ACS chemical biology

Article

Discovery of cryptic largimycins in Streptomyces reveals novel biosynthetic avenues enriching the structural diversity of the leinamycin family

Adriana Becerril, Ignacio Pérez-Victoria, Suhui Ye, Alfredo F Braña, Jesus Martin, Fernando Reyes, Jose A. Salas, and Carmen Méndez

ACS Chem. Biol., **Just Accepted Manuscript** • DOI: 10.1021/acschembio.0c00160 • Publication Date (Web): 20 Apr 2020

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Discovery of cryptic largimycins in *Streptomyces* reveals novel biosynthetic avenues enriching the structural diversity of the leinamycin family

Adriana Becerril,[†,‡,§] Ignacio Pérez-Victoria,[#,§] Suhui Ye, [†,‡] Alfredo F. Braña, [†,‡] Jesús Martín, [#] Fernando Reyes, [#] José A. Salas[†,‡] and Carmen Méndez*[,†,‡]

[†]Departamento de Biología Funcional e Instituto Universitario de Oncología del Principado de Asturias (I.U.O.P.A), Universidad de Oviedo, Oviedo, Spain

[‡]Instituto de Investigación Sanitaria de Asturias (ISPA), Oviedo, Spain

[#]Fundación MEDINA, Centro de Excelencia en Investigación de Medicamentos Innovadores en Andalucía, Armilla, Granada, Spain.

[§]These authors contributed equally

* **Correspondence:** Carmen Méndez (cmendezf@uniovi.es)

ABSTRACT

Largimycins are hybrid nonribosomal peptide-polyketides that constitute a new group of metabolites in the leinamycin family of natural products displaying unique structural features such as containing an oxazole instead of a thiazole ring or being oxime ester macrocycles, unprecedented in Nature, rather than macrolactams. Their discovery in *Streptomyces argillaceus* and *Streptomyces canus* has relied on the activation of two homologous silent gene clusters by overexpressing a transcriptional activator and cultivating in specific media. The proposed biosynthesis of largimycins includes the key action of the oxidoreductase LrgO, responsible for the formation of the oxime group involved in macrocyclization, and two putative cryptic biosynthetic steps consisting in chlorination of L-Thr by the NRPS loading module and incorporation of an olefinic exomethylene group by LrgJ PKS. The discovery of largimycins uncovers novel biosynthetic avenues employed by Nature to enrich the structural diversity of leinamycins and provides tools for combinatorial biosynthesis.

Streptomyces bacteria are a prolific source of natural products displaying some kind of bioactivity.[1] Many of these specialized metabolites are polyketides (PKs) and nonribosomal peptides (NRPs), including hybrid PK-NRPs. Leinamycin (LNM) is a remarkable example of such bioactive hybrid natural products.[2] It is characterized by a 1,3-dioxo-1,2-dithiolane moiety spirofused to thiazole-containing macrolactam ring (Figure 1a), and displays antitumor activity exerted through a novel mode of action rendering DNA damage.[2,3] The LNM macrolactam ring is synthesized by a hybrid nonribosomal peptide synthethase (NRPS)/acyltransferase (AT)-less type I polyketide synthase (PKS) featuring a novel pathway for PK β-alkylation and a didomain of unknown function (DUF)-cystein lyase domain (SH), which incorporates sulfur in the PK chain.[4-6] Since its discovery thirty years ago,[2] LNM has constituted the only known member of this class of compounds until the very recent discovery of two other members, the guangnanmycins (GNMs) and the weishanmycins (WSMs) (Figure 1b).[7]

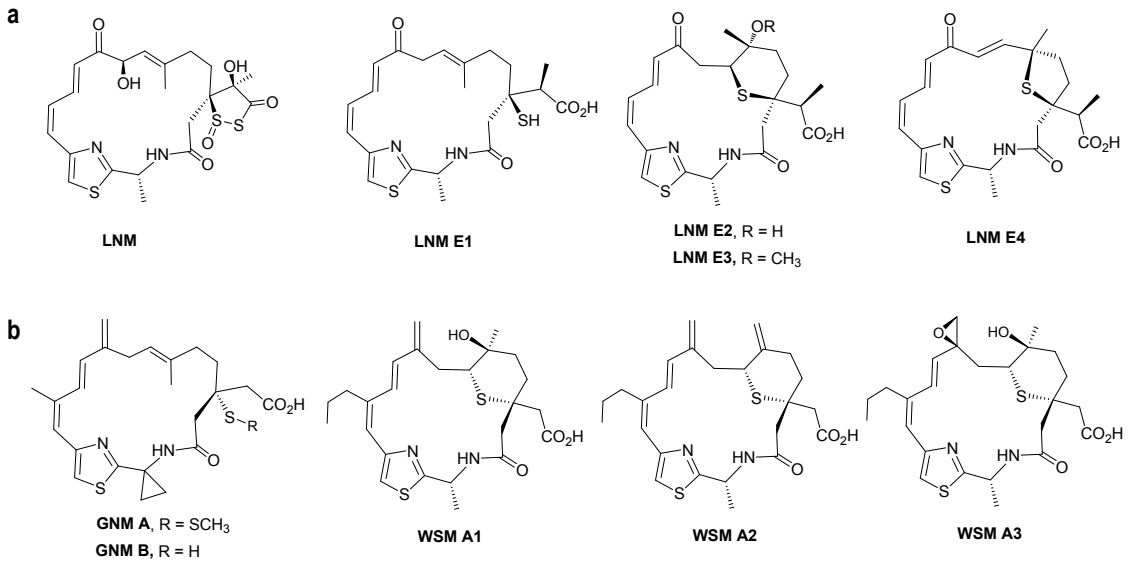


Figure 1. Structure of known leinamycin metabolites. a) leinamycins (LNMs). b) guangnanmycins (GNMs) and weishanmycins (WSMs).

Genome mining of *Streptomyces* species has untapped their enormous potential to produce specialized metabolites, based on the plethora of biosynthetic gene clusters (BGCs) identified.[8] However, most of these clusters are not (or only poorly) expressed under standard laboratory cultivation conditions. Several strategies have been developed to induce the expression of such silent BGCs, including the genetic manipulation and/or the modification of growth conditions, in order to identify their encoded metabolites.[9-11] Recently, genome mining of *Streptomyces argillaceus* ATCC 12956 has identified 31 putative BGCs.[12] In addition to the BGC for the antitumor mithramycin that had been previously cloned and characterized,[13] five cryptic BGCs

were identified and activated coding for compounds previously unknown to be produced by this microorganism including argimycins P, antimycins, carotenoids, desferrioxamine and germicidins.[12,14] Herein we report the identification of another cryptic BGC (*lrg*) coding for a hybrid NRPS/AT-less Type I PKS, its activation by combining two strategies, and the identification and structure elucidation of their encoded compounds, the largimycins (LRGs) that constitute a new group of compounds in the LNM family of natural products due to their unique structural features, ultimately expanding the known biosynthetic repertoire employed by Nature to generate diversity in this family.

RESULTS AND DISCUSSION

Activation and characterization of compounds encoded by cluster *lrg*. Cluster 11 (Figure 2) corresponds to a hybrid NRPS/AT-less Type I PKS gene cluster that spans 101.42 kb and contains 62 open reading frames (*orf*). Most of the encoded proteins from *lrgT1* to *lrgC3* show similarities to those encoded by the LNM BGC (*lnm*) (between 34% and 66% of identical amino acids) (Table 1 and Supporting Information), suggesting that compounds encoded by cluster 11 might be structurally related to LNM.[2,15]

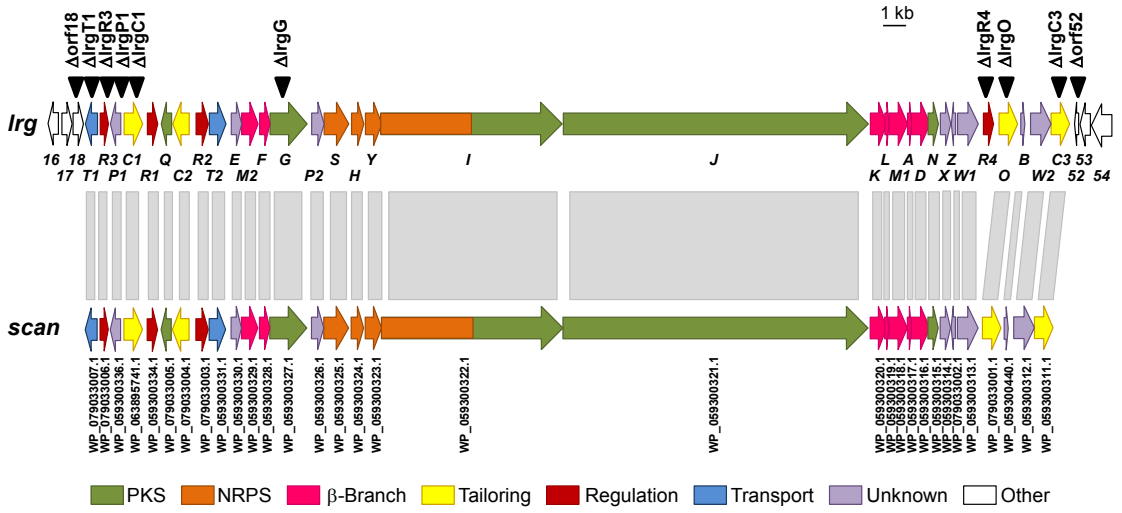


Figure 2. Genetic organization of cluster *lrg* from *Streptomyces argillaceus* and cluster *scan* from *Streptomyces canus*. Bars between clusters indicate homologous genes. Black triangles indicate those genes that have been inactivated in *S. argillaceus*. Genes are shown to scale.

