Genome mining for drug discovery: Activation of silent biosynthetic gene clusters

Project acronym: **GenoDrug**
Project no: EIB.10.023

European Research Area
Industrial Biotechnology

GenoDrug Final Report

Structure:

1. Background
2. Project aim and approach
3. Project partners
4. Progress and cooperations
5. Achievement of milestones and objectives
6. Problems encountered; proposed changes

GenoDrug Final Report

Structure:

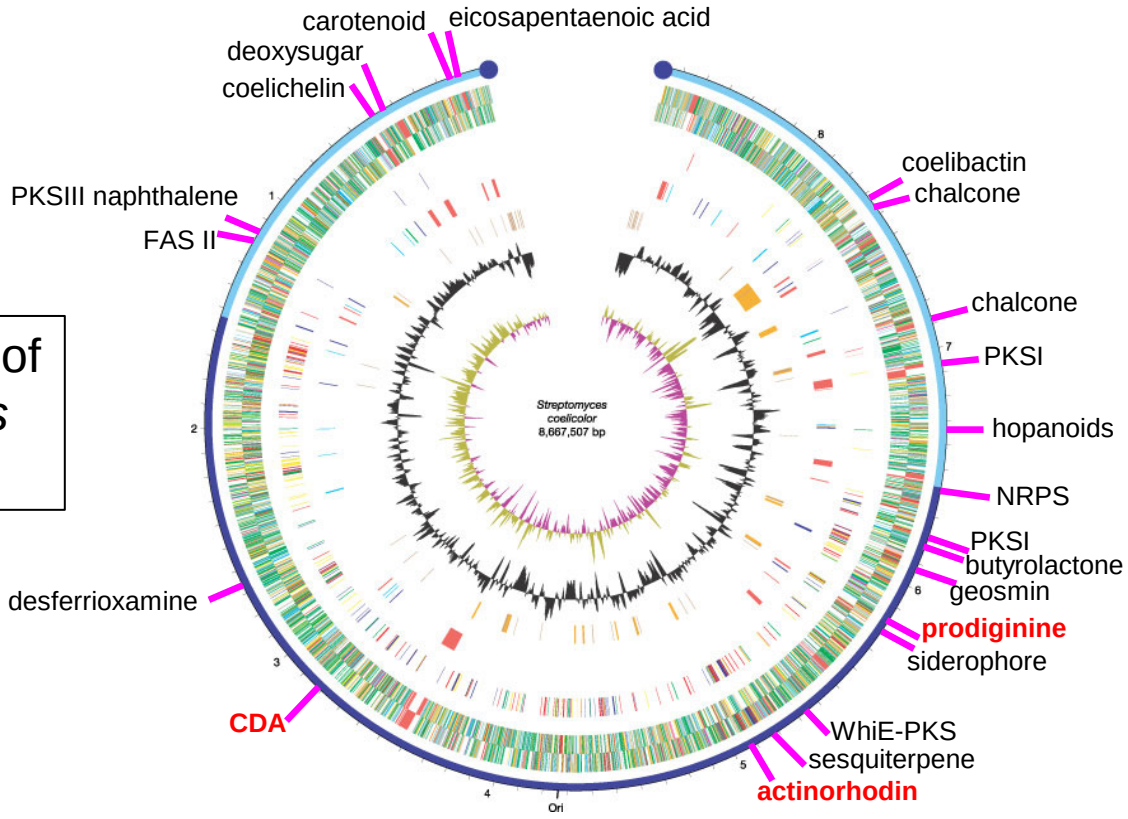
1. Background
2. Project aim and approach
3. Project partners
4. Progress and cooperations
5. Achievement of milestones and objectives
6. Problems encountered; proposed changes

GenoDrug Final Report: Background

***Streptomyces* strains contain numerous secondary metabolite biosynthetic gene clusters,**

much more than the number of products isolated from one strain

Genetic map of *Streptomyces coelicolor*



Modified according to ***Bentley et al., Nature 2002***

GenoDrug Final Report

Structure:

1. Background
2. Project aim and approach
3. Project partners
4. Progress and cooperations
5. Achievement of milestones and objectives
6. Problems encountered; proposed changes

GenoDrug Project: Aim

Aim: Exploit the unused potential

A new technology for drug discovery:
Activation of previously silent biosynthetic gene clusters of microbial genomes → Novel bioactive compounds

GenoDrug Project: Aim and Approach

Aim

A new technology for drug discovery:
Activation of previously silent biosynthetic gene clusters of microbial genomes → Novel bioactive compounds

General approach

1) Bioinformatic genome analysis of actinomycete strains ⇒ new gene clusters.

2) Development of strategies for the activation and expression of silent biosynthetic gene clusters using:

   a) global and pathway-specific regulators;

   b) introduction of artificial promoters;

   c) heterologous expression of gene clusters in genetically engineered host strains.

3) Production of new compounds from the engineered strains and testing for bioactivities, especially antibiotic and anticancer activities.

GenoDrug Final Report

Structure:

1. Background
2. Project aim and approach
3. Project partners
4. Progress and cooperations
5. Achievement of milestones and objectives
6. Problems encountered; proposed changes

GenoDrug Partners

Prof. Wohlleben Microbiology Tübingen University	 Germany	Bioinformatics Systems biology Genetic engineering
Prof. Zakrzewska-Czerwińska Biotechnology University of Wrocław	 Poland	Gene regulation studies Upregulation of expression
Prof. Heide Pharmaceutical Institute Tübingen University	 Germany	Gene cluster assembly Heterologous expression Combinatorial biosynthesis
Prof. Méndez Functional Biology University of Oviedo	 Spain	Precursor supply Genetic engineering Mutasynthesis
Dr. Ylihonko Galilaeus Oy (SME) Kaarina	 Finland	Strain optimization Fermentation Compound generation
Dr. Moris EntreChem SL (SME) Oviedo	 Spain	Structure elucidation Biological testing Toxicity testing

GenoDrug Final Report

Structure:

1. Background
2. Project aim and approach
3. Project partners
4. Progress and cooperations
5. Achievement of milestones and objectives
6. Problems encountered; proposed changes

Bioinformatic Genome Analysis (completed)

1. ***Streptomyces argillaceus*** ATCC 12956
2. ***Streptomyces albus*** J1074
3. ***Streptomyces collinus*** Tü 365
4. ***Amycolatopsis balhimycina*** DSM 5908
5. ***Catenulispora acidiphila*** DSM 44928
6. ***Amycolatopsis japonicum***
7. ***Streptomyces tendae*** Tü1028

Development of New Bioinformatic Tools: antiSMASH

Tilmann Weber

- **anti**biotics and **S**econdary **M**etabolites **A**nalysis **SH**ell
- Aim:
  - Integration of all available prediction methods for secondary metabolite biosynthesis genes into user friendly pipeline

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

K. Blin / T. Weber / W. Wohlleben

university of groningen

GBB

M. Medema / E. Takano / R. Breitling

UCSF University of California San Francisco

P. Cimermancic / M. Fischbach

http://antismash.secondarymetabolites.org/

Blin *et al*., Nucleic Acids Res., 2013

antiSMASH – Input Form

- [ ] DNA of Eukaryotic origin

Gene cluster types to search:

- [x] all
- [x] polyketides (type I)
- [x] polyketides (type II)
- [x] polyketides (type III)
- [x] nonribosomal peptides
- [x] terpenes
- [x] lantibiotics
- [x] bacteriocins
- [x] beta-lactams
- [x] aminoglycosides / aminocyclitols
- [x] aminocoumarins
- [x] siderophores
- [x] ectoines
- [x] butyrolactones
- [x] indoles
- [x] nucleosides
- [x] phosphoglycolipids
- [x] melanins
- [x] others

- [x] smCOG analysis for functional prediction and phylogenetic analysis of genes
- [x] Gene Cluster Blast Comparative Analysis
- [ ] Whole-genome BLAST results in EMBL output
- [x] Whole genome PFAM results in EMBL output

Senden

antiSMASH – Output Form

Select Gene Cluster:

Overview 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Identified secondary metabolite clusters

Cluster	Type	From	To
	The following clusters are from record unknown		
Cluster 1	Ectoine	397859	408248
Cluster 2	Other	1879080	1922856
Cluster 3	Nrps	2449130	2519199
Cluster 4	T1pks	2505679	2610944
Cluster 5	Other	2978921	3022823
Cluster 6	T4pks	3188097	3238424
Cluster 7	T1pks	3233337	3277536
Cluster 8	Lantipeptide	3305024	3328143
Cluster 9	Nrps	3387615	3459171
Cluster 10	Terpene	3925707	3946837
Cluster 11	Other	3948382	3992446
Cluster 12	Nrps	4069149	4120964
Cluster 13	Nrps	4111374	4163603
Cluster 14	Nrps	4379325	4441530
Cluster 15	Nrps	4530446	4587520
Cluster 16	T1pks	4713730	4760149
Cluster 17	Nrps-t1pks	4838508	4932058
Cluster 18	T1pks-t4pks	5121459	5172192
Cluster 19	Terpene	5278355	5299473
Cluster 20	Bacteriocin	5496314	5507129
Cluster 21	Terpene	5831519	5852448
Cluster 22	T1pks	6415623	6461442
Cluster 23	T3pks-nrps	6800521	6897743
Cluster 24	T1pks	6902710	6949003
Cluster 25	Amglyccycl	7571659	7618166
Cluster 26	Terpene	8232022	8254163
Cluster 27	Other	8286964	8327650

Blin *et al.*, Nucleic Acids Res., 2013

antiSMASH Results
PKS / NRPS Domain Organisation

PKS/NRPS domain annotation

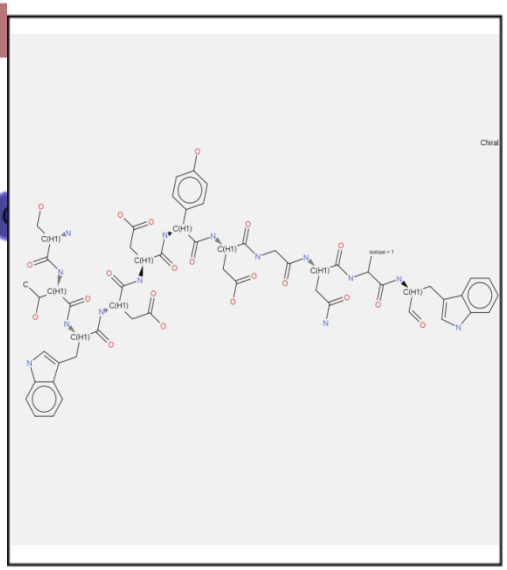
SCO3230 (nrps)

Domain AMP-binding3 (SCO3230)
Location: 2785-3184 AA
Predicted substrate: trp
-NRPSPredictor code: trp
-NRPSPredictor SVM: trp
-Minowa HMM: trp
NCBI BlastP on this domain

SCO3231 (glycopeptide nrps)

SCO3232 (nrps)

Predicted core structure



Monomers prediction: ser thr trp asp asp hpg asp gly asn nrp trp

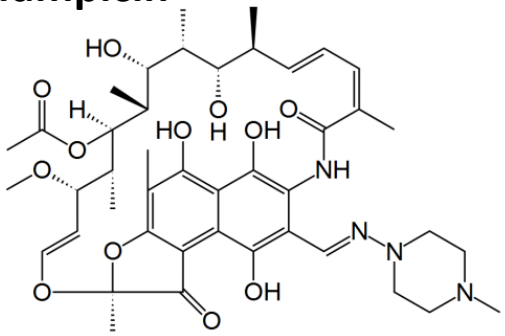
SCO3230:

NRPSPredictor code prediction, A1: ser
NRPSPredictor SVM prediction, A1: ser
Minowa prediction, A1: ser
Prediction details
NRPSPredictor code prediction, A2: thr
NRPSPredictor SVM prediction, A2: thr
Minowa prediction, A2: thr
Prediction details
NRPSPredictor code prediction, A3: trp
NRPSPredictor SVM prediction, A3: trp
Minowa prediction, A3: trp
Prediction details

Products of *Amycolatopsis* Strains

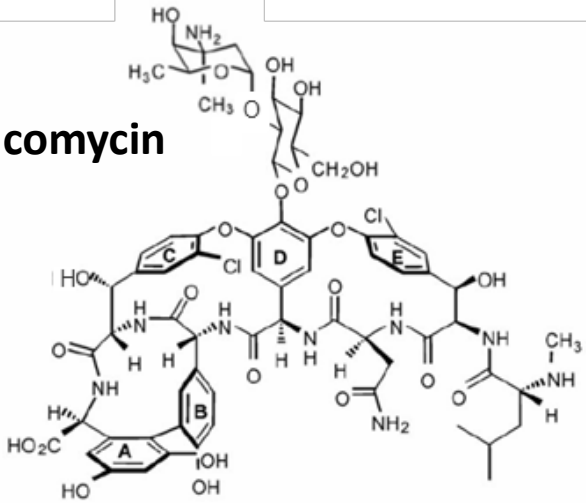
Amycolatopsis mediterranei

Rifampicin



Amycolatopsis orientalis

Vancomycin

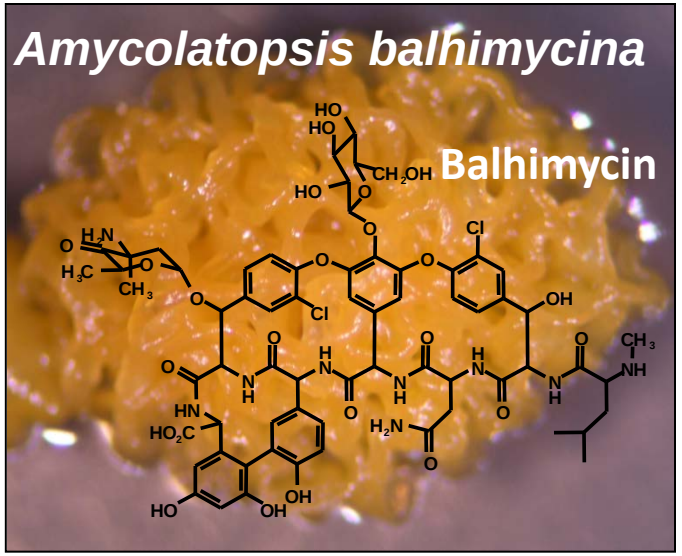


Amycolatopsis japonicum

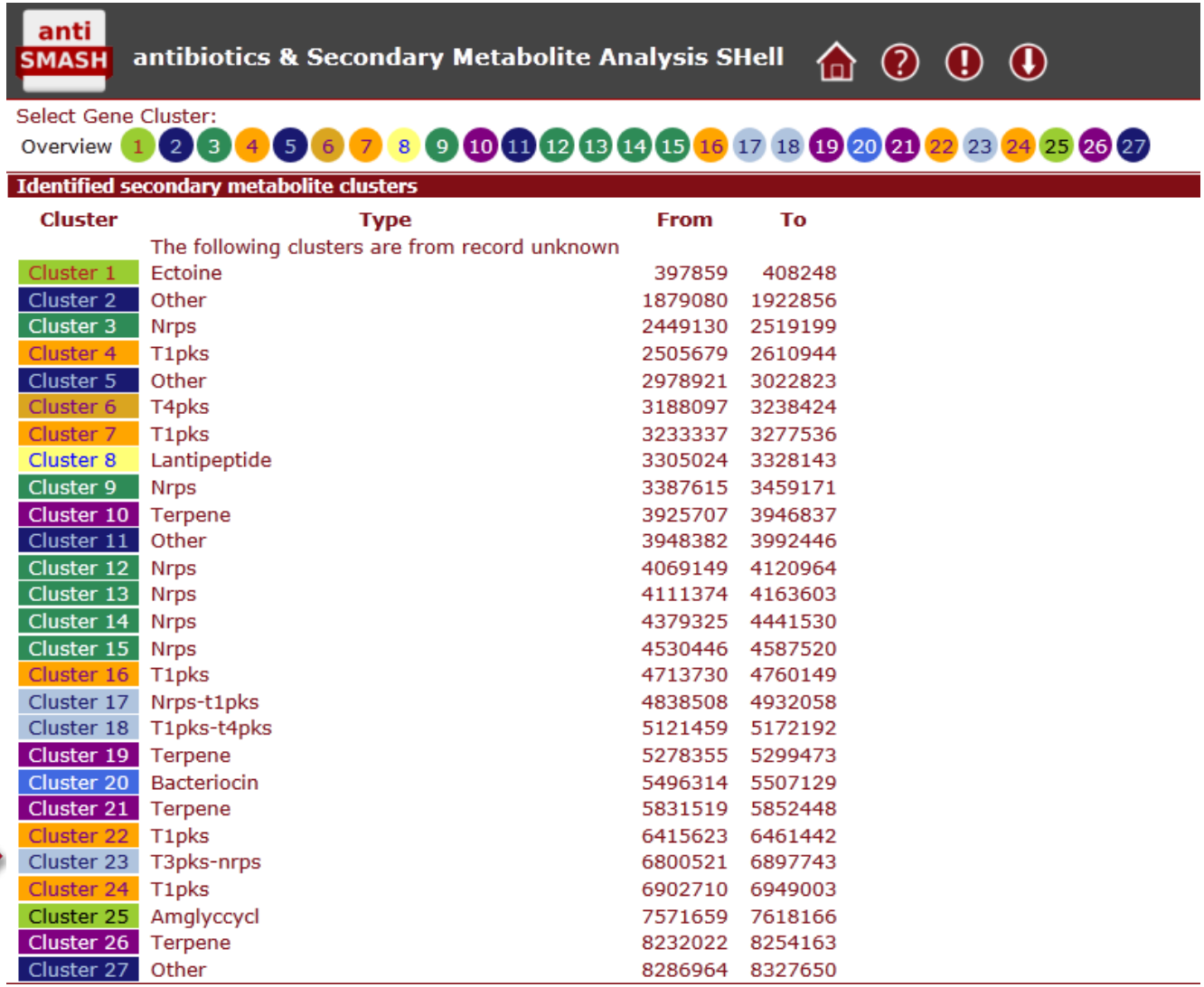
?

Amycolatopsis balhimycina

Balhimycin



Identification of a Glycopeptide Biosynthetic Gene Cluster in *A. japonicum*

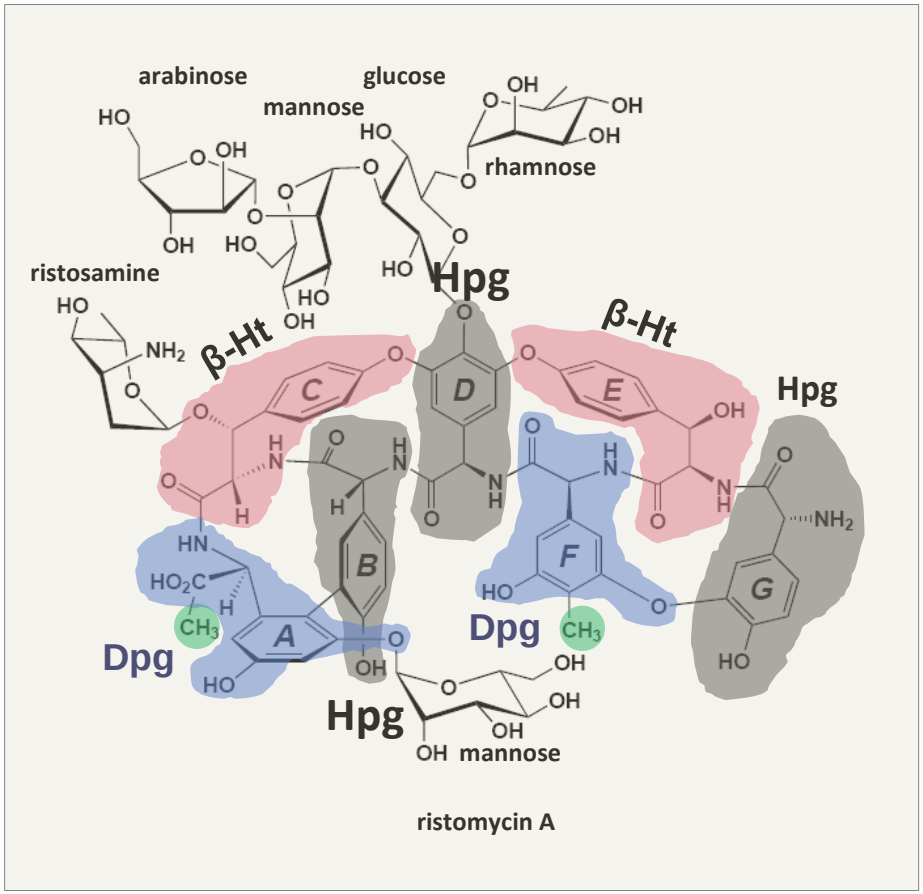


Identified secondary metabolite clusters

Cluster	Type	From	To
	The following clusters are from record unknown		
Cluster 1	Ectoine	397859	408248
Cluster 2	Other	1879080	1922856
Cluster 3	Nrps	2449130	2519199
Cluster 4	T1pks	2505679	2610944
Cluster 5	Other	2978921	3022823
Cluster 6	T4pks	3188097	3238424
Cluster 7	T1pks	3233337	3277536
Cluster 8	Lantipeptide	3305024	3328143
Cluster 9	Nrps	3387615	3459171
Cluster 10	Terpene	3925707	3946837
Cluster 11	Other	3948382	3992446
Cluster 12	Nrps	4069149	4120964
Cluster 13	Nrps	4111374	4163603
Cluster 14	Nrps	4379325	4441530
Cluster 15	Nrps	4530446	4587520
Cluster 16	T1pks	4713730	4760149
Cluster 17	Nrps-t1pks	4838508	4932058
Cluster 18	T1pks-t4pks	5121459	5172192
Cluster 19	Terpene	5278355	5299473
Cluster 20	Bacteriocin	5496314	5507129
Cluster 21	Terpene	5831519	5852448
Cluster 22	T1pks	6415623	6461442
Cluster 23	T3pks-nrps	6800521	6897743
Cluster 24	T1pks	6902710	6949003
Cluster 25	Amglyccycl	7571659	7618166
Cluster 26	Terpene	8232022	8254163
Cluster 27	Other	8286964	8327650

