



## 非核糖体多肽 Surugamides 生物合成基因簇镶嵌式结构的解析

许春敏<sup>1</sup>, 林霄<sup>3</sup>, 李蕾<sup>2</sup>, 王淑萍<sup>2</sup>, 杨帆<sup>2</sup>, 徐石海<sup>3</sup>, 邓海<sup>2</sup>, 周永军<sup>2</sup>, 林厚文<sup>1,2\*</sup>

<sup>1</sup> 江西中医药大学, 江西 南昌 330004

<sup>2</sup> 上海交通大学医学院附属仁济医院药学部, 上海 200127

<sup>3</sup> 暨南大学药学院, 广东 广州 510632

**摘要:**【目的】多肽化合物 Surugamides (sgm)生物合成基因簇包含 4 个非核糖体多肽合酶(NRPS)基因 *surA–D*, 负责 2 个 NRPS 生物合成途径。已有报道确认 *surA* 基因与 Surugamide A 产物相关, 而 *surB* 基因与 sgm F 产物相关, 但对 *surC* 和 *surD* 基因功能的归属尚没有实验证据。本工作拟在之前研究的基础上进一步确认 *surA* 和 *surD* 负责 Surugamide A 产物生物合成, 为基因工程改造 Surugamides 生物合成途径以及研究其 NRPS 蛋白之间的识别机制提供理论依据。【方法】从海绵中分离放线菌并通过 16S rRNA 基因序列比对分析其分类单元。通过在线数据库 antiSMASH 分析基因组序列, 发现天然产物生物合成基因簇。通过 UPLC-Q-TOF-MS 和 <sup>13</sup>C NMR 鉴定化合物结构。把构建完成的同源重组双交换质粒导入链霉菌宿主后筛选基因缺失或替换突变株。【结果】从胃甲海绵来源链霉菌 *S. albidoflavus* LHW3101 基因组中发现了 Surugamides 生物合成基因簇, 确认了该菌株发酵产物中的化合物 sgm A 和 sgm F。构建了 *surB* 和 *surC* 基因同时缺失的突变株 RJ9, 发现 RJ9 不再产 sgm F 而仍然产 Surugamide A。在缺失突变 *surB* 和 *surC* 基因的同时在 *surD* 基因前引入了组成型强启动子 *ermEp\**, 结果发现 RJ9 产 Surugamide A 水平是野生型菌株的约 2 倍。【结论】确认了 *surB* 和 *surC* 基因与 sgm A 产物无关。在 *surD* 基因前引入强启动子后显著提高了 Surugamide A 的产量, 提示 *surD* 基因与 sgm A 产物相关, 结合已报到 *surA* 基因与 Surugamide A 产物相关的证据, 进一步确认了 *surA* 和 *surD* 基因负责 Surugamide A 生物合成的推论。

**关键词:** 微生物天然产物, 非核糖体多肽合酶, 生物合成基因簇

微生物来源非核糖体多肽(nonribosomal peptide, NRP)类化合物是重要的药物发现源头, 例如目前临床使用的免疫抑制剂环孢霉素(cyclosporin)、抗生素万古霉素(vancomycin)、达

基金项目: 国家自然科学基金(31670096, 31628001, 31600014)

\*通信作者。Tel/Fax: +86-21-68383339; E-mail: franklin67@126.com

收稿日期: 2018-01-23; 修回日期: 2018-04-01; 网络出版日期: 2018-04-20

托霉素(daptomycin)和青霉素(penicillin)以及 2015 年发现的抑制所有金黄色葡萄球菌和结核分支杆菌变种的新型抗生素 teixobactin 均属于微生物来源的 NRP 类化合物<sup>[1-2]</sup>。微生物来源 NRP 类化合物是通过非核糖体肽合成酶<sup>[3]</sup>(nonribosomal peptide synthetase, NRPS)介导生物合成的<sup>[3]</sup>。在 NRPS 生物合成过程中酶催化功能域模块化和装配线排列的特点使得对其工程化改造和目标产物预测成为可能<sup>[4]</sup>。NRPS 含有的多个延伸模块(module)呈装配线模式依次传递延伸中的肽链,每个模块含有基本的功能域负责肽链合成,包括合成肽键的缩合功能域(condensation)、活化并加载氨基酸的腺苷酰化功能域(adenylation)以及肽酰基载体蛋白(peptide carrier protein/thiolation),在 NRPS 装配线末端通常含有负责肽链释放的硫酯酶<sup>[5]</sup>。此外在延伸模块中还可能存在一些非必需的修饰功能域,如负责氨基酸构型转化的差向异构酶(epimerase)功能域和在氨基上引入甲基的氮-甲基转移酶(N-methyltransferase)功能域。相比于化学合成多肽化合物需要昂贵的基因保护试剂和有毒溶剂,微生物生物合成多肽仅需要廉价的发酵培养基,可规模化生产并降低对环境的污染,因此发现并研究微生物来源 NRPS 生物合成途径对开发药用活性多肽类化合物具有重要价值。