To identify and characterize those cryptic compounds several approaches were sequentially followed: (i) Generation of a mutant in the *trans*-AT *lrgG* (see Supporting Information). UPLC analyses of broth extracts from this mutant (*S. argillaceus* ΔlrgG) and the wild type strain grown

in R5A medium did not show any differential peak, indicating that cluster 11 was silent (data not shown); (ii) Activation of cluster 11 by overexpressing/deleting putative transcriptional regulators.

Table 1. Predicted functions of genes in the *Streptomyces argillaceus lrg* gene cluster

Gene product	aa	Proposed function	Most similar protein (acc. number)	Identical aa (%)	Similar protein Lnm/Syr	Identical aa (%)
ORF18	202	Hypothetical protein	KUN87089.1	94		
LrgT1	509	MFS transporter	WP_100570177.1	94	LnmY	40
LrgR3	205	TetR family transcriptional regulator	WP_100570178.1	93		
LrgP1	191	Hypothetical protein	WP_100570179.1	94		
LrgC1	452	Cytochrome P450	WP_100570180.1	93		
LrgR1	222	Crp/Fnr family transcriptional regulator	WP_100570182.1	96	LnmO	43
LrgQ	250	Phosphopantetheinyl transferase	WP_100570183.1	91		
LrgC2	433	Cytochrome P450	WP_100570184.1	97	LnmA/LnmZ	36/36
LrgR2	252	Crp/Fnr family transcriptional regulator	WP_100570185.1	90	LnmO	42
LrgT2	437	Sodium:proton exchanger	WP_100570186.1	95		
LrgE	299	Hypothetical protein	WP_100570187.1	94	LnmE/LnmH	48/34
LrgM2	410	Hydroxymethylglutaryl-CoA synthase	WP_100570188.1	97	LnmM	45
LrgF	263	Enoyl-CoA hydratase	WP_100570189.1	97	LnmF	53
LrgG	798	AT-less acyltransferase/oxidoreductase	WP_100570343.1	91	LnmG	64
LrgP2	316	Hypothetical protein	WP_100570190.1	97	LnmH/LnmE	43/43
LrgS	599	NRPS (A-PCP)	WP_100570191.1	95	SyrB1	46
LrgH	314	Halogenase	WP_100570344.1	97	SyrB2	67
LrgY	391	Aminoacyltransferase	WP_100570192.1	94	SyrC	43
LrgI	4288	hybrid NRPS/AT-less Type I PKS	WP_100570193.1	91	LnmI	51
LrgJ	7340	AT-less type I PKS	WP_100570194.1	90	LnmJ	48
LrgK	330	acyltransferase/decarboxylase	WP_100570195.1	93	LnmK	54
LrgL	87	Acyl carrier protein	WP_100570196.1	94	LnmL	66
LrgM1	412	Hydroxymethylglutaryl-CoA synthase	WP_100570197.1	93	LnmM	63
LrgA	80	Acyl carrier protein	WP_100570198.1	97		
LrgD	423	Beta-ketoacyl-ACP synthase	WP_100570199.1	95		
LrgN	253	Type II thioesterase	WP_059300315.1	90	LnmN	54
LrgX	239	N-acetylglucosaminyl deacetylase	WP_100570201.1	93	LnmX	56
LrgZ	132	Hypothetical protein	WP_079252697.1	95	LnmZ'	34
LrgW1	487	4-coumarate-CoA ligase	WP_100570203.1	93	LnmW	44
LrgR4	201	TetR family transcriptional regulator	WP_100570204.1	88		
LrgO	395	FAD-dependent oxidoreductase	WP_100570205.1	94		
LrgB	121	Glyoxalase/bleomycin resistance protein	WP_100570206.1	93		
LrgW2	494	4-coumarate-CoA ligase	WP_100570207.1	96	LnmW	42
LrgC3	414	Cytochrome P450	PJM93929.1	96	LnmA	36
ORF52	93	Hypothetical protein	WP_033213698.1	75		
ORF53	179	Hypothetical protein	WP_100570209.1	82		
ORF54	489	Amino acid permease	WP_020126812.1	93		

A, adenilation domain; AT, acyltransferase domain; ACP, acyl carrier protein; MFS, major facilitator superfamily; NRPS, non-ribosomal peptide synthetase; PCP, peptidyl carrier protein; PKS, polyketide synthase**.** Lnm and Syr, leinamycin and syringomycin proteins.

Cluster 11 contains 6 genes for putative regulatory proteins, four of which located within the DNA region showing homology to the *lnm* BGC (Table 1). The putative activators *lrgR1* and *lrgR2* were independently expressed into *S. argillaceus* wild type strain generating *S. argillaceus* WT-R1 and *S. argillaceus* WT-R2 strains respectively, and mutants in the putative repressors *lrgR3* (*S. argillaceus* ΔlrgR3) and in *lrgR4* (*S. argillaceus* ΔlrgR4) were generated (see Supporting Information). Analyses of the metabolite profiles of these strains did not show any differential peak (data not shown) indicating that overexpression and/or deletion of single regulatory genes was not sufficient to activate cluster 11; (iii) Media screening. Composition of culture media greatly influences production of secondary metabolites.[16] Twenty-nine culture media, including the LNM production media LF1 and LF2,[2] were screened for activating cluster 11 (see Supporting Information). In two of them (SM19 and SM30) several differential peaks were detected in *S. argillaceus* WT-R2, being their intensity higher with SM30 medium (Figure 3a). This result indicated that the newly biosynthesized compounds are produced by *S. argillaceus* only when *lrgR2* is overexpressed and the recombinant strain is cultivated in specific culture media. The involvement of cluster 11 in their production was confirmed by overexpressing *lrgR2* into ΔlrgG mutant, which showed that the key peaks detected in *S. argillaceus* WT-R2 were not detected in *S. argillaceus* ΔlrgG-R2 (Figure 3a), confirming that these compounds are encoded by cluster 11. Improvement of production (about 75%) of the major compound (peak **b** in Figure 3a) could be achieved substituting glucose by molasses from rice, and this modified medium (SM30a) was chosen for all experiments thereafter.

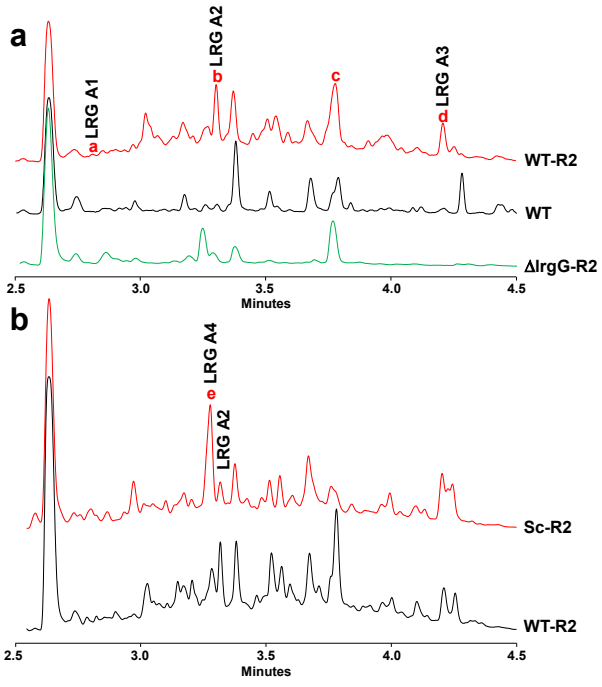


Figure 3. Production of largimycins by *S. argillaceus* and *S. canus*. UPLC chromatograms (330 nm) of ethyl acetate extracts of (a) *S. argillaceus* WT, *S. argillaceus* WT-R2, and *S. argillaceus* ΔlrgG-R2 strains cultivated in SM30; (b) *S. argillaceus* WT-R2 and *S. canus* WT-R2 (Sc-R2) strains cultivated in SM30a. Peaks with letters correspond to those largimycins (LRGs) selected for chemical characterization. Peaks for LRG A1, LRG A2, LRG A3 and LRG A4 are indicated.

The induced compounds in *S. argillaceus* WT-R2 were named largimycins (LRGs). The three main compounds detected in ethyl acetate extracts (peaks **b**-**d** in Figure 3a) were purified together with the compound from peak **a**, which turned out to be produced in higher amounts after solid-phase extraction of the scaled-up broth. Compounds from peaks **a** and **b** were named LRG A1 (**1**) and LRG A2 (**2**) respectively (Figure 4a). Their structures were elucidated (see Supporting Information) after high-resolution mass spectrometry (HRMS) and 1D (^{1}H) and 2D NMR spectroscopy, further assisted by comparison with the reported NMR data of known LNMs. Detailed analysis of the key correlations observed in the COSY and HMBC spectra (Figure 4b) combined with the determined molecular formulae rendered the full connectivity of **1** and **2** confirming the structural relation with known LNMs. The *Z* stereochemistry of the oxime double bond was established on the basis of the observed δ_C for C-18 which would be 2-4 ppm smaller in case of the *E* geometric isomers,[17] as indicated by comparison with model compounds and empirically-based prediction of ^{13}C NMR chemical shifts (Figure S9).[18,19] Their relative configuration was unambiguously determined based on the observed $^{3}J_{HH}$ and the key NOESY correlations (Figure 4c), assisted by molecular modelling employing 3D structural models generated after the reported X-ray structure of LNM E2 (Figure S9).[20] Finally, the absolute configurations at C-3, C-18 and the Cα of the S-conjugated N-acetyl-cysteine (CysNAc) units were assigned to be the same as those found in LNM, L-Thr and mycothiol respectively, on the basis of biosynthetic and phylogenetic arguments later explained,[7] thus providing the full absolute stereochemistry of **1** and **2** (Figure 4a). Although chemical instability issues hampered a complete spectroscopic characterization of compounds corresponding to peaks **c** and **d** (Figure 3a), fortunately in the case of compound from peak **d**, named LRG A3 (**3**), it was possible at least to unequivocally assign $C_{23}H_{23}ClN_2O_8S$ as its molecular formula (see Supporting Information). Based on this, its shared biosynthetic origin with **1** and **2** and the reported structure for LNM E4,[20] we have proposed a tentative chemical structure for **3** (Figure 4a).

