NRPS genes or only the first one *cpp*B. The analysis of the extracts of the heterologous hosts clearly demonstrated that the *cppB* is responsible of the biosynthesis of BE-18257 antibiotics and *cppM* is responsible of the synthesis of pentaminomycins. Not all the pentaminomycins were detected when expressed in the heterologous strain *Streptomyces albidoflavus* J1074 (pentaminomycins B, F and G were not detected). Pentaminomycins F and G contain the unusual amino acid 3-(2-pyridyl)-alanine (2-Pal), also named 2'-aza-L-phenylalanine (Carretero-Molina *et al.,* 2021). This amino acid was previously isolated from cultures of *Streptomyces* sp. SF2538 (Iwata, 1988). However, its presence within a peptide was reported only for the first time in these pentaminomycins, highlighting the structural novelty of these compounds. It may be assumed that amino acid 2-Pal is biosynthesized by the strain CA-170360, as it is not present in the medium, and then recruited by the biosynthetic pathway of pentaminomycins. Although nothing is known about the biosynthesis of this amino acid (Carretero-Molina *et al.,* 2021; Huang *et al.,* 2013), it is possible that *Streptomyces albidoflavus* J1074 does not possess the genes necessary for the synthesis of the amino acid 2-Pal and therefore is unable to produce pentaminomycins F and G.

One of the peculiarities of this *cpp* cluster is the lack of canonical thioesterase domains in the C-termini of both NRPS genes, that should be involved in the cyclization and release of the pentapeptides. The penicillin-binding protein-like protein gene (*cpp*A) identified upstream the gene *cppB* is a member of the recently described family of peptidyl NRP offloading cyclases included in the β-lactamase superfamily (Kuranaga *et al.,* 2018; Matsuda *et al.,* 2019a; 2019b) that act *in trans* to assemble peptide chains. This stand-alone enzyme could be involved in the release of peptide chains of both BE-18257 antibiotics and pentaminomycins from the PCP domains and in its macrocyclization to generate the cyclic peptides, as previously described previously (Zhou *et al.,* 2019; Kaweewan *et al*., 2020). The generation of a *cppA* knockout in this work has confirmed that this PBP-type TE is responsible of the release and cyclization of the BE-18257 antibiotics and pentaminomycins, as none of the cyclic pentapeptides were synthesized in the knockout while the biosynthesis of some of the pentaminomycins was restored in the complemented strain. This agrees with the *in vitro* studies recently published by Matsuda *et al*

(2021) where they demonstrated the macrocyclization activity of CppA (called PenA in their study). All these data validate CppA as a new member of the PBP-type TEs.

The production of BE-18257 antibiotics and pentaminomycins A and E could not be restored after complementation of the Δ*cppA* mutant with a high-copy number plasmid (pEM4T) harboring the gene *cpp*A under the control of strong constitutive promoter (*ermE**p). Only pentaminomycins C/H and D restored their production after complementation. Imbalance between copy-number and strength of promoters has been suggested previously to explain why complementation cannot ensure the production of all the compounds. Bilyk *et al*. (2016) failed to restore the production of grecocycline in a mutant by complementation of the native *greM2* under the control of the strong synthetic 21p promoter into the pUWL vector. The lack of production of grecocycline was suggested as the result of an unbalanced expression of *greM2* and the rest of the genes in the *gre* cluster from different plasmid construction. It has already been shown that the balanced expression of genes encoding enzymes from a biosynthetic gene cluster is critical for production levels (Du *et al.,* 2012).

Matsuda *et al* (2021) also compared CppA (PenA) with the recently described SurE, a stand-alone enzyme present in the surugamide BGC. Surugamides A-E are cyclic octapeptides with four D-amino acids from the marine-derived *Streptomyces* sp. JAMM992 (Takada *et al.,* 2013). Surugamide F is a decapeptide described later by Ninomiya *et al* (2016). The biosynthetic *sur* cluster is formed by four NRPS genes (*surA-D*) where *surA* and *surD* are responsible of the biosynthesis of surugamides A-E while *surB* and *surC* synthesizes surugamide F (Ninomiya *et al.,* 2016). This *sur* cluster also lacks domains for the typical peptide chains termination but it contains a PBP-type protein (*surE*) encoded upstream the NRPS genes responsible of the macrocyclization of surugamides A-E (Kuranaga *et al.,* 2018; Matsuda *et al.,* 2019a; 2019b; Thankachan *et al.,* 2019). This is very similar to what happens in the *cpp* cluster where a unique BGC encodes the biosynthetic routes for two different peptides, which are released and cyclized by a single penicillin-binding protein. The main difference between the two PBP-type TEs lies in the distinct preferences for the substrate chain length, the native substrates for CppA are pentapeptides, whereas those for SurE are octa- and decapeptides (Matsuda *et al.,* 2021).

This offloading PBP-type cyclase has been also reported in other NRPS pathways such as those of desotamide (Fazal *et al.,* 2020), ulleungmycin (Son *et al.,* 2017), noursamycin (Mudalungu *et al.,* 2019), curacomycin (Kaweewan *et al.,* 2017) or mannopeptimycin (Magarvey *et al.,* 2006). Enzyme-mediated peptide cyclization has been an intense researched area for many years, particularly the enzymatic cyclizations of medium and large sized peptides. However, the use of enzymes to cyclize small peptides, such as tetra- and pentapeptides, is considerably less characterized (Mandalapu *et al.,* 2018; Sardar *et al.,* 2015) as small peptides are beyond the scope of the currently available, well-known peptide cyclases, such as sortase A (Pishesha *et al.,* 2018) or butelase (Nguyen *et al.,* 2016), whose substrates normally have ten or more residues (Schmidt *et al.,* 2017).

Both BE-18257 antibiotics and pentaminomycins A-H contain tryptophan in their structures. The genome of *S. cacaoi* CA-170360 also contains some genes related to tryptophan biosynthesis downstream from the second NRPS gene (*cppM*), as it was described by Hwang *et al.* (2020) for strain GG23. These genes share a high homology with the previously described

genes for tryptophan biosynthesis (Radwanski *et al.,* 1995). Hwang proposed that these compounds may share the Trp biosynthetic genes to incorporate this amino acid to their structures (Hwang *et al.,* 2020). However, the results from this study clearly show that those genes are not required to incorporate Trp in the cyclic pentapeptides, since they were not included in the fragment cloned into pCPP1 and in pCPP2, and the pentaminomycins and BE-18257 were still produced. This indicates that the tryptophan, as well as the rest of the amino acids, are obtained from the primary metabolism amino acid pool.

The novel family of pentaminomycins A-H are also characterized by the presence of N^5-hydroxyarginine, a post-modular modification that would finalize their biosynthesis. The hydroxylation of hydrocarbon C-H bonds is a very difficult transformation. This reaction is catalyzed by a variety of metalloenzymes, among which the most diverse are many members of the cytochrome P450 family (Ortiz de Montellano, 2010). The *cpp* cluster includes two ORFs (*cppI* and *cppJ*) encoding cytochrome P450 enzymes, which have been suggested to be involved in the N-hydroxylation of arginine to form 5-OH-Arg in pentaminomycins (Kaweewan *et al.,* 2020; Hwang *et al.,* 2020). The pathway also contains regulatory genes and other genes of unknown function.

Characterization of the cacaoidin BGC

At the beginning of this work in 2017, the analysis of the whole genome sequence of *S. cacaoi* with antiSMASH (v4.0.1) could not predict any BGC as candidate for the cacaoidin biosynthesis, despite of the extensive lanthipeptide-specific predictive capabilities of the algorithm for the lanthipeptide class, core peptide cleavage functions, tailoring reactions, and Lan/MeLan bridges (Blin *et al.,* 2014; 2017). The current improved version of antiSMASH (v6.0.1) (Blin *et al.,* 2021) can effectively predict the cacaoidin BGC as a class V-lanthipeptide in the genome of *S. cacaoi* CA-170360. The cacaoidin cluster could not be identified when other RiPP-predictive specialized tools, such as BAGEL4, PRISM and RiPPMiner, were also used, suggesting that the identification of novel RiPPS NPs using genome mining is still a challenge (Kloosterman *et al.,* 2020).

Since the available bioinformatic tools were unable to identify the cacaoidin BGC, the C-terminal sequence of cacaoidin (Thr-Ala-Ser-Trp-Gly-Cys) was used as the query in a tBLASTn using the whole genome sequence to find a 162 bp Open Reading Frame (ORF) containing this sequence that helped to elucidate the final sequence of the peptide (Ortiz-Lopez *et al.,* 2020). Considering the structure of cacaoidin, a putative 30 Kb BGC (*cao* cluster) was identified containing 27 ORFs. The proposed functions were assigned after BLAST analysis and HHpred secondary structure prediction, enabling to putatively assign each ORF to a role in the PTMs of cacaoidin core peptide involving the AviMeCys ring and lanthionine formation, the terminal *N,N*-methylation, the incorporation of D-amino acids, the disaccharide biosynthesis and the tyrosine glycosylation.

Interestingly, neither the leader nor the core peptide sequences showed high sequence similarity with any other lanthipeptide or linaridin and no homologous genes of known dehydratases or cyclases commonly found in the four classes described for lanthipeptides nor in

the class of linaridins could be identified in this region. Moreover, the similarity with the recently described lexapeptide BGC (*lxm*) (Xu *et al.,* 2020) contributed as well to assigning some of the functions of the *cao* ORFs. Lexapeptide is a new class V lanthipeptide produced by *Streptomyces rochei* Sal35 which shares several structural similarities with cacaoidin, including the N-terminal *N,N*-dimethyl-alanine, C-terminal AviMeCys ring and the presence of D-Ala.

CaoD is a putative cypemycin decarboxylase CypD homologue containing a conserved phosphopantothenoylcysteine (PPC) synthetase/decarboxylase domain. CaoD shows little sequence similarity with CypD and LanD enzymes, both belonging to the HFCD protein family and involved in the catalysis of the oxidative decarboxylation of the C-terminal cysteine residue in the presence of a flavin cofactor (Ding *et al.,* 2018a; Sit *et al.,* 2011). The presence of the PPC domain supports the potential role of CaoD in the oxidative decarboxylation. In addition to this, CaoD shares a 54.8 % similarity with LxmD, an enzyme involved in the decarboxylation of lexapeptide in its C-terminal Cys residue during the formation of the AviMeCys ring (Xu *et al.,* 2020), suggesting the same role for CaoD.

The formation of lanthionine rings is accomplished by different dehydratases and cyclases depending on the lanthipeptide class (Arnison *et al.,* 2013). In class I, a dehydratase (LanB) generates the Dha and Dhb and a cyclase (LanC) adds the Cys thiol. In class II, a single modification enzyme (LanM) is involved that contains an N-terminal dehydratase domain and a C-terminal LanC-like domain. In classes III and IV lanthionine rings are produced also by a single enzyme, called LanKC for class III and LanL for class IV. Both enzymes show an N-terminal phospho-Ser/phosphor-Thr lyase domain, a central kinase-like domain and a C-terminal cyclase domain which contains Zn-binding ligands only in LanL (Repka *et al.,* 2017). Surprisingly none of the ORFs present in the *cao* cluster showed any homology with the previously described LanC, LanM, LanKC or LanL proteins.

In the lexapeptide biosynthesis, the lanthionine moiety is formed by a three-component lanthionine synthetase including the standalone monofunctional proteins LxmK, LxmX and LxmY, which are responsible for the Ser/Thr dehydration and cyclization to form Dha and Dhb, respectively, and the installation of the lanthionine ring (Xu *et al.,* 2020). LxmK is a putative kinase which phosphorylates Ser/Thr residues on the LxmA precursor peptide. Then LxmX, which is a phosphotransferase homolog, acts as a phosphoserine/phosphothreonine lyase to remove the phosphate groups, generating dehydro-amino acids. Meanwhile, the C-terminal Cys is decarboxylated by LxmD. Finally, LxmY is a cyclase that catalyzes the Michael-type addition of the Cys22 thiol group to the β-carbon of Dha19 to generate the lanthionine ring, and likely the addition of decarboxylated Cys38 to Dhb33 to generate AviMeCys (Xu *et al.,* 2020).

These proteins present a certain degree of homology with some proteins encoded in the *cao* cluster: LxmK is a putative kinase that shows a 49.6 % similarity with Cao9, and LxmX and LxmY are unknown proteins that have 40.5 % and 44.9 % similarities with Cao14 and Cao7, respectively. Interestingly, LxmY and Cao7 contain a HopA1 conserved domain (PFAM17914) that has been described in the HopA1 effector protein from *Pseudomonas syringae* (Park *et al.,* 2015) and was shown to directly bind the Enhanced Disease Susceptibility 1 (EDS1) complex in *Arabidopsis thaliana,* activating the immune response signaling pathway. Future research is still

needed to determine the function of this protein that can only be tentatively proposed as potential new type of lanthionine synthetase.

According to these data, Cao7-9-14 would also work as the three-component lanthionine synthetase: the putative kinase Cao9 would phosphorylate Ser/Thr residues on the CaoA precursor peptide, Cao14 would remove the phosphate groups and generate dehydro-amino acids, and Cao7 would catalyze a Michael-type addition of the Cys5 thiol group to the β-carbon of Dha1, generating the lanthionine bridge, and likely the addition of previously decarboxylated-Cys23 by CaoD to Dhb18, forming AviMeCys.

The N-terminal Ala dimethylation of cacaoidin is a characteristic of the family of linaridins. The prototypical member of linaridins is cypemycin (Claesen and Bibb, 2010), and its *N,N*-methylation is carried out by the S-adenosylmethionine (SAM)-dependent methyltransferase CypM (Ma *et al*., 2020; Mo *et al.,* 2017; Claesen and Bibb, 2011) and shows homology with the methyltransferase LxmM that catalyzes the lexapeptide N-terminal dimethylation (Xu *et al.,* 2020). Within the *cao* cluster, *cao4* encodes a putative O-methyltransferase containing the conserved Methyltransf_2 domain. However, despite the low homology found between Cao4 and CypM (15.2 % similarity) and LxmM (26.9 % similarity), the three proteins share the AdoMet_MTases superfamily domain, characteristic of the SAM-dependent methyltransferases. No other CypM or LxmM homologues have been found in the genome of strain CA-170360. Many class I lanthipeptide clusters from actinobacteria contain an O-methyltransferase, generically known as LanS. Two types of LanS enzymes have been described: $LanS_A$, which incorporates β-amino acid isoaspartate (Acedo *et al.,* 2019) and $LanS_B$, which methylates the C-terminal carboxylate of a RiPP precursor (Huo *et al.,* 2019). Cao4 shows very low homology with both types of LanS proteins. Thus, its role in the *N,N*-methylation has been confirmed in this work by the generation of a *cao4*-knockout and it will be discussed below.

The unusually high content of D-amino acids in cacaoidin is another of the characteristic features in its structure. D-amino acids provide a wide variety of properties to lanthipeptides such as resistance to proteolysis, induction of bioactivity or structural conformation (Yang *et al.,* 2015). However, only L-amino acids can be added by the ribosomal machinery, so the way to introduce D-stereocenters into lanthipeptides is modifying the genetically encoded L-Ser and L-Thr into Dha and Dhb, which are subjected to a diastereoselective hydrogenation to finally incorporate D-Ala and D-Abu, respectively (Repka *et al.,* 2017; Huo *et al.,* 2019). This reaction is carried out by dehydrogenases generically called LanJ (Repka *et al.,* 2017), which are divided in two classes, zinc-dependent dehydrogenases ($LanJ_A$) and flavin-dependent dehydrogenases ($LanJ_B$). $LanJ_B$ is able to reduce both Dha and Dhb, whereas $LanJ_A$ can only hydrogenate Dha. To date, only two $LanJ_B$ enzymes have been characterized, $CrnJ_B$ and $BsjJ_B$, involved in the biosynthesis of carnolysin (Lohans *et al.,* 2014) and bicereucin (Huo and van der Donk, 2016), respectively. Recently, another flavin-dependent oxidoreductase ($LahJ_B$) has been described in the putative lanthipeptide biosynthetic gene cluster *lah* (Huo *et al.,* 2019).

Xu *et al*. (2020) reported LxmJ as the first example of the $LanJ_C$ class in lanthipeptides. LxmJ is a novel $F_{420}H_2$-dependent reductase that catalyzes the stereospecific reduction of Dha28 to D-Ala28 in the mature lexapeptide. Within the *cao* BGC, Cao12 shows a 58.6 % similarity with LxmJ, and both proteins contain a $F_{420}H_2$-dependent oxidoreductase (MSMEG_4879 family)

conserved domain. This suggests that Cao12 could be also involved in the incorporation of D-amino acids (four D-Ala and one D-Abu) in the cacaoidin structure and be part of the new of $LanJ_C$ family. It has been demonstrated that LxmJ is unable to reduce Dhb when a S28T point mutation is introduced in LxmA (Xu *et al.,* 2020). However, as it was shown for $BsjJ_B$ in bicereucin biosynthesis (Huo and van der Donk, 2016), the reduction may be dependent of the position and/or sequence, so the incorporation of D-Abu in cacaoidin could also be performed by Cao12.

The disaccharide moiety β-6-deoxygulopyranosyl-(1→3)-α-rhamnopyranoside of cacaoidin has not been previously reported in any other natural product (Ortiz-López *et al.,* 2020). Four proteins are required for the synthesis of α-L-rhamnose: a Glucose-1-phosphate thymidylyltransferase (RmlA), a dTDP-D-glucose 4,6-dehydratase (RmlB), a dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase (RmlC) and a dTDP-4-keto-6-deoxy-L-mannose reductase (RmlD), although the corresponding genes do not have to be necessarily clustered (Girauld and Naismith, 2000).RmlA, RmlB and RmlD homologues are encoded in the *cao* cluster (*cao6*, *cao5* and *cao3*, respectively), missing a RmlC homologue; however, a *rmlC* gene and additional *rmlA, rmlB* and *rmlD* homologous genes are located outside the cacaoidin cluster in the genome of the strain CA-170360, confirming that this strain CA-170360 is able to synthesize L-rhamnose.

Bleomycin, tallysomycin and zorbamycin incorporate NDP-L-gulose or NDP-6-deoxy-L-gulose to their structures, and their biosynthesis was used as reference to look for the presence of similar proteins being encoded in our genome (Chen *et al.,* 2020; Galm *et al.,* 2011). Despite no homologues being found in the cacaoidin BGC, some protein homologues in the genome were identified such as a D-glycero-beta-D-manno-heptose 1-phosphate adenylyltransferase homolog, a NAD-dependent epimerase/dehydratase, a GDP-mannose 4,6-dehydratase and dTDP-glucose 4,6-dehydratase. However, as none of them were clustered, no conclusions could be made for the β-l-6-deoxy-gulose biosynthesis.

In the heterologous host *Streptomyces albidoflavus* J1074 the original cacaoidin with the β-6-deoxygulopyranosyl-(1→3)-α-rhamnopyranoside disaccharide was biosynthesized. The three genes required for the synthesis of L-rhamnose present in the *cao* BGC may be responsible of the generation of the rhamnose. Furthermore, in the genome of *S. albidoflavus* J1074, additional biosynthetic rhamnose genes are also found clustered, which might as well produce the rhamnose required for cacaoidin's disaccharide. On the other hand, no homologues from the bleomycin, tallysomycin or zorbamycin BGCs were found in the *S. albidoflavus* genome. However, as the biosynthetic pathway to produce gulose has not been described yet, it was assumed that the strain possesses the necessary genes for the generation of the gulose moiety.

The disaccharide α-L-rhamnose-β-L-6-deoxy-gulose of cacaoidin is O-linked to the phenolic oxygen of the aromatic ring of the tyrosine residue (Ortiz-López *et al.,* 2020). While asparagine N-glycosylation and serine, threonine or hydroxyproline O-glycosylation have been reported in many natural glycopeptides (Varki *et al.,* 1999), the O-glycosylation of tyrosine is not common. In prokaryotes, the O-glycosylation of tyrosine residues has been reported in the S-layer of the cell envelope of *Paenibacillus alvei*, *Thermoanaerobacter thermohydrosulfuricus* and *Thermoanaerobacterium thermosaccharolyticum* strains. In *P. alvei* CCM 2051^T, a polymeric branched polysaccharide is O-glycosidically linked *via* an adaptor to specific tyrosine residues of the S-layer protein SpaA by the O-oligosaccharyl:protein transferase WsfB (Messner *et al.,* 1995).

This protein is encoded within the *slg* cluster, that carries the genes necessary for the biosynthesis of this glycan chain. The *cao* cluster lacks an homologue of WsfB, so a candidate that O-glycosylates the tyrosine residue of cacaoidin coud not be proposed.