Blin *et al.*, Nucleic Acids Res., 2013

„Cluster 23“ Encodes the Biosynthesis of a Ristomycin-like Glycopeptide



ristomycin A

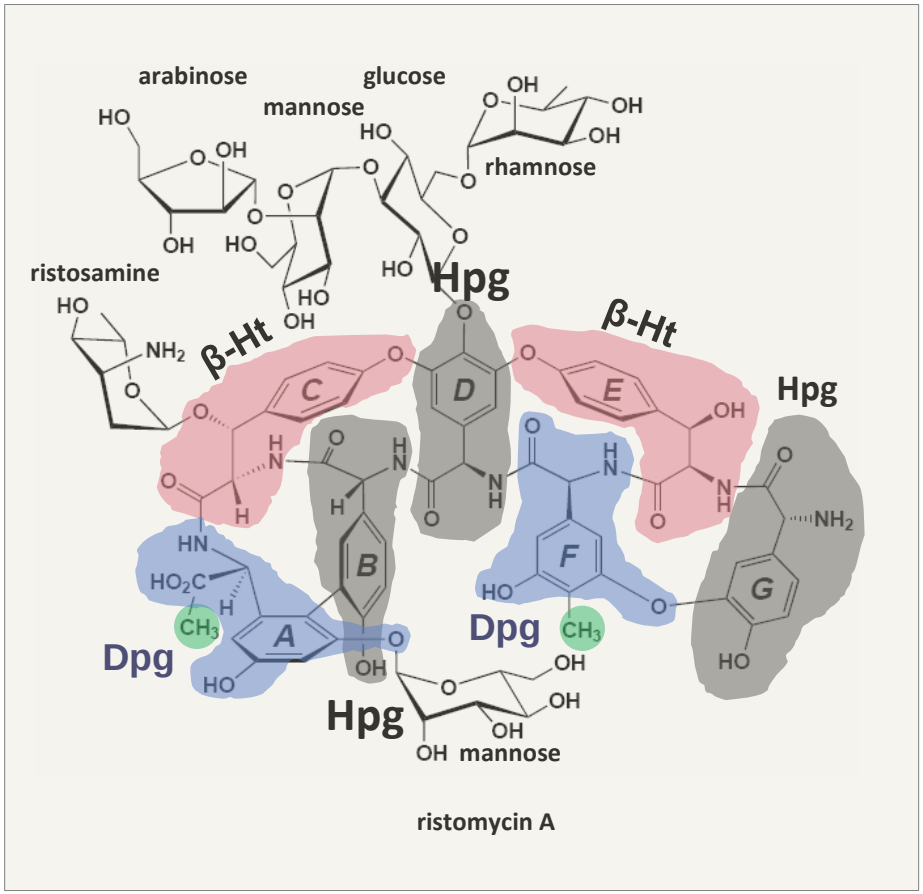
NRPS:

- 7 x aromatic amino acids
  - 3 x Hpg
  - 2 x β-Ht
  - 2 x Dpg

Tailoring enzymes:

- 4 x P450 monooxygenases
- 0 x acetyltransferase
- 6 x glycosyltransferase
- 2 x methyltransferase

„Cluster 23“ Encodes the Biosynthesis of a Ristomycin-like Glycopeptide



NRPS:

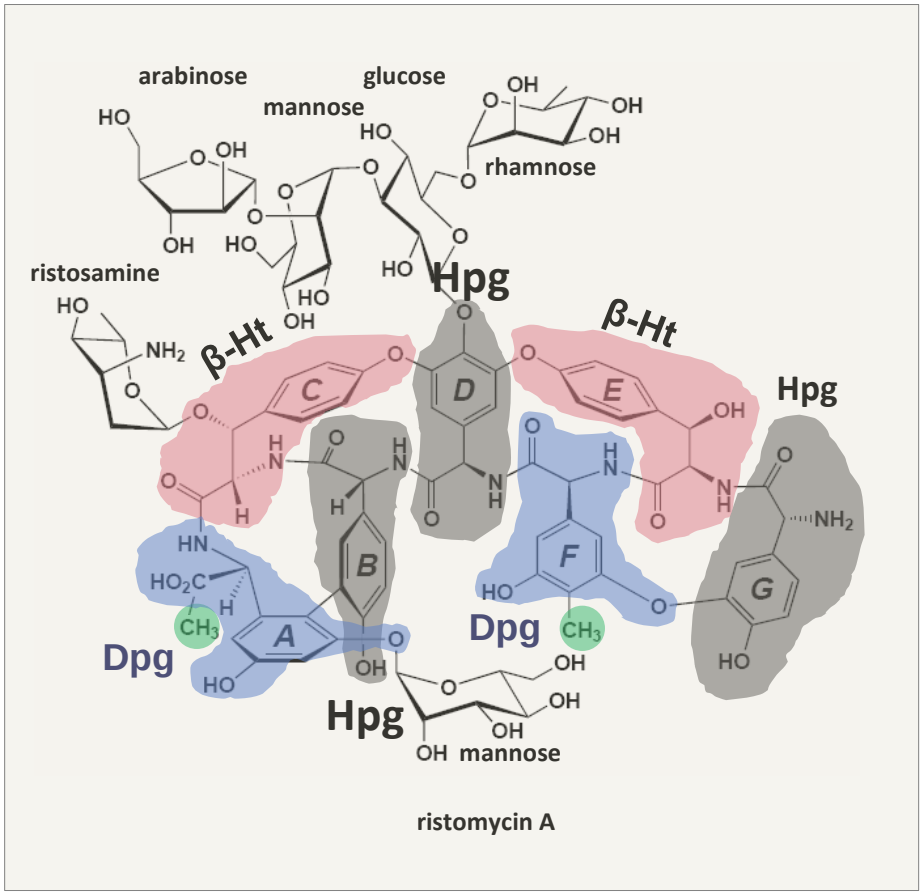
- **7 x aromatic amino acids**
  - **3 x Hpg**
  - **2 x β-Ht**
  - **2 x Dpg**

Tailoring enzymes:

- **4 x P450 monooxygenases**
- **0 x acetyltransferase**
- **6 x glycosyltransferase**
- **2 x methyltransferase**

- Cluster 23 is the first example of a type III glycopeptide gene cluster
- Ristomycin is a component of a diagnostic kit to screen for von Willebrand disease (vWD) and Bernard-Soulier syndrome (BSS)
- Market price: 100 mg: 560 €

„Cluster 23“ Encodes the Biosynthesis of a Ristomycin-like Glycopeptide



NRPS:

- 7 x aromatic amino acids
  - 3 x Hpg
  - 2 x β-Ht
  - 2 x Dpg

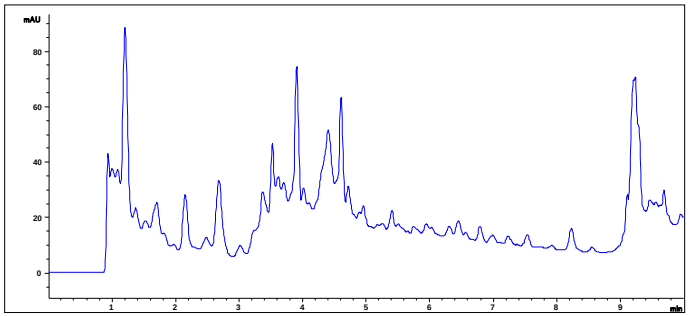
Tailoring enzymes:

- 4 x P450 monooxygenases
- 0 x acetyltransferase
- 6 x glycosyltransferase
- 2 x methyltransferase

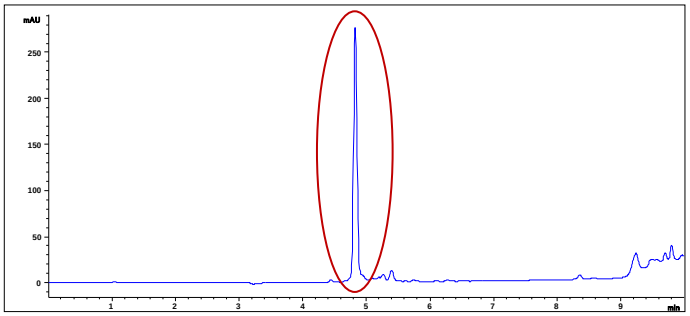
But the wild type does not produce a detectable amount of ristomycin

Heterologous Expression of the Regulator Gene *bbr* in *Amycolatopsis japonicum*

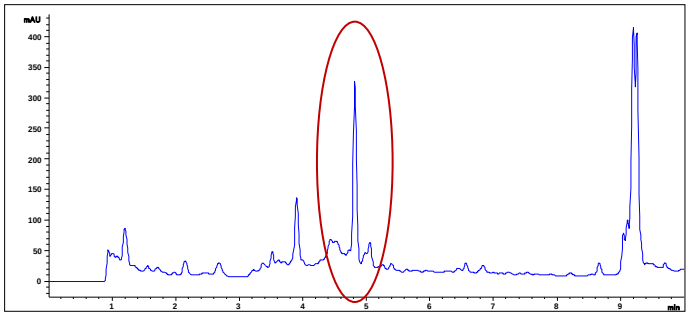
A. japonicum WT
→ Below detection limit

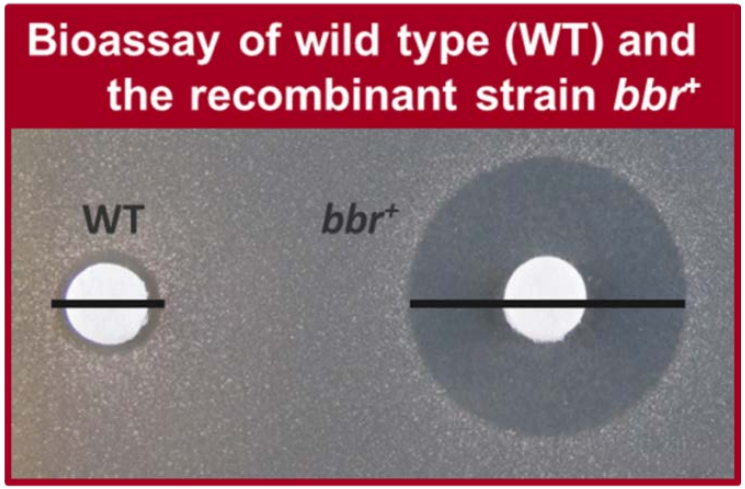


Ristomycin A standard



A. japonicum *bbr*⁺



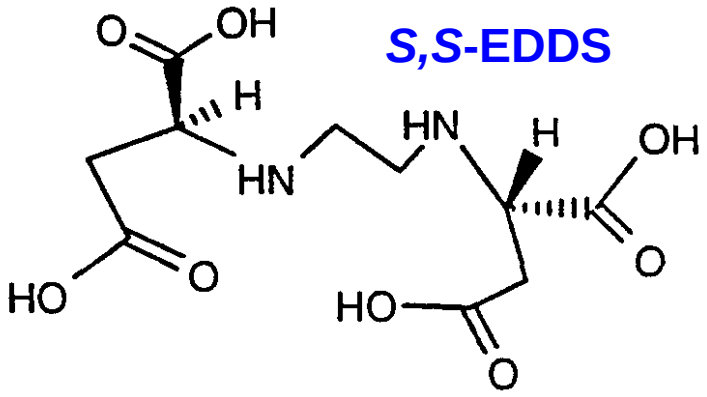


Bacillus subtilis as indicator strain

→ *Amycolatopsis japonicum bbr*⁺ produces 50 mg/l of a ristomycin-like glycopeptide in shaking flasks and 200 mg/l in the fermenter