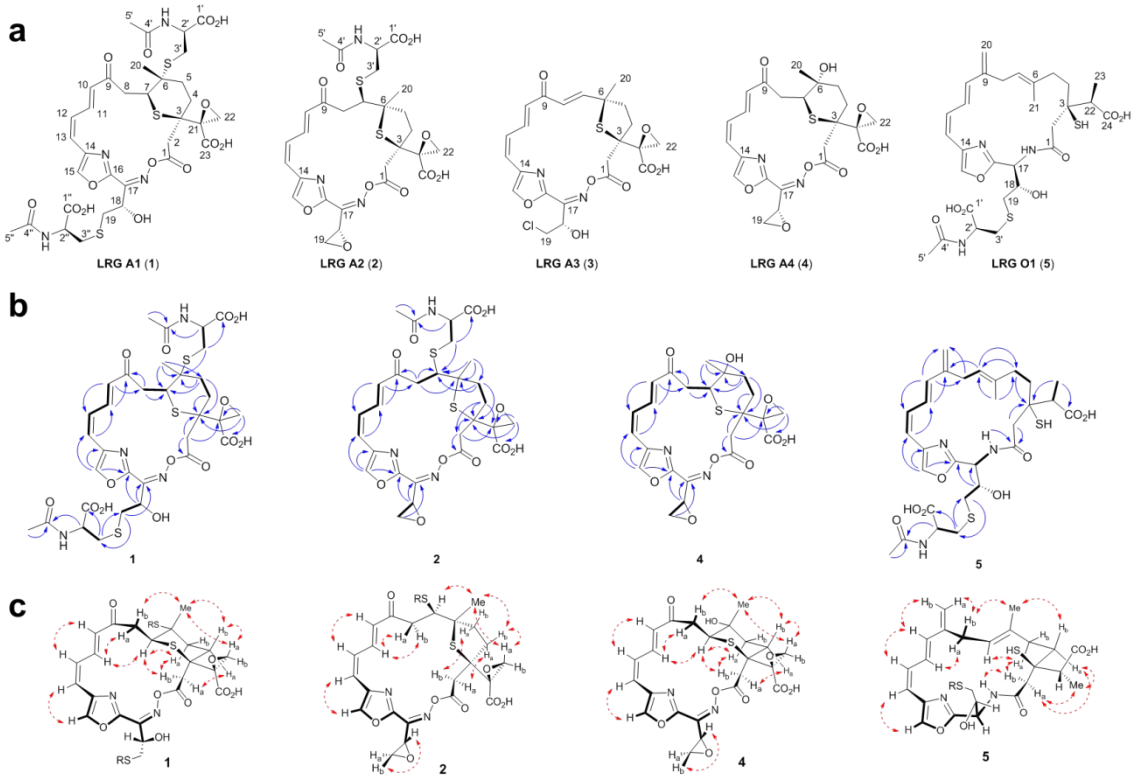


Figure 4. Structure of largymicins (LRGs). a) LRG A1 (**1**), A2 (**2**), A3 (**3**), A4 (**4**) and O1 (**5**). b) Key COSY correlations (bold bonds) and ^{1}H to ^{13}C HMBC correlations (blue arrows) determining the connectivity of LRGs. c) Key NOESY correlations (red arrows) employed alongside $^{3}J_{HH}$ to determine the relative configuration of LRGs.

LRGs thus show some structural similarity to three analogs of LNM (Figure 1a).[20] As LNMs E2 and E3, **1** contains a tetrahydrothiopyran ring embedded within the macrocycle while **2** contains a macrocycle-embedded tetrahydrothiophene ring, as found in LNM E4. However, LRGs display unique structural features compared to known LNMs: (i) they contain a 19-membered macrolactone ring closed through an unprecedented oxime ester bond instead of an 18-membered macrolactam ring. Very interestingly, natural oxime esters have been described previously only in some vibralactoximes of fungal origin,[21] and LRGs remarkably represent the first time that a macrocyclic natural product cyclized through an oxime ester is reported; (ii) they have an oxazole aromatic ring instead of a thiazole heterocycle; (iii) they contain one (in **2**) or two (in **1**) S-conjugated CysNAc moieties that are absent in LNMs; (iv) the C-3 side chain substituent contains an epoxide also absent in known LNMs; (v) the side chain at C-17 also differs from the displayed by known LNMs. Therefore, LRGs constitute a new group of oxazole-containing

macrocyclic oxime esters within the LNM family of natural products. Since cluster 11 was shown to encode LRGs it was renamed as cluster *lrg*.

Identification of *lrg* homologous BGCs in other strains. Characterization of LRG A4. Bioinformatics analysis of *lrg* gene products showed that Lrg proteins were highly similar (above 85% identity) to proteins encoded by clusters *M1013* from *Streptomyces* M1013, *scan* from *S. canus* ATCC 12647, and also to a recently identified *CB01373* cluster in *Streptomyces* CB01373 (Table 1).[7] All these clusters showed high synteny among them, being more remarkable that between *lrg* and clusters *M1013* and *scan* (Figure 2). Clusters *scan* and *CB01373* have been grouped into clade VII in a recent classification of LNM-type BGCs based on a phylogenetic analysis of the DUF–SH didomains.[7] According to these similarities, clusters *lrg* and *M1013* would constitute two new members of this clade. The phylogenetic classification of *lrg* automatically renders the C-3 absolute configuration of LRGs as being equal to that found for LNMs, since members of clade VII and clade I (to which the known LNMs producers belong) share the same stereochemistry at that chiral center.[7] Furthermore, members of clade VII are also predicted to select L-Thr as starting amino acid in the biosynthesis of their encoded compounds.[7] Since L-Thr is the precursor, as later explained, of the functional groups at C-18/C-19 observed in **1**-**3**, the absolute stereochemistry of C-18 in LRGs matches that of the hydroxylated methine in L-Thr.

In order to determine if some of those strains also produce LRGs, *S. canus* was cultivated under the same conditions mentioned above. Since no LRG were detected, the activator *lrgR2* was overexpressed using pEM4ATc-R2. Production of LRGs by the resultant recombinant strain *S. canus* WT-R2 was readily detected (Figure 3b). The main compound produced by this strain (corresponding to peak **e** in Figure 3b) was isolated and named LRG A4 (**4**). The structure of **4** was determined (see Supporting Information) after extensive HRMS and 1D (^{1}H) and 2D NMR analyses, likewise assisted by comparison with the NMR data of **1**, **2** and the reported for known LNMs. Its connectivity was established based on the determined molecular formula and the observed key correlations in the COSY and HMBC spectra (Figure 4b). The *Z* stereochemistry of the oxime double bond and the whole relative configuration were determined as described for **1** and **2** (Figures 4c and S8). Since the cluster *scan* and *lrg* belong to the same clade VII,[7] the absolute configurations of the chiral centers at C-3 and C-18 were assigned to be the same as **1**, thus providing the full absolute stereochemistry of **4** (Figure 4a).

Discovery of LRG A4 confirmed that cluster *scan* also encodes LRGs and suggests that clusters *M1013* and *CB01373* most probably also direct the biosynthesis of LRGs in their corresponding strains.

Limits of *lrg* biosynthesis gene cluster. Except *lrgR4*, genes between *lrgT1* and *lrgC3* in the *lrg* BGC are also preserved in *M1013* and *scan* clusters. Upstream *lrgT1* and downstream *lrgC3* this homology is lost (Table 1 and Figure 2). Therefore, these two genes were tentatively proposed as the limits of the *lrg* gene cluster. To confirm it, genes located at both ends of the cluster were sequentially inactivated, followed by overexpressing *lrgR2* into the generated mutants (and also into the ΔlrgR3 mutant mentioned above), and by analysis of LRGs production (see Supporting Information). Strains with deletions in *orf18*, *lrgT1*, *lrgR3, lrgP1* and *orf52* still produced LRGs, indicating that these genes were not essential for LRGs production in the tested conditions. However, production of **2** was abolished in mutants ΔlrgC1-R2 and ΔlrgC3-R2, indicating that *lrgC1* and *lrgC3* were essentials for LRGs production. According to these data and taking in consideration that genes *lrgT1*, *lrgR3* and *lrgP1* are also conserved in *M1013*, *scan* and in *CB01373* BGCs, we propose *lrgT1* as the left limit and *lrgC3* as the right limit of *lrg* BGC (Figure 2).