O-glycosylation of tyrosine residues is quite uncommon in natural products. Up to date, very few natural products undergo a tyrosine O-glycosylation. The first natural products containing a tyrosine O-glycosylation described were the lipoglycopeptide antibiotics mannopeptimycins, produced by *Streptomyces hygroscopicus,* which contain an O-linked di-mannose in a reaction catalyzed by the mannosyltransferases MppH and MppI (He *et al.,* 2002). Zhang *et al* (2009) isolated cycloaspeptide F from a *Cordyceps*-colonizing isolate of the fungus *Isaria farinosa*, which includes an O-glycosylated tyrosine with glucose. Gausemycins are cyclic lipoglycopeptides from *Streptomyces roseoflavus* INA-Ac-5812 and are the first natural products containing an O-glycosylation with a pentose, arabinose, mediated by the putative glycosyltransferase family 2 protein Orf17 from *gau* BGC (Tyurin *et al.,* 2021). The most recently described natural product containing an O-glycosylated tyrosine are the two novel lipoglycopeptides desmamides A and B isolated from the endophytic cyanobacterium *Desmonostoc muscorum* LEGE 12446. They are decapeptides featuring an aldopentose pyranose ring O-linked to the tyrosine residue (Freitas *et al.,* 2022). Interestingly, no glycosyltransferase-encoding genes in the *dsm* BGC or in its genomic neighborhood are found. Also, it is remarkable to cite the ramoplanin, which contains an O-glycosylation in the non-proteinogenetic amino acid Hpg (4-hydroxyphenylglycine), structurally similar to tyrosine. The *cao* cluster lacks homologues of the enzymes implied in the O-glycosylation tyrosine residues of the previous described natural products, but contains three glycosyltransferases (GTs) (Cao8, Cao16 and Cao24). The great majority of sugars present in bioactive natural products are deoxysugars and their incorporation to the corresponding aglycon is controlled by glycosyltransferases enzymes (Luzhetskyy *et al.,* 2007). These enzymes require an activated deoxysugar donor, normally an NDP or MPN-deoxysugar, to transfer to an acceptor aglycon (Lairson *et al.,* 2008), including polyketides, nonribosomal peptides, indolocarbazole or aminocoumarin among others. Most GTs catalyze O-glycosidic bond formation between the sugar and its natural aglycone (Salcedo *et al.,* 2016).

The mechanism for the regio- and stereo-specific transfer of the sugars can occur via the inverting or retaining mechanism, which also defines the stereo-chemical outcome at the anomeric center (α- or β-glucosides). The inverting mechanism follows a single displacement mechanism by a nucleophilic attack of the acceptor on the C-1 of the sugar donor inverting the anomeric stereochemistry (Schmid *et al.,* 2016). For retaining GTs, different mechanisms have been proposed, although the exact mechanism remains a controversial issue in the field of glycobiology (Schuman *et al.,* 2013).

Despite the presence of three GTs in the cacaoidin BGC, only the two deoxy sugar moieties L-rhamnose and L-6-deoxy-gulose, in the form of a disaccharide, are detected in the structure. Based on the inverting or retaining behavior of the glycosyltransferases, it was first proposed that both Cao8 and Cao16, belonging to the inverting GT-2 family, might work cooperatively to attach the *α*-L-rhamnose unit, while Cao24, belonging to the retaining GT-4 family, would incorporate the β-L-6-deoxygulose unit (Ortiz-López *et al.,* 2020). The three GTs were proposed to work in a cooperative manner to achieve effective glycosylation (Ortiz-López

et al., 2020). However, as it will be discussed below, in this work it has been demonstrated that only Cao8 and Cao16 are involved in the glycosylation of cacaoidin.

RiPPs Recognition Elements (RRE) are structurally similar conserved precursor peptide-binding domain present in the majority of known prokaryotic RiPP modifying enzymes and are usually responsible for the leader peptide recognition (Burkhart *et al*, 2015). These RREs are related to the small peptide chaperone PqqD, involved in the biosynthesis of pyrroloquinoline quinone (PQQ) (Klinman and Bonnot, 2014), which reportedly binds to PqqA (precursor peptide) to do its function (Latham *et al*, 2015). HHPred was used to search for PqqD-like domains in the proteins encoded in the *cao* cluster, but no RREs were found, suggesting the possibility of alternative leader peptide recognition domains that are unrelated to the already known RREs (Burkhart *et al*, 2015). As homology detection algorithms will become more accurate and more sequences will become available, additional RREs will be found. In fact, the identification of an RRE within the protease StmE, involved in lasso peptide streptomonomicin (STM) biosynthesis, and an ocin_ThiF_like cyclodehydratase (TOMM F) protein from TOMM (Thiazole/Oxazole-Modified Microcin) biosynthetic gene clusters, allowed to assign its non-previously proposed function (Burkhart *et al.,* 2015).

The leader peptide processing is another key step in the post-translational modification of RiPPs (Repka *et al.,* 2017). The N-terminal leader peptide plays a role in targeting the unmodified precursor by the posttranslational modifying enzymes, in the secretion of the peptide and in keeping the modified pre-peptide inactive (Plat *et al.,* 2011). The enzymes responsible for the removal of the leader peptide depend on the type of lanthipeptide. Class I lanthipeptides are exported by the ABC transporter LanT and their leader peptides are cleaved by the serine protease LanP (Zhang *et al.,* 2012). In class II, both secretion and cleavage are performed by a unique enzyme with a conserved N-terminal cysteine protease domain, called $LanT_P$ (Knerr and van der Donk, 2012).

In the cacaoidin cluster, Cao15 encodes a putative Zn-dependent peptidase belonging to the M16 peptidase family that may be involved in the leader peptide processing. Recently, it has been reported that the leader peptide of the class III lanthipeptide NAI-112 (Sheng *et al.,* 2020;) is removed by a bifunctional Zn-dependent M1-class metalloprotease, AplP, that first cleaves the N-terminal segment of the leader peptide as an endopeptidase, and subsequently removes the remaining leader sequence through its aminopeptidase activity (Addlagatta *et al.,* 2006). Leader peptide removal in class III lanthipeptides does not have a general mechanism. In fact, in labyrinthopeptins and curvopeptins, an endopeptidase is involved in the partial N-terminal segment removal of the leader peptide and the remaining overhang is progressively trimmed off by an additional aminopeptidase (Chen *et al.,* 2019; Krawczyk *et al.,* 2013). In other cases, such as flavipeptin, a designated prolyl oligopeptidase (POP) is involved in the cleavage of the leader peptide of modified precursor peptides at the C-terminal of a Pro residue, although it is not clear if a second aminopeptidase is needed to complete the leader peptide removal (Völler *et al.,* 2013). Class IV lanthipeptides often lack a designated protease to cleave the leader peptide, but it has been reported that some of them might also use AplP homologs (Chen *et al.,* 2019). When AplP and Cao15 were compared, both proteins showed low homology degree (17.2 % identity, 25.4 % similarity). On the other hand, peptidases LxmP1 and LxmP2 from lexapeptide BGC also

belong to the M16 peptidase family and share 25.4 % and 30.4 % similarity with Cao15, respectively. LxmP1 has an inactive M16 peptidase domain and LxmP2 an active domain. It has been suggested that LxmP1 and LxmP2 may form a heterodimer to remove the leader peptide from LxmA (Xu *et al.,* 2020). Since Cao15 harbors both the active and the inactive M16 peptidase domains, Cao15 has been proposed as the cacaoidin leader peptidase.

Besides, three ABC transporters were found in the *cao* pathway (Cao11, Cao18 and Cao19) that were proposed as responsible for the export and self-resistance of cacaoidin, have also been reported in other lanthipeptides. In addition to the active processing of the leader sequence coupled to active transport, two non-universal immunity strategies have been adopted by strains producing class I and II lanthipeptides, comprising a lipoprotein (LanI) and an ABC transporter (LanFEG). ABC transporters LanFEG prevent lantibiotic toxicity by active transporting the peptides to the extracellular space where they are sequestering by LanI immunity proteins (Geiger *et al.,* 2019). A self-immunity mechanism has not been deeply studied for class III and IV lanthipeptides but, as in the case of cacaoidin, it has also been proposed that ABC transporters could play a role in the self-resistance of the producer strains by the secretion of the drug through the cell membrane and inhibiting the growth of the producing strain to prevent the going back of the compound to the cell (Méndez and Salas, 2001).

The regulation of *cao* gene expression seems to be under the control of different classes of regulators. Five transcriptional regulators are found including two putative transcriptional activators (CaoR1 and CaoR5) and three putative transcriptional repressors (CaoR2, CaoR3 and CaoR4). Further studies of the regulation of lanthipeptide biosynthesis will clarify their role in the production of the antibiotic.

The function of the remaining five genes identified in the *cao* cluster (*cao17, cao21, cao23, cao25* and *cao26*) still cannot be proposed.

Identification of putative class V lanthipeptide BGCs in other strains

Up to date, only three BGCs encoding class V lanthipeptides have been described: cacaoidin (Román-Hurtado *et al,* 2021a), pristinin A3 (Kloosterman *et al.,* 2020) and lexapeptide (Xu *et al,* 2020). Before pristinin A and lexapeptide BGCs were described, a BLAST search performed with the *cao* BGC against the whole genome sequence database of GenBank did not find any cluster homologous to that of cacaoidin. However, the analysis of the below-threshold scores obtained after using CaoA and Cao7 as independent queries, found additional pathways that could encode new class V lanthipeptides in different actinomycete strains. All of these clusters share the presence of a HopA1 domain-containing protein, a $F_{420}H_2$-dependent oxidoreductase, a CypD-related protein, a Zn-dependent or S9 peptidase and a putative phosphotransferase. Furthermore, most of them have also homologues of the O-methyltransferase Cao4 containing a Methyltransf_2 domain, which seems to be characteristic of the O-methyltransferases present in the putative class V lanthipeptides BGCs. These preliminary data suggested the existence of a minimal set of genes required to ensure the core structural features of a class V lanthipeptides and a broader distribution of potential BGCs encoding new members of this RiPP class. The publication of the genome mining algorithm

decRiPPter (Data-driven Exploratory Class-independent RiPP TrackER) (Kloosterman *et al.,* 2020) confirmed these preliminary data, since it led to the identification of 42 new candidate novel RiPP families that could not be found by existing programs. One of these candidates was identified as the pristinin A3 BGC (*spr*) and classified as a class V lanthipeptide. Using the proteins SprH3 and SprPT from the *spr* cluster, which are homologous to the proteins Cao9 and Cao7, respectively, 195 orthologs were found in *Streptomyces,* 219 in other Actinobacteria and 161 in Firmicutes (Kloosterman *et al.,* 2020).

Heterologous expression of the *cao* cluster: identification of cacaoidin-2O

The putative cacaoidin BGC was cloned and expressed in the heterologous hosts *Streptomyces albidoflavus* J1074 and *Streptomyces coelicolor* M1152 and M1154 to confirm that these 27 genes included in the *cao* cluster were sufficient for the biosynthesis of the antibiotic. While no cacaoidin or derivate production was detected in the *Streptomyces coelicolor* hosts, in *Streptomyces albidoflavus* J1074, cacaoidin was detected in very small amounts compared to the wild type strain after 13 days of fermentation (0.14 mg/L titer) confirming that the fragment cloned contained all the information to ensure the synthesis of cacaoidin. In addition, a new variant with molecular formula $C_{107}H_{162}N_{24}O_{30}S_2$ was produced in a higher extent. The isolation and chemical characterization of this new compound confirmed that it lacked two atoms of oxygen (cacaoidin-2O), one in each sugar, which led to propose that in the heterologous host the same aglycon is incorporating a disaccharide different from the β-6-deoxygulopyranosyl-(1→3)-*α*-rhamnopyranoside of the original cacaoidin, and thus, the glycosylation pattern changes from the original *S. cacaoi* to the heterologous *S. albidoflavus* strain.

The enrichment study performed with L-rhamnose in the production medium R2YE was performed to confirm if these conditions could favor the balance towards the production of cacaoidin in the heterologous host. The wild-type *Streptomyces cacaoi* was confirmed to be able to produce cacaoidin in similar amounts in all the R2YE formulations (in the presence or absence of sucrose, glucose or rhamnose). This strain seems to be able to use similarly the three sugars as carbon source, and in the presence of rhamnose the production of cacaoidin was slightly higher, which can be supported by the fact that one of the sugars of the disaccharide of cacaoidin is α-rhamnopyranoside (Ortiz-López *et al.,* 2020). Interestingly, on the other hand, the heterologous host *Streptomyces albidoflavus* J1074 was only able to produce cacaoidin-2O, and wild-type cacaoidin in a lesser extent, in the R2YE formulation with sucrose, regardless of the presence of glucose of rhamnose. This heterologous host might be using only the sucrose as carbon source to biosynthesize the new sugars for the disaccharide of cacaoidin-2O.

Substrate flexibility is an interesting feature of some sugar biosynthesis and GT enzymes. This is especially remarkable in GTs which show different degrees of promiscuity towards the acceptor and donor molecules. The heterologous expression of deoxysugar biosynthesis and GTs genes have allowed the generation of interesting derivatives from diverse natural products. This phenomenon of glycodiversification in a heterologous host can be observed in other heterologously expressed natural products.

The biosynthetic gene cluster of nogalamycin, *snog* (pSnogaori), was cloned from *Streptomyces nogalater* ATCC 27451 to *Streptomyces albidoflavus* J1074 (Claesson *et al.,* 2012). Heterologous expression of the cosmid pSnogaori resulted in the production of nogalamycin F, where the L-nogalamine at the 1-position was swapped by an L-olivose sugar, which hinted the relaxed donor-substrate specificity on the part of the glycosyltransferase SnogD, which was surprising given that ElmGT was not able to previously transfer TDP-4'-keto-L-rhamnose, which is more resembled to its natural substrate (Nybo *et al.,* 2012). Based on that, cacaoidin's aglycon might also be incorporating TDP-L-olivose to its structure in the heterologous host, as the expected molecular mass would match the molecular mass of cacaoidin-2O. These results could be explained due to the endogenous TDP-L-olivose biosynthesis in *S. albidoflavus* J1074 (Brown *et al.,* 2020). Furthermore, the major drawback in the heterologous expression of the *snog* cluster is that the low production titers due to insufficient catalytic efficiency of the glycosyltransferases toward noncognate substrates, as observed with the nogalamycin aglycon L-olivose that is incorporated instead of L-nogalamine (Brown *et al.,* 2020). This may also explain the low production titers of cacaoidin-2O in the heterologous strain with a different disaccharide from the original strain and it might be translated in an inefficient catalytic activity of the GTs.

S. albidoflavus J1074 has been shown to produce previously cryptic compounds following genetic manipulation techniques and changes in the cultivation conditions, such as carotenoids, antimycins, candicidins, indigoidine and the glycosylated compounds paulomycins A, B and E (Olano *et al.,* 2014; Myronovskyi *et al.,* 2014). The gene cluster *plm* involved in paulomycins biosynthesis contains all genes necessary for the biosynthesis of L-paulomycose (González *et al.,* 2016). An intermediate in the biosynthesis of L-olivose, NDP-4-keto-L-olivose (Lombó *et al.,* 2009), might be generated by the activity of several enzymes from the pathway (Plm21, Plm20, Plm13, Plm40 and Plm42). This intermediate might then suffer an *O*-methylation, most probably performed by Plm41, and the incorporation of a two-carbon side chain at C4 to generate L-paulomycose (González *et al.,* 2016). *S. albidoflavus* J1074 might be leveraging all the genes from the *plm* BGC for the biosynthesis of NDP-4-keto-L-olivose, which is used for the generation of nogalamycin F and probably the new cacaoidin-2O. However, further studies are needed to draw conclusions.

As happens with SnogD in the biosynthesis of nogalamycin, the glycosyltransferases in the *cao* cluster seem to be flexible with the substrate they incorporate to the core peptide. There are another examples of substrate flexibility in glycosyltransferases. The glycosyltransferase ElmGT contains a 2,3,4-tri-O-methyl-α-L-rhamnose sugar appendage O-glycosydically linked at 8-position of the core of elloramycin (Decker *et al.,* 1995). ElmGT also exhibits remarkable flexibility toward the transfer of many TDP-deoxysugar moieties (Lombó *et al.,* 2004) as the heterologous expression of the biosynthetic gene cluster in *Streptomyces* sp. strains led to the production of the new analogs with novel deoxysugar moieties (Nybo *et al* 2012). The heterologous expression of the *elm* cluster in *Streptomyces lividans*, along with the pKOL construct expressing the *oleSEVWL* genes from the TDP-L-oleandrose pathway for the formation of TDP-4-keto-L-olivose, resulted in the generation of the new analog 8-demethyl-8-(a'-keo)-α-L-olivosyltetracenomycin C (Brown *et al.,* 2020), which supports the substrate flexibility of ElmGT toward TDP-4-keto-L-sugars, a fact which is unusual among natural products glycosyltransferases. The glycosyltransferase of doxorubicin, Akns, also showed moderate substrate promiscuity toward

noncanonical TDP-deoxysugar donor substrates when heterologously expressed in *Streptomyces venezuelae* along several constructions encoding heterologous TDP-deoxysugar biosynthetic pathways and generating glycosylated derivatives of doxorubicin (Han *et al.,* 2011).

Grecocyclines, members of the angucycline antibiotics, were isolated from *Streptomyces* sp. Acta 1362 and contains a disaccharide side chain and an additional amino sugar attached to the aglycon (Paulutat *et al.,* 2010). As it happens with cacaoidin, this natural product also changed its glycosylation pattern when it was heterologously expressed in *S. albidoflavus* J1074 (Bilyk *et al.,* 2016). In this case, although no glycosylated grecocycline was produced, several known angucyclines, such as rabelomycin, dehydrorabelomycin and tetrangulol, were observed, which are non-glycosylated intermediates formed during the biosynthesis of angucycline polyketides (Luzhetskyy *et al.,* 2005; Lombó *et al.,* 2009).

From the new heterologous cacaoidin-2O, the generation of the aglycon was achieved by its controlled hydrolysis, incubating at 4 °C with 35.6 % ACN/H_2O containing 0.1 % TFA for several days. Interestingly, the aglycon could not be obtained from the original cacaoidin using either of the methods previously described. This might be explained by the difference of sugars forming the disaccharide and its configuration in the linkage to the tyrosine residue. The disaccharide of the original cacaoidin and its configuration seems to be much more stable and resistant to hydrolysis than the one present in the heterologous cacaoidin-2O. However, further studies need to be performed to draw any conclusion.

Obtention of knockout mutants in the *cao* cluster

The role of Cao4 in the N-methylation of cacaodin-20 was studied by the generation of a *cao4*-knockout in the heterologous strain CA-300429. In this *cao4*-knockout cacaoidin-2O was not detected, but the $M+2H^+ + NH_4^+$ ion at *m/z* 773.3843, indicative of the molecular formula $C_{105}H_{158}N_{24}O_{30}S_2$, corresponding with the demethylated cacaoidin-2O was detected. The role of Cao4 in the dimethylation was confirmed when the knockout was complemented and the production of cacaoidin-2O was reverted. This demethylated cacaoidin-2O was also detected in the heterologous host and the complemented strain, along the cacaoidin-2O, but in a much lesser extent. It is likely that these strains are producing this cacaoidin-2O variant as an intermediate which has not been post-translationally modified yet by Cao4. These results confirmed, thereby, the role of Cao4 in the *N,N*-methylation of the N-terminal alanine. Furthermore, a comparative analysis between Cao4 and other O-methyltransferases present in putative class V lanthipeptide BGCs showed that all of them contained a Methyltransf_2 domain, which is not present in the methyltransferases of other classes of lanthipeptides or other O-methyltransferases. This suggests the existence of a new class of methyltransferases containing the characteristic Methyltransf_2 domain and present only in class V lanthipeptide BGCs, responsible for the *N,N*-methylation in these compounds.

The generation of knockouts of the three glycosyltransferases demonstrated that only Cao8 and Cao16 are involved in cacaoidin biosynthesis. In both *cao8* and *cao16* knockout strains, CA-301537 and CA-303085, the biosynthesis of cacaoidin became disrupted as no cacaoidin-2O was detected. The complementation of these knockout strains with the corresponding genes

resulted in the retrieval of the production of cacaoidin-2O, suggesting that the biosynthesis of the compound requires the involvement of both glycosyltransferases.

Interestingly, the lack of detection of monoglycosylated or aglycon variants of cacaoidin in the knockouts, suggested that cacaoidin might be glycosylated with a disaccharide to be efficiently produced. In eukaryotic organisms, O-glycosylations have been described to provide protection against proteolytic degradation and thermal disruption (Patsos and Corfield, 2009). An example can be found in the low-density lipoprotein LDL receptors of the cell surface, which requires O-glycans to avoid proteolytic cleavage of their extracellular domains. The transport of the transferrin receptor (TfR) between the endosomes and the cell surface also presents this protection against proteolysis due to the O-glycosylation of threonine 104 (Patsos and Corfield, 2009). Another recently described example is the fibroblast growth factor 23 (FGF23), secreted by osteoblasts and osteocytes as an important regulator of phosphate homeostasis, whose O-glycosylation is shown to prevent its cleavage (Takashi *et al.,* 2019; Goth *et al.,* 2018) The removal of the sugar moieties from these O-glycans results in the loss of this protection. Thus, site-specific O-glycosylation seems to ensure the general protection from proteolytic degradation. However, although most of the examples show that O-glycosylation decrease cleavage efficiency, there are several cases where O-glycosylation seems to enhance cleavage or even change the cleavage site (Goth *et al.,* 2018). Although this protection effect by O-glycosylation has been only described to date in eukaryotic cells, the O-glycosylation in cacaoidin seems to provide a similar benefit as no aglycon is detected in the *cao8* and *cao16* knockouts. In prokaryotes, the knowledge of the glycosylation is still very limited. Further studies will clarify this protective role of the O-glycosylation in cacaoidin.