Under negotiation with company XYZ for commercialisation

Amycolatopsis japonicum, a Talented Producer of Bioactive Compounds

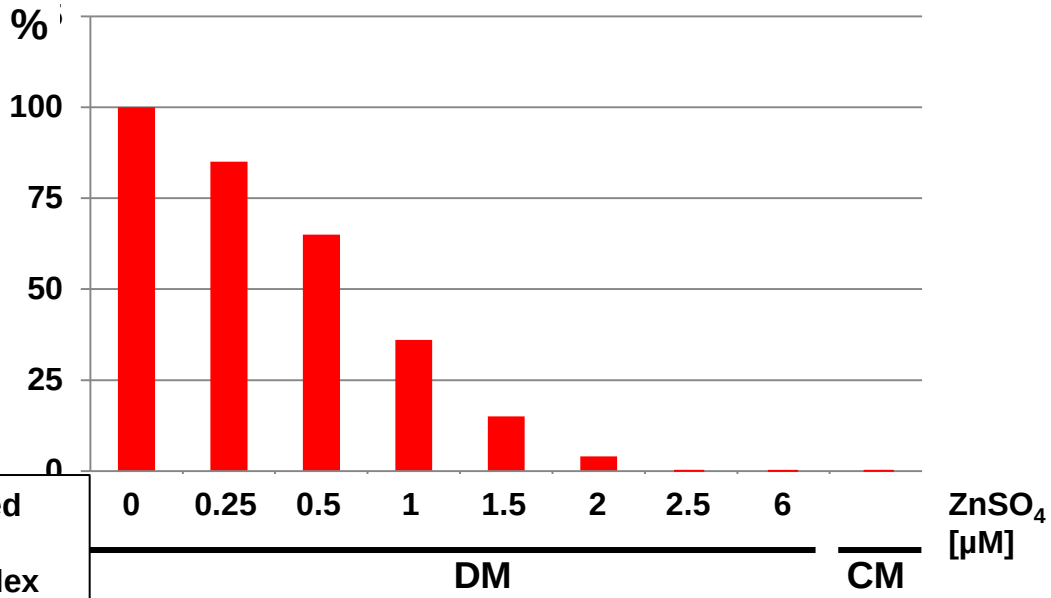


***S,S*-EDDS**

EDDS = Ethylene-diamine-*N*,*N*′-disuccinic acid

***Amycolatopsis japonicum* produces *S,S*-EDDS**

which has similar chelating properties as EDTA, but which is biodegradable

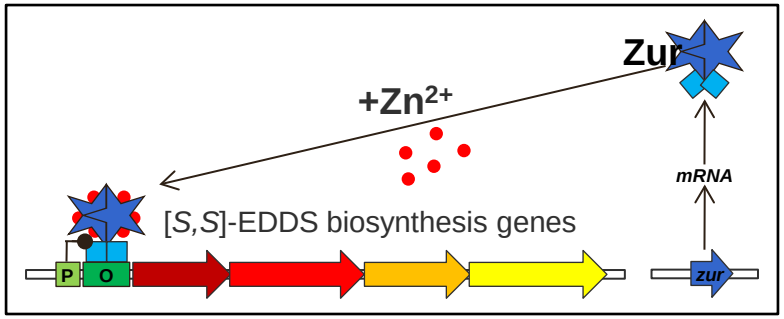


DM = Defined medium
CM = Complex medium

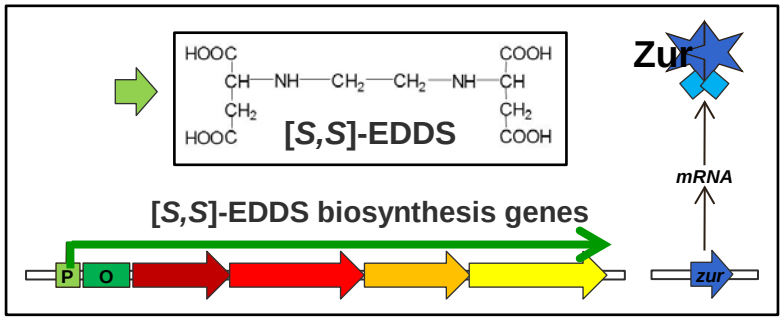
Problem:
Synthesis is strongly Zn-repressed and biosynthesis was unknown

EDDS-Production in *A. japonicum* Is Repressed in the Presence of Zinc

In the presence of Zn: Transcription is blocked



In the absence of Zn: Transcription is possible

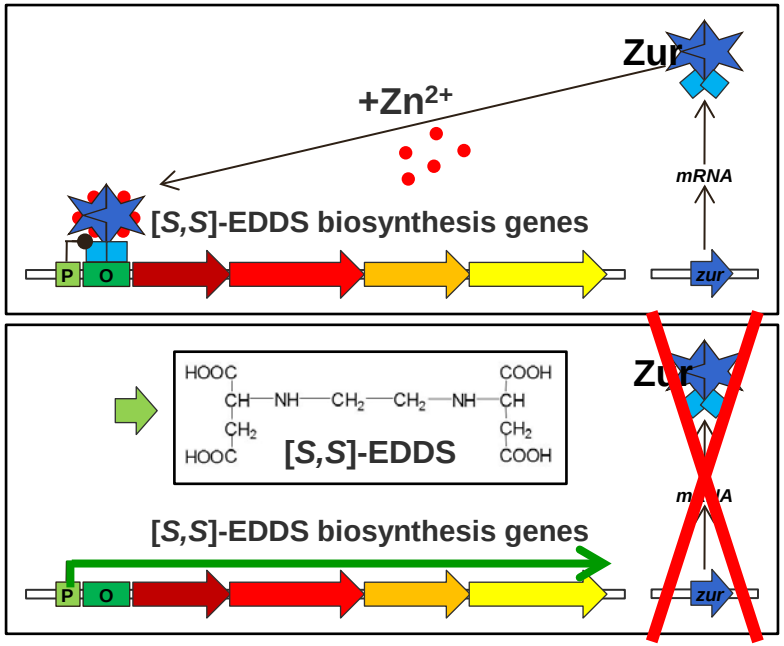


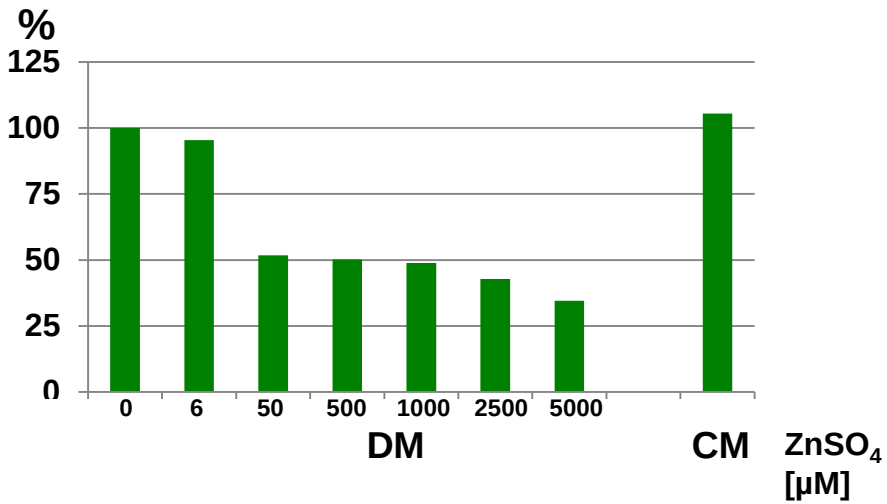
- **Identification of the Zn-dependent regulator**
- **Identification of the EDDS biosynthetic gene cluster**

An A. japonicum Δzur Mutant Produces EDDS in High Yield in Complex Media

In the presence of Zn: Transcription is blocked

In the absence of Zur: Transcription is possible





Solution

- **Identification of the Zn-dependent regulator**
- **Identification of the EDDS biosynthetic gene cluster**
- **Elimination of the Zn-repression**
- **EDDS-production in complex media in the presence of Zn**

Patent Application: P 53563 DE

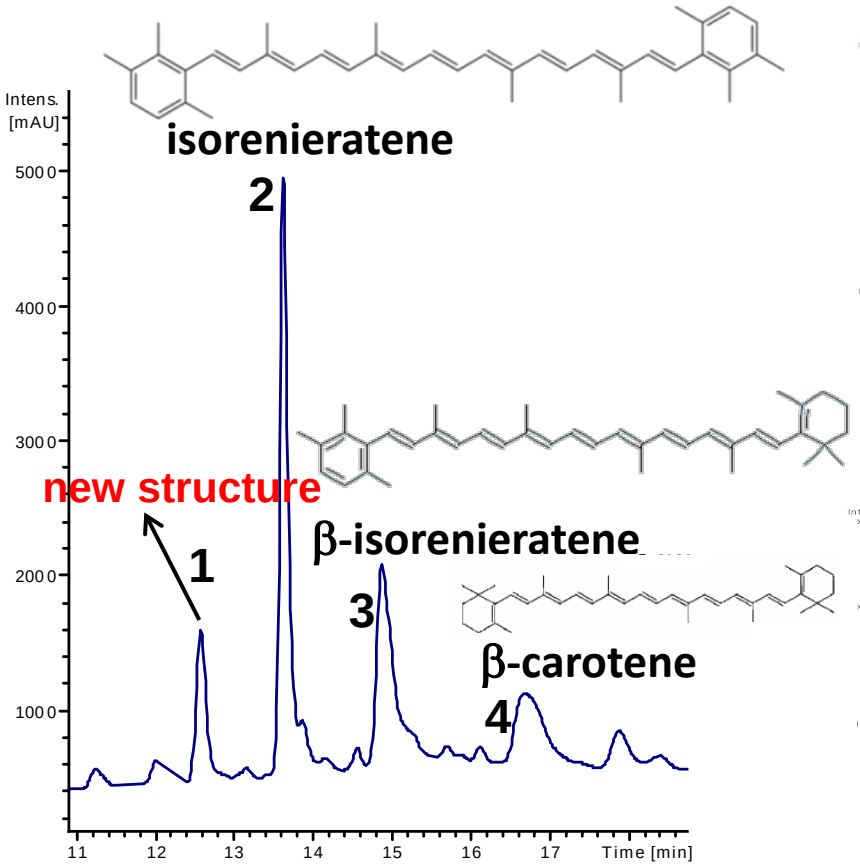
Heterologous Expression of Cosmids from *Streptomyces collinus* in *Streptomyces albus*

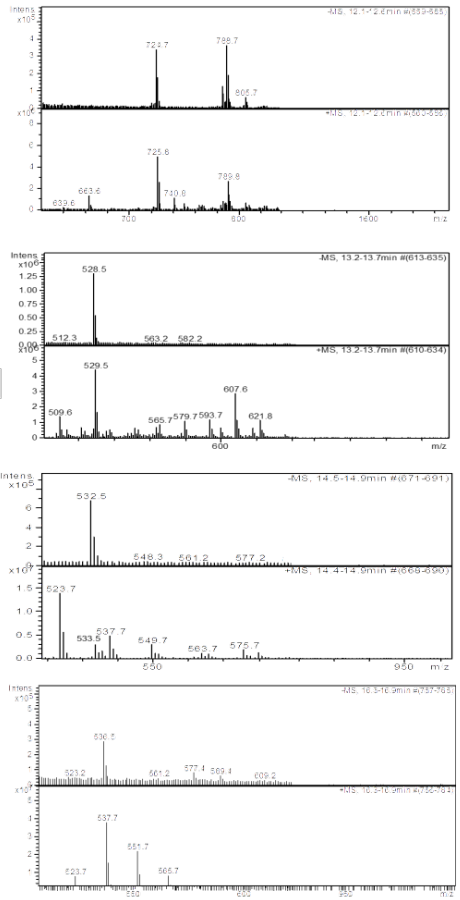
Cluster 1	Cluster 4	Cluster 7	Cluster 10	Cluster 16	Cluster 18	Cluster 19	Cluster 24
lantibiotic	carotenoid	t1pks-nrps	t3pks	t2pks	t1pks-nrps	nrps	t1pks

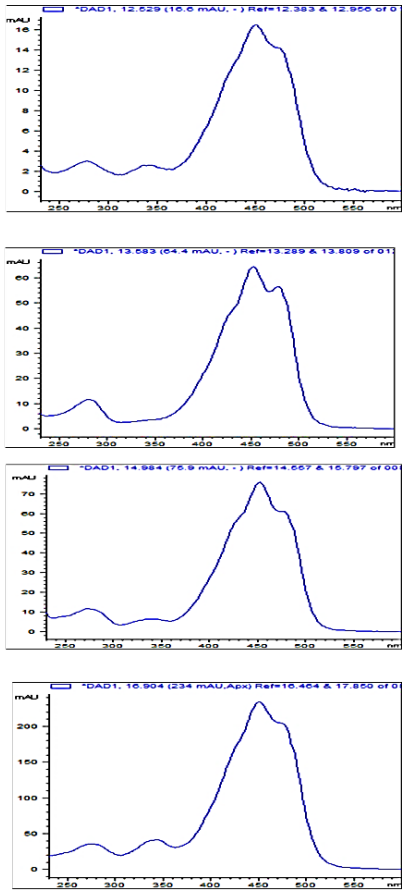
„Cluster 4“ induces visible morphological changes in *S. albus*