我们从胃甲海绵(*Dactylospongia elegans*)中分离获得了链霉菌 *S. albidoflavus* LHW3101,从其发酵产物中发现了环八肽化合物 Surugamide A (sgm A)和链状十肽化合物 Surugamide F (sgm F),基因组分析发现了该化合物对应的 NRPS 类型生物合成基因簇。该基因簇包含 4 个 NRPS 基因 A-D,前期研究基于缺失突变 A 和 B 基因的结果,推测 A 和 D 基因负责 sgm A 的合成,而位于 A 和 D 基因之间的 B 和 C 基因负责 sgm F 的合成<sup>[6-7]</sup>。

本研究通过构建 B 和 C 基因的缺失突变株以及为 D 基因引入强启动子,进一步明确了 sgm A 生物合成的必需基因,为基因工程改造 sgm 生物合成途径以及研究其 NRPS 蛋白之间的识别机制提供了研究基础。

## 1 材料和方法

### 1.1 材料

**1.1.1 主要试剂和仪器:** 超保真 DNA 聚合酶为 High-Fidelity PCR Master Mix(NEB Phusion<sup>®</sup>), 普通 DNA 聚合酶为 2×fast *Taq* PCR Master Mix (上海柏根生物), 质粒提取试剂盒购自上海捷瑞生物工程有限公司, PCR 片段回收试剂盒为 Zymoclean Gel DNA Recovery Kit, 限制性内切酶均为 NEB 生产, U-Clone Master Mix for DNA Cloning 试剂盒购自百诺威科技(武汉)有限公司, 抗生素均购自生工生物工程(上海)股份有限公司, 色谱级乙腈和甲醇为德国默克公司产品, 分析纯试剂乙酸乙酯和甲醇购自上海化学试剂有限公司。液质联用(LC-MS)为 Waters HPLC-ACQUITY QDa, 高分辨质谱为 Waters UPLC-QTOF-MS/MS, 中压色谱系统购自三泰科技(常州)有限公司。振荡摇床(ZQZY-70BS)为上海知楚仪器有限公司生产。

**1.1.2 培养基:** ISP2(g/L): 酵母提取物 5, 麦芽提取物 2, 葡萄糖 2, 海盐 18。M1(g/L): 淀粉 10, 酵母提取物 4, 蛋白胨 2, 琼脂粉 18, 磷酸二氢钾 2, 海盐 18, 25 μg/mL 萘啶酮酸, 50 μg/mL 放线菌酮。LB (g/L): 胰蛋白胨 10, 酵母提取物 5, 氯化钠 5, 葡萄糖 2。SFM (g/L): 低温黄豆饼粉 20, 甘露醇 20, 琼脂粉 20, pH 7.2, MgCl<sub>2</sub> 10 mmol/L。TSBY (g/L): 大豆胰蛋白胨 30, 酵母提取物 5, 蔗糖 103, 海盐 18, pH 7.2。

**1.1.3 菌株和质粒:** *Streptomyces albidoflavus* LHW3101 由本实验室分离获得, *Streptomyces albus* J1074 为通用模式菌株, RJ9 为 *S. albidoflavus* LHW3101 来源的基因缺失突变株。 *E. coli* DH10B 用于质粒构建, *E. coli* ET12567/pUZ8002 用于双亲结合转移。质粒 pRJ26 为本工作构建, 用于获得基因缺失突变株 RJ9。质粒 pYH7 携带 apramycin(apr) 抗性(Apr<sup>R</sup>)基因, 是大肠杆菌-放线菌穿梭质粒, 用于构建基因缺失突变株<sup>[8]</sup>。质粒 pIB139 用于 PCR 模板, 扩增组成型强启动子 *ermEp*<sup>\*[9]</sup>。