On the other hand, the production of cacaoidin-2O was not affected in the *cao24* knockout strain CA-300649, and the compound was produced normally as in the heterologous host *S. albidoflavus* J1074/pCAO CA-300429. This result suggested that Cao24 is not involved in the biosynthesis of cacaoidin and may even not be part of the *cao* BGC.

Cooperative glycosyltransferases have been described in the biosynthesis of many other natural products. In the cluster responsible of the biosynthesis of saquayamycins (*sqn*) there are three GTs (SqnG1, SqnG2 and SqnG3). Salem *et al* (2017) proved that SqnG1 and SqnG2 work in a cooperative way to generate all the saquayamycin variants while SqnG3 might act as a chaperone to assist the glycosylating activity of SqnG1/G2. Also, the macrolactams sipanmycins A and B are synthesized by the *sip* cluster, a PKS type I-oligosaccharide class that contains four GT genes (*sipS4*, *sipS9*, *sipS14* and *sipS15*). Malmierca *et al* (2018) described that the four GTs in the *sip* cluster are required for the transfer of the disaccharide containing two amino sugars, with SipS15 and SipS4 working cooperatively to transfer of D-xylosamine, and SipS9 and SipS14 working together cooperatively to transfer of D-sipanose. Apparently, none of the GTs in each pair can replace the activity of the other. This is exactly what happens in cacaoidin biosynthesis: Thus, Cao8 and Cao16 are also cooperative glycosyltransferases as both seem to be required for the tyrosine O-glycosylation with the disaccharide and one of them cannot replace the activity of the other, since independent inactivation of the two GTs led to the cessation of the biosynthesis of cacaoidin in the knockout strains, where no aglycon or monoglycosylated variant are either detected.

Two different mechanisms can be proposed to link a disaccharide to the aglycon molecule: by the sequential and successive addition of the two monosaccharides by independent GTs or by independent disaccharide formation by one of the GT and its transfer to the aglycon by the second GT. Following the first hypothesis, the independent inactivation of the GTs would generate a monoglycosylated intermediate and an aglycon. On the other hand, according to the second hypothesis, the independent inactivation of the GT would only generate the aglycon. As the independent inactivation of both GT genes generated two cacaoidin-nonproducing strains, where neither the aglycon nor the monoglycosylated intermediate were detected, the two GTs seem to work cooperatively to establish the linkage of a previously formed disaccharide with the tyrosine residue. In the antitumor drug mithramycin from *Streptomyces argillaceus*, a similar process was described. Mithramycin is a polyketide containing two deoxysugars chains, a disaccharide consisting in two D-olivoses and a trisaccharide composed of D-olivose, D-oliose and D-mycarose. The *mtm* cluster contains two glycosyltransferases, MtmGI and MtmGII, presumably responsible of the transfer of the D-olivose disaccharide to the aglycon. The independent inactivation of either MtmGI or MtmGII produces compounds that lack the D-olivose disaccharide, suggesting that both glycosyltransferases are involved in its transfer (Fernández *et al.,* 1998).

In addition to cacaoidin, only two glycosylated RiPPs produced by bacteria have been described so far. On the one hand, glycocins, post-translationally glycosylated bacteriocins whose sugar moieties are linked to the side chains of either Cys, Ser or Thr residues by the S-glycosyltransferase enzymes (Norris *et al.,* 2016) generating very unusual S-linkages between the residues and glucose. They are usually “glycoactives” and the sugar moieties are essential for the antimicrobial activity. Several glycocins have been identified to date, such as sublancin 168, produced by *Bacillus subtillis* with glucosylation in cysteine residues (Oman *et al.,* 2011); glycocin F, produced by *Lactobacillus plantarum* and with two N-acetylglucosamine moieties, one β-O-linked to a serine residue and the other S-linked to a C-terminal cysteine (Stepper *et al.,* 2011); ASM1, produced by *Lactobacillus plantarum* and homologous to glycocin F (Hata *et al.,* 2010); thurandacin, encoded by *Bacillus thuringiensis* and considered a hypothetical natural product not isolated from culture with β-S-glucosylations in cysteine and serine residues (Wang *et al.,* 2014a); and recently pallidocin, produced by *Aeribacillus pallidus* and with a cysteine residue S-glucosylated (Kaunietis *et al.,* 2019).

Another glycosylated RiPP is NAI-112, a class III lanthipeptide produced by an *Actinoplanes* sp. strain (Iorio *et al.,* 2013) with antinociceptive/antiallodynic activity. The molecule binds to phosphate-containing lipids and blocks pain sensation by decreasing levels of lysophosphatidic acid in the TRPV1 pathway (Tocchetti *et al.,* 2021). NAI-112 carries a 6-deoxyhexose moiety N-linked to a tryptophan residue by the Trp(N)-glycosyltransferase AplG (Sheng *et al.,* 2020). This glycosyltransferase is specific of Trp amino acids and required the presence of both labionin and methyl-labionine in the C- and N- terminus of NAI-112, respectively, for its performance. Both Lab/MeLan rings are generated by the unique class III lanthipeptide synthetase AplKC. However, low homologies were found between these enzymes and the GTs present in the cacaoidin cluster.

The *cpp* and the *cao* BGCs are species-specific trait of *S. cacaoi*

These two *cpp* and *cao* BGCs have been shown to be the responsible for the biosynthesis of BE-18257 A-C and pentaminomycins A-H, and cacaoidin, respectively. A tblastn search of the CppB and CppM protein sequences from the *cpp* BGC and the precursor peptide CaoA sequence from the *cao* BGC against both nucleotide and Whole Genome Sequence (WGS) databases from NCBI showed that both *cpp* and *cao* clusters were only present in the genomes of strains belonging to *Streptomyces cacaoi* species: *S. cacaoi* NHF165, *S. cacaoi* DSM 40057, *S. cacaoi* subs. *cacaoi* NRRL-1220, *S. cacaoi* OABC16, *Streptomyces* sp. NRRL S-1868, *Streptomyces* sp. NRRL F-5053 and *S. cacaoi* NBRC 12748, apart from our strain *Streptomyces cacaoi* CA-170360. No other pentaminomycin- or cacaoidin-related peptides or pathways were found in the genomes of the phylogenetically closest *Streptomyces* relatives nor even the other *S. cacaoi* subs. *asoensis* strains. This indicates that both *cpp* and *cao* clusters are strain-specific BGCs, that are exclusive and highly conserved within members of this species *Streptomyces cacaoi* subs. *cacaoi* and is an excellent example of the biosynthesis of a specialized metabolite that could be used as a species-specific trait (Seipke, 2015), reflecting that chemical novelty and richness can be found at the strain level and that the analysis of the genomes of closely related strains constitutes a promising approach for the identification of novel BGCs (Vicente *et al.,* 2018).

Actinomycete strains belonging to the same species are likely to possess similar secondary metabolites. The analysis of BGCs in *Nocardia brasiliensis* and *Salinispora* species have shown that most BGCs are shared within each species, with strain-specific ones being relatively limited. Komani *et al* (2018) showed that *S. diastaticus* subsp. *ardesiacus* strains TP-A0882 and NBRC 15402T share an almost identical set of BGCs, while *S. coelicofavus* strains NBRC 15399^{T} and ZG0656 shared their own similar set of gene clusters. They also found that most of the gene clusters in S*. rubrogriseus* NBRC 15455^{T} were present also in *S. coelicolor* (correctly classifed as *Streptomyces violaceoruber*) A3 (Hatano *et al.,* 1994). As the sequence similarities in these regions were very high (>93 %), it might be considered plausible that the two strains NBRC 15455^{T} and A3 may actually be the same species (Komani *et al.,* 2018). Choudoir *et al* (2018) reported that while strains within a species can share a core set of BGCs, the number of accessory BGCs within a given species can be quite larger, with each strain having its own repertoire of strain-specific BGCs. However, the majority of these strain-specific BGCs remains uncharacterized and lack similarity to BGCs documented in the public databases, as happened with cacaoidin BGC before lexapeptide BGC was described.

F. Conclusions

I. The genome of *Streptomyces cacaoi* CA-170360, which comprises 8.7 Mb and a GC content of 73.2 %, contains 33 biosynthetic gene clusters putatively involved in the biosynthesis of different secondary metabolites. Some of the compounds encoded by these clusters, such as cacaoidin, pentaminomycins A-H, BE-18257 antibiotics, actinonin, naseseazines, puromycin, nocardamine and terragine E, were detected in culture fermentations following an OSMAC approach.

II. The heterologous expression of the *cpp* biosynthetic gene cluster confirmed its involvement in the biosynthesis of BE-18257 antibiotics and pentaminomycins A-H. The independent cloning of the *cppB* NRPS genes demonstrated that *cppB* is the responsible of the production of the BE-18257 compounds while *cppM* is involved in the biosynthesis of pentaminomycins A-H. The heterologous expression also confirmed that the tryptophan biosynthetic genes located downstream *cppM* are not necessary for the biosynthesis of the cyclic pentapeptides.

III. The penicillin-binding protein-like CppA is a member of the peptidyl NRP offloading cyclases family included in the β-lactamase superfamily and it is involved in the release and cyclization of the cyclic pentapeptides, as demonstrated with the Δ*cppA* knockout.

IV. The heterologous expression of the *cao* BGC proved that the 27 genes contained in the cluster are sufficient for the biosynthesis of cacaoidin. The cluster includes the decarboxylase CaoD, a three component lanthionine synthetase comprised by the standalone monofunctional proteins Cao7, Cao9 and Cao14, the O-methyltransferase, the $F_{420}H_2$-dependent oxidoreductase Cao12, three glycosyltransferases (Cao8, Cao16 and Cao24), the Zn-dependent peptidase Cao15, some rhamnose-biosynthetic proteins, ABC transporters and other regulatory and unknown proteins.

V. A new variant of cacaoidin is produced in the heterologous host *S. albidoflavus* J1074/pCAO, together with the wild-type cacaoidin. This variant, named cacaoidin-2O, has a disaccharide lacking two oxygen atoms, one in each sugar, and, unlike the original cacaoidin, can be hydrolyzed to yield the aglycon under acid conditions.

VI. The *Δcao4* knockout confirmed the role of the O-methyltransferase Cao4 in the *N,N*-methylation of the N-terminal alanine of cacaoidin. Cao4 is a new class of methyltransferase containing the characteristic Methyltransf_2 domain and is present in several class V lanthipeptide BGCs.

VII. The *Δcao8, Δcao16* and *Δcao24* knockouts demonstrated that only the glycosyltransferases Cao8 and Cao16 are involved in the O-glycosylation of the tyrosine residue in cacaoidin, working cooperatively to join the previously formed disaccharide. These two glycosyltransferases show substrate flexibility as the disaccharides of the original cacaoidin and cacaoidin-2O are different. Cao24 does not seem to be involved in cacaoidin biosynthesis.

VIII. Both *cao* and *cpp* BGCs are highly conserved strain-specific clusters exclusively present in the genomes of strains belonging to the *Streptomyces cacaoi* subsp. *cacaoi* species.

F. Conclusiones

I. El genoma de la cepa *Streptomyces cacaoi* CA-170360, que tiene 8.7 Mb y un contenido en GC del 73.2 %, contiene 33 rutas biosintéticas potencialmente involucradas en la biosíntesis de diferentes metabolitos secundarios. Algunas de estas rutas son las responsables de la producción de cacaoidina, pentaminomicinas A-H, antibióticos BE-18257, actinonina, naseseazinas, puromicina, nocardamina and terragina E, compuestos detectados en caldos de fermentación.

II. La expression heteróloga de la ruta *cpp* confirmó su participación en la biosíntesis de los antibióticos BE-18257 y las pentaminomicinas A-H. La clonación independiente del gen *cppB* demostró que esta NRPS es la responsable de la producción de los BE-18257 mientras que la NRPS *cppM* participa en la biosíntesis de las pentaminomicinas A-H. LA expression heteróloga también confirmó que los genes de biosíntesis del triptófano situados aguas bajo del gen *cppM* no son necesarios para la biosíntesis de las dos familias de ciclopentapéptidos.

III. La proteína CppA, similar a las proteínas de unión a penicilina, pertenece a la familia de ciclasas liberadoras de péptidos no ribosomales, incluída en la superfamilia de las β-lactamasas, y está involucrada en la liberación y ciclación de las dos familias de ciclopentapéptidos, como ha quedado demostrado con el mutante por reemplazamiento génico Δ*cppA*.

IV. La expresión heteróloga de la ruta *cao* confirmó que los 27 genes presentes en la misma son suficientes para la biosíntesis de cacaoidina. Dicha ruta incluye la descarboxilasa CaoD, una lantionina sintetasa formada por las tres proteínas monofuncionales Cao7, Cao9 y Cao14, la O-metiltransferasa Cao4, la oxidorreductasa dependiente de $F_{420}H_2$ Cao12, tres glicosiltransferasas (Cao8, Cao16 y Cao24), la peptidasa dependiente de Zn Cao15, algunas de las proteínas implicadas en la biosíntesis de ramnosa, transportadores ABC y otros genes reguladores y proteínas de función desconocida.

V. El hospedador heterólogo *S. albidoflavus* J1074/pCAO produce, junto con la cacaoidina original, una nueva variante, llamada cacaoidina-2O, que tiene un disacárido distinto, cuyos monosacáridos poseen cada uno un átomo menos de oxígeno. A diferencia de la cacaoidina original, esta variante puede hidrolizarse bajo condiciones ácidas para generar el aglicón.

VI. El mutante por reemplazamiento génico *Δcao4* confirmó el papel de la O-metiltransferasa Cao4 en la doble metilación del residuo de alanina en el extremo amino-terminal de la cacaoidina. Cao4 es una nueva clase de metiltransferasa que contiene el dominio característico Methyltransf_2 y se encuentra en muchos de las rutas de biosíntesis de lantipéptidos de clase V.

VII. Los mutantes por reemplazamiento génico *Δcao8, Δcao16* y *Δcao24* confirmaron que únicamente las tres glicosiltransferasas, Cao8 y Cao16, están implicadas en la O-glicosilación del residuo de tirosina de la cacaoidina, y que actúan de forma cooperativa

para unir el disacárido, previamente ensamblado. Estas dos glicosiltransferasas muestran cierta flexibilidad de sustrato, ya que los disacáridos que unen en la cacaoidina original y la cacaoidina-2O son diferentes. La glicosiltransferasa Cao24 parece no tener ninguna función en la biosíntesis de cacaoidina.

VIII. Tanto *cao* como *cpp* son rutas biosintéticas altamente conservadas y específicas de cepa, ya que sólo se encuentran en los genomas de cepas pertenecientes a la especie *Streptomyces cacaoi* subsp. *cacaoi*.

G. Abbreviation List

2-Pal	3-(2-pyridyl)-alanine
A	Adenylation domain
ABC	ATP-Binding Cassette
ACN	Acetonitrile
ACP	Acyl carrier protein
Am	Apramycin
AMDIS	Automated Mass Spectral Deconvolution and Identification System
AMP	Adenosin Monophosphate
AntiSMASH	Antibiotics and Secondary Metabolite Analysis Shell
Ap	Ampicillin
ATCC	American Type Culture Collection
ATP	Adenosin Triphosphate
AviCys	S-[(Z)-2-aminovinyl]-D-cysteine
AviMeCys	S-[(Z)-2-aminovinyl]-D-(3-methyl) cysteine
BAC	Bacterial Artificial Chromosome
BAGEL	Bayesian Analysis of Gene Essentiality
BGC	Biosynthetic Gene Cluster
BLAST	Basic Local Alignment Search Tool
C	Condensation domain
CAMPR	Collection of Anti-Microbial Peptides
Cao	cacaoidin
CATCH	Cas9-Assisted Targeting of CHromosome segments
CFU	Colony Forming Unit
CIAP	Calf Intestinal Alkaline Phosphatase
CID	Collision-induced dissociation
Cm	Chloramphenicol
Cpp	Cyclic pentapeptides
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	Crispr RNA
Cy	Cyclization domain
DEPC	Diethyl pyrocarbonate
Dha	2,3-dehydroalanine
Dhb	(Z)-2,3-dehydrobutyrine
DiPaC	Direct Pathway Cloning
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
DNP	Dictionary of Natural Products
dTDP	Deoxythymidine Diphosphate
E	Epimerization domain
EDTA	Ethylenediaminetetraacetic acid
EIC	Extracted Ion Chromatogram
$FADH_2$	1,5-dihydro-Flavin Adenine Dinucleotide
FMN	Flavin Mononucleotide
FT	Fourier Transform

GDP	Guanosine diphosphate
GNPS	Global Natural Products Social Molecular Networking
GT	Glycosyltransferase
HDAC	Histone Deacetylase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HFCD	Homo-Oligomeric containing Cys-Decarboxylase
HHPred	Homology detection & structure prediction by HMM-HMM comparison
HiFi	High Fidelity
HMM	Hidden Markov Model
HPLC	High Performance Liquid Chromatography
HTS	High-Throughput Screening
HTTM	Horizontally Transferred TransMembrane domain
ICNP	International Code of Nomenclature of Prokaryotes
ISR	Integrase-mediated Site-specific Recombination system
KAc	Potassium Acetate
Km	Kanamycin
Lab	Labionine
LAL	Large ATP-binding regulators of the LuxR family
Lan/MeLan	Lanthionine/MethylLantionine
LAP	Linear azol(in)e-containing peptide
LC-DAD-MS	Liquid Chromatography-Photodiode Array Detector-Mass Spectrometer
LC-HRESIMS	Liquid Chromatography-High Resolution Electrospray Ionization Mass Spectrometry
LC-HRMS	Liquid Chromatography-High Resolution Mass Spectrometry
LC-MS	Liquid Chromatography-Mass Spectrometry
LLHR	Linear-Linear Homologous Recombination
LPNS	List of Prokaryotic names with Standing in Nomenclature
m/z	Mass/Charge
MEK	Methyl Ethyl Ketone
MiBiG	Minimun Information about Biosynthetic Gene Clusters
MIC	Minimum Inhibitory Concentration
MPLC	Medium Pressure Liquid Chromatography
MRSA	Methicillin-Resistant *Staphylococcus aureus*
MS/MS	Tandem Mass Spectrometry
MSH	Melanocyte Stimulating Hormone
MT	Methyltransferase
NAD	Nicotinamide Adenine Dinucleotide
NADPH	Nicotinamide-Adenine Dinucleotide Phosphate
Nal	Nalidixic acid
NaOH	Sodium Hydroxide
NCBI	National Center for Biotechnology Information
NDP	Nucleoside Diphosphate
NGS	Next Generation Sequencing
NIST	National Institute of Standards and Technology

NMR	Nuclear Magnetic Resonance
NP	Natural Product
NRP	Non-Ribosomal Peptide
NRPS	Non-Ribosomal Peptide Synthetase
O/N	Overnight
OD	Optical Density
ORF	Open Reading Frame
OSMAC	One Strain Many Compounds
Ox	Oxidase domain
PAC	Phage Artificial Chromosome
PAM	Protospacer Adjacent Motif
PBP	Penicillin-Binding Protein
PBS	Phosphate-Buffered Saline
PCP	Peptidyl Carrier Protein
PCR	Polymerase Chain Reaction
PK	Polyketide
PKS	Polyketide Synthase
PMFS	Phenylmethylsulfonyl Fluoride
POP	Prolyl-oligopeptidase
PPC	Phosphopantothenoylcysteine
PPTase	Phosphopantetheinyl Transferase
PRISM	Prediction Informatics for Secondary Metabolites
PTM	Post-Translational Modification
R	Reductase domain
REE	Rare Earth Elements
Rif	Rifampycin
RiPP	Ribosomally Synthesized and Post-translationally modified Peptide
RNA	Ribonucleic acid
RNAse	Ribonuclease
RP-HPLC	Reverse-phase High Performance Liquid Chromatography
RRE	RiPP Recognition Element
RS	Recognition Sequence
SAM	S-Adenosyl Methionine
SARP	*Streptomyces* Antibiotic Regulatory Protein
SDR	Short-Chain Dehydrogenase
SDS	Sodium Dodecyl Sulfate
sgRNA	single guide RNA
SMRT	Single Molecule Real Time
TAE	Tris-Acetate-EDTA
TAR	Transformation-Associated Recombination
tBLASTn	BLAST of translated nucleotide databases using a protein query.
TDP	Thymidine diphosphate
TE	Thioesterase
TE	Tris-EDTA
TFA	Trifluoroacetic Acid

Tm	Melting Temperature
TOMM	Thiazole/Oxazole-Modified Microcin
TOF	Time of Flight
tracrRNA	trans-activating crispr RNA
Tsr	Thiostrepton
UV	Ultraviolet
UV/Vis	Ultraviolet/Visible
WGS	Whole Genome Sequencing
WT	Wild-type

H. References

1. **Acedo, J. Z., Bothwell, I. R., An, L., Trouth, A., Frazier, C. and van der Donk, W. A**. (2019). O-methyltransferase-mediated incorporation of a β-amino acid in lanthipeptides. *J. Am. Chem. Soc*. **141**:16790–16801. doi:10.1021/jacs.9b07396.