Rückert *et al*., J. Biotechnol., 2014

Carotenoid Production in *S. albus*

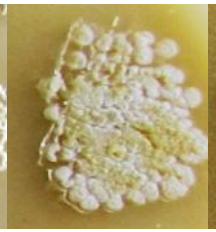
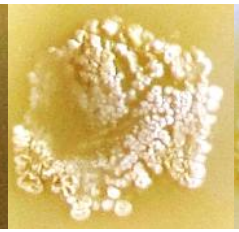
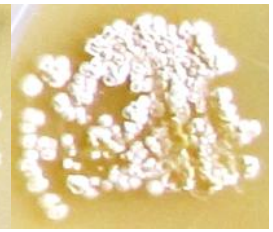






Heterologous expression of "Cluster 4" resulted in the production of four carotenoids

Heterologous Expression of Cosmids from *Streptomyces collinus* in *Streptomyces albus*

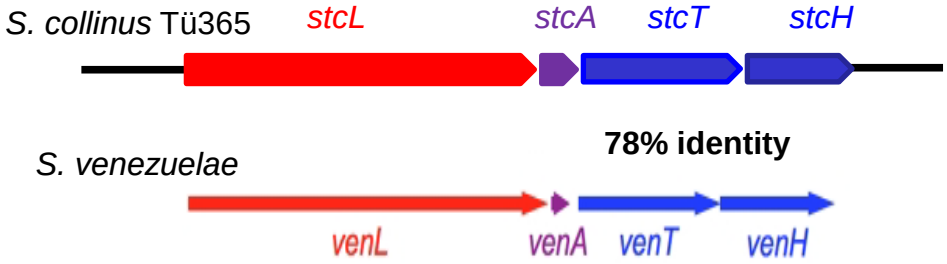
Cluster 1	Cluster 4	Cluster 7	Cluster 10	Cluster 16	Cluster 18	Cluster 19	Cluster 24
							
lantibiotic	carotenoid	t1pks-nrps	t3pks	t2pks	t1pks-nrps	nrps	t1pks

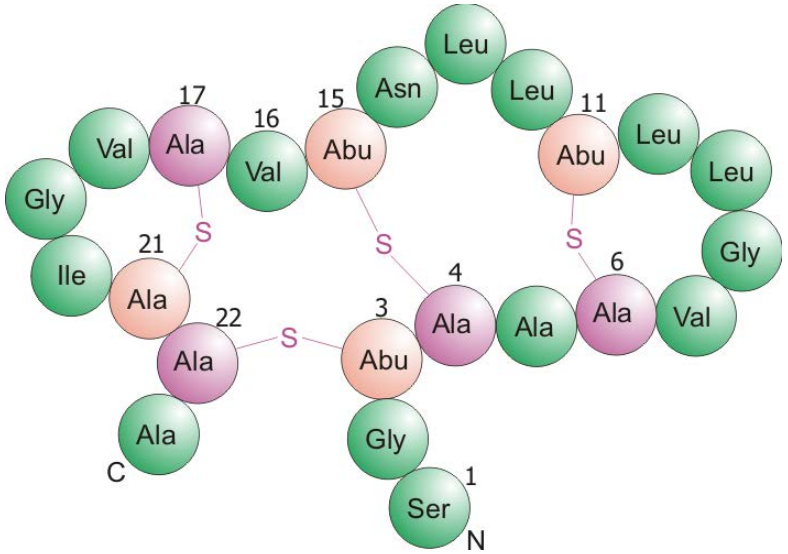
„Cluster 1“ is predcited to encode a new lantibiotic

Rückert *et al*., J. Biotechnol., 2014

Heterologous Expression of „Cluster 1“ in *S. coelicolor* M1146

„Cluster1“ lantibiotic: Streptocollin





Streptocollin

Structure elucidation in cooperation with
Martin Jasyk & Roderich Süßmuth, TU Berlin

```
VenA   MENHDIELLAHLHALPETDPFGVDGAPFAA    TCECVGLLTLLNTVCIGISCA
SclA   MENHDLDLLARLHALPETDPVGVDGEAFAN    TCACVGLLTLLNSVCIGITC
StcA   MENQTLELLAHLHALPETDPETDPVDFDGASYSGTCACVGLLTLLNTVCVGISCA
```

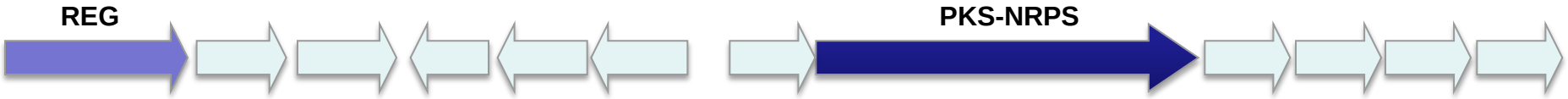
Characterisation of the first isolated lanthipeptide belonging to the type IV class

Iftime *et al*., in preparation

Selection of Bioinformatically Identified Gene Clusters from *S. albus* and *S. argillaceus*

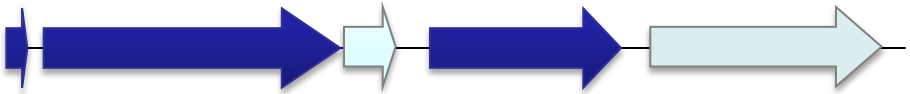
Cluster 9 (NRPS)

Cluster 26 (NRPS-PKS)



Cluster 1701 (PKS)

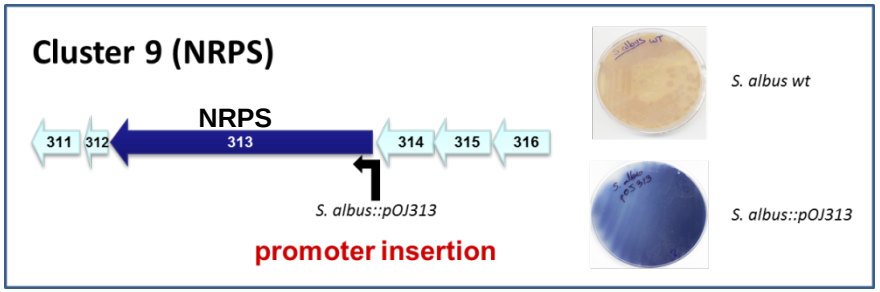
Cluster 909 (lanthipeptide)

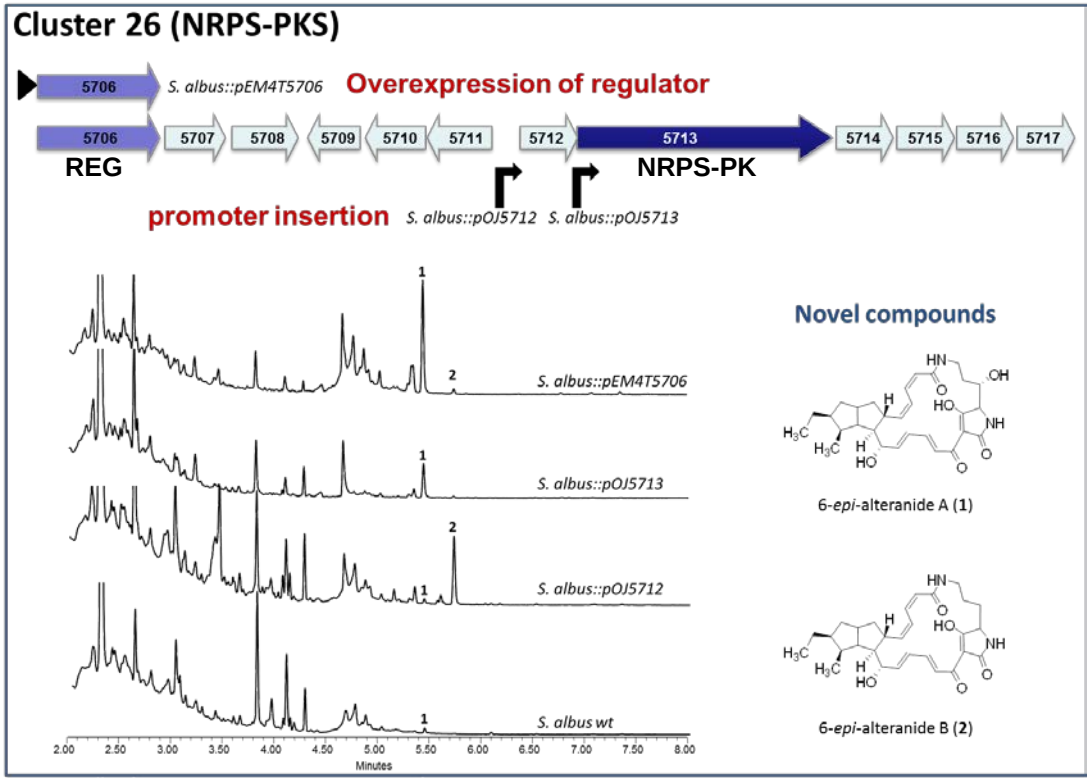


Cluster 25 (lanthipeptide)

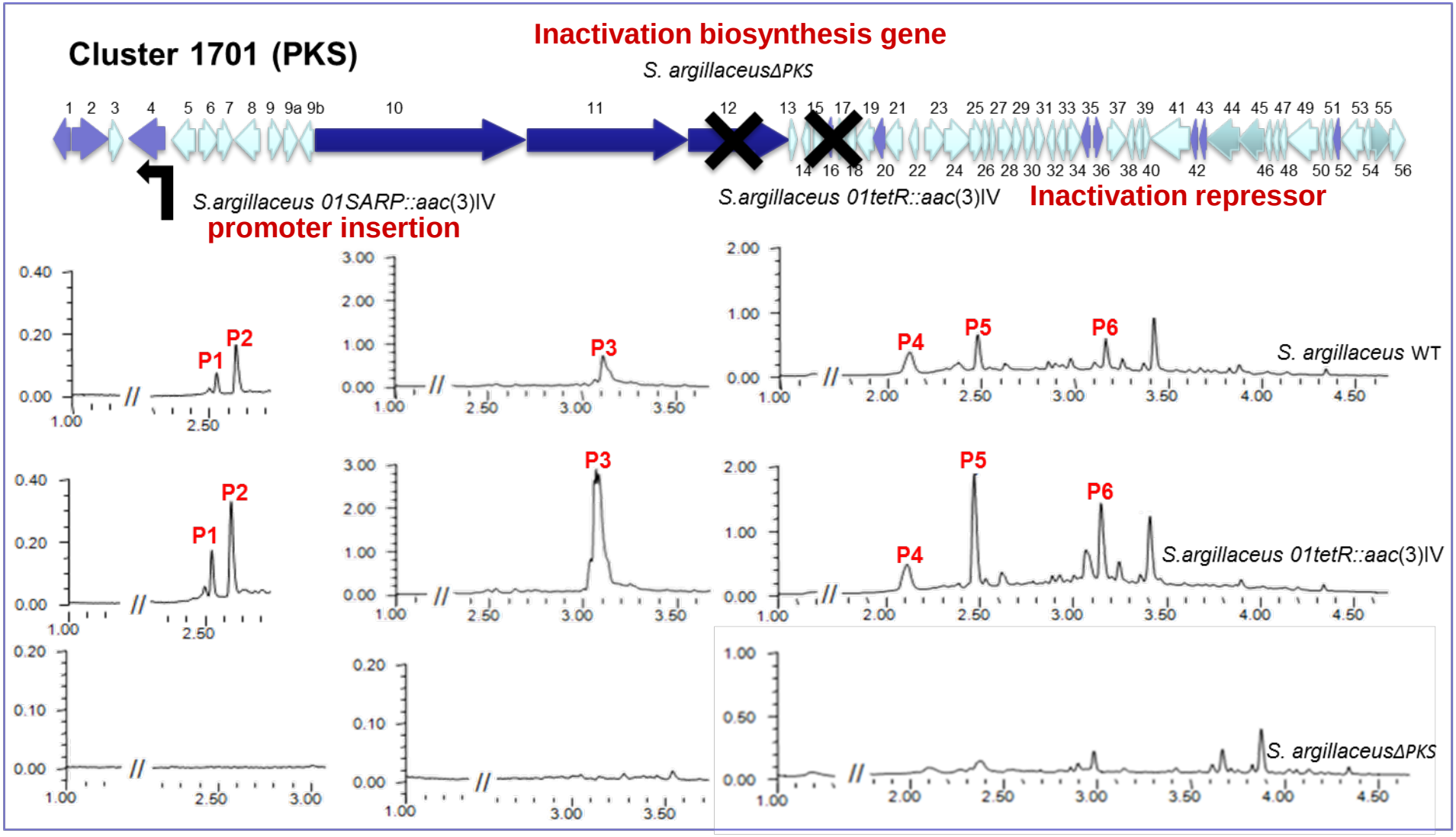
Cluster 1705 (other)