**1.1.4 PCR 引物和模板:** 用于 PCR 扩增同源重组交换臂的引物为: L-surBC-S (5'-GACTGATCAAG GCGAATACTTCAGAGGCTGGCGTACGTCCTG TCC-3'), L-surBC-A (5'-ACACTCGCATGCATACT AGAGAATCCCCCTGGGGCTGGTACCTG-3'), R-surBC-S (5'-TGCCGGTTGGTAGGATCCACATAT GACCACGCCCTCGTCGCGGGA-3'), *ermP-sur-S* (5'-AGGGGGATTCTCTAGTATGCATGCGAGTGT C-3')和 *ermP-sur-A* (5'-TGGTCATATGTGGATCCT ACCAACCGGCAC-3')。验证突变株 RJ9 的引物为 *surBC-2S* (5'-ATGTGGTGCATCGACAGCAGCAG-3') 和 *surBC-2A* (5'-TCAACTCCAACGGCAAGGTG AC-3')以及 *ermP-sur-S* 与 *ermP-sur-A*。用于菌种鉴定 16S rRNA 基因扩增的引物为 27F (5'-AGAGTT TGATCCTGGCTCAG-3'), 1492R (5'-GGTTACCTT GTTACGACTT-3')。用于筛选突变株的基因组 DNA 快速提取方法为: 用牙签从单菌落上挑取菌丝体于 PCR 管, 加入 20  $\mu$ L 10% Chelex<sup>®</sup> 100 Resin (Bio-Rad), 100  $^{\circ}$ C 加热 20 min, 取上清 1  $\mu$ L 作为 PCR 模板。

## 1.2 海绵共附生放线菌分离和鉴定

取胃甲海绵 1 cm $\times$ 1 cm 方块, 用 1.2% 无菌海盐水洗 3 次后充分研磨后加入 10 mL 无菌海盐

水得到样品液, 依次做 10<sup>-1</sup>、10<sup>-2</sup> 和 10<sup>-3</sup> 梯度稀释, 各取 200  $\mu$ L 分别涂布于 M1 培养基平板上, 28  $^{\circ}$ C 培养 14 d, 选取放线菌生长表型的菌落在 ISP2 平板上划线获得单菌落, 从单菌落上取菌丝体提取基因组 DNA 作为模板, 用引物 27F 和 1492R 扩增 16S rRNA 基因, 通过琼脂糖胶电泳分离和回收 1.6 kb 的 PCR 片段并进行测序, 序列输入在线数据库 Ezbiocloud 分析菌株与已发现菌株的相似度。

## 1.3 克隆构建

以 LHW3101 菌株基因组 DNA 为模板 PCR 扩增同源交换臂, 用引物 L-surA-S 和 L-surA-A 扩增获得 1657 bp 左臂, 用引物 R-surA-S 和 R-surA-A 扩增获得 1638 bp 右臂。以质粒 pIB139 为模板, PCR 扩增包含组成型强启动子 *ermEp*<sup>\*</sup> 的 212 bp 片段, 引物为 *ermP-sur-S* 和 *ermP-sur-A*。用 *EcoR* I 和 *Hind* III 双酶切质粒 pYH7 后, 通过琼脂糖凝胶电泳回收 7086 bp 载体片段。通过 Gibson DNA assembly 的方法<sup>[10]</sup>连接以上 4 个片段。反应条件为: 5  $\mu$ L U-Clone Master Mix、5  $\mu$ L DNA 片段(包括约 50  $\mu$ g 载体片段, 每个插入片段 100–200  $\mu$ g)、50  $^{\circ}$ C 反应 60 min。反应结束后取 5  $\mu$ L 反应液通过钙转化转入 *E. coli* DH10B, 通过菌落 PCR 用引物 *surBC-2S* 和 *surBC-2A* 筛选目标质粒 pRJ26 的克隆。

## 1.4 基因缺失突变株筛选

先将质粒 pRJ26 通过钙转化导入 *E. coli* ET12567/pUZ8002 获得用于双亲接合转移的大肠杆菌。将大肠杆菌过夜培养液按 1/100 (*V/V*) 转接 LB 培养基, 培养 3–4 h 后取 1.5 mL 菌液离心 (6000 r/min, 30 s) 收集菌体, 通过重悬浮和离心的方法用 LB 培养基洗涤菌体 3 次, 然后把菌体悬浮于 150  $\mu$ L LB 培养基待用。从平板挑取链霉菌