2. **Addlagatta, A., Gay, L. and Matthews, B. W**. (2006). Structure of aminopeptidase N from *Escherichia coli* suggests a compartmentalized, gated active site. *Proc. Natl. Acad. Sci. USA*. **103**:13339-13344. doi: 10.1073/pnas.0606167103.

3. **Agrawal, P., Amir, S., Deepak, Barua, D. and Mohanty, D**. (2021). RiPPMiner-Genome: a web resource for automated prediction of crosslinked chemical structures of RiPPs by genome mining. *J. Mol. Biol*. **433**:166887. doi: 10.1016/j.jmb.2021.166887.

4. **Almeida, E. L., Kaur, N., Jennings, L. K., Carrillo-Rincón, A. F., Jackson, S. A., Thomas, O. P. and Dobson, A. D. W.** (2019). Genome mining coupled with OSMAC-based cultivation reveal differential production of surugamide A by the marine sponge isolate *Streptomyces* sp. SM17 when compared to its terrestrial relative *S. albidoflavus* J1074. *Microorganisms*. **7**:394. doi: 10.3390/microorganisms7100394.

5. **Ames, B. D., Korman, T. P., Zhang. W., Smith, P., Vu, T., Tang, Y. and Tsai, S. C.** (2008). Crystal structure and functional analysis of tetracenomycin ARO/CYC: Implications for cyclization specificity of aromatic polyketides. *Proc. Natl. Acad. Sci. USA*. **105**:5349-5354. doi: 10.1073/pnas.0709223105.

6. **Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K. D., Fischbach, M. A., Garavelli, J. S., Göransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Müller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J., Rebuffat, S., Ross, R. P., Sahl, H. G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Süssmuth, R. D., Tagg, J. R., Tang, G. L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M. and van der Donk, W. A.** (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep*. **30**:108-160. doi: 10.1039/c2np20085f.

7. **Atanasov, A. G., Waltenberger, B., Pferschy-Wenzig, E. M., Linder, T., Wawrosch, C., Uhrin, P., Temml, V., Wang, L., Schwaiger, S., Heiss, E. H., Rollinger, J. M., Schuster, D., Breuss, J. M., Bochkov, V., Mihovilovic, M. D., Kopp, B., Bauer, R., Dirsch, V. M. and Stuppner, H.** (2015). Discovery and resupply of pharmacologically active plant-derived natural products: a review. *Biotechnol*. *Adv*. **33**:1582-1614. doi: 10.1016/j.biotechadv.2015.08.001.

8. **Ayobami, O., Brinkwirth, S., Eckmanns, T. and Markwart, R.** (2022). Antibiotic resistance in hospitals-acquired ESKAPE-E infections in low and lower-middle-income countries: a

systematic review and meta-analysis. *Emerg. Microbes. Infect*. **11**: 443-451. doi: 10.1080/22221751.2022.2030196.

9. **Baltz, R. H.** (2016). Genetic manipulation of secondary metabolite biosynthesis for improved production in *Streptomyces* and other actinomycetes. *J. Ind. Microbiol. Biotechnol*. **43**:343-370. doi: 10.1007/s10295-015-1682-x.

10. **Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Meier-Kolthoff, J. P., Klenk, H. P., Clément, C., Ouhdouch, Y. and van Wezel, G. P.** (2015). Taxonomy, physiology and natural products of Actinobacteria. *Microbiol. Mol. Biol. Rev*. **80**:1-43. doi: 10.1128/MMBR.00019-15.

11. **Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C. H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M. A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B. G., Parkhill, J. and Hopwood, D. A.** (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). Nature. **417**:141-147. doi: 10.1038/417141a.

12. **Berdy, J**. (2005). Bioactive microbial metabolites, a personal view. *J. Antibiot*. **58**:1-26. doi: 10.1038/ja.2005.1.

13. **Berdy, J.** (2012). Thoughts and facts about antibiotics: where we are now and where we are heading. ***J. Antibiot***. 65:385-395. doi: 10.1038/ja.2012.27.

14. **Besser, J., Carleton, H. A., Gernet-Smidt, P., Lindsey, R. L. and Trees, E**. (2018). Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clin. Microbiol. Infect*. **24**:335-341. doi: 10.1016/j.cmi.2017.10.013.

15. **Biecker, A. L., Liu, X., Thorson, J. S., Yang, Z. and Van Lanen, S. G.** (2019). Biosynthetic and synthetic strategies for assembling capuramycin-type *Antituberculosis* antibiotics. *Molecules*. **24**:433. doi: 10.3390/molecules24030433.

16. **Bilyk, O., Sekurova, O. N., Zotchev, S. B. and Luzhetskyy**. (2016). Cloning and heterologous expression of the grecocycline biosynthetic gene cluster. *PLOS ONE*. **11**. doi: 10.1371/journal.pone.0158682.

17. **Birnboim, H. C. and Doly, J.** (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids. Res*. **7**:1513-1523. doi: 10.1093/nar/7.6.1513.

18. **Bister, B., Bischoff, D., Ströbele, M., Riedlinger, J., Reicke, A., Wolter, F., Bull, A. T., Zähner, H., Fiedler, H. P. and Sussmüth, R. D**. (2004). Abyssomicin C-A polycyclic antibiotic from a marine *Verrucosispora* strain as an inhibitor of the p-aminobenzoic acid/tetrahydrofolate biosynthesis pathway. *Angew. Chem. Int*. **43**:2574-2576. doi: 10.1002/anie.200353160.

19. **Blin, K., Kazempour, D., Wohlleben, W. and Weber, T.** (2014). Improved lanthipeptide detection and prediction for antiSMASH. *PLOS ONE*. **9**. doi: 10.1371/journal.pone.0089420.

20. **Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van wezel, G. P., Medema, M. and Weber, T**. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res*. **49**:29-35. doi: 10.1093/nar/gkab335.

21. **Blin, K., Shaw, S., Kautsar, S. A., Medema, M. H. and Weber, T.** (2021). The antiSMASH database version 3: increased taxonomic coverage and new query features for modular enzymes. *Nucleic Acids Res.* **49:**D639–D643. doi: 10.1093/nar/gkaa978.

22. **Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., Suarez Duran, H. G., de Los Santos, E. L. C., Kim, H. U., Nave, M., Dickschat, J. S., Mitchell, D. A., Shelest, E., Breitling, R., Takano, E., Lee, S. Y., Weber, T. and Medema, M. H.** (2017). antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res*. **45:**W36-W41. doi: 10.1093/nar/gkx319.

23. **Bode, H. B., Bethe, B., Höfs, R. and Zeeck, A.** (2002). Big Effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. *Chembiochem*. **3**:619–627. doi: 10.1002/1439-7633(20020703)3:7<619::AID-CBIC619>3.0.CO;2-9.

24. **Bonet, B., Teufel, R., Crusemann, M., Ziemert, N. and Moore, B. S.** (2015). Direct capture and heterologous expression of *Salinispora* natural product genes for the biosynthesis of enterocin. *J. Nat. Prod.* **78**:539–542. doi: 10.1021/np500664q.

25. **Braña, A. F., Méndez, C., Díaz, L. A., Manzanal, M. B. and Hardisson, C.** (1986). Glycogen and trehalose accumulation during colony development in *Streptomyces antibioticus. J. Gen. Microbiol*. **132**:1319-1326. doi: 10.1099/00221287-132-5-1319.

26. **Brockhausen, I., Ho, B., Liu, B., Lau, K., Szarek, W. A., Wang, L. and Feng, L.** (2008). Characterization of two β-1,3-glucosyltransferases from *Escherichia coli* serotypes O56 and O152. *J. Bacteriol*. **190**:4922-4932. doi: 10.1128/JB.00160-08.

27. **Brown, K. V., Wandi, B. N., Metsä-Ketelä, M. and Nybo, S. E.** (2020). Pathway engineering of anthracyclines: blazing trails in natural product glycodiversification. *J. Org. Chem*. **85**:12012-12023. doi: 10.1021/acs.joc.0c01863.

28. **Burkhart, B. J., Hudson, G. A., Dunbar, K. L. and Mitchell, D. A**. (2015). A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nat. Chem. Biol*. **11**:564-570. doi: 10.1038/nchembio.1856.

29. **Caetano, T., Krawczyk, J. M., Mösker, E., Süssmuth, R. D. and Mendo, S**. (2011). Heterologous expression, biosynthesis and mutagenesis of type II lantibiotics from *Bacillus licheniformis* in *Escherichia coli*. *Chem. Biol*. **18**:90-100. doi: 10.1016/j.chembiol.2010.11.010.

30. **Cândido, E., de Barros, E., Cardoso, M. H. and Franco, O. L**. (2019). Bacterial cross-resistance to anti-infective compounds. Is it a real problem? *Curr. Opin. Microbiol*. **48**: 76-81. doi: 10.1016/j.coph.2019.05.004.

31. **Cannell, R. J. P.** (1998). How to approach the isolation of a natural product. En *Methods in Biotechnology*, Vol. **4**: Natural Products Isolation. Humana Press Inc. Totowa, New Jersey, USA.

32. **Carretero-Molina, D., Ortiz-López, F. J., Martín, J., González, I., Sánchez-Hidalgo, M., Román-Hurtado, F., Díaz, C., de la Cruz, M., Genilloud, O. and Reyes, F.** (2021). Pentaminomycins F and G, Nonribosomal Peptides Containing 2-Pyridylalanine. *J Nat Prod.* **84**:1127-1134. doi: 10.1021/acs.jnatprod.0c01199.

33. **Carretero-Molina, D., Ortiz-López, F. J., Martín, J., Oves-Costales, D., Díaz, C., de la Cruz, M., Cautain, B., Vicente, F., Genilloud, O. and Reyes, F.** (2020). New Napyradiomycin Analogues from *Streptomyces* sp. Strain CA-271078. *Mar. Drugs*. **18**:22. doi:10.3390/md18010022.

34. **Castiglione, F., Lazzarini, A., Carrano, L., Corti, E., Ciciliato, I., Gastaldo, L., Candiani, P., Losi, D., Marinelli, F., Selva, E. and Parenti, F**. (2008). Determining the structure and mode of action of microbisporicin, a potent lantibiotic active against multiresistant pathogens. *Chem. Biol*. **15**:22-31. doi: 10.1016/j.chembiol.2007.11.009.

35. **Chater, K. F. and Wilde, L. C.** (1976). Restriction of a bacteriophage of *Streptomyces albus* G involving endonuclease Sall. *J. Bacteriol*. **128**:644-650. doi: 10.1128/jb.128.2.644-650.1976.

36. **Chatterjee, C., Paul, M., Xie, L. and van der Donk.** (2005). Biosynthesis and mode of action of lantibiotics. *Chem. Rev*. **105**:633-683. doi: 10.1021/cr030105v.

37. **Chen, H., Cui, J., Wang, P., Wang, X. and Wen, J.** (2020). Enhancement of bleomycin production in *Streptomyces verticillus* through global metabolic regulation of N-acetylglucosamine and assisted metabolic profiling analysis. *Microb. Cell. Fact*. **19**:32. doi:10.1186/s12934-020-01301-8.

38. **Chen, J. and Xie, J.** (2011). Role and regulation of bacterial LuxR-like regulators. *J. Cell Biochem*. **112**:2694–2702, doi:10.1002/jcb.23219.

39. **Chen, S., Xu, B., Chen, E., Wang, J., Lu, J., Donadio, S., Ge, H. and Wang, H.** (2019). Zn-dependent bifunctional proteases are responsible for leader peptide processing of class III lanthipeptides. *Proc. Natl. Acad. Sci. USA*. **116**:2533-2538. doi: 10.1073/pnas.1815594116.

40. **Chervin, J., Stierhof, M., Tong, M. H., Peace, D., Hansen, K. O., Urgast, D. S., Andersen, J. H., Yu, Y., Ebel, R., Kyeremeh, K., Paget, V., Cimpan, G., Wyk, A. V., Deng, H., Jaspars, M. and Tabudravu, J. N.** (2017). Targeted dereplication of microbial natural products by High Resolution MS and predicted LC retention time. *J. Nat. Prod*. **80**:1370-1377. doi: 10.1021/acs.jnatprod.6b01035.

41. **Choudoir, M. J., Pepe-Ranney, C. and Buckley, D. H.** (2018). Diversification of secondary metabolite biosynthetic gene clusters coincides with lineage divergence in *Streptomyces*. *Antibiotics*. **7**:12. doi: 10.3390/antibiotics7010012.

42. **Claesen, J. and Bibb, M.** (2010). Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides. *Proc. Natl. Acad. Sci. USA*. **107**:16297–16302, doi:10.1073/pnas.1008608107.

43. **Claesen, J. and Bibb, M. J.** (2011). Biosynthesis and regulation of grisemycin, a new member of the linaridin family of ribosomally synthesized peptides produced by *Streptomyces griseus* IFO 13350. *J. Bacteriol*. **193**:2510-2516, doi: 10.1128/JB.00171-11.

44. **Claessen, D., de Jong, W., Dijkhuizen, L. and Wösten, H. A. B.** (2006). Regulation of *Streptomyces* development: reach for the sky. *Trends Microbiol*. **14**:313-319. doi: 10.1016/j.tim.2006.05.008.

45. **Claesson, M., Siitonen, V., Dobritzsch, D., Metsä-Keytelä, M. and Schneider, G.** (2012). Crystal structure of the glycosyltransferase SnogD from the biosynthetic pathway of nogalamycin in *Streptomyces nogalater. FEBS J*. **279**:3251-3263. doi: 10.1111/j.1742-4658.2012.08711.x.

46. **Cohn, F.** (1875). Untersuchungen uber Bacterian. II. *Beitrage z. Biol. d. Pflanzen*. **1**:141-207.

47. **Cuthbertson, L., and Nodwell, J. R.** (2013). The TetR family of regulators. *Microbiol. Mol. Biol. Rev*. **77**:440–475, doi:10.1128/MMBR.00018-13.

48. **Dai, J., Han, R., Xu, Y., Li, N., Wang, J, and Dan, W.** (2020). Recent progress of antibacterial natural products: future antibiotics candidates. *Bioorg. Chem*. **101**:103922. doi: 10.1016/j.bioorg.2020.103922.

49. **Dang, T., and Süssmuth, R. D.** (2017). Bioactive peptide natural products as lead structures for medicinal use. *Acc. Chem. Res.* **50**:1566-1576. doi: 10.1021/acs.accounts.7b00159.

50. **De Simeis, D. and Serra, S**. (2021). Actinomycetes: a never-ending source of bioactive compounds-an overview on antibiotics production. *Antibiotics*. **10**:483. doi: 10.3390/antibiotics10050483.

51. **Decker, H., Rohr, J., Motamedi, H., Zähner, H. and Hutchinson, C. R. R.** (1995). Identification of *Streptomyces olivaceus* Tü 2553 genes involved in the production of the polyketide elloramycin. *Gene*. **166**:121-126. doi: 10.1016/0378−1119(95)00573-7.

52. **Ding, W., Mo, T., Mandalapu, D. and Zhang, Q.** (2018a). Substrate specificity of the cypemycin decarboxylase CypD. *Synth. Syst. Biotechnol*. **3**:159-162. doi: 10.1016/j.synbio.2018.09.002.

53. **Ding, W., Yuan, N., Mandalapu, D., Mo, T., Dong, S. and Zhang, Q**. (2018b). Cypemycin decarboxylase CypD is not responsible for aminovynil-cysteine (AviCys) ring formation. *Org. Lett.* **20**:7670-7673. doi: 10.1021/acs.orglett.8b03380.

54. **Du, D., Wang, L., Tian, Y., Liu, H., Tan, H. and Niu, G.** (2015). Genome engineering and direct cloning of antibiotic gene clusters via phage ΦBT1 integrase-mediated site-specific recombination in *Streptomyces*. *Sci. Rep*. **5**:8740. doi: 10.1038/srep08740.

55. **Du, J., Yuan, Y., Si, T., Lian, J. and Zhao, H.** (2012). Customized optimization of metabolic pathways by combinatorial transcriptional engineering. *Nucleic Acids Res*. **18**:e142. doi: 10.1093/nar/gks549.

56. **Du, L., Sánchez, C., Chen, M., Edwards, D. J. and Shen, B.** (2000). The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chem. Biol*. **7**:623-642. doi: 10.1016/s1074-5521(00)00011-9.

57. **Dyke, K. G., Jevons, M. P. and Parker, M. T.** (1966). Penicillinase production and instrinsic resistance to penicillins in *Staphylococcus aureus*. *Lancet*. **1**:835-838. doi: 10.1016/s0140-6736(66)90182-6.

58. **Ehmann, D. E., Gehring, A. M. and Walsh, C. T.** (1999). Lysine biosynthesis in *Saccharomyces cerevisiae*: mechanism of alpha-aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5. *Biochemistry*. **38**:6171-6177. doi: 10.1021/bi9829940.

59. **Fazal, A., Webb, M. E. and Seipke, R. F.** (2020). The Desotamide Family of Antibiotics. *Antibiotics*. **9**:452. doi: 10.3390/antibiotics9080452.

60. **Feling, R. H., Buchanan, G. O., Mincer, T. J., Kauffman, C. A., Jensen, P. R. and Fenical, W**. (2003). Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora. Angew. Chem. Int.* **42**:355-357. doi: 10.1002/anie.200390115.

61. **Fernández, E., Weissbach, U., Sánchez-Reillo, C., Braña, A. F., Méndez, C., Rohr, J. and Salas J. A.** (1998). Identification of two genes from *Streptomyces argillaceus* encoding glycosyltransferases involved in transfer of a disaccharide during biosynthesis of the antitumor drug mithramycin. *J. Bacteriol*. **180**:4929-4937. doi: 10.1128/JB.180.18.4929-4937.1998.

62. **Fernández, M. and Sánchez, J.** (2002). Nuclease activities and cell death processes associated with the development on surface cultures of *Streptomyces antibioticus* ETH7451. *Microbiology*. **148**:405-412. doi: 10.1099/00221287-148-2-405.

63. **Fernández-Lozano, M. J., Remsing, L. L., Quirós, L. M., Braña, A. F., Fernández, E., Sánchez, C., Méndez, C., Rohr, J. and Salas, J. A**. (2000). Characterization of two polyketide methyltransferases involved in the biosynthesis of the antitumor drug mithramycin by *Streptomyces argillaceus. J. Biol. Chem*. **275**:3065-3074. doi: 10.1074/jbc.275.5.3065.

64. **Fiedler, H. P., Bruntner, C., Riedlinger, J., Bull, A. T., Knutsen, G., Goodfellow, M., Jones, A., Maldonado, L., Pathom-aree, W., Beil, W., Schneider, K., Keller, S. and Sussmüth, R. D.** (2008). Proximicin A, B and C, novel aminofuran antibiotic and anticancer compounds isolated from marine strains of the actinomycete *Verrucosispora. J. Antibiot*. **61**:158-163. doi: 10.1038/ja.2008.125.

65. **Flärdh, K**. (2003). Growth polarity and cell division in *Streptomyces*. *Curr. Opin. Microbiol*. **6**:564-571. doi: 10.1016/j.mib.2003.10.011.

66. **Flett, F., Mersinias, V. and Smith, C. P**. (1997). High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting *Streptomycetes*. *FEMS Microbiol. Lett*. **155**:223-229. doi: 10.1111/j.1574-6968.1997.tb13882.x.

67. **Fontana, A., Iturrino, L., Corens, D. and Crego, A. L**. (2020). Automated open-access liquid chromatography high resolution mass spectrometry to support drug discovery projects. *J. Pharm. Biomed. Anal*. **178**:112908. doi: 10.1016/j.jpba.2019.112908.

68. **Freitas, S., Castelo-Branco, R., Wenzel-Storjohann, A., Vasconcelos, V. M., Tasdemir, D. and Leao, P. N**. (2022). Structure and biosynthesis of desmamides A-C, lipoglycopeptides from the endophytic cyanobacterium *Desmonostoc muscorum* LEGE 12446. *ChemRxiv*. doi: 10.26434/chemrxiv-2022-q069.

69. **Fu, C. Y., Tang, M. C., Peng, C., Li, L., He, L. L., Liu, W. and Tang, G. L**. (2009). Biosynthesis of 3-hydroxy-5-methyl-o-methyltyrosine in the saframycin/safracin biosynthesis pathway. *J. Microbiol. Biotechnol*. **19**:439-446. doi: 10.4014/jmb.0808.484.

70. **Fu, J., Bian, X., Hu, S., Wang, H., Huang, F., Seibert, P. M., Plaza, A., Xia, L., Müller, R., Stewart, A. F. and Zhang, Y.** (2012). Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol*. **30**:440-446. doi: 10.1038/nbt.2183.