Analysis of the largimycin biosynthetic gene cluster and proposed biosynthesis pathway. The *lrg* cluster spans 69.81 kb and encompasses thirty three genes (Figure 2, Table 1). Based on the structure of LRGs and on bioinformatic analysis, a pathway for LRGs biosynthesis is proposed (Figure 5).

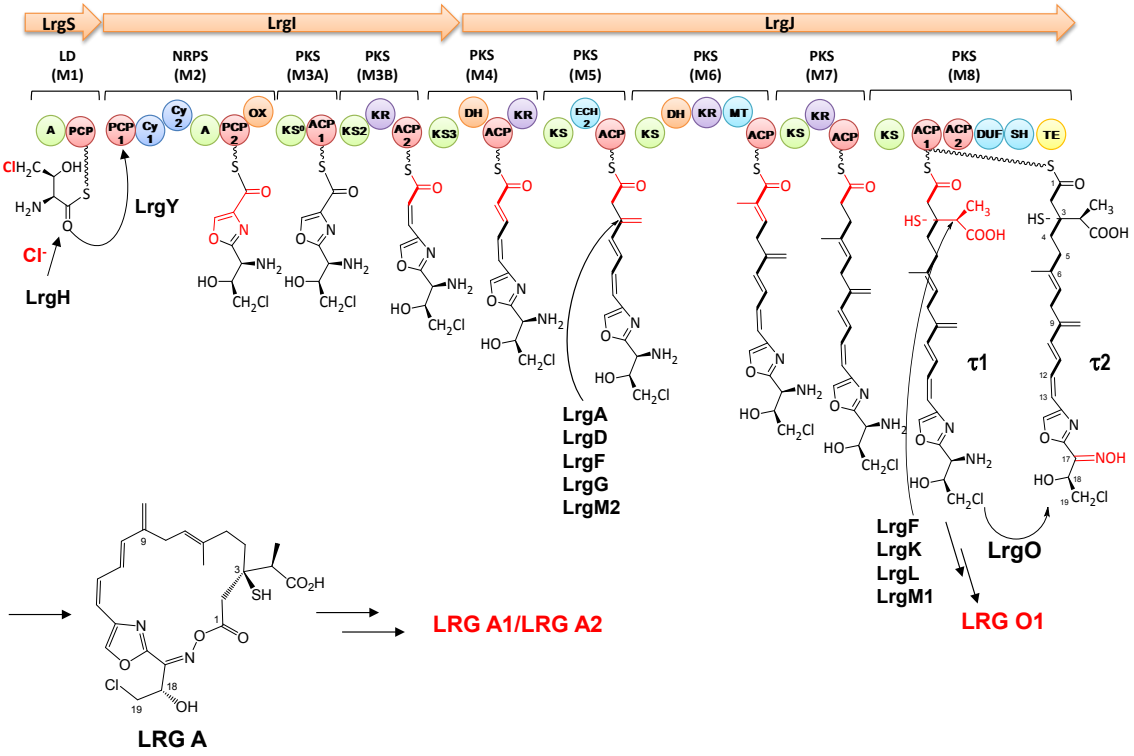


Figure 5. Proposed biosynthesis pathway for largimycins.

The biosynthesis would start by synthesizing the peptide moiety by the NRPS LrgI**.** This NRPS lacks a typical loading module. This role could be fulfilled by several discrete proteins (LrgS, LrgH and LrgY) that show similarity to SyrB1, SyrB2 and SyrC respectively, from the syringomycin BGC of *Pseudomonas syringae*.[22] LrgS is a didomain enzyme that contains an adenylation (A) domain at the N-terminus and a peptidyl carrier protein (PCP) domain at the C-terminus. The specificity-conferring codes for the A domain (DFWNVGMVH and LQTHFDFSVWEGNQVFGGEVNMYGITETTVHVTA),[23,24] predict L-Thr as substrate and are identical to those identified in WP_059300325.1 and in CB01373_Q from *scan* and *CB01373* BGCs, respectively. That prediction was confirmed for CB01373_Q by *in vitro* assays indicating that members of clade VII,[7] which also includes LrgS, recognize L-Thr as substrate, a specificity in agreement with the chemical structure of LRGs. LrgH shows similarity to nonheme Fe^{II} α-ketoglutarate- and O_2-dependent aliphatic halogenases such as SyrB2, which chlorinates L-Thr tethered to the PCP domain of SyrB1.[25] LrgY is a member of the α/β-hydrolase family to which SyrC belongs, which shuttles L-threonyl/4-Cl-L-threonyl moieties *in trans* between the PCP domain of SyrB1 and that located at the ninth module of SyrE NRPS.[26] According to these similarities LrgSHY would constitute the loading module (Module 1) of NRPS LrgI (Figure 5): LrgS would activate and load L-Thr; then, this amino acid tethered to the PCP domain of LrgS would be halogenated by LrgH and transferred by LrgY to the hybrid NRPS/PKS LrgI. Although **1**, **2** and **4** are not halogenated compounds, they contain a CysNAc-S-conjugate substituent at C-19 (**1**) or an epoxide at C-18/C-19 (**2** and **4),** which most likely derive from the corresponding halohydrin motif of the 4-Cl-L-threonyl precursor unit in the biosynthesis (Figure 6).

The identified molecular formula for **3** ($C_{23}H_{23}ClN_2O_8S$) indeed provides a direct evidence for halogenation during LRGs biosynthesis. Since **2** and **4** do not contain a chlorine atom but the chlorination of the tethered L-Thr is apparently required for the formation of the epoxide functional group observed in the starting amino acid unit, the halogenation event may be considered 'cryptic' within the overall biosynthetic pathway leading to these LRGs. Such a biosynthetic route resembles the reported 'cryptic' chlorination step catalyzed by the halogenase CmaB in the coronatine biosynthesis pathway, which is required for the formation of the cyclopropyl ring observed in the non-chlorinated isolated compound.[27] Several other BGCs related to *lnm* BGC and grouped into clades IV, V, VIII, XIV, XV and XVIII,[7] also encode a SyrB1, SyrB2 and SyrC like proteins, which could fulfill the same roles in their respective pathways than those herein proposed for LrgSHY.

The N-terminus of LrgI corresponds to a NRPS (Module 2, Figure 5). It contains a domain

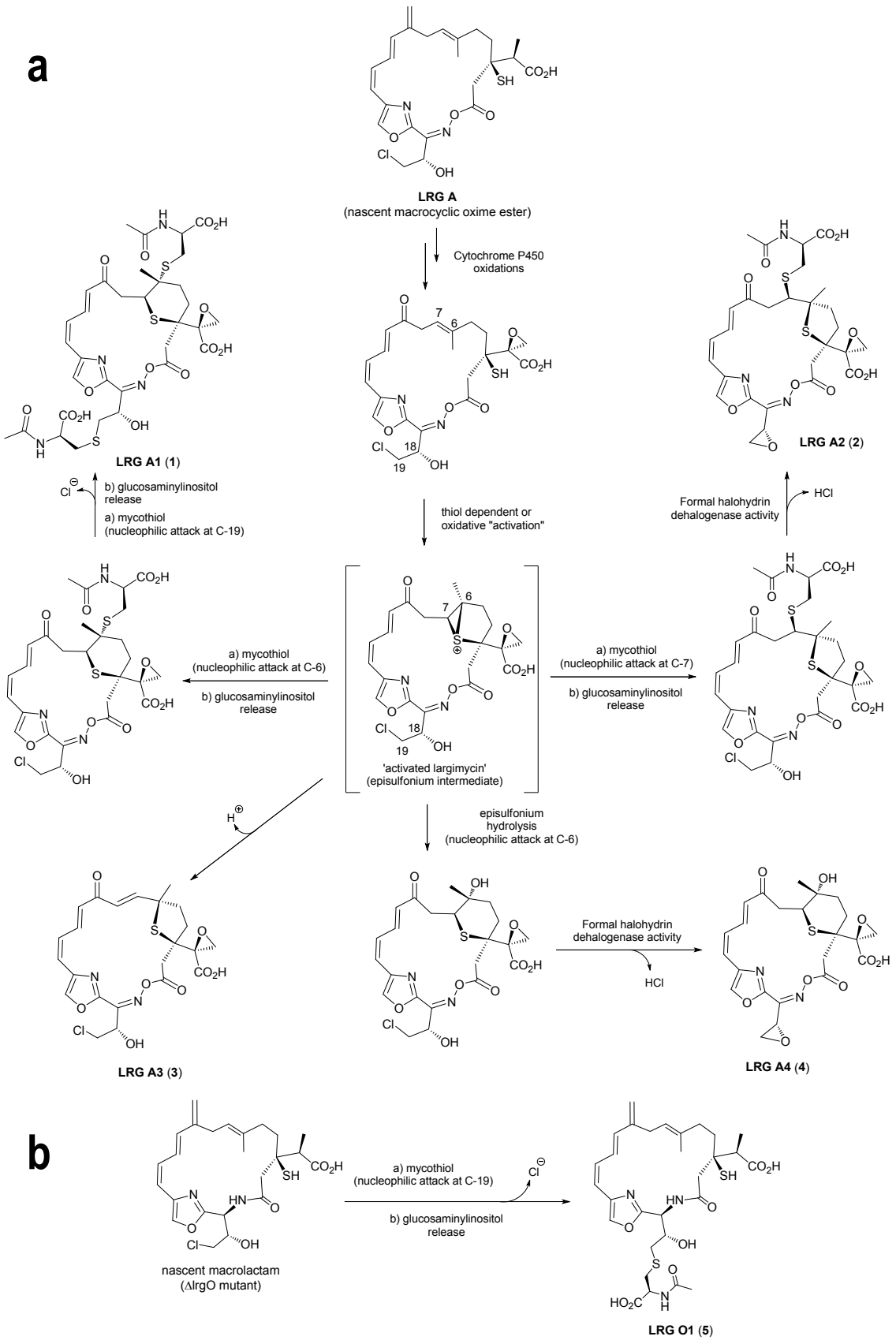


Figure 6. a) Proposed biosynthetic origin of S-conjugated *N*-acetyl-cysteine moieties and C-18/C-19 oxirane rings in largimycins involving the key episulfonium intermediate; b) Proposed biosynthetic origin of LRG O1.