LHW3101 的单菌落接种 TSBY 培养基, 30 °C、220 r/min 培养 36 h 后取 1 mL 菌液离心 (10000 r/min, 1 min) 收集菌体, 用 LB 培养基洗涤菌体 2 次后把菌体悬浮于 150  $\mu$ L LB 培养基待用。把大肠杆菌和链霉菌的菌液混和后涂布在 SFM 平板 (10 mmol/L  $Mg^{2+}$ ), 30 °C 培养 12 h 后用 1 mL 抗生素水溶液覆盖平板, 保证水吹干后平板抗生素浓度为 10  $\mu$ g/mL apramycin 和 25  $\mu$ g/mL 萘啶酮酸。覆盖后的平板于 30 °C 培养 4 d 后可见  $Apr^R$  单菌落。挑取  $Apr^R$  单菌接种 SFM 平板 (10  $\mu$ g/mL apramycin) 进行抗性复证, 选取抗性正确的单菌落在 SFM 平板上转接 3 次, 收集孢子并梯度稀释孢子悬浊液后涂布在 SFM 平板, 30 °C 培养 3 d 后挑取单菌落筛选  $Apr^S$  菌株, 提取基因组 DNA 作为 PCR 模板, 用引物 surBC-2S 和 surBC-2A 筛选双交换突变株以及引物对 ermP-sur-S 和 ermP-sur-A 筛选双交换突变株, 通过对 PCR 产物进行测序确认突变株。

### 1.5 链霉菌发酵及化合物提取

少量液体发酵及化合物提取: 从 SFM 平板挑取单菌落接种到 10 mL TSBY 培养基, 用 100 mL 弹簧三角瓶, 30 °C、220 r/min 培养 36 h, 取 1 mL 菌液作为种子同样条件发酵 2 d。取 0.6 mL 发酵液于 1.5 mL 离心管, 加入 0.6 mL 乙酸乙酯, 加热 40 °C 并 1000 r/min 振荡 30 min, 萃取液 14000 r/min 离心 2 min, 收集有机相于 1.5 mL 离心管, 待乙酸乙酯挥发完, 加入 400  $\mu$ L 甲醇溶解样品, 经 14000 r/min 离心 5 min 后做 LC-MS 分析, 上样量为 20  $\mu$ L。

大量液体发酵及化合物提取: 将少量液体发酵种子液按 10% (*V/V*) 接种到 150 mL TSBY 培养基, 用 500 mL 弹簧三角瓶在相同条件下培养 28 h 后得到二级种子, 再按同样比例接种到 2 L 弹簧三角瓶, 每瓶装 500 mL TSBY 培养基, 同样条件

发酵 5 d。发酵液用等体积的乙酸乙酯萃取 3 次, 减压旋蒸萃取液后收集浸膏。

### 1.6 化合物 Surugamides 色谱分离和纯化

将 16 g 浸膏经过凝胶色谱柱 Sephadex LH-20 进行初步分离, 洗脱溶剂为甲醇。LC-MS 分析跟踪目标馏分, 对目标馏分再进行中压液相色谱分离, ODS 色谱柱 (15  $\mu$ m, 100  $\text{\AA}$ , 20 g), 甲醇梯度洗脱条件: 0–30 min (60%→80%), 30–60 min (80%), 60–90 min (80%→100%), 流速 10 mL/min。LC-MS 分析跟踪目标馏分, 对目标馏分再通过制备型 HPLC 做最终纯化, 洗脱溶剂为乙腈: 水 (0.1% 甲酸)=43:57, 色谱柱为 Phenomenex Venus Prep C<sub>18</sub> (5  $\mu$ m, 250 mm×10 mm), 流速 3 mL/min。

### 1.7 LC-MS 及 NMR 分析条件

常规 HPLC 分析条件为: A 相为水 (0.1% 甲酸), B 相为乙腈 (MeCN); MeCN 洗脱梯度为 0–17 min (30%→69.7%), 17.0–17.1 min (69.7%→100%), 17.1–21.0 min (100%), 21.0–21.1 min (100%→30%), 21.1–26.0 min (30%); 流速 0.7 mL/min; 检测波长 190 nm–400 nm; 柱温 25 °C, 色谱柱为 Waters Xbridge C<sub>18</sub> (5  $\mu$ m, 250 mm×4.6 mm)。低分辨质谱为 Waters ACQUITY QDa, 正离子模式。UPLC 流动相为 MeCN 和 0.1% 甲酸水溶液, 洗脱梯度为 0–0.2 min (5% MeCN), 0.2–5.0 min (100% MeCN), 5–7 min (100% MeCN), 7.0–7.1 min (5% MeCN), 7.1–9.0 min (5% MeCN), 色谱柱为 Acquity UPLC BEH C<sub>18</sub> (1.7  $\mu$ m, 2.1 mm×50 mm)。ESI-Q-TOF 高分辨质谱分析采用正离子模式。核磁数据在 Bruker Avance III 600 MHz 核磁共振仪上采集, 核磁样品溶剂 DMSO-*d*<sub>6</sub>。