Activation of Silent Gene Clusters from *S. albus*





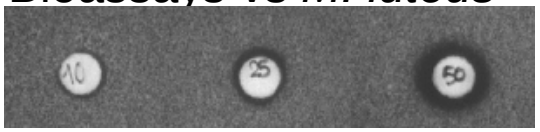
Activation of the Silent Gene Cluster 1701 from *S. argillaceus*



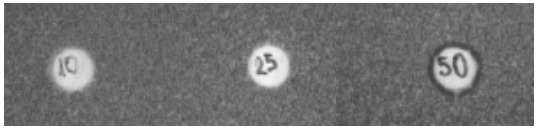
Bioassay of Novel Compounds Encoded by Cluster 1701 of *S. argillaceus*

Bioassays vs *M. luteus*

Nigrifactin (P3)



Argimycin I (P1) + Argimycin II (P2)



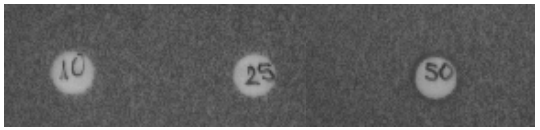
Argimycin IV (P4)

No activity observed

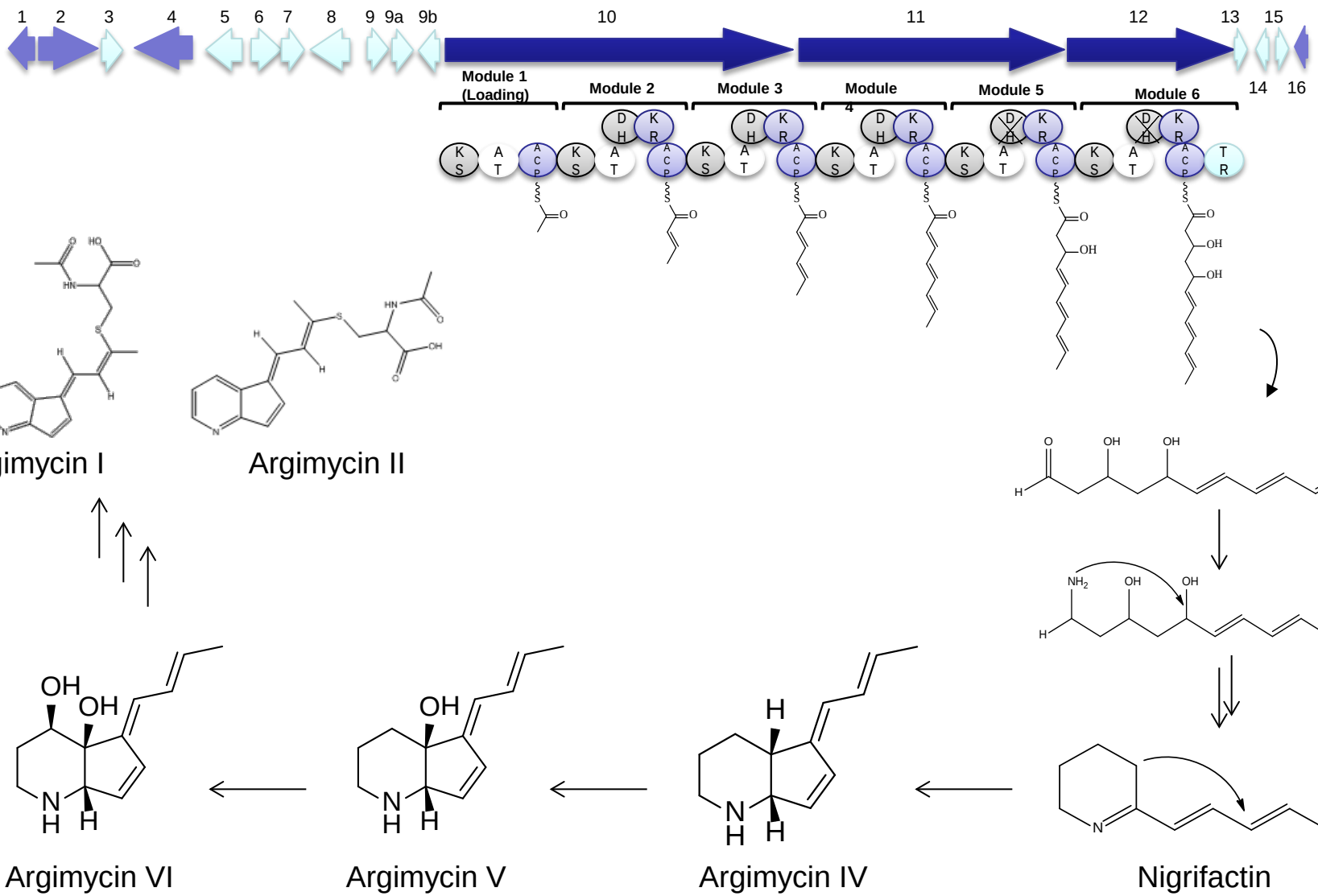
Argimycin V (P5)

No activity observed

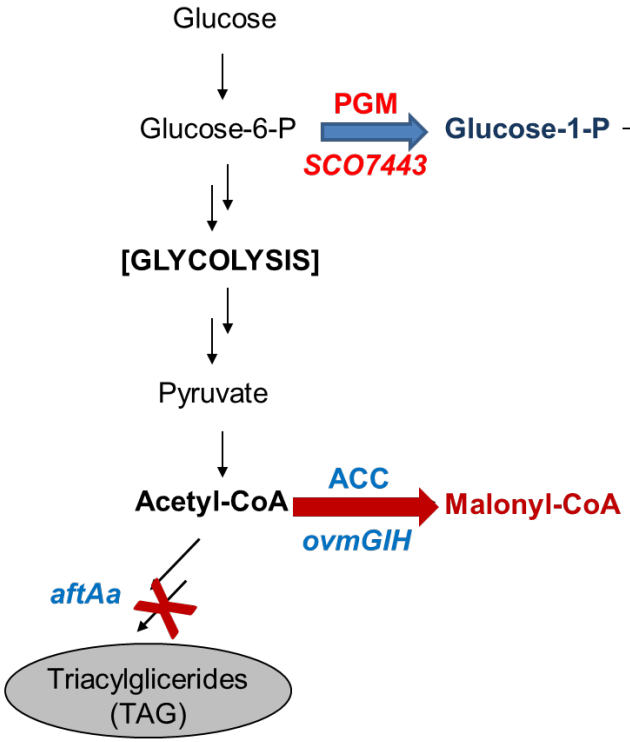
Argimycin VI (P6)

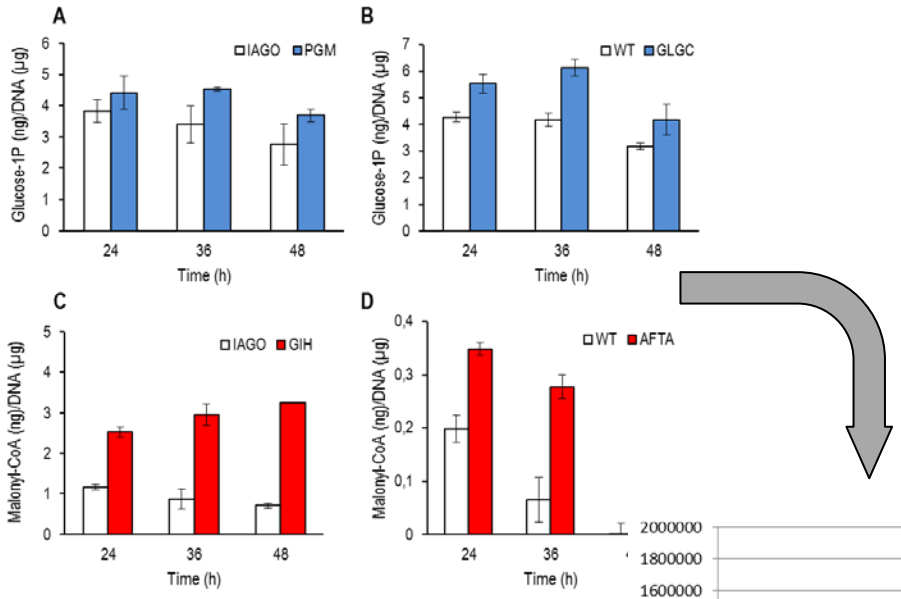


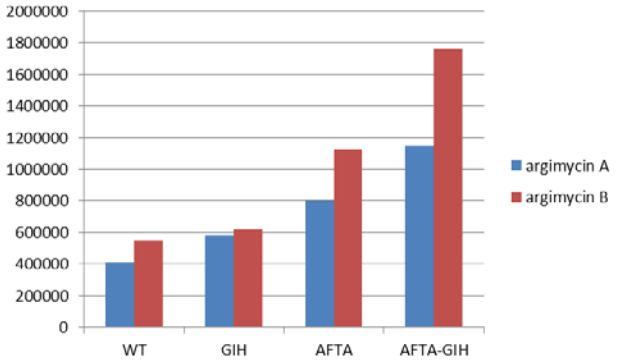
Proposed Biosynthesis Pathway for Argimycins



Increasing the Intracellular Pool of Glucose-1-phosphate and Malonyl-CoA in *S. argillaceus* to Increase Production of Argimycins



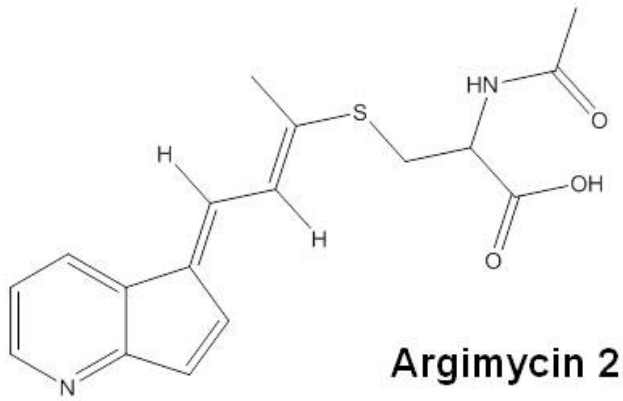




Zabala et al., Metab. Engineer., 2013

Entrechem SL

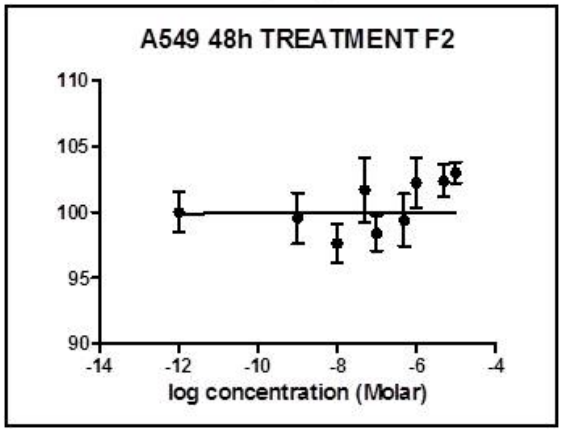
Evaluation of Argimycin 2 (Cytotoxicity Study)