organization (PCP1-Cy1-Cy2-A-PCP2-Ox) with two-tandem cyclization (Cy) and one oxidation (Ox) domain, found in some modules involved in thiazole and oxazole ring formation, such as the hybrid NRPS/AT-less Type I PKSs of *lnm*-type gene clusters.[7] Cy domains are responsible for the condensation, cyclization and dehydration events that lead to the formation of thiazoline and oxazoline rings. Tandem Cy domains are found in some NRPS such as the VibF from vibriobactin BGC, Cy2 carrying out the condensation step and Cy1 doing the cyclization/dehydration steps.[28] Ox domains catalyze the oxidation of thiazoline- and oxazoline-*S*-enzyme intermediates into the corresponding thiazole- and oxazole-*S*-enzymes.[29] The presence of this Ox domain in LrgI is consistent with the oxazol ring found in LRGs. The substrate specificity-conferring codes of LrgI A domain do not match any known A domain, but it is more similar to those recognizing L-Cys in LnmI and in other *lnm*-type BGCs. Noticeable, the LrgI A domain and its homologous counterparts in *scan*, *M1013*, and *CB01373* BGCs, share identical codes containing conserved amino acids that differ from those recognizing L-Cys (Table S2). Given the chemical structure of LRGs, the oxazole ring would be synthesized by the condensation of Thr and Ser residues. According to this, we propose that the A domains of LrgI and WP_059300322.1 from *scan* would specify for Ser.

The C-terminus of LrgI corresponds to an AT-less Type I PKS. This together with the AT-less Type I PKS LrgJ would constitute the PKS megasynthase responsible for the biosynthesis of the PK chain of LRGs. This PKS is organized into six modules (Figure 5): LrgI contains Modules 3A and 3B, and the β-ketoacyl synthase (KS) domain of Module 4 that is splitted between LrgI and LrgJ; LrgJ also contains Modules 5 to 8. The PKS LrgI shows a domain organization similar to LnmI from *lnm* BGC:[15] KS1-Acyl Carrier Protein (ACP)1-KS2-Ketoreductase (KR)-ACP2-KS3. KS2 and KS3 contain the canonical catalytic triad CHH that is necessary for decarboxylative condensation,[30] but KS1 has a mutated triad (CAH). In LnmI both KS1 and KS2 domains are essential for LNM biosynthesis but playing different roles: KS1 transfers the PCP-tethered peptidyl intermediate of NRPS Module 2 to the ACP1 of PKS Module 3, and KS2 catalyzes the decarboxylative condensation event between peptidyl-*S*-ACP1 and malonyl-*S*-ACP2.[31] Therefore, KS1-like domains correspond to domains in non-elongating modules, which are frequently found in AT-less Type I PKSs.[32] According to this, KS1 (renamed KS^0)-ACP1 of LrgI would constitute a non-elongating module (Module 3A), while KS2-KR-ACP2 would be the first elongating module (Module 3B) for LRGs PK biosynthesis (Figure 5). Like LnmI,[15] Module 3B of LrgI also lacks a Dehydratase (DH) domain that would be required for double bond formation between C-12 and C-13.

The AT-less Type I PKS LrgJ shows the following domain composition (Figure 5): Module 4 (DH-ACP-KR; the corresponding KS domain KS3 is located at LrgI), Module 5 (KS-ECH2-ACP), Module 6 (KS-DH-KR-MT-ACP), Module 7 (KS-KR-ACP) and Module 8 (KS-ACP1-ACP2-DUF-SH-TE). This organization is identical to that of PKS GnmT from the guangnanmycin BGC,[7] but differs from that of LnmJ at Modules 5 to 7:[15] LrgJ lacks the ACP1 domain at Module 6, the DH domain at Module 7 and it contains an Enoyl-CoA hydratase (ECH2) domain at Module 5. Members of the ECH2 family of enzymes catalyze a decarboxylation step to afford a β-alkylating intermediate during incorporation of β-branch groups to natural products.[32,33] They can exist as freestanding proteins or can be located as domains within larger multidomain proteins as it occurs in GnmT and LrgJ. The existence of this ECH2 domain at Module 5 together with the presence of genes for so called "HMG cassettes" (see below) in the *lrg* BGC suggests the incorporation of an olefinic exomethylene group at C-9 during the biosynthesis of LRGs. Module 6 contains a methyltransferase (MT) domain, which agrees with the presence of a methyl group substituent at C-6 in LRGs. Like all *lnm*-like BGCs, Module 8 contains a KS, two tandem ACP domains, a DUF domain, a SH domain and a thioesterase domain (TE). Runs of ACP domains are often found in modules involved in β-alkylation and are supposed to be an anchor point for the "HMG proteins" that are responsible for introducing a β-branch.[33] In a similar way as in LNM biosynthesis,[6] β-alkylation should occur at the C-3 position of the β-ketoacyl-*S*-ACP intermediate attached to Module 8 of LrgJ, and before the installation of the SH group by DUF-SH domains. These DUF-SH didomains are present in all BGCs for LNM family of compounds identified so far,[7] and are involved in sulfur incorporation into the PK backbone: the DUF domain adding L-Cys to the PK, and the SH domain catalyzing the C-S bond cleavage that originates the C-3 sulfur.[6] Phylogenetic analyses of DUF and SH domains from *lnm*-like BGCs showed they tend to cluster into two groups: one from pathways that use propionyl-*S*-ACP as the predicted substrate and the other from pathways using acetyl-*S*-ACP.[7] The LrgJ-DUF-SH didomain is more similar to those that putatively use propionyl-*S*-ACP as substrate, which is in accordance with the alkyl branch at C-3 displayed by LRGs. As already mentioned, the DUF-SH didomain in the *lrg* gene cluster is phylogenetically classified in clade VII, the same as the homologous didomains in the *scan* and *CB01373* clusters,[7] automatically indicating that LRGs and LNMs share the same absolute stereochemistry at C-3 since sulfur incorporation proceeds in both clades I (LNMs) and VII (LRGs) with the same stereoselectivity.

Biosynthesis of the PK chain requires an AT in each elongation step. However, neither LrgI nor LrgJ contain AT domains (AT-less PKSs). This function would be fulfilled by LrgG that shows high similarity to the *trans*-AT LnmG from LNM pathway and is predicted to accept malonyl-CoA

as substrate,[34] in agreement with the chemical structure of LRGs. The *lrg* BGC also contains a phosphopantetheinyl transferase encoding gene (*lrgQ*) that could be involved in transferring 4'-phosphopanteteine to ACPs of PKS and to PCPs of NRPS,[35] and a Type II TE (*lrgN*) that could play an editing role eliminating aberrant growing chains from the PKS.[36]

Considering all information mentioned above and taking into account the chemical structure of LRGs, a putative tethered full-length peptide-PK intermediate **τ1** would result from the action of LrgSHYGIJ, "HMG proteins" (see below) and some enzymatic activities required to generate a C12-C13 *Z* double bond and to generate a methylene at C-5. All *lnm*-like BGCs contain a Module 3B that also lack the DH domain, but even so all LNM-like compounds described so far show the C12-C13 *Z* double bond. Formation of the fully saturated ethylene unit corresponding to the C4-C5 bond would require the existence of a DH and an enoyl reductase (ER) domain at Module 7 of LrgJ, which are missing (Figure 5). Interestingly, GNMs also contain a saturated ethylene unit at those positions and similarly to LrgJ the corresponding module also lacks DH and ER domains.[7] Therefore, some of these functions could be supplied by the action of domains at other modules or provided *in trans* by unknown enzymes. Both features have been described in other AT-less PKSs.[32] Formation of the tethered intermediate **τ2** would be followed by its release through a cyclization, catalyzed by the TE domain of Module 8 of LrgJ, via an unprecedented ester bond involving the oxime hydroxy group, generating the putative nascent macrolactone intermediate LRG A (Figure 5). This cyclization would require a previous conversion of the amino terminal group of **τ1** into an oxime group to render the putative intermediate **τ2**. The *lrg* cluster contains *lrgO* that encodes a putative NAD(P)/FAD-dependent oxidoreductase similar to ClmM (44% identical amino acids) from the the collismycin BGC,[37] which is involved in the formation of an oxime functional group. To determine if LrgO plays a similar role in the formation of the oxime group of LRGs, its coding gene was inactivated (see Supporting Information). After expressing *lrgR2* into this mutant, cultivation of the resultant strain (ΔlrgO-R2) showed that previous LRGs production was abolished and some new peaks were detected (Figure 7).