### 1.8 基因组测序及生物信息学分析

通过细菌基因组 DNA 抽提试剂盒 (天根) 提

取和纯化样本 DNA, 利用超声波将样本 DNA 随机打断, 通过末端修复、连接接头, 选择 200–350 bp 的 DNA 文库片段, 使用 Agilent 2100 系统对文库进行质检, 然后上机测序。测序平台为 Illumina HiSeq Xten, 测序方法为双端测序, 最终产生大于 4G 的原始 FASTQ 数据。通过 FASTQC 进行数据质控, Trimmomatic 去除接头和质量差的碱基, 保留长度大于 40 bp 的序列, 使用 SPAdes 3.6 软件进行基因组 *de novo* 拼接, 获得 scaffold 片段长度大于 200 kb。用软件 Prokka 进行基因预测和注释。通过在线服务器 Ezbiocloud (<https://www.ezbiocloud.net/dashboard>) 分析 16S rRNA 序列, 通过次级代谢产物生物合成基因簇预测网站 antiSMASH (<http://antismash.secondarymetabolites.org/>) 预测基因组次级代谢产物生物合成基因簇, 通过 NRPS-PKS 在线分析网站 (<http://www.nii.res.in/nrps-pks.html>) 分析 NRPS 大蛋白的催化功能域构成。Surugamides 生物合成基因簇 DNA 序列已提交 GenBank 数据库, 序列编号 MH070261。

## 2 结果和分析

### 2.1 从海绵中分离获得 *S. albidoflavus* LHW3101

将来自中国南海永兴岛的胃甲海绵

(*Dactylosporgia elegans*) 粉末悬浊液梯度稀释后涂平板培养 10–30 d, 共获得放线菌生长表型的单菌落 32 个。通过在线服务器 Ezbiocloud 对 16S rRNA 基因序列进行比对分析, 发现其中一株菌与 *S. albidoflavus* DSM 40455T 相似度达到 100%<sup>[11]</sup>, 命名该菌株为 *S. albidoflavus* LHW3101。LC-MS 分析 LHW3101 菌株发酵产物后发现其次级代谢产物较为丰富, 所以对其进行了基因组扫描测序。基因组序列经生物信息学分析和基因注释后发现了 5 个 NRPS 类型基因簇。

### 2.2 非核糖体多肽 Surugamides 生物合成基因簇分析

将 LHW3101 基因组序列递交 AntiSMASH 数据库进行比对分析, 发现一个包含 18 个肽链延伸功能模块的大型 NRPS 基因簇, 该基因簇与来源于 *S. albus* J1074 的 Surugamides (sgm) 生物合成基因簇几乎完全相同, 2 个基因簇结构和基因大小完全相同, 仅有 2 个 NRPS 基因存在微小差别 (99% 的相同度) (图 1, 表 1)。前期研究发现 sgm 基因簇应包括 2 个 NRPS 生物合成途径, 推测负责环八肽化合物 sgm A 生物合成的基因是 *surA* 和 *surD*, 而嵌入在这 2 个基因之间的 *surB* 和 *surC* 基因可能负责链状十肽化合物 sgm F 的生物合成<sup>[6–7]</sup>。两个 NRPS 生物合成体系均没有后修饰步骤和硫酯酶功能域。



图 1. 链霉菌 LHW3101 和 J1074 来源的 Surugamides 生物合成基因簇结构相同

Figure 1. The same gene organizations are found for the Surugamides gene clusters derived respectively from *S. albidoflavus* LHW3101 and *S. albus* J1074.

表 1. 链霉菌 LHW3101 和 J1074 来源 Surugamides 基因簇蛋白序列比较分析

Table 1. The protein similarity analysis of Surugamides biosynthetic gene clusters derived respectively from *S. albidoflavus* LHW3101 and *S. albus* J1074

ORF name	aa (LHW3101 or J1074)	Identity/%	Predicted function
SurR	138	100	Gent family DNA-binding transcriptional regulator
ORF1	368	100	Membrane protein
SurE	451	100	Beta-lactamase
SurA	5726	99	NRPS (A-T, C-A-T-E, C-A-T-E, C-A-T)
SurB	4256	100	NRPS (A-T, C-A-T-E, C-A-T, C-A-T)
SurC	7691	100	NRPS (C-A-T, C-A-T-E, C-A-T, C-A-T-E, C-A-T, C-A-T-E)
SurD	4114	99	NRPS (C-A-T, C-A-T-E, C-A-T-E)
ORF2	452	99	Drug resistance transporter
ORF3	203	100	TetR family transporter regulator