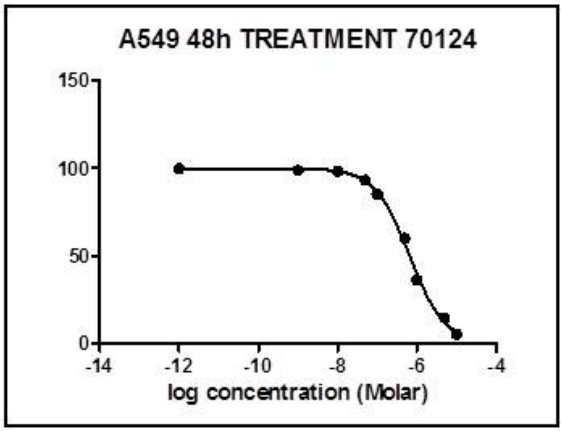
Argimycin 2

Argimycin 2 showed no cytotoxic activity against A459 lung cancer cells

Argimycin 2



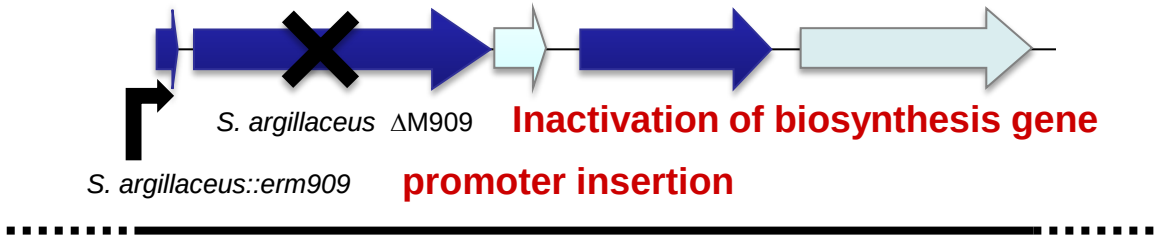
Control



Patent Application: in preparation

Strategies to Activate Other Silent Gene Clusters from *S. argillaceus*

Cluster 909 (lanthipeptide)



Heterologous expression

Cluster 25 (lanthipeptide)

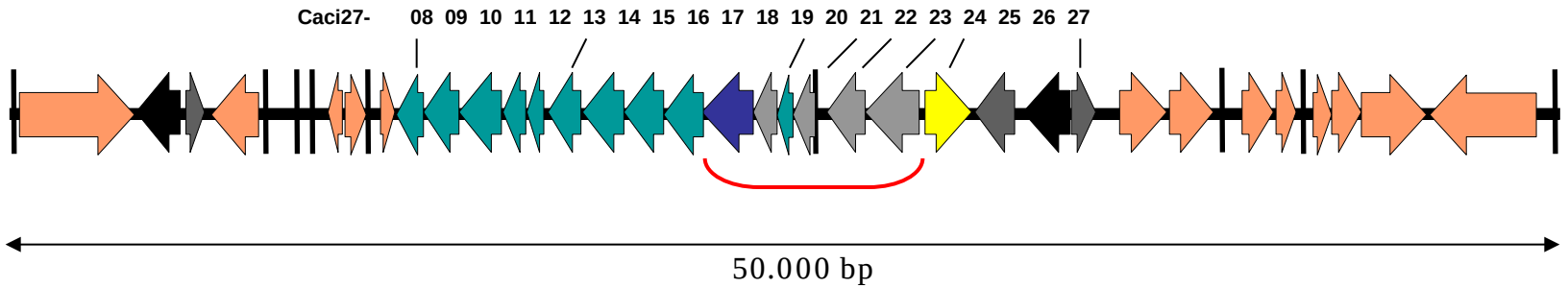
promoter insertion

Cluster 1705 (other)

Overexpression of activators

Deletion of biosynthesis genes

Bioinformatic Identification of a Gene Cluster for an Aminocoumarin Antibiotic in *Catenulispora acidiphila*



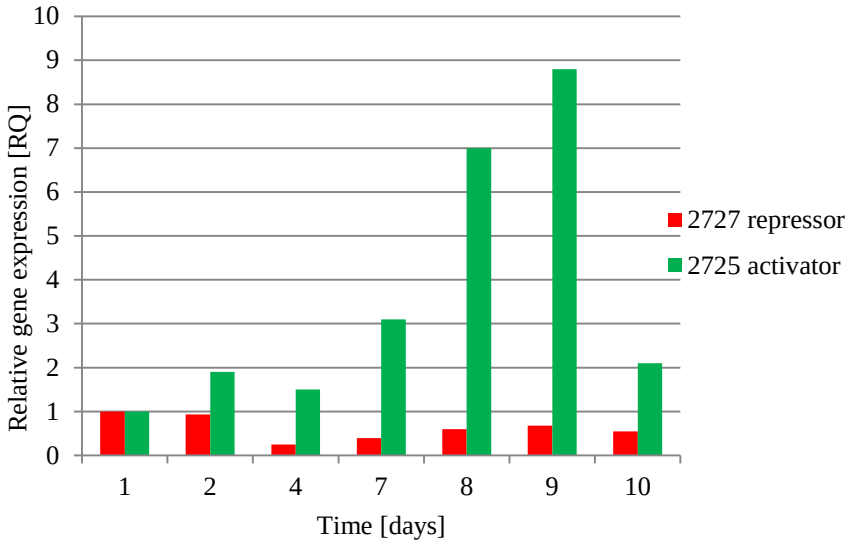
Caci2717: CloL-like
Caci2718: CloK-like
Caci2719: hypothetical protein
Caci2720: CloJ-like
Caci2721: CloY-like
Caci2722: CloI-like
Caci2723: CloH-like

Aminocoumarin moiety
+ amide bond
+ halogenase

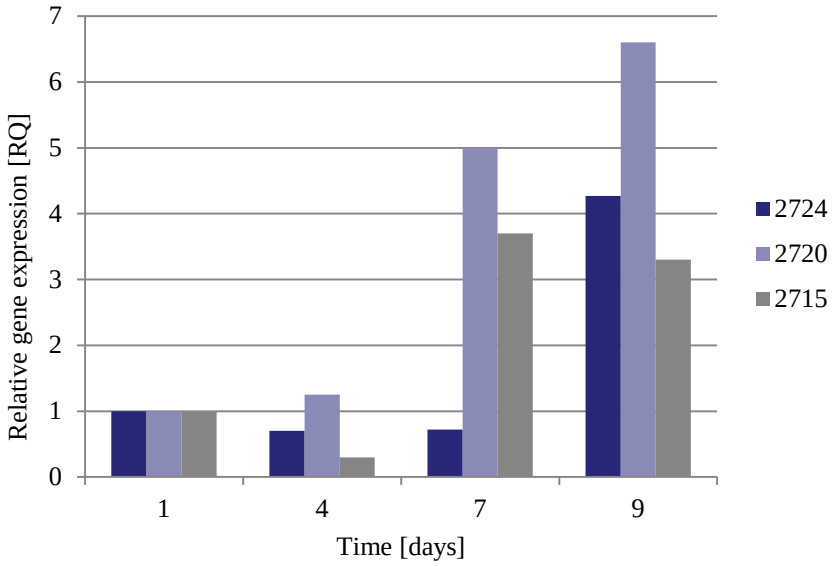
Caci2724: FADH-dependent halogenase (similar to *clohal*)

Gene Cluster Transcriptional Profiles

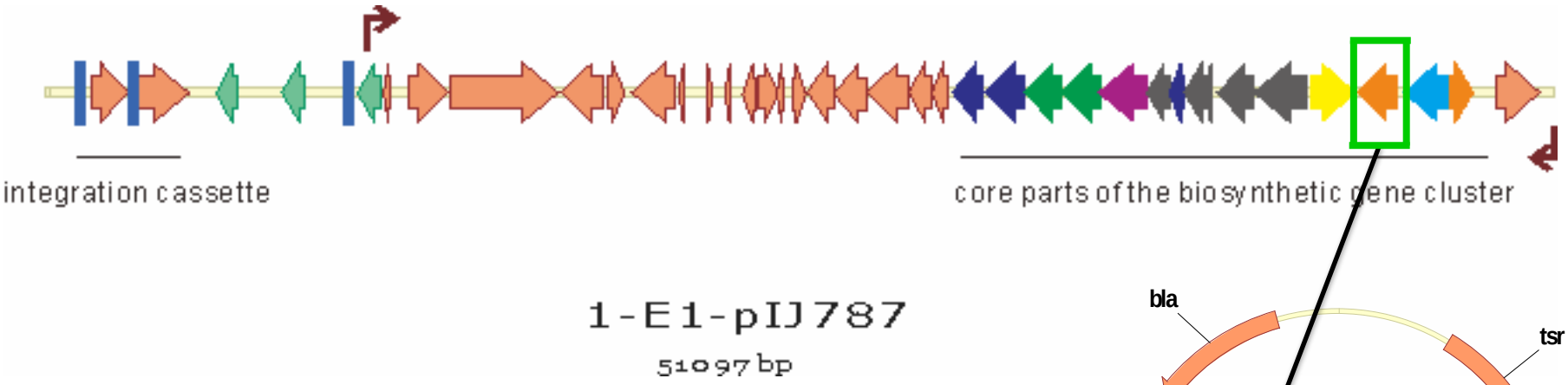
Different profiles of expression of genes encoding an activator and a repressor during culture growth

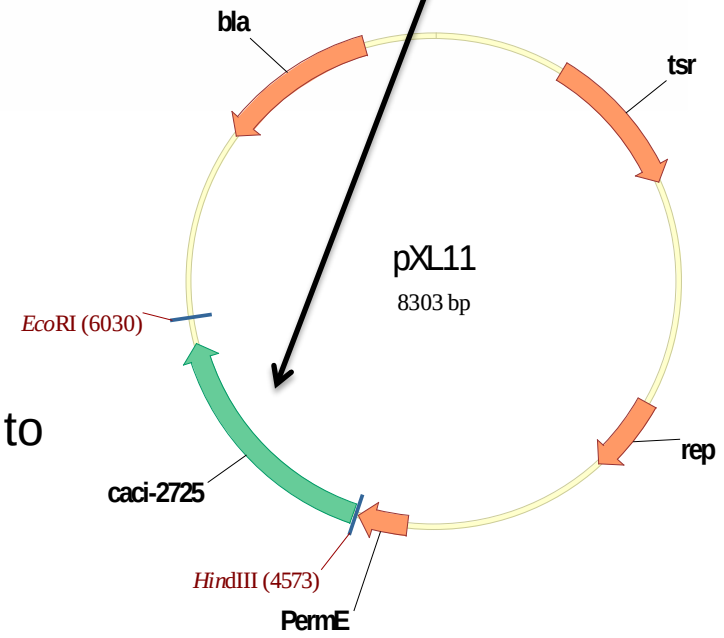


Expression of cluster genes is correlated with the expression of the activator gene