71. **Galm, U., Wendt-Pienkowski, E., Wang, L., Huang, S. X., Unsin, C., Tao, M., Coughlin, J. M. and Shen, B.** (2011). Comparative analysis of the biosynthetic gene clusters and pathways for three structurally related antitumor antibiotics: Bleomycin, tallosomycin and zorbamycin. *J. Nat. Prod*. **74**, 526–536. doi:10.1021/np1008152.

72. **Gao, B. and Gupta, R. S.** (2012). Phylogenetic framework and molecular signatures for the main clades of the phylum Actinobacteria. *Microbiol. Mol. Biol. Rev*. **76**:66-112. doi: 10.1128/MMBR.05011-11.

73. **Gehring, A. M., Mori, I., Perry, R. D. and Walsh, C. T.** (1998). The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis*. *Biochemistry*. **37**:11637-11650. doi: 10.1021/bi9812571.

74. **Geiger, C., Korn, S. M., Häsler, M., Peetz, O., Martin, J., Kötter, P., Morgner, N. and Entian, K. D.** (2019). LanI-Mediated Lantibiotic Immunity in *Bacillus subtilis*: Functional Analysis. *Appl. Environ. Microbiol*. **85**:e00534-19. doi: 10.1128/AEM.00534-19.

75. **Genilloud, O.** (2017). Actinomycetes: still a source of novel antibiotics. *Nat. Prod. Rep*. **34**:1203. doi: 10.1039/c7np00026j.

76. **Genilloud, O.** (2018). Mining actinomycetes for novel antibiotics in the omics era: Are we ready to exploit this new paradigm? *Antibiotics.* **7**:85. doi:10.3390/antibiotics7040085.

77. **Georgiou, M. A., Dommaraju, S. R., Guo, X., Mast, D. H. and Mitchell, D. A.** (2020). Bioinformatic and reactivity-based discovery of linaridins. *ACS Chem. Biol*. **15**:2976-2985. doi: 10.1021/acschembio.0c00620.

78. **Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. and Smith, H. O.** (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature*. **6**:343-345. doi: 10.1038/nmeth.1318.

79. **Girauld, M. F. and Naismith, J. H.** (2000). The rhamnose pathway. *Curr. Opin. Struct. Biol*. **19**:687-696. doi: 10.1016/s0959-440x(00)00145-7.

80. **Goethals, K., Vereecke, D., Jaziri, M., Van Montagu, M. and Holsters, M.** (2001). Leafy gall formation by *Rhodococcus fascians*. *Annu. Rev. Phytopathol*. **39**:27–52. doi: 10.1146/annurev.phyto.39.1.27.

81. **Gómez-Escribano, J. P. and Bibb, M. J.** (2011). Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol*. **4**:207-2015. doi: 10.1111/j.1751-7915.2010.00219.x.

82. **Gómez-Escribano, J. P., Alt, S., Bibb, M.** (2016). Next generation sequencing of actinobacteria for the discovery of novel natural products. *Mar. Drugs*. **14**:78. doi: 10.3390/md14040078.

83. **González, A., Rodríguez, M., Braña, A. F., Méndez, C., Salas, J. A. and Olano, C.** (2016). New insights into paulomycin biosynthesis pathway in *Streptomyces albus* J1074 and generation of novel derivatives by combinatorial biosynthesis. *Microb. Cell. Fact*. **15**:56. doi: 10.1186/s12934-016-0452-4.

84. **González-Cerón, G.; Miranda-Olivares, O. J. and Servín-González, L.** (2009). Characterization of the methyl-specific restriction system of *Streptomyces coelicolor* A3 (2) and of the role played by laterally acquired nucleases. *FEMS Microbiol Lett*. **301**:35-43. doi: 10.1111/j.1574-6968.2009.01790.x.

85. **Goodfellow, M., Kämpfer, P., Busse, H. J., Trujillo, M. E., Suzuki, K., Ludwig, W. and Whitman, W. B.** (2012). Bergey's Manual of systematic bacteriology. Second Edition. Volumen five: The Actinobacteria, Part A and B. Springer New York Dordrecht Heidelberg London. doi: 10.1007/978-0-387-68233-4.

86. **Goth, C. K., Vakhrushev, S. Y., Joshi, H. J., Clausen, H. and Schjoldager, K. T.** (2018). Fine-tuning limited proteolysis: a major role for regulated site-specific O-glycosylation. *Trends Biochem. Sci*. **43**:269-284. doi: 10.1016/j.tibs.2018.02.005.

87. **Goto, Y., Li, B., Claesen, J., Shi, Y., Bibb, M. J. and van der Donk, W. A.** (2010). Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. *PLOS Biol.* **8**:e1000339. doi: 10.1371/journal.pbio.1000339.

88. **Grant-Mackie, E. S., Williams, E. T., Harris, P- W. R. and Brimble, M. A.** (2021). Aminovinyl cysteine containing peptides: a unique motif that imparts key biological activity. *JACS Au*. **1**:1527-1540. doi: 10.1021/jacsau.1c00308.

89. **Greunke, C., Duell, E. R., D'Agostino, P. M., Glöckle, A., Lamm, K. anf Gulder, T. A. M.** (2018). Direct pathway cloning (DiPac) to unlock natural product biosynthetic potential. *Metab. Eng*. **47**:334-345. doi: 10.1016/j.ymben.2018.03.010.

90. **Hacker, C., Glinski, M., Hornbogen, T., Doller, A. and Zocher, R.** (2000). Mutational analysis of the N-methyltransferase domain of the multifunctional enzyme enniatin synthetase. *J. Biol. Chem.* **10**:40. doi: 10.1074/jbc.M002614200.

91. **Hammami, R., Zouhir, A., Lay, C. L., Hamida, J, B. and Fliss, I.** (2010). BACTIBASE second release: a database and tool platform for bacteriocin characterization. *BMC. Microbiol*. **10**:22. doi: 10.1186/1471-2180-10-22.

92. **Han, A. R., Park, J. W., Lee, M. K., Ban, Y. H., Yoo, Y. J., Kim, E. J., Kim, E., Kim, B. G., Sohng, J. K. and Yoon, Y. J.** (2011). Development of a *Streptomyces venezuelae*-based combinatorial biosynthetic system for the production of glycosylated derivatives of doxorubicin and its biosynthetic intermediates. *Appl. Environ. Microbiol*. **77**:4912-4923. doi: 10.1128/AEM.02527-10.

93. **Hardisson, C., Manzanal, M. B., Salas, J. A. and Suárez, J. E.** (1978). Fine structure, physiology and biochemistry of arthrospore germination in *Streptomyces antibioticus. J. Gen. Microbiol*. **105**:203-214. doi: 10.1099/00221287-105-2-203.

94. **Hata, T., Tanaka, R. and Ohmomo, S**. (2010). Isolation and characterization of plantaricin ASM1: a new bacteriocin produced by *Lactobacillus plantarum* A-1. *Int. J. Food Microbiol*. **137**:94-99. doi: 10.1016/j.ijfoodmicro.2009.10.021.

95. **Hatano, K., Tamura, T. and Nishii, T.** (1994). Taxonomic status of *Streptomyces coelicolor* A3(2) and *Streptomyces lividans* 66. *Actinomycetological*. **8**:47-50. doi: 10.3209/saj.8_47.

96. **Havarstein, L. S., Diep, D. B. and Nes, I. F.** (1995). A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concominant with export. *Molec. Microbiol*. **16**:229-240. doi: 10.1111/j.1365-2958.1995.tb02295.x.

97. **He, H., Williamson, R. T., Shen, B., Graziani, E. I., Yang, H. Y., Sakya, S. M., Petersen, P. J. and Carter, G. T.** (2002). Mannopeptimycins, novel antibacterial glycopeptides from *Streptomyces hygroscopicus*, LL-AC98. *J. Am. Chem. Soc*. **124**:9729-9736. doi: 10.1021/ja020257s.

98. **Hetrick, K. J. and van der Donk, W. A**. (2017). Ribosomally synthesized and post-translationally modified peptide natural product discovery in the genomic era. *Curr. Opin. Chem. Biol.* **38**:36-44. doi: 10.1016/j.cbpa.2017.02.005.

99. **Hirsch, C. F. and Ensign, J. C.** (1978). Some properties of *Streptomyces viridochromogenes* spores. *J. Bacteriol*. **134**:1056-1063. doi: 10.1099/00221287-131-4-833.

100. **Horna, D. H., Gómez, O., Olano, C., Palomino-Schätzlein, M., Pineda-Lucena, A., Carbajo, R. J., Braña, A. F. and Méndez C, Salas J. A.** (2011). Biosynthesis of the RNA polymerase inhibitor streptolydigin in *Streptomyces lydicus*: tailoring modification of 3-methyl-aspartate. *J. Bacteriol*. **193**:2647-2651. doi: 10.1128/JB.00108-11.

101. **Hozzein, W. H., Rabie, W. and Ali, M. I. A.** (2011). Screening the Egyptian desert actinomycetes as candidates for new antimicrobial compounds and identification of a new desert *Streptomyces* strain. *Afr. J. Biotechnol*. **10**:2295-2301. doi:10.5897/AJB10.1973.

102. **Huang, S. X., Lohman, J. R., Juang, T. and Shen, B.** (2013). A new member of the 4-methylideneimidazole-5-one-containing aminomutase family from the enediyne kedarcidin biosynthetic pathway. *Proc. Natl. Acad. Sci. USA*. **110**:8069-8074. doi: 10.1073/pnas.1304733110.

103. **Hubert, J., Nuzillard, J.M. and Renault, J. H.** (2017). Dereplication strategies in natural product research: How many tools and methodologies behind the same concept? *Phytochem. Rev*. **16**:55-95. doi: 10.1007/s11101-015-9448-7.

104. **Hug, J. J., Bader, C. D., Remskar, M., Cirnski, K. and Müller, R.** (2018). Concepts and methods to access novel antibiotics from actinomycetes. *Antibiotics*. **7**:44. doi: 10.3390/antibiotics7020044.

105. **Huo, L. and van der Donk, W. A.** (2016). Discovery and characterization of bicereucin, an unusual D-amino acid-containing mixed two-component lantibiotic. *J. Am. Chem. Soc*. **138**:5254–5257. doi:10.1021/jacs.6b02513.

106. **Huo, L., Zhao, X., Acedo, J. Z., Estrada, P., Nair, S. K. and van der Donk, W. A.** (2019). Characterization of a dehydratase and methyltransferase in the biosynthesis of a ribosomally-synthesized and post-translationally modified peptide in *Lachnospiraceae*. *Chembiochem*. **21**:190–199, doi:10.1002/cbic.201900483.

107. **Hwang, K. S., Kim, H. U., Charusanti, P., Palsson, B. O. and Lee, S. Y.** (2014). Systems biology and biotechnology of *Streptomyces* species for the production of secondary metabolites. *Biotechnol. Adv.* **32**:255-268. doi: 10.1016/j.biotechadv.2013.10.008.

108. **Hwang, S., Le, L. T. H. L., Shin, J., Shin, J., Lee, M. J. and Oh, D. C.** (2020). Pentaminomycins C–E: Cyclic pentapeptides as autophagy inducers from a mealworm beetle gut bacterium. *Microorganisms*. **8**:1390. doi:10.3390/microorganisms8091390.

109. **Igarashi, Y., Iida, T., Yoshida, R. and Furumai, T.** (2002). Pteridic acids A and B, novel plant growth promoters with auxin-like activity from *Streptomyces hygroscopicus* TP-A0451. *J. Antibiot*. **55**:764-767. doi: 10.7164/antibiotics.55.764.

110. **Iorio, M., Sasso, O., Maffioli, S. I., Bertorelli, R., Monciardini, P., Sosio, M., Bonezzi, F., Summa, M., Brunati, C., Bordoni, R., Corti, G., Tarozzo, G., Piomelli, D., Reggiani, A. and Donadio, S.** (2013). A glycosylated, labionin-containing lanthipeptide with marked antinociceptive activity. *ACS Chem. Biol*. **9**:398-404. doi: 10.1021/cb400692w.

111. **Iwata, M.** (1988). B-(2-Pyridyl)-l-α-alanine from *Streptomyces* sp. *Sci. Rep. Meiji. Seika. Kaisha*. **27**:63-66.

112. **Iyer, L. M., Koonin, E. V. and Aravind, L.** (2001). Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. *Proteins*. **43**:134-144. doi: 10.1002/1097-0134(20010501)43:2<134::AID-PROT1025>3.0.CO;2-I.

113. **Jang, J. P., Hwang, G. J., Kwon, M. C., Ryoo, I. J., Jang, M., Takahashi, S., Ko, S. K., Osada, H., Jang, J. H. and Ahn, J. S.** (2018). Pentaminomycins A and B, hydroxyarginine-containing

cyclic pentapeptides from *Streptomyces* sp. RK88-1441. *J. Nat. Prod*. **81**:806-810. doi: 10.1021/acs.jnatprod.7b00882.

114. **Jiang, W. and Zhu, T. F.** (2016). Targeted isolation and cloning of 100-Kb microbial genomic sequences by Cas9-assisted targeting of chromosome segments. *Nat. Protoc*. **11**:960-975. doi: 10.1038/nprot.2016.055.

115. **Jiang, W., Zhao, X., Gabrieli, T., Lou, G., Ebenstein, Y. and Zhu, T. F.** (2015). Cas9-assisted targeting of chromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat. Commun*. **6**:8101. doi: 10.1038/ncomms9101.

116. **Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. and Madden, T. L.** (2008). NCBI BLAST: a better web interface. *Nucleic. Acids Res*. **36**:W5–W9. doi: 10.1093/nar/gkn201.

117. **Jones, C., Allsopp, L., Horlick, J., Kulasekara, H. and Filloux, A.** (2013). Subinhibitory Concentration of kanamycin induces the *Pseudomonas aeruginosa* type VI secretion system. *PLOS ONE*. **8**:e81132. doi: 10.1371/journal.pone.0081132.

118. **Kaunietis, A., Buivydas, A., Citavicius, D. J. and Kuipers, O. P**. (2019). Heterologous biosynthesis and characterization of a glycocin from a thermophilic bacterium. **10**:1115. *Nat. Commun*. doi: 10.1038/s41467-019-09065-5.

119. **Kautsar, S. A., Blin, K., Shaw, S., Navarro-Muñoz, J. C., Terlouw, B. R., van der Hooft, J. J. J., van Santen, J. A., Tracanna, V., Suárez-Durán, H. G., Andreu, V. P., Selem-Mojica, N., Alanjary, M., Robinson, S. L., Lund, G., Epstein, S. C., Sisto, A. C., Charkoudian, L. K., Collemre, J., Linington, R. G., Weber, T. and Medema, M. H.** (2020). MIBiG 2.0: A repository for biosynthetic gene clusters of known function. *Nucleic. Acids Res*. **48**:454–458. doi:10.1093/nar/gkz882.

120. **Kawai, K., Wang, G., Okamoto, S. and Ochi, K.** (2007). The rare earth, scandium, causes antibiotic overproduction in *Streptomyces* spp. *FEMS. Microbiol. Lett*. **274**:311-315. doi: 10.1111/j.1574-6968.2007.00846.x.

121. **Kaweewan, I., Hemmi, H., Komaki, H. and Kodani, S.** (2020). Isolation and structure determination of a new antibacterial peptide pentaminomycin C from *Streptomyces cacaoi* subsp. *cacaoi*. *J. Antibiot*. **73**:224–229. doi:10.1038/s41429-019-0272-y.

122. **Kaweewan, I., Komaki, H., Hemmi, H. and Kodani, S.** (2017). Isolation and structure determination of new antibacterial peptide curacomycin based on genome mining. *Asian J. Org. Chem*. **6**:1838–1844. doi:10.1002/ajoc.201700433.

123. **Keijser, B. J. F., van Wezel, G. P., Canters, G. W., Kieser, T. and Vijgenboom, E.** (2000). The ram-dependence of *Streptomyces lividans* differentiation is bypassed by copper. *J. Mol. Microbiol. Biotechnol*. **2**:565–574. doi: 10.1159/000071269.

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

125. **Klinman, J. P. and Bonnot, F.** (2014). Intrigues and intricacies of the biosynthetic pathways for the enzymatic quinocofactors: PQQ, TTQ, CTQ, TPQ, and LTQ. *Chem. Rev*. **114**:4343-4365. doi: 10.1021/cr400475g.

126. **Kloosterman, A. M., Cimermancic, P., Elsayed, S. S., Du, C., Hadjithomas, M., Donia, M. S., Fischbach, M. A., van Wezel, G. P. and Medema, M. H.** (2020). Expansion of RiPP biosynthetic space through integration of pan-genomics and machine learning uncovers a novel class of lanthipeptides. *PLOS Biol*. **18:** e3001026. doi:10.1371/journal.pbio.3001026.

127. **Kloosterman, A. M., Medema, M, M. H. and van Wezel, G. P.** (2021). Omics-based strategies to discover novel classes of RiPP natural products. *Curr. Opin. Biotech*. **69**:60-67. doi: 10.1016/j.copbio.2020.12.008.

128. **Knerr, P. J. and van der Donk, W. A.** (2012). Discovery, biosynthesis, and engineering of lanthipeptides. *Annu. Rev. Biochem*. **81**:479-505. doi: 10.1146/annurev-biochem-060110-113521.

129. **Komani, H., Sakurai, K., Hosoyama, A., Kimura, A., Igarashi, Y. and Tamura, T.** (2018). Diveristy of nonribosomal peptide synthetase and polyketide synthase gene clusters among taxonomically close *Streptomyces* strains. *Sci. Rep*. **8**:6888. doi: 10.1038/s41598-018-24921-y.

130. **Komatsu, M., Komatsu, K., Koiwai, H., Yamada, Y., Kozone, I., Izumikawa, M., Hashimoto, J., Takagi, M., Omura, S., Shin-ya, K., Cane, D. E. and Ikeda, H.** (2013). Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. *ACS Synth. Biol*. **2**:384–396. doi: 10.1021/ sb3001003.

131. **Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H. and Phillippy, A. M.** (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. **27**:722-736. doi: 10.1101/gr.215087.116.

132. **Krawczyk, B., Völler, G. H., Völler, J., Ensle, P. and Süssmuth, R. D.** (2012). Curvopeptin: a new lanthionine-containing class III lantibiotic and its co-substrate promiscuous synthetase. *Chembiochem*. **13**:2065-2071. doi: 10.1002/cbic.201200417.

133. **Krawczyk, J. M., Völler, G. H., Krawczyk, B., Kretz, J., Brönstrup, M. and Süssmuth, R. D.** (2013). Heterologous expression and engineering studies of labyrinthopeptins, class III lantibiotics from *Actinomadura namibiensis. Chem. Biol*. **20**:111–122. doi: 10.1016/j.chembiol.2012.10.023.

134. **Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K.** (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol*. **35:**1547-1549. doi:10.1093/molbev/msy096.

135. **Kuranaga, T., Matsuda, K., Sano, A., Kobayashi, M., Ninomiya, A., Takada, K., Matsunaga, S. and Wakimoto, T.** (2018). Total synthesis of the nonribosomal peptide surugamide B and identification of a new offloading cyclase family. *Angew. Chem. Int.* **57**:9447-9451. doi: 10.1002/anie.201805541.

136. **Labeda, D. P.** (1987). Transfer of the Type Strain of *Streptomyces erythraeus* (Waksman 1923) Waksman and Henrici 1948 to the genus *Saccharopolyspora* Lacey and Goodfellow 1975 as *Saccharopolyspora erythraea* sp. *nov*., and designation of a neotype strain for *Streptomyces erythraeus. Int. J. Syst. Evol. Microbiol*. **37**:19-22. doi: 10.1099/00207713-37-1-19.

137. **Labeda, D. P., Doroghazi, J. R., Ju, K. S. and Metcalf, W. W.** (2014). Taxonomic evaluation of *Streptomyces albus* and related species using multilocus sequence analysis and proposals to emend the description of *Streptomyces albus* and describe *Streptomyces pathocidini* sp. nov. *Int. J. Syst. Evol. Microbiol*. **64**:894-900. doi: 10.1099/ijs.0.058107-0

138. **Lachance, H., Wetzel, S., Kumar, K. and Waldmann, H.** (2012). Charting, navigating and populating natural product chemical space for drug discovery. *J. Med. Chem*. **55**:5989-6001. doi: 10.1021/jm300288g.

139. **Lairson, L. L., Henrissat, B., Davies, G. J. and Withers, S. G.** (2008). Glycosyltransferases: structure, functions and mechanisms. *Annu. Rev. Biochem*. **77**:521-555. doi: 10.1146/annurev.biochem.76.061005.092322.

140. L**atham, J. A., Iavarone, A. T., Barr, I., Juthani, P. V. and Klinman, J. P.** (2015). PqqD is a novel peptide chaperone that forms a ternary complex with the radical S-adenosylmethionine protein PqqE in the pyrroloquinoline quinone biosynthetic pathway. *J. Biol. Chem*. **290**:12908–12918. doi: 10.1074/jbc.M115.646521.