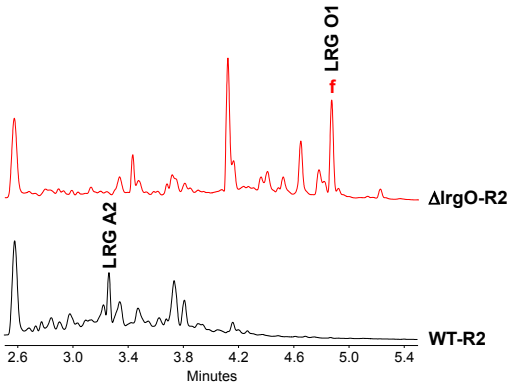


Figure 7. Production of largimycins by *S. argillaceus* ΔlrgO. UPLC chromatograms (300 nm) of ethyl acetate extracts of *S. argillaceus* WT-R2 and *S. argillaceus* ΔlrgO-R2 strains cultivated in SM30a. Peaks for LRG A2 and LRG O1 are indicated.

Compound from peak **f** was purified and chemically characterized. It corresponds to a novel compound named LRG O1 (**5**) (Figure 4a). The structure of **5** was likewise determined (see Supporting Information) after detailed HRMS and NMR spectroscopic analyses, further assisted by comparisons with the NMR data of **1**, **2**, **4**, LNM E1 and GNM B.[7,20] Its connectivity was derived from the determined molecular formula and the observed key correlations in the COSY and HMBC spectra (Figure 4b), while determination of the relative configuration relied on molecular modeling combined with analysis of the observed $^{3}J_{HH}$ and the key NOESY correlations (Figures 4c and S8). Based on its shared biosynthetic origin, the absolute configuration at C-3 and C-18 were assigned to be the same as **1**, while the membership of cluster *lrg* to the mentioned clade VII,[7] also allows assigning to C-22 the same absolute configuration determined for LNM E1,[20] and to C-17 the same stereochemistry as the α position of L-Thr, thus providing the full absolute stereochemistry of **5** (Figure 4a). Interestingly, **5** lacks the ketone at C-9 observed for **1**-**4** but rather displays an olefinic exomethylene group at that position, as described for GNMs and WSMs,[7] confirming the incorporation of this moiety during PK biosynthesis. Additionally, **5** also lacks the oxime functional group observed in **1**-**4**, thus being a macrolactam as are the known LNMs, GNMs and WSMs. This structural feature contrasts with the oxime ester closing of the macrocycle in **1**-**4** and ultimately demonstrates the essential role of LrgO in the oxidation of the amine group of the starting amino acid unit to render the oxime. Such oxidation before offloading the fully assembled tethered intermediate **τ2** by macrocyclization forming the oxime ester (Figure 5) resembles the case of the giant stambomycin macrolides,[38] where a cytochrome P450-catalyzed hydroxylation of the PK tethered chain is required for its release by macrolactonization in what was claimed to be a novel mechanism for macrolactone formation in PK antibiotic biosynthesis. In the case of **1**-**4** we are likewise dealing with a novel mechanism in secondary metabolism for offloading linear tethered precursors by macrocyclization via oxime ester formation. However, the macrolactam nature of **5** demonstrates that amine oxidation is not essential for successful LRG offloading by macrocyclization, as otherwise expected based on the structure of known LNMs. The structure of **5** (Figure 4a) is closely related to LNM E1 and GNM B (Figure 1),[7,20] being very similar to LRG A, the proposed nascent macrocyclic oxime ester intermediate of the LRG pathway (Figure 5).

Domain organization of Modules 5 and 8 suggests the incorporation of β-branches at C-9 and C-3 during PK elongation (Figure 5). β-branching requires a dedicated set of enzymes ("HMG cassettes") typically constituted by an ACP, an AT, a KS and a hydroxymethylglutaryl-CoA synthase (HCS), which introduce a β-branch into the growing PK chain that can be further dehydrated and decarboxylated by ECH proteins/domains.[33,39] The *lrg* BGC contains genes coding for discrete ACPs (*lrgL* and *lrgA*) and KS (*lrgD*), HCSs (*lrgM1* and *lrgM2*), a dehydratase ECH1 (*lrgF*) and a bifunctional AT/decarboxylase (DC) (*lrgK*) (Table 1). In addition, there is an ECH2 domain at Module 5. LrgKLM1 and LrgF are homologous to LnmKLM and LnmF respectively, from the *lnm* BGC,[15] which have been shown to participate in the attachment of a propionyl branch at C-3 during LNM biosynthesis.[5] A similar role could be fulfilled by LrgKLM1 and LrgF in the biosynthesis and attachment of the β-alkyl branch at C-3 of LRGs (Figure 5). The existence of additional genes encoding HCS (LrgM2), ACP (LrgA) and KS (LrgD) proteins, together with the presence of the ECH2 domain at Module 5, suggests the incorporation of an olefinic exomethylene group at C-9 (Figure 5). That group is present in **5** (Figure 4a). Intriguingly, **1** and **2** contain a keto group at that position, which suggests that the olefinic exomethylene group is modified after being incorporated to the PK chain and that the introduction of this group is a "cryptic" step in the biosynthetic pathway.

Several biosynthesis steps should occur between the putative nascent oxime macrolactone intermediate LRG A and the final compounds **1** and **2** (Figure 5). The main differences between **1** and **2** and the proposed nascent intermediate LRG A are the existence of a saturated single bond between C-6 and C-7; a keto group at C-9; an epoxide at the C-3 branch; a C-17 side chain containing an S-conjugated CysNAc moiety (**1**) or an epoxide (**2**); a tetrahydrothiopyran ring system between C-3 and C-7 (in **1**) or a tetrahydrothiophene heterocycle between C-3 and C-6 (**2**); and one CysNAc-S-conjugate substituent at C-6 (**1**) or at C-7 (**2**) (Figure 3a). Three cytochromes P450 encoding genes (*lrgC1*, *lrgC2* and *lrgC3*), identified within *lrg* cluster (Figure 2, Table 1), could be involved in epoxide formation at the C-3 side chain and in the conversion of the exomethylene group at C-9 into a keto group. Both epoxidation and dealkylation reactions have been shown to be carried out by cytochrome P450s.[40] In this work we have shown that *lrgC1* and *lrgC3* are required for LRGs biosynthesis (see above). The reported isolation of WSM A3, which contains an epoxide at C-9/C-22, alongside WSMs A1 and A2, which display an olefinic double bond at those positions,[7] suggests a possible oxidative dealkylation mechanism involving a first epoxidation step to explain the conversion of the olefinic exomethylene at C-9 into the corresponding ketone observed in **1**, **2** and **4**.

There are several genes (*lrgE*, *lrgP2*, *lrgW1*, *lrgW2*, *lrgX* and *lrgZ*) that have homolog counterparts in *lnm* BGC and related clusters, and some others (*lrgB* and *lrgP1*) that do not have homologs in *lnm* BGC but they are conserved in several *lnm*-type BGCs, including those for GNM and WSM biosynthesis.[7] This suggests that the encoded proteins might play common functions in the biosynthesis of this family of compounds. All of these *lnm* homologous genes have been shown to be essential for LNM production,[15] but only in the case of LnmE a possible role has been proposed in sulfur incorporation required to form the 1,3-dioxo-1,2-dithiolane moiety of LNM.[20] We have shown that inactivation of *lrgP1* does not affect LRG production (see above). It is tempting to propose that all or some of these genes might contribute to the formation of **1** and **2**, which could thus be envisaged deriving from LRG A via an oxidative mechanism involving a putative episulfonium ion intermediate, in analogy to LNM and LNM E1 as already also proposed for WSMs (Figure 6).[7,20]

LRGs **1** and **2** contain two and one S-conjugated CysNAc moieties, respectively. On the one hand, the CysNAc-S-conjugate substituent observed at C-6 or C-7 is proposed to be incorporated by nucleophilic attack of the CysNAc thiol group, from a mycothiol molecule, over those carbons (C-6 for **1** and C-7 for **2**) in the putative episulfonium ion intermediate derived from the nascent LRG A by an oxidation process (Figure 6). On the other hand, the CysNAc-S-conjugate substituent observed at C-19 in **1** and in **5** is proposed to be incorporated by nucleophilic attack of the same thiol group over the chlorinated C-19 in the halohydrin functionality of the starting 4-Cl-L-Thr unit (Figure 6). Thus, these LRGs would be shunt products generated by non-enzymatic events based on the inherent nucleophilicity of thiols, as it has been proposed for the biosynthesis of other compounds containing S-conjugated CysNAc adducts,[12,41] and could be considered derived from a LRG detoxification mechanism based on CysNAc S-alkylation by the episulfonium intermediate. This would provide an alternative mechanism for self-inmmunity in actinomycete producers of natural products of the LNM family additional to the inherent self-resistance based on the prodrug nature of LNM itself, claimed to be reductively activated in the presence of thiols after cellular uptake.[42] As suggested by Maier et al. (2014),[41] S-conjugated CysNAc most probably originates from mycothiol, the major low molecular weight thiol in actinomycetes, which plays important antioxidant and detoxification roles in these microorganisms.[43,44] After mycothiol conjugation with the corresponding metabolite precursor (episulfonium and chlorinated LRG), the amide bond between the cysteine and the GlcNAc moieties of mycothiol would be hydrolyzed by the action of an specialized amidase (Mca) releasing the corresponding mercapturic acids, the CysNAc-S-conjugated metabolites (**1**, **2**), plus glucosaminylinositol which is recycled for further mycothiol biosynthesis (Figures 6 and S9). This

biosynthetic origin of the CysNAc moieties automatically establishes the L absolute configuration at C_α of these amino acid units. Interestingly, the absence of any S-conjugated CysNAc substituent in **4**, the main largymycin produced by *S. canus*, suggests that the inherent cellular levels of mycothiol are much higher in *S. argillaceus*, something compatible with the variability of mycothiol content among actinomycetes.[44]