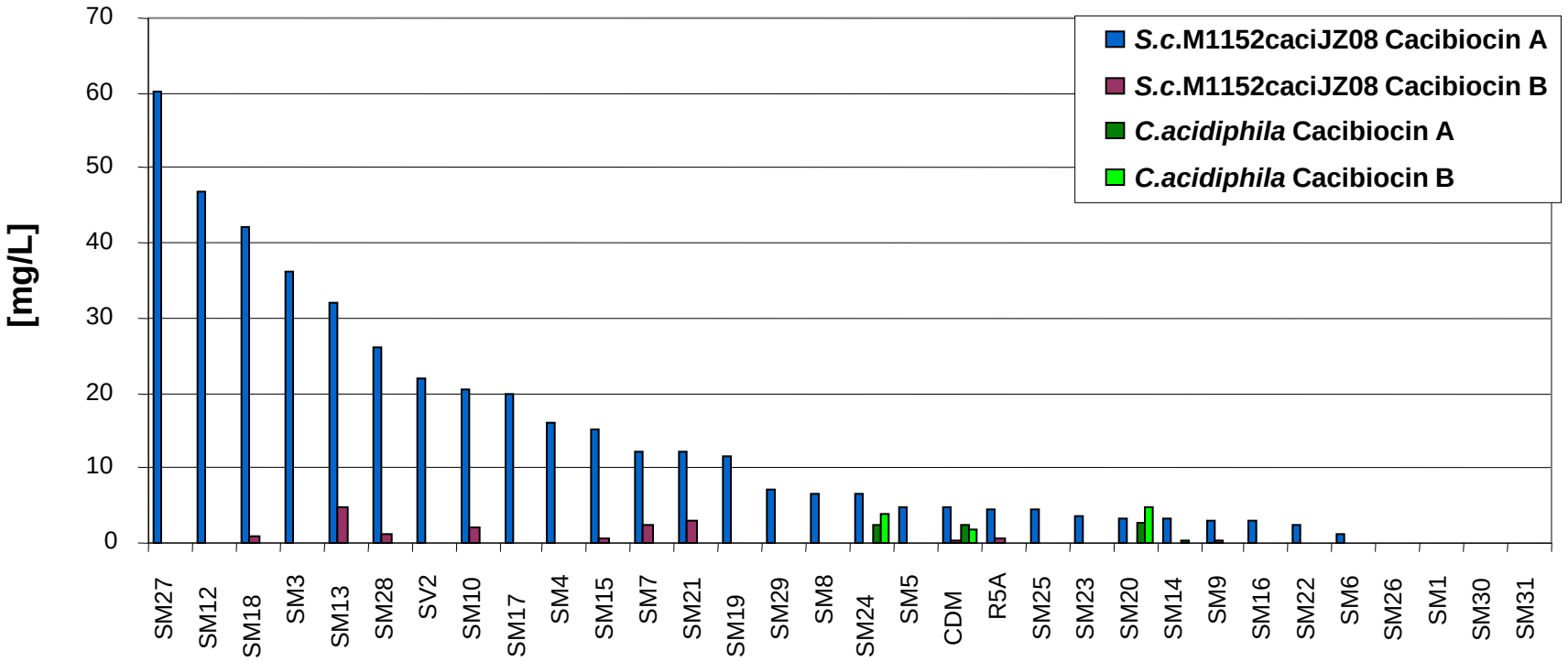
Activation of Heterologous Expression by Overexpression of the *luxR* Regulatory Gene





caci2725, a *luxR* family regulatory gene belonging to the predicted cacibiocin cluster was cloned into pUWL201 under control of the *ermE** promoter

Screening of Different Media for Cacibiocin Production



→ Production of 60 mg/L cacibiocin A in *S.coelicolor* M1152(caciJZ08)

→ Production of 4.6 mg/L cacibiocin B in *Catenulispora acidiphila*

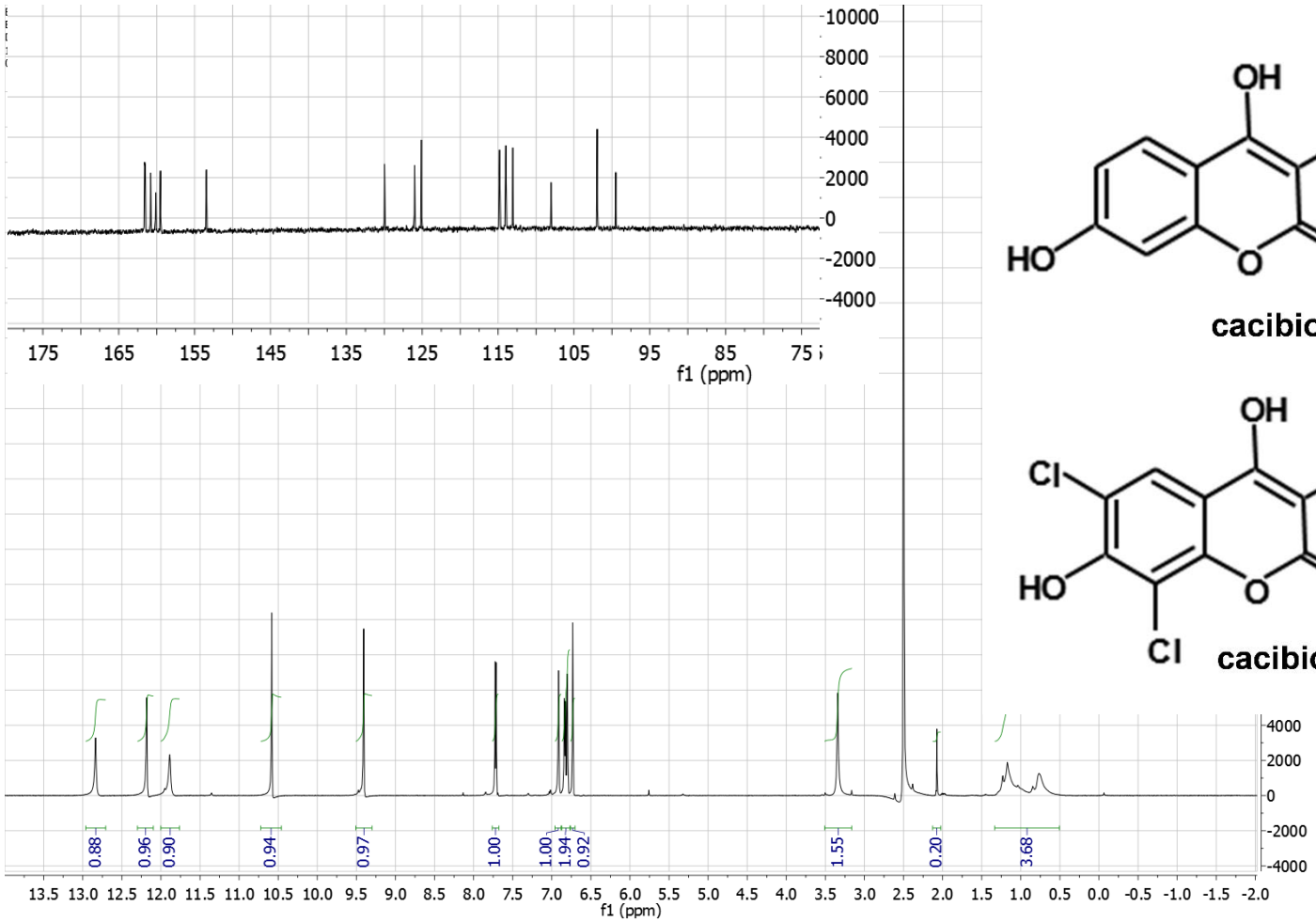
→ Different optimal media for *C.acidiphila* and the heterologous host

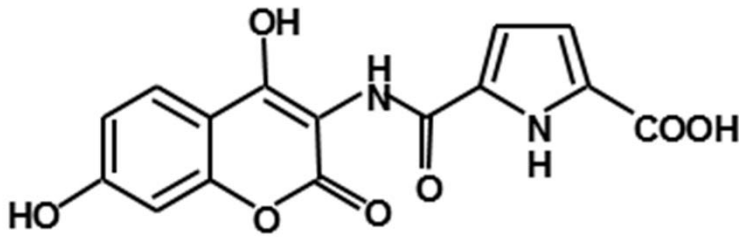
Fermentation of the Engineered Aminocoumarin Producer Strain

100 liter fermentation
of *S. coelicolor* M1152 10E2int pXL11
at Galilaeus Oy

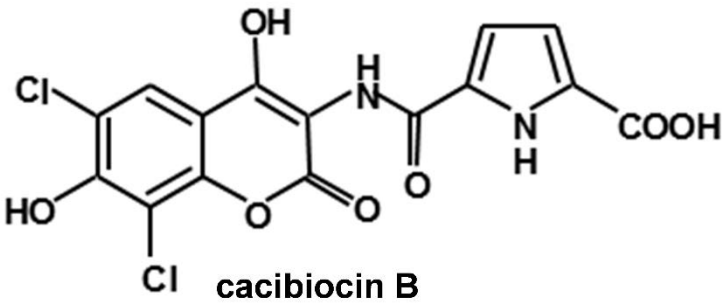
≈ 120 mL extract
cacibiocin A and B
(new natural products)

Structure Elucidation of Cacibiocin A and B





cacibiocin A



cacibiocin B

Patent application: 10 2013 021828.4

New Industrially Relevant Products/Processes from GenoDrug

Argimycins/ Nigrifactin:	**New antibiotics**
Cacibiocins:	**New aminocoumarin**
Carotenoid:	**New isorenieratene (?) derivative**
EDDS:	**Biodegradable chelator: High yield production in complex media**
Ristomycin:	**Production strain for a high value component of a diagnostic kit**
Streptocollin:	**First type IV lantibiotic**

+ many silent gene clusters and products thereof, awaiting further characterisation

Patent application Commercialisation

GenoDrug Final Report

Structure:

1. Background
2. Project aim and approach
3. Project partners
4. Progress and cooperations
5. Achievement of milestones and objectives
6. Problems encountered; proposed changes

Achievement of Milestones

		Project month
M 1	Bioinformatic identification of clusters	6
M 2	Bioinformatic identification of regulators	6
M 3	Transcription profiles of gene clusters	9
M 4	Metabolite profiles under different growth conditions	9
M 5	Techniques for overexpression/deletion of regulatory genes	15
M 6	Gene clusters with introduced artificial promoters constructed	15
M 7	Heterologous expression constructs generated	15
M 8	Engineering stategies for increasing precursor supply established	18
M 9	Expression of previously silent clusters	18
M 10	Chemical identification of novel compounds	21
M 11	New compounds from different chemical classes by fermentation	24
M 12	Structure elucidation of new compounds	27
M 13	Pharmacological testing of new compounds initiated	27

Achievement of Deliverables

Deliverables		Month
D 1	Report of Co-ordinator at kick-off seminar	1
D 2	DNA sequences biosynthetic gene clusters of antibiotics	6
D 3	Improved bioinformatic technologies	12
D 4	Report on transcription profiles of gene clusters of interest	12
D 5	Strains containing heterologous gene clusters	18
D 6	Gene clusters with introduced artificial promoters	18
D 7	Mid-term progress report	18
D 8	Strains manipulated with respect to regulatory genes	24
D 9	Strains manipulated with respect to precursor supply	24
D 10	Strains expressing new bioactive compounds	24
D 11	Successful genetic strategies for the activation of gene clusters	24
D 12	Technologies for metabolic engineering to increase production	24
D 13	Technologies for upscaling	24
D 14	New bioactive compounds for drug discovery programs	24
D 15	Chemical structures of new bioactive compounds	30
D 16	Report on bioactivities and toxicities	36
D 17	Process patents	36
D 18	Substance patents	36
D 19	Scientific publications	36
D 20	Final progress report	36

GenoDrug Final Report

Structure:

1. Background
2. Project aim and approach
3. Project partners
4. Progress and cooperations
5. Achievement of milestones and objectives
6. Problems encountered; changes

no major problems or changes

EIB.10.023 GenoDrug
Genome mining for drug discovery:
Activation of silent biosynthetic gene clusters

Finland

Germany

Poland

Spain

Thank you!