141. **Laureti, L., Song, L., Huang, S., Corre, C., Leblond, P., Challis, G. L. and Aigle, B.** (2011). Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens. Proc. Nat. Acad. Sci. USA*. **108**:6258–6263. doi: 10.1073/pnas.1019077108.

142. **Lee, H. and van der Donk, W. A.** (2022). Macrocyclization and backbone modification in RiPP biosynthesis. *Annu. Rev. Biochem*. **91**. doi: 10.1146/annurev-biochem-032620-104956.

143. **Li, Y., Kong, L., Shen, J., Wang, Q., Liu, Q., Yang, W., Deng, Z. and You, D.** (2019). Characterization of the positive SARP family regulator PieR for improving piericidin A1 production in *Streptomyces piomogeues* var. *hangzhouwanensis. Synth. Syst. Biotechnol*. **4**:16–24. doi: 10.1016/j.synbio.2018.12.002.

144. **Little, R. F. and Hertweck, C.** (2022). Chain release mechanisms in polyketide and nonribosomal peptide biosynthesis. *Nat. Prod. Rep*. **39**:163. doi: 10.1039/d1np00035g.

145. **Liu, H., Jiang, H., Haltli, B., Kulowski, K., Muszynska., Feng, X., Summers, M., Young, M., Graziani, E., Koehn, F., Carter, G. T. and He, M.** (2009). Rapid cloning and heterologous expression of the meridamycin biosynthetic cluster using a versatile *Escherichia coli-Streptomyces* artificial chromosome vector, pSBAC. *J. Nat. Prod*. **72**:389-395. doi: 10.1021/np8006149.

146. **Liu, Z., Zhao, Y., Huang, C. and Luo, Y.** (2021). Recent advances in silent gene cluster activation in *Streptomyces*. *Front. Bioeng. Biotechnol*. **9**:632230. doi: 10.3389/fbioe.2021.632230.

147. **Liu, X., Tang, J., Wang, L. and Lio, R.** (2019a). Mechanism of CuO nanoparticles on stimulating production of actinorhodin in *Streptomyces coelicolor* by transcriptional analysis. *Sci Rep*. **9**:11253. doi: 10.1038/s41598-019-46833-1.

148. **Liu, X., Tang, J., Wang, L. and Giesy, J. P.** (2019b). Al_2O_3 nanoparticles promote secretion of antibiotics in *Streptomyces coelicolor* by regulating gene expression through the nano effect. *Chemosphere*. **226**:687–695. doi: 10.1016/j.chemosphere.2019.03.156.

149. **Liu, X., Tang, J., Wang, L., Liu, Q. and Liu, R.** (2019c). A comparative analysis of ball-milled biochar, graphene oxide, and multi-walled carbon nanotubes with respect to toxicity induction in *Streptomyces*. *J Environ. Manage*. **243**:308–317. doi: 10.1016/j.jenvman.2019.05.034.

150. **Locci, R., Baldacci, E. and Petrolini-Baldan, B.** (1969). The genus *Streptoverticillium*. A taxonomic study. *J. Microbiol*. **17**:1-60.

151. **Lohans, C. T., Li, J. L. and Vederas, J. C.** (2014). Structure and biosynthesis of carnolysin, a homologue of enterococcal cytolysin with D-amino acids. *J. Am. Chem. Soc*. **136**:13150–13153. doi:10.1021/ja5070813.

152. **Lombó, F., Abdelfattah, M. S., Brana, A. F., Salas, J. A., Rohr, J. and Mendez, C.** (2009). Elucidation of oxygenation steps during oviedomycin biosynthesis and generation of derivatives with increased antitumor activity. *Chembiochem*. **10**:296-303. doi: 10.1002/cbic.200800425.

153. **Lombó, F., Gibson, M., Greenwell, L., Braña, A. F., Rohr, J., Salas, J. A. and Méndez, C.** (2004). Engineering biosynthetic pathways for deoxysugars: branched-chain sugar pathways and derivatives from the antitumor tetracenomycin. *Chem. Biol*. **11**:1709-1718. doi: 10.1016/j.chembiol.2004.10.007.

154. **Luzhetskyy, A., Taguchi, T., Fedoryshyn, M., Durr, C., Wohlert, S. E., Novikov, V. and Bechthold, A.** (2005). LanGT2 catalyzes the first glycosylation step during landomycin A biosynthesis. *Chembiochem*. **6**:1406-1410. doi: 10.1002/cbic.200500018.

155. **Luzhetskyy, A., Weiss, H., Charge, A., Welle, E., Linnenbrink, A., Vente, A. and Bechthold, A.** (2007). A strategy for cloning glycosyltransferase genes involved in natural products biosynthesis. *Appl. Microbiol. Biotechnol*. **75**:1367-1375. doi: 10.1007/s00253-007-0950-8.

156. **Ma, S. and Zhang, Q.** (2020). Linaridin natural products. *Nat. Prod. Rep*. **37**:1152–1163, doi:10.1039/C9NP00074G.

157. **MacNeil, D. J., Gewain, K. M., Ruby, C. L., Dezeny, G., Gibbons, P. H. and MacNeil, T.** (1992). Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene*. **111**:61-8. doi: 10.1016/0378-1119(92)90603-m.

158. **Magarvey, N. A., Haltli, B., He, M., Greenstein, M. and Hucul, J. A.** (2006). Biosynthetic pathway for mannopeptimycins, lipoglycopeptide antibiotics active against drug-resistant Gram-Positive pathogens. *Antimicrob. Agents Chemother*. **50**:2167–2177. doi:10.1128/aac.01545-05.

159. **Malmierca, M. G., Pérez-Victoria, I., Martín, J., Reyes, F., Méndez, C., Olano, C. and Salas, J. A.** (2018). Cooperative involvement of glycosyltransferases in the transfer of amino sugars during the biosynthesis of the macrolactam sipanmycin by *Streptomyces* sp. strain CS149. *App. Environ. Microbiol*. **84**:e01462-18. doi: 10.1128/AEM.01462-18.

160. **Mandalapu, D., Ji, X., Chen, J., Guo, C., Liu, W. Q., Ding, W., Zhou, J. and Zhang, Q.** (2018). Thiosterase-mediated synthesis of teixobactin analogues: Mechanism and substrate specificity. *J. Org. Chem*. **83**:7271-7275. doi: 10.1021/acs.joc.7b02462.

161. **Manteca, A. and Sánchez, J.** (2009). *Streptomyces* development in colonies and soils. *Appl. Environ. Microbiol*. **75**:2920-2924. doi: 10.1128/AEM.02288-08.

162. **Manteca, A., Claessen, D., López-Iglesias, C. and Sánchez, J.** (2007). Aerial hyphae in surface cultures of *Streptomyces lividans* and *Streptomyces coelicolor* originate from viable segments surviving an early programmed cell death event. *FEMS Microbiol. Lett*. **274**:118-125. doi: 10.1111/j.1574-6968.2007.00825.x.

163. **Mao, D., Okada, B. K., Wu, Y., Xu, F., and Seyedsayamdost, M. R.** (2018). Recent advances in activating silent biosynthetic gene clusters in bacteria. *Curr. Opin. Microbiol*. **45**:156–163. doi: 10.1016/j.mib.2018.05.001.

164. **Martens, E. and Demain, A. L.** (2017). The antibiotic resistance crisis, with a focus on the United States. *J. Antibiot*. **70**:520-526. doi: 10.1038/ja.2017.30.

165. **Martín, J. F.** (2004). Phosphate control of the biosynthesis of antibiotics and other secondary metabolites is mediated by the PhoR-PhoP system: an unfinished story. *J. Bacteriol.* **186**:5197-5201. doi: 10.1128/JB.186.16.5197-5201.2004.

166. **Martín, J. F. and Demain, A. L.** (1980). Control of antibiotic biosynthesis. *Microbiol. Rev*. **44**:230-251. doi: 10.1128/mr.44.2.230-251.1980.

167. **Martín, J. F. and Liras, P.** (1989). Organization and expression of genes involved in the biosynthesis of antibiotics and other secondary metabolites. *Annu. Rev. Microbiol*. **43**:173-206. doi: 10.1146/annurev.mi.43.100189.001133.

168. **Martín, J., Crespo, G., González-Menéndez, V., Pérez-Moreno, G., Sánchez-Carrasco, P., Pérez-Victoria, I., Ruiz-Pérez, L. M., González-Pacanowska, D., Vicente, F., Genilloud, O., Bills, G. F. and Reyes, F.** (2014). MDN-0104, an antiplasmodial betaine lipid from *Heterospora chenopodii. J. Nat. Prod.* **77**:2118–2123 doi: 10.1021/np500577v.

169. **Martín, J., Da Sousa, T. S., Crespo, G., Palomo, S., González, I., Tormo, J. R., de la Cruz, M., Anderson, M., Hill, R. T., Vicente, F., Genilloud, O. and Reyes, F.** (2013). Kocurin, the true structure of PM 181104, an anti-MRSA thiazolyl peptide from the marine-derived bacterium *Kocuria palustris. Mar. Drugs*. **11**:387–398. doi: 10.3390/md11020387.

170. **Matsuda, K., Fujita, K. and Wakimoto, T.** (2021). PenA, a penicillin-binding protein type thiosterase specialized for small peptide cyclization. *J. Ind. Microbiol. Biotech*. **48**. doi: 10.1093/jimb/kuab023.

171. **Matsuda, K., Kobayashi, M., Kuranaga, T., Takada, K., Ikeda, H., Matsunaga, S. and Wakimoto, T.** (2019a). SurE is a trans-acting thiosterase cuclizing two distinct non-ribosomal peptides. *Org. Biomol. Chem*. **17**:1058-1061. doi: 10.1039/c8ob02867b.

172. **Matsuda, K., Kuranaga, T. and Wakimoto, T.** (2019b). A new cyclase family catalyzing head-to-tail macrolactamization of nonribosomal peptides. *J. Syn. Org. Chem. Jpn*. **77**:1106-1115. doi: :10.5059/yukigoseikyokaishi.77.1106.

173. **Mazodier, P., Petter, R. and Thompson, C.** (1989). Intergeneric conjugation between *Escherichia coli* and *Streptomyces* species. *J. Bacteriol*. **171**:3583-3585. doi: 10.1128/jb.171.6.3583-3585.1989.

174. **McCarthy, A. J. and Williams, S. T.** (1992). Actinomycetes as agents of biodegradation in the environment- a review. *Gene*. **115**:189-192. doi: 10.1016/0378-1119(92)90558-7.

175. **McCormick, M. H., McGuire, J. M., Pittenger, G. E., Pittenger, R. C. and Stark, W. M.** (1955). Vancomycin, a new antibiotic. I. Chemical and biologic properties. *Antibiot. Annu*. **3**:606-611.

176. **McErlean, M., Overbay, J. and van Lanen, S.** (2019). Refining and expanding nonribosomal peptide synthetase function and mechanism. *J. Ind. Microbiol. Biotechnol*. **46**:493–513. doi: 10.1007/s10295-018-02130-w.

177. **McGuire, J. H., Bunch, R. L., Anderson, R. C., Boaz, H. E., Flynn, E. H., Powell, H. W., & Smith, J. W.** (1952). Ilotycin, a new antibiotic. *Antibiot. Chemother*. **2**:281-283.

178. **Méndez, C. and Salas, J. A.** (2001). The role of ABC transporters in antibiotics-producing organisms: drug secretion and resistance mechanisms. *Res. Microbiol*. **152**:341-350. doi: 10.1016/s0923-2508(01)01205-0.

179. **Menéndez, N., Nur-e-Alam, M., Fischer, C., Braña, A. F., Salas, J. A., Rohr, J. and Méndez, C.** (2006). Deoxysugar transfer during chromomycin A3 biosynthesis in *Streptomyces griseus* subsp. *griseus*: new derivatives with antitumor activity. *Appl. Environ. Microbiol*. **72**:167-177. doi: 10.1128/AEM.72.1.167-177.2006.

180. **Messner, P., Christian, R., Neuninger, C. and Schulz, G.** (1995). Similarity of "core" structures in two different glycans of tyrosine-linked eubacterial S-layer glycoproteins. *J. Bacteriol.* **177**:2188–2193. doi: 10.1128/jb.177.8.2188-2193.1995.

181. **Miao, V., Coëffet-LeGal, M. F., Brian, P., Brost, R., Penn, J., Whiting, A., Martin, S., Ford, R., Parr, I., Bouchard, M., Silva, C. J., Wrigley, S. K. and Baltz, R. H.** (2005). Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology*. **151**:1507-1523. doi: 10.1099/mic.0.27757-0.

182. **Miethke, M., Pieroni, M., Weber, T., Brönstrup, M., Hammann, P., Halby, L., Arimondo, P. B., Glaser, P., Aigle, B., Bode, H. B., Moreira, R., Li, Y., Luzhetskyy, A., Medema, M. H., Pernodet, J. L., Stadler, M., Tormo, J. R., Genilloud, O., Truman, A. W., Weissman, K. J., Takano, E., Sabatini, S., Stegmann, E., Brötz-Oesterhelt, H., Wohlleben, W., Seemann, M., Empting, M., Hirsch, A. K. H., Loretz, B., Lehr, C. M., Titz, A., Herrmann, J., Jaeger, T., Alt, S., Hesterkamp, T., Winterhalter, M., Schiefer, A., Pfarr, K., Hoerauf, A., Graz, H., Graz, M., Lindvall, M., Ramurthy, S., Karlén, A., van Dongen, M., Petkovic, H., Keller, A., Peyrane, F., Donadio, S., Fraisse, L., Piddock, L. J. V., Gilbert, I. H., Moser, H. E. and Müller, R.** (2021). Towards the sustainable discovery and development of new antibiotics. *Nat. Rev. Chem*. **5**:726-749. doi: 10.1038/s41570-021-00313-1.

183. **Miller, J. H.** (1972). Experiments in molecular genetics. Cold Spring Harbor Laboratory. New York.

184. **Minami, Y., Yoshida, K., Azuma, R., Urakawa, A., Kawauchi, T., Otani, T., Komiyama, K. and Omura, S.** (1994). Structure of cypemycin, a new peptide antibiotic. *Tetrahedron Lett.* **35**:8001-8004. doi: 10.1016/0040-4039(94)80033-2.

185. **Mitousis, L., Thoma, Y. and Musiol-Kroll, E. M.** (2020). An update on molecular tools for genetic engineering of actinomycetes-The source of important antibiotics and other valuable compounds. *Antibiotics*. **9**:494. doi: 10.3390/antibiotics9080494.

186. **Mo, T., Liu, W. Q., Ji, W., Zhao, J., Chen, T., Ding, W., Yu, S and Zhang, Q.** (2017). Biosynthetic insights into linaridin natural products from genome mining and precursor peptide mutagenesis. *ACS Chem. Biol*. **12**:1484–1488, doi:10.1021/acschembio.7b00262.

187. **Montalbán-López, M., Scott, T. A., Ramesh, S., Rahman, I. R., van Heel, A. J., Viel, J. H., Bandarian, V., Dittmann, E., Genilloud, O., Goto, Y., Grande-Burgos, M. J., Hill, C., Kim, S., Koehnke, J., Latham, J. A., Link, A. J., Martínez, B., Nair, S. K., Nicolet, Y., Rebuffat, S., Sahl, H. G., Sareen, D., Schmidt, E. W., Schmitt, L., Severinov, K., Süssmuth, R. D., Truman, A. W., Wang, H., Weng, J. K., van Wezel, G. P., Zhang, Q., Zhong, J., Piel, J., Mitchell, D. A., Kuipers, O. P. and van der Donk, W. A.** (2021). New developments in RiPP discovery, enzymology and engineering. *Nat. Prod. Rep*. **38**:130-239. doi: 10.1039/d0np00027b.

188. **Mudalungu, C. M., Von Törne, W. J., Voigt, K., Rückert, C., Schmitz, S., Sekurova, O. N., Zotchev, S. B. and Süssmuth, R. D.** (2019). Noursamycins, chlorinated cyclohexapeptides identified from molecular networking of *Streptomyces noursei* NTR-SR4. *J. Nat. Prod*. **82**:1478–1486. doi: 10.1021/acs.jnatprod.8b00967.

189. **Müller, W. M., Schmiederer, T., Ensle, P., Süssmuth, R. D.** (2010). In vitro biosynthesis of the prepeptide of type III lantibiotic labyrinthopeptin A2 including formation of a C-C bond as a post-translational modification. *Angew. Chem. Int*. **49**:2436-2440. doi: 10.1002/anie.200905909.

190. **Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H.** (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold. Spring. Harb. Symp. Quant. Biol*. **51**:263-273. doi: 10.1101/sqb.1986.051.01.032.

191. **Antimicrobial Resistance Collaborators.** (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*. **399**:12-18. doi: 10.1016/S0140-6736(21)02724-0.

192. **Myronovskyi, M., Tokovenko, B., Brötz, E., Rückert, C., Kalinowski, J. and Luzhetskyy, A.** (2014). Genome rearrangements of *Streptomyces albus* J1074 lead to the carotenoid gene cluster activation. *Appl. Microbiol. Biotechnol*. **98**:795-806. doi: 10.1007/s00253-013-5440-6.

193. **Nah, H. J., Woo, M. W., Choi, S. S. and Kim, E. S.** (2015). Precise cloning and tandem integration of large polyketide biosynthetic gene cluster using *Streptomyces* artificial chromosome system. *Microb. Cell. Fact*. **14**:140. doi: 10.1186/s12934-015-0325-2.

194. **Nakajima, S., Niiyama, K., Ihara, M. and Suda, H.** (1991). Endothelin-binding inhibitors, BE-18257A and BE-18257B II. Structure determination. *J. Antibiot*. **44**:1348-1356. doi: 10.7164/antibiotics.44.1348.

195. **Nan, J., Brostromer, E., Liu, X. Y., Kristensen, O. and Su, X. D.** (2009). Bioinformatics and structural characterization of a hypothetical protein from *Streptococcus* mutans: implication of antibiotic resistance. *PLOS ONE*. **4**:e7245. doi: 10.1371/journal.pone.0007245.

196. **Nathan, C. and Cars, O**. (2014). Antibiotic resistance-Problems, progress and prospects. *N. Engl. J. Med*. **371**:1761-1763. doi: 10.1056/NEJMp1408040.

197. **Neethu, K., Karthik, L. and Li, Z.** (2022). Nonribosomally and Ribosomally Synthesized Bioactive Peptides (NRPS and RiPPs) from Actinobacteria. In: Karthik, L. (eds) Actinobacteria. Springer, Singapore. doi: 10.1007/978-981-16-5835-8_5.

198. **Newman, D. J. and Cragg, G. M.** (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod*. **83**:770-803. doi: 10.1021/acs.jnatprod.9b01285.

199. **Nguyen, C, T., Dhakal, D., Pham, V. T. T., Nguyen, H. T. and Sohng, J. K.** (2020). Recent advances in strategies for activation and discovery/characterization of cryptic biosynthetic gene clusters in *Streptomyces*. *Microorganisms*. **8**:616. doi: 10.3390/microorganisms8040616.

200. **Nguyen, G. K. T., Qiu, Y., Cao, Y., Hemu, X., Liu, C. F. and Tam, J. P.** (2016). Butelase-mediated cyclization and ligation of peptides and proteins. *Nat. Protoc*. **11**:1977-1988. doi: 10.1038/nprot.2016.118.

201. **Ninomiya, A., Katsuyama, Y., Kuranaga, T., Miyazaki, M., Nogi, Y., Okada, S., Wakimoto, T., Ohnishi, Y., Matsunaga, S. and Takada, K.** (2016). Biosynthetic gene cluster for surugamide A encompasses an unrelated decapeptide, surugamide F. *Chembiochem*. **17**:1709-1712. doi: 10.1002/cbic.201600350.

202. **Norris, G. E. and Patchett, M. L.** (2016). The glycocins: in a class of their own. *Curr. Op. Struct. Biol*. **40**:112-119. doi: 10.1016/j.sbi.2016.09.003.

203. **Novella, I. S., Barbés, C. and Sánchez, J.** (1992). Sporulation of *Streptomyces antibioticus* ETHZ7451 in submerged culture. *Can. J. Microbiol*. **38**:769-773. doi: 10.1139/m92-125.

204. **Nybo, S. E., Shabaan, K. A., Kharel, M. K., Sutardjo, H., Salas, J. A., Méndez, C. and Rohr, J.** (2012). Ketoolivosyl-tetracenomycin C: a new ketosugar bearing tetracenomycin reveals new insight into the substrate flexibility of glycosyltransferase ElmGT. *Bioorg. Med. Chem. Lett*. **22**:2247-2250. doi: 10.1016/j.bmcl.2012.01.094.

205. **Ochi, K., Tanaka, Y. and Tojo, S.** (2014). Activating the expression of bacterial cryptic genes by *rpoB* mutations in RNA polymerase or by rare earth elements. *J. Ind. Microbiol. Biotechnol*. **41**:403-414. doi: 10.1007/s10295-013-1349-4.