Bearing in mind the cryptic halogenation step of the tethered L-Thr starting unit catalyzed by the action of the LrgH halogenase, it can be proposed that formally a halohydrin dehalogenase enzymatic activity is responsible for the formation of the oxirane ring at C-18/C-19 observed in **2** and **4**. However, known halohydrin dehalogenases are examples of postindustrial evolution of bacterial enzymes to acquire activity toward the degradation of halogenated xenobiotics.[45] As expected, no homolog of halohydrin dehalogenases was found either in the *lrg* cluster nor elsewhere in *S. argillaceus* genome. Thus, the discovery of these oxirane-containing LRGs reveals an unprecedented type of halohydrin dehalogenase activity herein proposed for first time in secondary metabolite biosynthesis. The enzyme responsible must catalyze the formation of the epoxide by a mechanism formally equivalent, from the point of view of the substrate, to that of actual halohydrin dehalogenases but likely involving different type and arrangement of key residues for interaction with the halohydrin motif and deprotonation of the halohydrin hydroxyl group. At this stage, it cannot be defined whether such an enzyme acts over a tethered linear precursor or over released chlorinated macrocycles. However, the isolation of **1** and **5**, containing S-conjugated CysNAc at C-19, strongly suggests a freestanding chlorinated macrocycle as the most probable substrate. Future identification of the gene involved in the formation of the oxirane ring from its halohydrin precursor will provide very valuable information and the access to eventual investigations on the biotechnological potential of a new kind of halohydrin dehalogenase activity.

CONCLUSION

Largimycins, encoded by the silent *lrg* gene cluster in *S. argillaceus* and the homologous cryptic *scan* cluster in *S. canus,* are new and structurally unique members of the leinamycin family of natural products. These secondary metabolites contain an oxazol ring rather than the thiazole found in known LNMs and they are also the first naturally occurring macrocycles closed via an oxime ester bond ever reported. The flavin-dependent oxidoreductase LrgO responsible for the formation of the oxime group through which macrocyclization in LRGs takes place has been successfully identified shedding light on an unprecedented mechanism in secondary metabolism for offloading linear tethered precursors by oxime ester based macrocyclization. Very

interestingly, the proposed biosynthesis of LRGs involves two cryptic steps: i) chlorination of L-Thr by the NRPS loading module and ii) incorporation of an olefinic exomethylene in the growing PK chain. The oxirane ring observed in the starting amino acid unit of LRGs is proposed to be derived from a formal halohydrin dehalogenase activity over the cryptic 4-Cl-L-Thr moiety which remarkably represents the first time that such enzymatic transformation is proposed in secondary metabolism. Future work will be required to identify the intriguing protein displaying such formal halodydrin dehalogenase activity. That enzyme and the oxime forming *N*-oxidase LrgO herein identified deserve, in our view, further exploration based on a possible potential as biocatalysts. The discovery of cryptic LRGs reveals novel and creative biosynthetic avenues employed by Nature to enrich the structural diversity of the LNM family of natural products and provides exciting genes which expand the toolbox to generate new 'non-natural' analogues by combinatorial biosynthesis which would be very valuable to further investigate structure-activity relationship in the fascinating LNM family.

ASSOCIATED CONTENT

Supporting Information

This material is available free of charge via the internet.

Strains, culture conditions, plasmids and DNA manipulations; Generation of mutants; Plasmid constructs for gene expression; UPLC Analysis and Purification of largimycins; Spectroscopic analysis of largimycins and molecular modelling; Structure elucidation of largimycins; Alignments of Lrg proteins with Lnm and Syr proteins; Confirmation of *Streptomyces argillaceus* mutants; UPLC analysis of *S. argillaceus* WT-R2 cultivated in different media; UPLC analyses of extracts from *S. argillaceus* mutants; Spectroscopic data of LRG A1-A4 and LRG O1; Energy-minimized molecular models of LRG A1-A4 and LRG O1; Determination of the oxime double bond stereochemistry; Mechanism of formation of a CysNAc-S-conjugate after nucleophilic attack of mycothiol over an alkyl chloride; Oligonucleotides used for PCR; Comparison of specifity-conferring codes of adenilation domains of hybrid NRPS/PKS from leinamycin (*lnm*)-type gene clusters; ^{1}H NMR and ^{13}C NMR data of LRG A1, LRG A2, LRG A4 and LRG O1.

Accession Codes

The sequence of *Streptomyces argillaceus lrg* gene cluster has been deposited at European Nucleotide Archive (EBI-ENA) under the accession number LR131959.1 and at MIBiG under the accession number BGC0001853.

AUTHOR INFORMATION

Corresponding author

*E-mail: cmendezf@uniovi.es

ORCID

Carmen Méndez: 0000-0003-2729-841X

Ignacio Pérez-Victoria: 0000-0002-4556-688X

Jesús Martín: 0000-0001-7487-2790

Fernando Reyes: 0000-0003-1607-5106

AUTHOR CONTRIBUTIONS

C.M. and J.A.S. conceived and designed the project; A.B. and S.Y., conducted experiments; A.B. and A.F.B. carried out compound purifications; I.P.V., J.M. and F.R. performed chemical characterization of compounds; C.M. wrote the manuscript and I.P.V. contributed to preparing the final version of it. All authors read and approved the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by grants from the Spanish Ministry of Economy and Competitiveness, MINECO (Grants BIO2014-56752-R and PIM2010EEI-00752 to C.M.), and by the grant "Apoyo a grupos de excelencia", Principado de Asturias-FEDER (FC-15-GRUPIN14-014). A.B. and S.Y. were recipient of predoctoral fellowships from MINECO. The NMR spectrometer used in this work was purchased via a grant for scientific and technological infrastructure from the Ministerio de Ciencia e Innovación (PCT-010000-2010-4). We thank C. Olano for helping in managing DNA sequences.

REFERENCES

(1) Newman, D.J. and Cragg, G.M. (2016). Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod. 79*, 629–661.

(2) Hara, M., Asano, K., Kawamoto, I., Takiguchi, T., Katsumata, S., Takahashi, K., and Nakano, H. (1989). Leinamycin, a new antitumor antibiotic from *Streptomyces*: producing organism, fermentation and isolation. *J. Antibiot. (Tokyo) 42*, 1768-17674.

(3) Viswesh, V., Gates, K., and Sun, D. (2010). Characterization of DNA damage induced by a natural product antitumor antibiotic leinamycin in human cancer cells. *Chem. Res, Toxicol.* 23, 99-107.

(4) Cheng, Y.Q., Tang, G.L., and Shen, B. (2003). Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. U. S. A*. *100*, 3149-3154.

(5) Huang, Y., Huang, S.X., Ju, J., Tang, G., Liu, T., and Shen, B. (2011). Characterization of the *InmKLM* genes unveiling key intermediates for β-alkylation in leinamycin biosynthesis. *Org. Lett. 13*, 498-501.

(6) Ma, M., Lohman, J.R., Liu, T., and Shen, B. (2015). C-S bond cleavage by a polyketide synthase domain. *Proc. Natl. Acad. Sci. U. S. A. 112*, 10359-10364.

(7) Pan, G., Xu, Z., Guo, Z., Hindra, Ma, M., Yang, D., Zhou, H., Gansemans, Y., Zhu, X., Huang, Y., Zhao, .LX., Jiang, Y., Cheng, J., Van Nieuwerburgh, F., Suh, J.W., Duan, Y., and Shen**,** B. (2017). Discovery of the leinamycin family of natural products by mining actinobacterial genomes. *Proc. Natl. Acad. Sci. U. S. A. 114*, E11131-E11140.

(8) Nett, M., Ikeda, H., and Moore, B.S. (2009). Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep. 26*, 1362-1384.

(9) Olano, C., Méndez, C., and Salas, J.A. (2014). Strategies for the design and discovery of novel antibiotics using genetic engineering and genome mining. In Antimicrobial compounds: current strategies and new alternatives, T.G. Villa and P. Veiga-Crespo, eds. (Berlin, Springer). pp. 1-25.

(10) Rutledge, P.J. and Challis, G.L. (2015). Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol*. *13*, 509-523.

(11) Sekurova, O.N., Schneider, O., and Zotchev, S.B. (2019). Novel bioactive natural products from bacteria via bioprospecting, genome mining and metabolic engineering. *Microb. Biotechnol*. *12*, 828-844.

(12) Ye, S., Molloy, B., Braña, A.F., Zabala, D., Olano, C., Cortés, J., Morís, F., Salas, J.A., and Méndez, C. (2017). Identification by genome mining of a type I polyketide gene cluster from *Streptomyces argillaceus* involved in the biosynthesis of pyridine and piperidine alkaloids argimycins P. *Front. Microbiol*. *8*, 194.

(13) Méndez, C., González-Sabín, J., Morís, F., and Salas, J.A. (2015). Expanding the chemical diversity of the antitumoral compound mithramycin by combinatorial biosynthesis and biocatalysis: the quest for mithralogs with improved therapeutic window. *Planta Med. 81*, 1326-1338.

(14) Becerril, A., Álvarez, S., Braña, A.F., Rico, S., Díaz, M., Santamaría, R.I., Salas, J.A., and Méndez, C. (2018). Uncovering production of specialized metabolites by *Streptomyces argillaceus*: activation of cryptic biosynthesis gene clusters using nutritional and genetic approaches. *PLoS One 13*, e0198145.

(15) Tang, G.L., Cheng, Y.Q., and Shen, B. (2004). Leinamycin biosynthesis revealing unprecedented architectural complexity for a hybrid polyketide synthase and nonribosomal peptide synthetase. *Chem. Biol. 11*, 33-45.

(16) Bode, H.B., Bethe, B., Hofs, R., and Zeeck, A. (2002). Big effects from small changes: possible ways to explore nature's chemical diversity. *Chembiochem 3*, 619-627.

(17) Hawkes, G.E., Herwig, K., and Roberts, J.D. (1974). Nuclear magnetic resonance spectroscopy. Use of carbon-13 spectra to establish configurations of oximes. *J. Org. Chem*. *39*, 1017-1028.

(18) Kajiro, H., Mitamura, S., Mori, A., and Hiyama, T. (1999). A practical synthesis of (1*S*,2*R*)-1-amino-2-indanol, a key component of an HIV protease inhibitor, indinavir. *Bull. Chem. Soc. Jpn*. *72*, 1093-1100.

(19) Elyashberg, M.E., Blinov, K.A., and Williams, A.J. (2009). The application of empirical methods of ^{13}C NMR chemical shift prediction as a filter for determining possible relative stereochemistry. *Magn. Reson. Chem*. *47*, 333-341.

(20) Huang, S.X., Yun, B.S., Ma, M., Basu, H.S., Church, D.R., Ingenhorst, G., Huang, Y., Yang, D., Lohman, J.R., Tang, G.L., Ju, J., Liu, T., Wilding, G., and Shen**,** B. (2015). Leinamycin E1 acting as an anticancer prodrug activated by reactive oxygen species. *Proc. Natl. Acad. Sci. U. S. A. 112*, 8278-8283.

(21) Chen, H.P., Zhao, Z.Z., Li, Z.H., Dong, Z.J., Wei, K., Bai, X., Zhang, L., Wen, C.N., Feng, T., and Liu, J.K. (2016). Novel natural oximes and oxime esters with a vibralactone backbone from the basidiomycete *Boreostereum vibrans*. *ChemistryOpen 5*, 142-149.

(22) Zhang, J.H., Quigley, N.B., and Gross, D.C. (1995). Analysis of the *syrB* and *syrC* genes of *Pseudomonas syringae* pv. *syringae* indicates that syringomycin is synthesized by a thiotemplate mechanism. *J. Bacteriol*. *177*, 4009-4020.

(23) Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., and Huson, D.H. (2005). Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res*. *33*, 5799-5808.

(24) Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol*. *6*, 493-505.

(25) Vaillancourt, F.H., Yin, J., and Walsh, C.T. (2005). SyrB2 in syringomycin E biosynthesis is a nonheme Fe^{II} alpha-ketoglutarate- and O2-dependent halogenase. *Proc. Natl. Acad. Sci. U. S. A. 102*, 10111-10116.

(26) Singh, G.M., Vaillancourt, F.H., Yin, J., and Walsh, C.T. (2007). Characterization of SyrC, an aminoacyltransferase shuttling threonyl and chlorothreonyl residues in the syringomycin biosynthetic assembly line. *Chem. Biol. 14*, 31-40.

(27) Vaillancourt, F.H., Yeh, E., Vosburg, D.A., O'Connor, S.E., and Walsh, C.T. (2005). Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature 436*, 1191-1194.

(28) Marshall, C.G., Hillson, N.J., and Walsh, C.T. (2002). Catalytic mapping of the vibriobactin biosynthetic enzyme VibF. *Biochemistry 41*, 244-250.

(29) 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.

(30) Xu, W., Qiao, K., and Tang, Y. (2013). Structural analysis of protein–protein interactions in type I polyketide synthases. *Crit. Rev. Biochem. Mol. Biol*. *48*, 98–122.

(31) Huang, Y., Tang, G.L., Pan, G., Chang, C.Y., and Shen, B. (2016). Characterization of the ketosynthase and acyl carrier protein domains at the LnmI nonribosomal peptide synthetase-polyketide synthase interface for leinamycin biosynthesis. *Org. Lett*. *18*, 4288-4291.

(32) Helfrich, E.J. and Piel, J. (2016). Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep*. *33*, 231-316.

(33) Calderone, C.T., Kowtoniuk, W.E., Kelleher, N.L., Walsh, C.T., and Dorrestein, P.C. (2006). Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-*S*-carrier proteins in the pksX pathway of *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A*. *103*, 8977-8982.

(34) Yadav, G., Gokhale, R.S., and Mohanty, D. (2003). Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol*. *328*, 335–363.

(35) Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C., and Walsh, C.T. (1996). A new enzyme superfamily- the phosphopantetheinyl transferases. *Chem. Biol*. *3*, 923-936.

(36) Heathcote, M.L., Staunton, J., and Leadlay, P.F. (2001). Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem. Biol. 8*, 207–220.

(37) Garcia, I., Vior, N.M., González-Sabín, J., Braña, A.F., Rohr, J., Moris, F., Méndez, C., and Salas, J.A. (2013). Engineering the biosynthesis of the polyketide-nonribosomal peptide collismycin A for generation of analogs with neuroprotective activity. *Chem. Biol*. *20*, 1022-1032.

(38) Song, L., Laureti, L., Corre, C., Leblond, P., Aigle, B., and Challis, G.L. (2014). Cytochrome P450-mediated hydroxylation is required for polyketide macrolactonization in stambomycin biosynthesis. *J Antibiot (Tokyo). 67*, 71-76.

(39) Simunovic, V. and Müller, R. (2007). 3-hydroxy-3-methylglutaryl-CoA-like synthases direct the formation of methyl and ethyl side groups in the biosynthesis of the antibiotic myxovirescin A. *Chembiochem 8*, 497-500.

(40) Rudolf, J.D., Chang, C.Y., Ma, M., and Shen, B. (2017). Cytochromes P450 for natural product biosynthesis in Streptomyces: sequence, structure, and function. *Nat. Prod. Rep. 34*, 1141-1172.

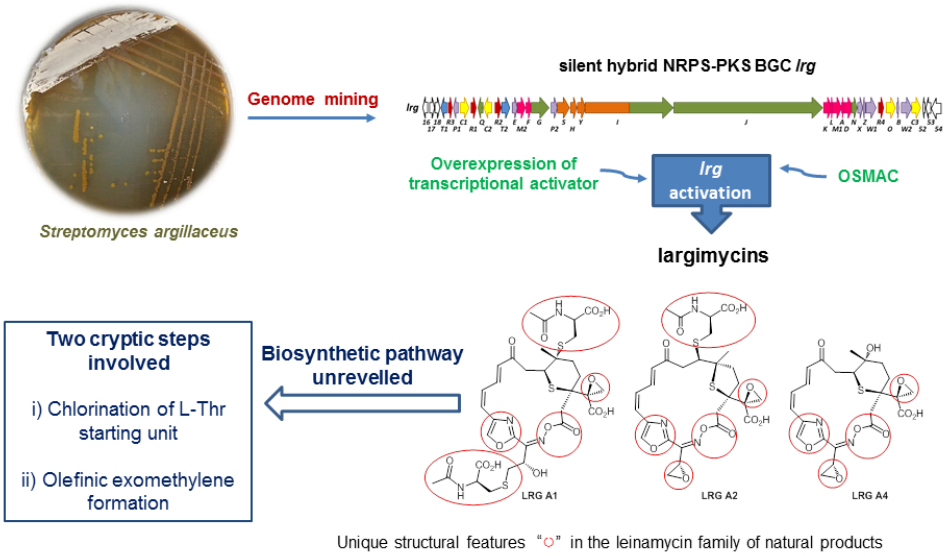
(41) Maier, S., Pflüger, T., Loesgen, S., Asmus, K., Brötz, E., Paululat, T., Zeeck, A., Andrade, S., and Bechthold, A. (2014). Insights into the bioactivity of mensacarcin and epoxide formation by MsnO8. *Chembiochem 15*, 749-756.

(42) Asai, A., Hara, M., Kakita, S., Kanda, Y., Yoshida, M., Saito, H., and Saitoh, Y. (1996). Thiol-mediated DNA alkylation by the novel antitumor antibiotic leinamycin. *J. Am. Chem. Soc*. *118*, 6802-6803.

(43) Imber, M., Pietrzyk-Brzezinska, A.J., and Antelmann, H. (2019). Redox regulation by reversible protein S-thiolation in Gram-positive bacteria. *Redox Biol. 20*, 130-145.

(44) Jothivasan, V.K., and Hamilton, C.J. (2008). Mycothiol: synthesis, biosynthesis and biological functions of the major low molecular weight thiol in actinomycetes. *Nat. Prod. Rep*. *25*, 1091-1117.

(45) Schallmey, M., Floor, R.J., Szymanski, W., and Janssen, D.B. (2012). Hydrolysis and reverse hydrolysis: halohydrin dehalogenases. In Comprehensive chirality, Carreira, E.M. and Yamamoto, H. eds. (Amsterdam, Elsevier), pp. 143–155.



Unique structural features "○" in the leinamycin family of natural products

254x190mm (96 x 96 DPI)