206. **Okada, B. K., Seyedsayamdost, M. R.** (2017). Antibiotic dialogues: induction of silent biosynthetic gene clusters by exogenous small molecules. *FEMS Microbiol. Rev*. **41**:19-33. doi: 10.1093/femsre/fuw035.

207. **Olano, C., García, I., González, A., Rodríguez, M., Rozas, D., Rubio, J., Sánchez-Hidalgo, M., Braña, A. F., Méndez, C. and Salas, J. A.** (2014). Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074. *Microb. Biotechnol*. **7**:242-256. doi: 10.1111/1751-7915.12116.

208. **Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pae, N. R. and Stahl, D. A.** (1986). Microbial ecology and evolution: a ribosomal RNA approach. *Annu. Rev. Microbiol*. **40**:337-365. doi: 10.1146/annurev.mi.40.100186.002005.

209. **Oman, T. J., Boettcher, J. M., Wang, H., Okalibe, X. N. and van der Donk, W. A.** (2011). Sublancin is not a lantibiotic but an S-linked glycopeptide. *Nat. Chem. Biol*. **7**:78-80. doi: 10.1038/nchembio.509.

210. **Ortega, M. A. and van der Donk, W. A.** (2016). New insights into the biosynthetic logic of ribosomally synthesized and post-translationally modified peptide natural products. *Cell Chem. Biol.* **23**:31-44. doi: 10.1016/j.chembiol.2015.11.012.

211. **Ortega, M. A., Velásques, J. E., Garg, N., Zhang, Q., Joyce, R. E., Nair, S. K. and van der Donk, W. A.** (2014). Substrate specificty of the lanthipeptide peptidase ElxP and the oxidoreductase ElxO. *ACS Chem. Biol*. **9**:1718-1725. doi: 10.1021/cb5002526.

212. **Ortiz de Montellano, R. P.** (2010). Hydrocarbon hydroxylation by cytochrome P450 enzymes. *Chem. Rev*. **110**:932-948. doi: 10.1021/cr9002193.

213. **Ortiz-López, F. J., Carretero-Molina, D., Sánchez-Hidalgo, M., Martín, J., González, I., Román-Hurtado, F., de la Cruz, M., García-Fernández, S., Reyes, F., Deisinger, J. P., Müller, A., Schneider, T. and Genilloud, O.** (2020). Cacaoidin, first member of the new lanthidin RiPP family. *Angew. Chem. Int*. **59**:12654-12658. doi: 10.1002/anie.202005187.

214. **Pan, R., Bal, X., Chen, J., Zhang, H. and Wang, H.** (2019). Microbe secondary metabolites using OSMAC strategy: a literature review. *Front. Microbiol*. **10**:294. doi: 10.3389/fmicb.2019.00294.

215. **Pan, Y., Lu, C., Dong, H., Yu, L., Liu, G. and Tan, H.** (2013). Disruption of *rimP-SC*, encoding a ribosome assembly cofactor, markedly enhances the production of several antibiotics in *Streptomyces coelicolor. Microb. Cell Fact*. **12**:65. doi: 10.1186/1475-2859-12-65.

216. **Parenti, F., Beretta, G., Berti, M. and Arioli, V.** (1978). Teichomycins, new antibiotics from *Actinoplanes teichomyceticus* nov. sp. *J. Antibiot*. **31**:276-283. doi: 10.7164/antibiotics.31.276.

217. **Park, Y., Shin, I. and Rhee, S.** (2015). Crystal structure of the effector protein HopA1 from *Pseudomonas syringae. J. Struct. Biol*. **189**:276–280, doi: 10.1016/j.jsb.2015.02.002.

218. **Parte, A. C., Sardà Carbasse, J., Meier-Kolthoff, J. P., Reimer, L. C. and Göker, M.** (2020). List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* **70:**5607-5612. doi: 10.1099/ijsem.0.004332.

219. **Patsos, G. and Corfield, A.** (2009). O-Glycosylation: structural diversity and functions. pp.111-137. *Wiley-VCH*. Weinheim, Germany.

220. **Paulutat, T., Kulik, A., Hausmann, H., Karagouni, A. D., Zinecker, H., Imhoff, J. F. and Fiedler, H. P.** (2010). Grecocyclines: new angucyclines from *Streptomyces* sp. Acta 1362. *Eur. J. Org. Chem*. **62**:445-452. doi: 10.1002/ejoc.201000054.

221. **Pérez-Victoria, I., Martín, J. and Reyes, F.** (2016). Combined LC/UV/MS and NMR strategies for the dereplication of marine natural products. *Planta Med*. **82**:857-871. doi: 10.1055/s-0042-101763.

222. **Piddock, L. J. V.** (2012). The crisis of no new antibiotics: what is the way forward? *Lancet*. *Infect. Dis*. **12**:249-253. doi: 10.1016/S1473-3099(11)70316-4.

223. **Pishesha, N., Ingram, J. R. and Ploegh, H. L.** (2018). Sortase A: a model for transpeptidation and its biological applications. *Annu. Rev. Cell Dev. Biol*. **34**:163-188. doi: 10.1146/annurev-cellbio-100617-062527.

224. **Plat, A., Kluskens, L. D., Kuipers, A., Rink, R. and Moll, G. N.** (2011). Requirements of the engineered leader peptide of nisin for inducing modification, export and cleavage. *Appl. Environ. Microbiol.* **77**:604-611. doi: 10.1128/AEM.01503-10.

225. **Pokhrel, R., Bhattarai, N., Baral, P., Gerstman, B. S., Park, J. H., Handfield, M. and Chapagain, P. P.** (2022). Lipid II binding and transmembrane properties of various antimicrobial lanthipeptides. *J. Chem. Theory Comput*. **18**:516-525. doi: 10.1021/acs.jctc.1c00666.

226. **Pradel, E., Parker, C. T. and Schnaitman, C. A.** (1992). Structures of the *rfaB*, *rfaI*, *rfaJ*, and *rfaS* genes of *Escherichia coli* K-12 and their roles in assembly of the lipopolysaccharide core. *J. Bacteriol*. **174**:4736-4745. doi: 10.1128/jb.174.14.4736-4745.1992.

227. **Quail, M. A., Kozarewa, I., Smith, F., Scally, A., Stephens, P. J., Durbin, R., Swerdlow, H. and Turner, D. J.** (2008). A large genome center's improvement to the Illumina sequencing system. *Nat. Methods*. **5**:1005-1010. doi: 10.1038/nmeth.1270.

228. **Radauer, C., Lackner, P. and Breiteneder, H.** (2008). The Bet v1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol. Biol*. **8**:286. doi: 10.1186/1471-2148-8-286.

229. **Radwanski, E. R. and Last, R. L.** (1995). Tryptophan biosynthesis and metabolism: biochemical and molecular genetics. *Plant Cell*. **7**:921-934. doi: 10.1105/tpc.7.7.921.

230. **Rateb, M. E., Zhai, Y., Ehrner, E., Rath, C. M., Wang, X., Tabudravu, J., Ebel, R., Bibb, M., Kyeremeh, K., Dorrestein, P. C. Hong, K., Jaspars, M. and Deng, H.** (2015). Legonaridin, a new member of linaridin RiPP from a Ghanaian *Streptomyces* isolate. *Org. Biomol. Chem*. **13**:9585-9592. doi: 10.1039/C5OB01269D.

231. **Read, A. F. and Woods, R. J.** (2014). Antibiotic resistance management. *Evol. Med. Public Health*. **2014**:147. doi: 10.1093/emph/eou024.

232. **Rebets, Y., Brözt, E., Tokovenko, B. and Luzhestskyy, A.** (2014). Actinomycetes biosynthetic potential: how to bridge in silico and in vivo? *J. Ind. Microbiol. Biotechnol*. **41**:387-402. doi: 10.1007/s10295-013-1352-9.

233. **Repka, L. M., Chekan, J. R., Nair, S. K. and van der Donk, W. A.** (2017). Mechanistic understanding of lanthipeptide biosynthetic enzymes. *Chem. Rev*. **117**:5457-5520. doi: 10.1021/acs.chemrev.6b00591.

234. **Rhoads, A. and Au, K. F.** (2015). PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics*. **13**:278-289. doi: 10.1016/j.gpb.2015.08.002.

235. **Richard, A. S., Zhang, A., Park, S. J. and Chloe, H.** (2015). Virion-associated phosphatidylethanolamine promotes TIM1-mediated infection by Ebola, dengue, and West Nile viruses. *Proc. Natl. Acad. Sci. USA*. **112**:14682-14687. doi: 10.1073/pnas.1508095112.

236. **Robbel, L. and Marahiel, M. A.** (2010). Daptomycin, a bacterial lipopeptide synthesized by a nonribosomal machinery. *J. Biol. Chem*. **285**:27501-27508. doi: 10.1074/jbc.R110.128181.

237. **Rodriguez, E., Hu, Z., Ou, S., Volchegursky, Y., Hutchinson, C. R. and McDaniel, R.** (2003). Rapid engineering of polyketide overproduction by gene transfer to industrially optimized strain. *J. Ind. Microbiol. Biotechnol*. **30**:480–488. doi:10.1007/s10295-003-0045-1.

238. **Rogers, L. A.** (1928). The inhibiting effect of *Streptococcus lactis* on *Lactobacillus bulgarius. J. Bacteriol*. **16**:321-325. doi: 10.1128/jb.16.5.321-325.1928.

239. **Román-Hurtado, F., Sánchez-Hidalgo, M., Martín, J., Ortiz-López, F. J. and Genilloud, O.** (2021a). Biosynthesis and heterologous expression of cacaoidin, the first member of the lanthidin family of RiPPs. *Antibiotics*. **10**:403. doi: 10.3390/antibiotics10040403.

240. **Román-Hurtado, F., Sánchez-Hidalgo, M., Martín, J., Ortiz-López, F. J., Carretero-Molina, D., Reyes, F. and Genilloud, O.** (2021b). One pathway, two cyclic Non-Ribosomal Pentapeptides: heterologous expression of BE-18257 antibiotics and pentaminomycins from *Streptomyces cacaoi* CA-170360. *Microorganisms*. **9**:135. doi: 10.3390/microorganisms9010135.

241. **Roth, P., Reijneveld, J., Gorlia, T., Dhermain, F., De Vos, F., Vanlancker, M., O'Callaghan, C., Le Rhun, E., van den Bent, M., Mason, W. and Weller, M.** (2019). P14.124 EORTC 1709/CCTG CE.8: A phase III trial of marizomib in combination with standard temozolomide-based radiochemotherapy versus standard temozolomide-based radiochemotherapy alone in patients with newly diagnosed glioblastoma. *Neuro. Oncol.* **21**:iii98. doi: 10.1093/neuonc/noz126.359.

242. **Rutledge, P. J. and Challis, G. L.** (2017). Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol*. **13**:509-523. doi: 10.1038/nrmicro3496.

243. **Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. and Arnheim, N.** (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*. **230**:1350-1354. doi: 10.1126/science.2999980.

244. **Saitou, N. and Nei, M.** (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. 4:406–425, doi: 10.1093/oxfordjournals.molbev.a040454.

245. **Salcedo R. G., Olano, C., Gómez, C., Fernández, R., Braña, A. F., Méndez, C., de la Calle, F. and Salas, J. A.** (2016). Characterization and engineering of the biosynthesis gene cluster for antitumor macrolides PM100117 and PM100118 from a marine actinobacteria: generation of a novel improved derivative. *Microb. Cell Fact*. **15**:44. doi: 10.1186/s12934-016-0443-5.

246. **Salem, S. M., Weidenbach, S. and Rohr, J.** (2017). Two cooperative glycosyltransferases are responsible for the sugar diversity of saquayamycins isolated from *Streptomyces* sp.KY 40-1. *ACS Chem. Biol*. **12**:2529-2534. doi: 10.1021/acschembio.7b00453.

247. **Sambrook, J. and Russell, D. W.** (2001). Molecular cloning. A laboratory manual. New York.

248. **Sánchez-Hidalgo, M., Martín, J. and Genilloud, O.** (2020). Identification and heterologous expression of the biosynthetic gene cluster encoding the lasso peptide humidimycin, a caspofungin activity potentiator. *Antibiotics.* **9**:67. doi: 10.3390/antibiotics9020067.

249. **Sánchez-Hidalgo, M., Pascual, J., de la Cruz, M., Martín, J., Kath, G. S., Sigmund, J. M., Masurekar, P., Vicente, F., Genilloud, O. and Bills, G. F.** (2012). Prescreening bacterial colonies for bioactive molecules with Janus plates, a SBS standard double-faced microbial culturing system. *Antonie van Leeuwenhoek*. **102**:361-374. doi: 10.1007/s10482-012-9746-7.

250. **Sanger, F., Nicklen, S. and Coulson, A. R.** (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. SCI. USA*. **74**:5463-5467. doi: 10.1073/pnas.74.12.5463.

251. **Sardar, D., Lin, Z. and Schmidt, E. W.** (2015). Modularity of RiPP enzymes enables designed synthesis of decorated peptides. *Chem. Biol*. **22**:907-916. doi: 10.1016/j.chembiol.2015.06.014.

252. **Schmid, J., Heider, D., Wendel, N. J., Sperl, N. and Sieber, V.** (2016). Bacterial glycosyltransferases: challenges and opportunities of a highly diverse enzyme class toward tailoring natural products. *Front. Microbiol*. **7**:182. doi: 10.3389/fmicb.2016.00182.

253. **Schmidt, M., Toplak, A., Quaedflieg, P., van Maarseveen, J. H. and Nuijens, T.** (2017). Enzyme-catalyzed peptide cyclization. *Drug. Discov. Today Technol*. **26**:11-16. doi: 10.1016/j.ddtec.2017.11.007.

254. S**chmoock, G., Pfennig, F., Jewiarz, J., Laubinger, W., Schauwecker, F. and Keller, U.** (2005). Functional cross-talk between fatty acid synthesis and nonribosomal peptide synthesis in quinoxaline antibiotic-producing streptomycetes. *J. Biol.Chem*. **280**:4339-4349. doi: 10.1074/jbc.M411026200.

255. **Schneider, T. L., Shen, B. and Walsh, C. T.** (2003). Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. *Biochemistry*. **42**:9722-9730. doi: 10.1021/bi034792w.

256. **Schuman, B., Evans, S. V. and Fyles, T. M.** (2013). Geometrical tributes of retaining glycosyltransferase enzymes favor an orthogonal mechanism. *PLOS ONE*. **8**:e71077. doi:10.1371/journal.pone.0071077.

257. **Schwarzer, D., Finking, R. and Marahiel, M. A.** (2003). Nonribosomal peptides: from genes to products. *Nat. Prod. Rep*. **20**:275-287. doi: 10.1039/B111145K.

258. **Seipke, R. F.** (2015). Strain-level diversity of secondary metabolism in *Streptomyces albus. PLOS ONE*. **1**:e0116457. doi: 10.1371/journal.pone.0116457.

259. **Shang, Z., Winter, J. M., Kauffman, C. A., Yang, I. and Fenical, W.** (2019). Salinipeptins: integrated genomic and chemical approaches reveal unusual D-amino acid-containing ribosomally synthesized and post-translationally modified peptides (RiPPs) from a great salt lake *Streptomyces* sp. *ACS Chem. Biol*. **14**:415-425. doi: 10.1021/acschembio.8b01058.

260. **Shao, Z. and Zhao, H.** (2012). DNA assembler: a synthetic biology tool for characterizing and engineering natural product gene clusters. *Methods Enzymol*. **517**:203–224. doi:10.1016/B9780-12-404634-4.00010-3.

261. **Shao, Z., Rao, G., Li, C., Abil, Z., Luo, Y. and Zhao, H.** (2013). Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. *ACS Synth. Biol*. **2**:662-669. doi: 10.1021/sb400058n.

262. **Sheng, W., Xu, B., Chen, S., Li, Y., Liu, B. and Wang, H.** (2020). Substrate tolerance of the biosynthetic enzymes of glycosylated lanthipeptide NAI-112. *Org. Biomol. Chem*. **18**:6095. doi: 10.1039/d0ob01215g.

263. **Shima, J., Hesketh, A., Okamoto, S., Kawamoto, S. and Ochi, K.** (1996). Induction of actinohodin production by *rpsL* (encoding ribosomal protein S12) mutations that confer streptomycin resistance in *Streptomyces lividans* and *Streptomyces coelicolor* A3(2*). J. Bacteriol.* **178**:7276-7284. doi: 10.1128/jb.178.24.7276-7284.1996.

264. **Silver, L. L.** (2011). Challenges of antibacterial discovery. *Clin. Microbiol. Rev*. **24**:71-109. doi: 10.1128/CMR.00030-10.

265. **Singer, A. C., Shaw, H., Rhodes, V. and Hart, A.** (2016). Review of antimicrobial resistance in the environment and its relevance to environmental regulators. *Front. Microbiol*. **7**:1728. doi: 10.3389/fmicb.2016.01728.

266. **Sit, C. S., Yoganathan, S. and Vederas, J. C.** (2011). Biosynthesis of aminovinyl-cysteine-containing peptides and its application in the production of potential drug candidates. *Acc. Chem. Res*. **44**:261-268. doi: 10.1021/ar1001395.

267. **Skinnider, M. A., Johnston, C. W., Gunabalasingam, M., Merwin, N. J., Kieliszek, A. M., MacLellan, R. J., Li, H., Ranieri, M. R. M., Webster, A. L. H., Cao, M. P. T., Pfeifle, A., Spencer, N., To, Q. H., Wallace, D. P., Dejong, C. A. and Magarvey, N. A.** (2020). Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nat. Commun*. **11**:6058. doi: 10.1038/s41467-020-19986-1.

268. **Slatko, B. E., Gardner, A. F. and Ausubel, F. M.** (2018) Overview of next-generation aequencing technologies. *Curr. Protoc. Mol. Biol*. **122**:e59. doi: 10.1002/cpmb.59.

269. **Son, S., Hong, Y. S., Jang, M., Heo, K. T., Lee, B., Jang, J. P., Kim, J. W., Ryoo, I. J., Kim, W. G., Ko, S. K., Kim, B. Y., Jang, J. H. and Ahn, J. S.** (2017). Genomics-driven discovery of chlorinated cyclic hexapeptides ulleungmycins A and B from a *Streptomyces* species. *J. Nat. Prod*. **80**:3025–3031. doi: 10.1021/acs.jnatprod.7b00660.

270. **Spellberg, B. and Gilbert, D. N**. (2014). The future of antibiotics and resistance: a tribute to a career of leadership by John Bartlett. *Clin. Infect. Dis*. **59**:71-75. doi: 10.1093/cid/ciu392.

271. **Spellberg, B., Srinivasan, A. and Chambers, H. F.** (2016). New societal approaches to empowering antibiotic stewardship. *JAMA*. **315**:1229-1230. doi: 10.1001/jama.2016.1346.

272. **Stachelhaus, T. and Walsh, C. T.** (2000). Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase. *Biochemistry*. **39**:5775-5787. doi: 10.1021/bi9929002.

273. **Stegmann, E., Albersmeier, A., Spohn, M., Gert, H., Weber, T., Wohlleben, W., Kalinowski, J. and Rückert, C.** (2014). Complete genome sequence of the actinobacterium *Amycolatopsis japonica* MG417-CF17^{T} (=DSM 44213^{T}) producing (S, S)-N, N'-ethylenediaminedisuccinic acid. J. *Biotechnol*. **189**:46-47. doi: 10.1016/j.jbiotec.2014.08.034.

274. **Stein, S. E.** (1999). An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Am. Soc. Mass. Spectrom*. **10**:770–781. doi: 10.1016/S1044-0305(99)00047-1.

275. **Steller, S., Vollenbroich, D., Leenders, F., Stein, T., Conrad, B., Hofemeister, J., Jacques, P., Thonart, P. and Vater, J.** (1999). Structural and functional organization of the fengycin synthetase multienzyme system from *Bacillus subtilis* B213 and A1/3. *Chem. Biol*. **6**:31-41. doi: 10.1016/S1074-5521(99)80018-0.

276. **Stepper, J., Shastri, S., Loo, T. S., Preston, J. C., Novak, P., Man, P., Moore, C. H., Havlícek, V., Patchett, M. L. and Norris, G. E.** (2011). Cysteine S-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Lett*. **585**:645-650. doi: 10.1016/j.febslet.2011.01.023.

277. **Süssmuth, R. D. and Mainz, A.** (2017). Nonribosomal peptide-synthesis-principles and prospects. *Angew. Chem. Int*. **56**:3770-3821. doi: 10.1002/anie.201609079.

278. **Takada, K., Ninomiya, A., Naruse, M., Sun, Y., Miyazaki, M., Nogi, Y., Okada, S. and Matsunaga, S.** (2013). Surugamides A-E, cyclic octapeptides with four D-amino acid residues, from a marine *Streptomyces* sp.: LC-MS-aided inspection of partial hydrolysates for the dintrinction of D- and L-amino acid residues in the sequence. *J. Org. Chem*. **78**:6746-6750. doi: 10.1021/jo400708u.

279. **Takahashi, Y. and Nakashima, T.** (2018). Actinomycetes, an inexhaustible source of naturally occurring antibiotics. *Antibiotics*. **7**:45. doi: 10.3390/antibiotics7020045.

280. **Takashi, Y., Kosako, H., Sawatsubashi, S., Kinoshita, Y., Ito, N., Tsoumpra, M. K., Nangaku, M., Abe, M., Matsuhisa, M., Kato, S., Matsumoto, T. and Fukumoto, S.** (2019). Activation of unliganded FGF receptor by extracellular phosphate potentiates proteolytic protection of FGF23 by its O-glycosylation. *Proc. Natl. Acad. Sci. USA*. **116**:11418-11427. doi: 10.1073/pnas.1815166116.

281. **Tamura, K.** (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol. Biol. Evol*. **9**:678–687. doi: 10.1093/oxfordjournals.molbev.a040752.

282. **Tanaka, Y., Kasahara, K., Hirose, Y., Murakami, K., Kugimiya. And Ochi, K.** (2013). Activation and products of the cryptic secondary metabolite biosynthetic gene clusters by rifampin resistance (*rpoB*) mutations in actinomycetes. *J. Bacteriol*. **195**:2959-2970. doi: 10.1128/JB.00147-13.

283. **Thankachan, D., Fazal, A., Francis, D., Song, L., Webb, M. E. and Seipke, R. F.** (2019). A trans-acting cyclase offloading strategy for nonribosomal peptide synthetases. *ACS Chem. Biol*. **14**:845-849. doi: 10.1021/acschembio.9b00095.

284. **Tintore, M., Vidal-Jordana, A. and Sastre-Garriga, J.** (2019). Treatment of multiple sclerosis-success from bench to bedside. *Nat. Rev. Neurol*. **15**:53-58. doi: 10.1038/s41582-018-0082-z.

285. **Tiwari, K. and Gupta, R. K.** (2012). Rare actinomycetes: a potential storehouse for novel antibiotics. *Crit. Rev. Biotechnol*. **32**:108-132. doi: 10.3109/07388551.2011.562482.

286. **Tiwari, K. and Gupta, R. K.** (2014). Bioactive metabolites from rare actinomycetes. *Stud. Nat. Prod. Chem*. **41**:419-512. doi: 10.1016/B978-0-444-63294-4.00014-0.

287. **Tocchetti, A., Iorio, M., Hamid, Z., Armirotti, A., Reggiani, A. and Donadio, S.** (2021). Understanding the mechanism of action of NAI-112, a lanthipeptide with potent antinociceptive activity. *Molecules*. **26**:6764. doi: 10.3390/molecules26226764.

288. **Tong, Y., Robertsen, H. L., Blin, K., Weber, T. and Lee, S. Y.** (2018) CRISPR-Cas9 Toolkit for Actinomycete Genome Editing. *Methods Mol. Biol*. **1671**:163-184. doi: 10.1007/978-1-4939-7295-1_11

289. **Trabelsi, I., Oves-Costales, D., Manteca, A., Genilloud, O., Altalhi, A. and Nour, M.** (2016). Antimicrobial activities of some actinomycetes isolated from different rhizospheric soils in Tunisia. *Curr. Microbiol*. **73**:220-227. doi: 10.1007/s00284-016-1053-5

290. **Tyurin, A. P., Alferova, V. A., Paramonov, A. S., Shuvalov, M. V., Kudryakova, G. K., Rogozhin, E. A., Zherebker, A. Y., Brylev, V. A., Chistov, A. A., Baranova, A. A., Biryukov, M. V., Ivanov, I. A., Prokhorenko, I. A., Grammatikova, N. E., Kravchenko, T. V., Isakova, E. B., Mirchink, E. P., Gladkikh, E. G., Svirshchevskaya, E. V., Mardanov, A. V., Beletsky, A. V., Kocharovskaya, M. V., Kulyaeva, V. V., Shashkov, A. S., Tsvetkov, D. E., Nifantiev, N. E., Apt, A. S., Majorov, K. B., Efimova, S. S., Ravin, N. V., Nikolaev, E. N., Ostroumova, O. S., Katrukha, G. S., Lapchinskaya, O. A., Dontsova, O. A., Terekhov, S. S., Osterman, I. A., Shenkarev, Z. O. and Korshun, V. A.** (2021). Gausemycins A, B: cyclic lipoglycopeptides from *Streptomyces* sp. *Angew. Chem. Int*. **60**:18694-18703. doi: 10.1002/anie.202104528

291. **van der Donk, W. A. and Nair, S. K.** (2014). Structure and mechanism of lanthipeptide biosynthetic enzymes. *Curr. Opin. Struct. Biol*. **29**:58-66. doi: 10.1016/j.sbi.2014.09.006

292. **van der Meer, J. R., Polman, J., Beerthuyzen, M. M., Siezen, R. J., Kuipers, O. P. and De Vos, W. M.** (1993). Characterization of the *Lactococcus lactis* nisin A operon genes *nisP*, encoding a subtilisin-like serine protease involved in precursor processing, and *nisR*, encoding a regulatory protein involved in nisin biosynthesis. *J. Bacteriol*. **175**:9. doi: 10.1128/jb.175.9.2578-2588.1993.

293. **van der Meij, A., Worsley, S. F., Hutchings, M. I. and van Wezel, G. P.** (2017). Chemical ecology of antibiotic production by actinomycetes. *FEMS Microbiol. Rev*. **41**: 392-416. doi: 10.1093/femsre/fux005.

294. **van Heel, A. J., de Jong, A., Song, C., Viel, J. H., Kok, J. and Kuipers, O. P.** (2018). BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res*. **46**:278-281. doi: 10.1093/nar/gky383.

295. **van Santen, J. A., Poynton, E. F., Iskakova, D., McMann, E., Alsup, T. A., Clark, T. N., Fergusson, C. H., Fewer, D. P., Hughes, A. H., McCadden, C. A., Parra, J., Soldatou, S., Rudolf, J. D., Janssen, E. M. L., Duncan, K. R. and Linington, R. G.** (2020). The Natural Products Atlas 2.0: a database of microbioally-derived natural products. *Nucleic Acids Res*. **50**:D1317-D1323. doi: 10.1093/nar/gkab941.

296. **Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G. and Marth, J.** (1999). Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

297. **Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. G. and van Sinderen, D.** (2007). Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol. Mol. Biol. Rev*. **71**:495-548. doi: 10.1128/MMBR.00005-07.

298. **Vicente, C. M., Thibessard, A., Lorenzi, J. N., Benhadj, M., Hôtel, L., Gacemi-Kirane, D., Lespinet, O., Leblond, P. and Aigle, B.** (2018). Comparative genomics among closely related *Streptomyces* strains revealed specialized metabolite biosynthetic gene cluster diversity. *Antibiotics*. **7**:86. doi: 10.3390/antibiotics7040086.

299. **Völler, G. H., Krawczyk, B., Ensle, P. and Süssmuth, R. D.** (2013). Involvement and unusual substrate specificity of a prolyl oligopeptidase in class III lanthipeptide maturation. *J. Am. Chem. Soc*. **135**:7426-7429. doi: 10.1021/ja402296m.

300. **Völler, G. H., Krawczyk, J. M., Pesic, A., Krawczyk, B., Nachtigall, J. and Süssmuth, R. D.** (2012). Characterization of new class III lantibiotics - Erythreapeptin, avermipeptin and griseopeptin from *Saccharopolyspora erythraea, Streptomyces avermitilis* and *Streptomyces griseus* demonstrates stepwise N-terminal leader processing. *Chembiochem*. **13**:1174-1183. doi: 10.1002/cbic.201200118.

301. **von Döhren, H. and Gräfe, U.** (1997). General aspects of the secondary metabolism. In Biotechnology: products of secondary metabolism. Eds. Rehm, H. J., Reed, J. VHC Weinhein. 1-55. doi: 10.1002/9783527620890.ch1.

302. **Waghu, F. H., Barai, R. S., Gurung, P. and Idicula-Thomas, S.** (2016). CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Res*. **4**:D1094-D1097. doi: 10.1093/nar/gkv1051.

303. **Waksman, S. A. and Henrici, A. T.** (1943). The nomenclature and classification of the actinomycetes. *J. Bacteriol*. **46**:337-341. doi: 10.1128/jb.46.4.337-341.1943.

304. **Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K. and Earl, A. M.** (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE*. **9**:e112963. doi: 10.1371/journal.pone.0112963.

305. **Waltenberger, B., Mocan, A., Smejkal, K., Heiss, E. H. and Atanasov, A. G.** (2016). Natural products to counteract the epidemic of cardiovascular and metabolic disorders. *Molecules*. **21**:807. doi: 10.3390/molecules21060807.

306. **Wang, F., Wei, W., Zhao, J., Mo, T., Wang, X., Huang, X., Ma, S., Wang, S., Deng, Z., Ding, W., Liang, Y. and Zhang, Q.** (2021). Genome mining and biosynthesis study of a type B linaridin reveals a highly versatile α-N-methyltransferase. *CCS Chemistry*. **3**:1049-1057. doi: 10.31635/ccschem.020.202000247.

307. **Wang, H. and van der Donk, W. A.** (2012). Biosynthesis of the class III lanthipeptide catenulipeptin. *ACS Chem. Biol*. **7**:1529-1535. doi: 10.1021/cb3002446.

308. **Wang, H., Li, Z., Jia, R., Yin, J., Li, A., Xia, L., Yin, Y., Müller, R., Fu, J., Stewart, A. F. and Zhang, Y.** (2018a). ExoCET: exonuclease *in vitro* assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. *Nucleic Acids Res.* **46**:2697. doi:10.1093/nar/gkx1296.

309. **Wang, H., Oman, T. J., Zhang, R., García de Gonzalo, C. V., Zhang, Q. and van der Donk, W. A.** (2014a). The glycosyltransferase involved in thurandacin biosynthesis catalyzes both O- and S-glycosylation. *J. Am. Chem. Soc*. **136**:84-87. doi: 10.1021/ja411159k.

310. **Wang, H., Yang, L., Wu, K. and Li, G.** (2014b). Rational selection and engineering of exogenous principle sigma factor (σHrdB) to increase teicoplanin production in an industrial strain of *Actinoplanes teichomyceticus. Microb. Cell Fact*. **13**:10. doi: 10.1186/1475-2859-13-10.

311. **Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W. T., Crüsemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderón, M., Kersten, R. D., Pace, L. A., Quinn, R. A., Duncan, K. R., Hsu, C. C., Floros, D. J., Gavilan, R. G., Kleigrewe, K., Northen, T., Dutton, R. J., Parrot, D., Carlson, E. E., Aigle, B., Michelsen, C. F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B. T., Gerwick, L., Liaw, C. C., Yang, Y. L., Humpf, H. U., Maansson, M., Keyzers, R. A., Sims, A. C., Johnson, A. R., Sidebottom, A. M., Sedio, B. E., Klitgaard, A., Larson, C., Boya P, C. A., Torres-Mendoza, D., Gonzalez, D. J., Silva, D. B., Marques, L. M., Demarque, D. P., Pociute, E., O'Neill, E. C., Briand, E., Helfrich, E. J. N., Granatosky, E. A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J. J., Zeng, Y., Vorholt, J. A., Kurita, K. L., Charusanti, P., McPhail, K. L., Nielsen, K. F., Vuong, L., Elfeki, M., Traxler, M. F., Engene, N., Koyama, N., Vining, O. B., Baric, R., Silva, R. R., Mascuch, S. J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P. G., Dai, J., Neupane, R., Gurr, J., Rodríguez, A. M. C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B. M., Almaliti, J., Allard, P. M., Phapale, P., Nothias, L. F., Alexandrov, T., Litaudon, M., Wolfender, J. L., Kyle, J. E., Metz, T. O., Peryea, T., Nguyen, D. T., VanLeer, D., Shinn, P., Jadhav, A., Müller, R., Waters, K. M., Shi, W., Liu, X., Zhang, L., Knight, R., Jensen, P. R., Palsson, B. O., Pogliano, K., Linington, R. G., Gutiérrez, M., Lopes, N. P., Gerwick, W. H., Moore, B. S., Dorrestein, P. C. and Bandeira, N.** (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Netwoking. *Nat. Biotechnol*. **34**:828-837. doi: 10.1038/nbt.3597.

312. **Wang, C., Huang, D. and Liang, S.** (2018b). Identification and metabolomic analysis of chemical elicitors for tacrolimus accumulation in *Streptomyces tsukubaensis. Appl. Microbiol. Biotechnol.* **102**:7541–7553. doi: 10.1007/s00253-018-9177-0.

313. **Weber, T., Welzel, K., Pelzer, S., Vente, A. and Wohlleben, W.** (2003). Exploiting the genetic potential of polyketide producing streptomycetes*. J. Biotechnol*. **106**:221-232. doi: 10.1016/j.jbiotec.2003.08.004.

314. **Weinstein, M. J., Luedemann, G. M., Oden, E. M., Wagman, G. H., Rosselet, J. P., Marquez, J. A., Coniglio, C. T., Charney, W., Herzog, H. L. and Black, J.** (1963). Gentamycin, a new antibiotic complex from *Micromonospora*. *J. Med. Chem*. **6**:463-464. doi: 10.1021/jm00340a034.

315. **Wenski, S. L., Thiengmag, S. and Helfrich, E. J. N.** (2022). Complex peptide natural products: biosynthetic principles, challenges and opportunities for pathway engineering. *Synth. Syst. Biotechnol*. **7**:631-647. doi: 10.1016/j.synbio.2022.01.007.

316. **Widdick, D. A., Dodd, H. M., Barraille, P., White, J., Stein, T. H., Chater, K. F., Gasson, M. J. and Bibb, M. J.** (2003). Cloning and engineering of the cinnamycin biosynthetic cluster from *Streptomyces cinnamoneus cinnamoneus* DSM 40005. *Proc. Natl. Acad. Sci. USA.* **100**:4316-4321. doi: 10.1073/pnas.0230516100.

317. **Willey, J. M. and van der Donk, W. A.** (2007). Lantibiotics: peptides of diverse structure and function. *Annu. Rev. Microbiol*. **61**-477-501. doi: 10.1146/annurev.micro.61.080706.093501.

318. **Wolf, S., Schmidt, S., Müller-Hannemann, M. and Neumann, S.** (2010). In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics*. **11**:148. doi: 10.1186/1471-2105-11-148, doi: 10.1186/1471-2105-11-148.

319. **Wood, H. E., Devine, K. M. and McConnell, D. J.** (1990). Characterization of a repressor gene (*xre*) and a temperature-sensitive allele from the *Bacillus subtilis* prophage, PBSX. *Gene.* **96**:83–88, doi:10.1016/0378-1119(90)90344-q.

320. **Woodruff, H. B.** (1980). Natural products from microorganisms. *Science*. **208**:1225-1229. doi: 10.1126/science.7375932.

321. **Xu, M. and Wright, G. D.** (2019). Heterologous expression-facilitated natural products' discovery in Actinomycetes. *J. Ind.Microbiol. Biotechnol*. **46**:415–431. doi: 10.1007/s10295-018-2097-2.

322. **Xu, M., Zhang, F., Cheng, Z., Bashiri, G., Wang, J., Hong, J., Wang, Y., Xu, L., Chen, X., Huang, S. X., Lin, S., Deng, Z. and Tao, M.** (2020). Functional genome mining reveals a novel class V lanthipeptide containing a D-amino acid introduced by an $F_{420}H_2$-dependent reductase. *Angew. Chem. Int.* **59**:18029–18035, doi:10.1002/anie.202008035.

323. **Xu, M., Wang, Y., Zhao, Z., Gao, G., Huang, S. X., Kang, Q., He, X., Lin, S., Pang, X., Deng, Z. and Tao, M.** (2016). Functional genome mining for metabolites encoded by large gene clusters through heterologous expression of a whole-genome bacterial artificial chromosome library in *Streptomyces* spp. *Appl. Environ. Microbiol*. **82**:57955805. doi:10.1128/AEM.01383-16.

324. **Yagüe, P., López-García, M. T., Rioseras, B., Sánchez, J. and Manteca, A.** (2013). Pre-sporulation stages of *Streptomyces* differentiation: state-of-the-art and future perspectives. *FEMS Microbiol. Lett*. **342**:79-88. doi: 10.1111/1574-6968.12128.

325. **Yamanaka, K., Reynolds, K. A., Kersten, R. D., Ryan, K. S., Gonzalez, D. J., Nizet, V., Dorrestein, P. C. and Moore, B. S.** (2014). Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. USA*. **115**:1957-1962. doi: 10.1073/pnas.1319584111.

326. **Yang, X. and van der Donk, W. A.** (2015). Post-translational introduction of D-alanine into ribosomally synthesized peptides by the dehydroalanine reductase NpnJ. *J. Am. Chem. Soc.* **137**:12426–12429. doi:10.1021/jacs.5b05207.

327. **Yates, K. R., Welsh, J., Udegbunam, N. O., Greenman, J., Maraveyas, A. and Madden, L. A.** (2012). Duramycin exhibits antiproliferative properties and induces apoptosis in tumour cells. *Blood. Coagul. Fibrinol*. **23**:396-401. doi: 10.1097/MBC.0b013e3283538875.

328. **Yoo, Y. J., Hwang, J-Y., Shin, H-L., Cui, H., Lee, J. and Yoon, Y. J.** (2015). Characterization of negative regulatory genes for the biosynthesis of rapamycin in *Streptomyces rapamycinicus* and its application for improved production. *J. Ind. Microbiol. Biotechnol*. **42**:125–135. doi: 10.1007/s10295-014-1546-9.

329. **Zhang, J. H., Chung, T. D. Y. and Oldenburg, K. R.** (1999). A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screen*. **4**:67–73. doi: 10.1177/108705719900400206.

330. **Zhang, Q. W., Lin, L. G. and Ye, W. C.** (2018). Techniques for extraction and isolation of natural products: a comprehensive review. *Chin. Med*. **13**:20. doi: 10.1186/s13020-018-0177-x.

331. **Zhang, Q., Yu, Y., Velásquez, J. E. and van der Donk, W. A.** (2012). Evolution of lanthipeptide synthetases. *Proc. Natl. Acad. Sci. USA*. **109**:18361-18366. doi: 10.1073/pnas.1210393109.

332. **Zhang, Y., Liu, S., Liu, H., Liu, X. and Che, Y.** (2009). Cycloaspeptides F and G, cyclic pentapeptides from a *Cordyceps*-colonizing isolate of *Isaria farinosa*. *J. Nat. Prod*. **72**:1364-1367. doi: 10.1021/np900205m.

333. **Zhao, Z., Shi, T., Xu, M., Brock, N. L., Zhao, Y. L., Wang, Y., Deng, Z., Pang, X. and Tao, M.** (2016). Hybrubins: bipyrrole tetramic acids obtained by crosstalk between a truncated undecylprodigiosin pathway and heterologous tetramic acid biosynthetic genes. *Org. Lett*. **18**:572–575. doi: 10.1021/acs.orglett.5b03609.

334. **Zhou, Y., Lin, X., Xu, C., Shen, Y., Wang, S. P., Liao, H., Li, L., Deng, H. and Lin, H. W**. (2019). Investigation of penicilin binding protein (PBP)-like peptide cyclase and hydrolase in surugamide non-ribosomal peptide biosynthesis. *Cell Chem. Biol*. **26**:737–744. doi: 10.1016/j.chembiol.2019.02.010.

335. **Zhou, Y., Murphy, A. C. Samborskyy, M., Prediger, P., Dias, L. C. and Leadlay, P. F.** (2015). Iterative mechanism of macrodiolide formation in the anticancer compound conglobatin. *Chem. Biol*. **22**:745-754. doi: 10.1016/j.chembiol.2015.05.010.

336. **Zhu, S., Duan, Y. and Huang, Y.** (2019). The application of ribosome engineering to natural product discovery and yield improvement in *Streptomyces. Antibiotics*. **8**:133. doi: 10.3390/antibiotics8030133.

337. **Zhuo, Y., Zhang, W., Chen, D., Gao, H., Tao, J., Liu, M., Gou, Z., Zhou, X., Ye, B. C., Zhang, S. and Zhang, L. X.** (2010). Reverse biological engineering of *hrdB* to enhance the

production of avermectins in an industrial strain of *Streptomyces avermitilis. Proc. Nat. Acad. Sci. USA.* **107**:11250–11254 doi: 10.1073/pnas.1006085107.

338. **Ziemert, M., Alanjary, M. and Weber, T.** (2016). The evolution of genome mining in microbes-a review. *Nat. Prod. Rep*. **33**:988-1005. doi: 10.1039/C6NP00025H.

339. **Zimmermann, L., Stephens, A., Nam, S-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N. and Alva, V.** (2017). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol*. **2836**:30587-30589. doi: 10.1016/j.jmb.2017.12.007.

340. **Zong, G., Fu, J., Zhang, P., Zhang, W., Xu, Y., Cao, G. and Zhang, R.** (2021). Use of elicitors to enhance or activate the antibiotic production in *Streptomyces*. *Crit. Rev. Biotechnol*. **27**:1-24. doi: 10.1080/07388551.2021.1987856.