可能负责 *sgm F* 生物合成的 NRPS 蛋白 SurB 和 SurC 共加载 10 个氨基酸(图 2-A), 包括 1 个色氨酸(起始模块)、2 个亮氨酸(第一和五延伸模块)、

3 个缬氨酸(第二、六和八延伸模块)、1 个苏氨酸(第三延伸模块)、2 个丙氨酸(第七和九延伸模块)以及加载于第 4 个延伸模块的  $\beta$ -氨基酸 3-amino-2

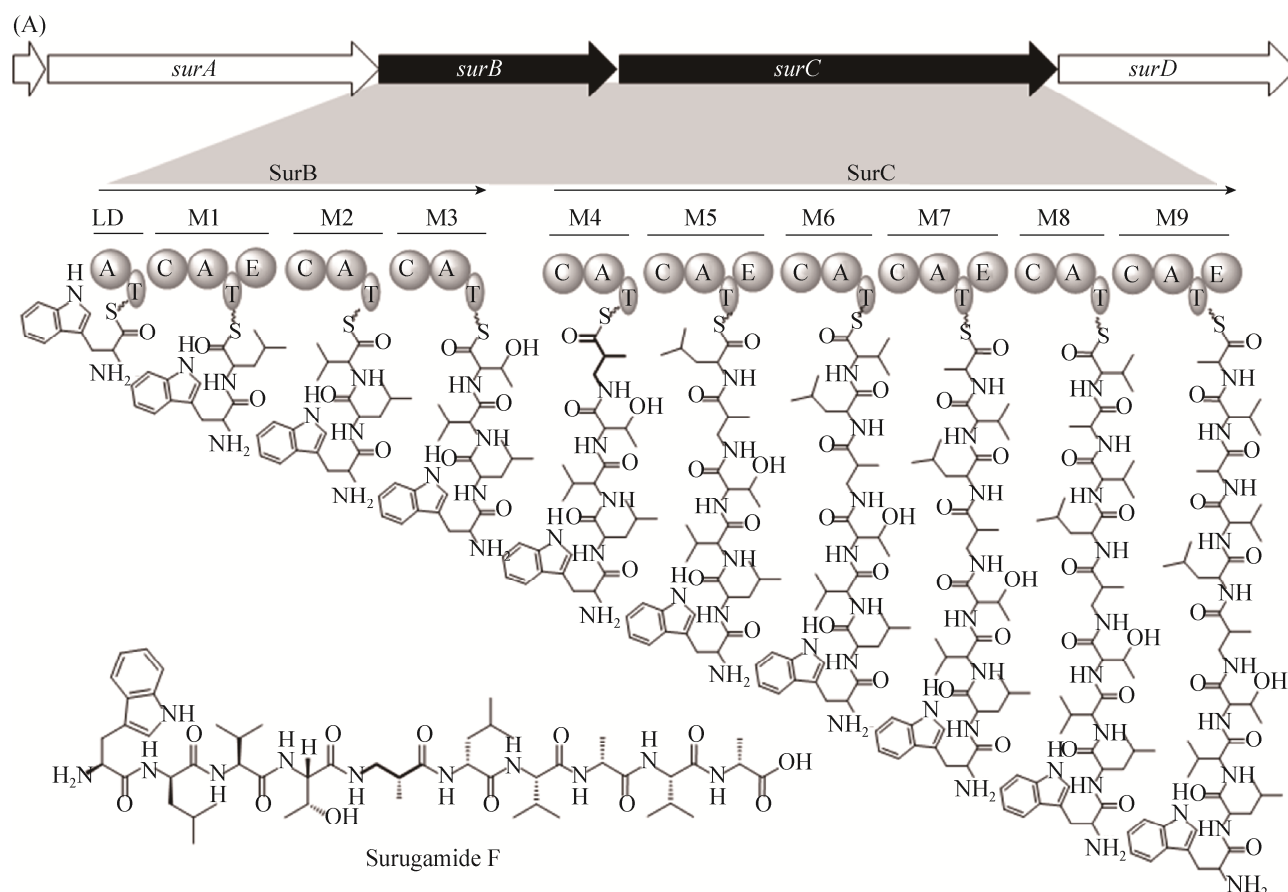
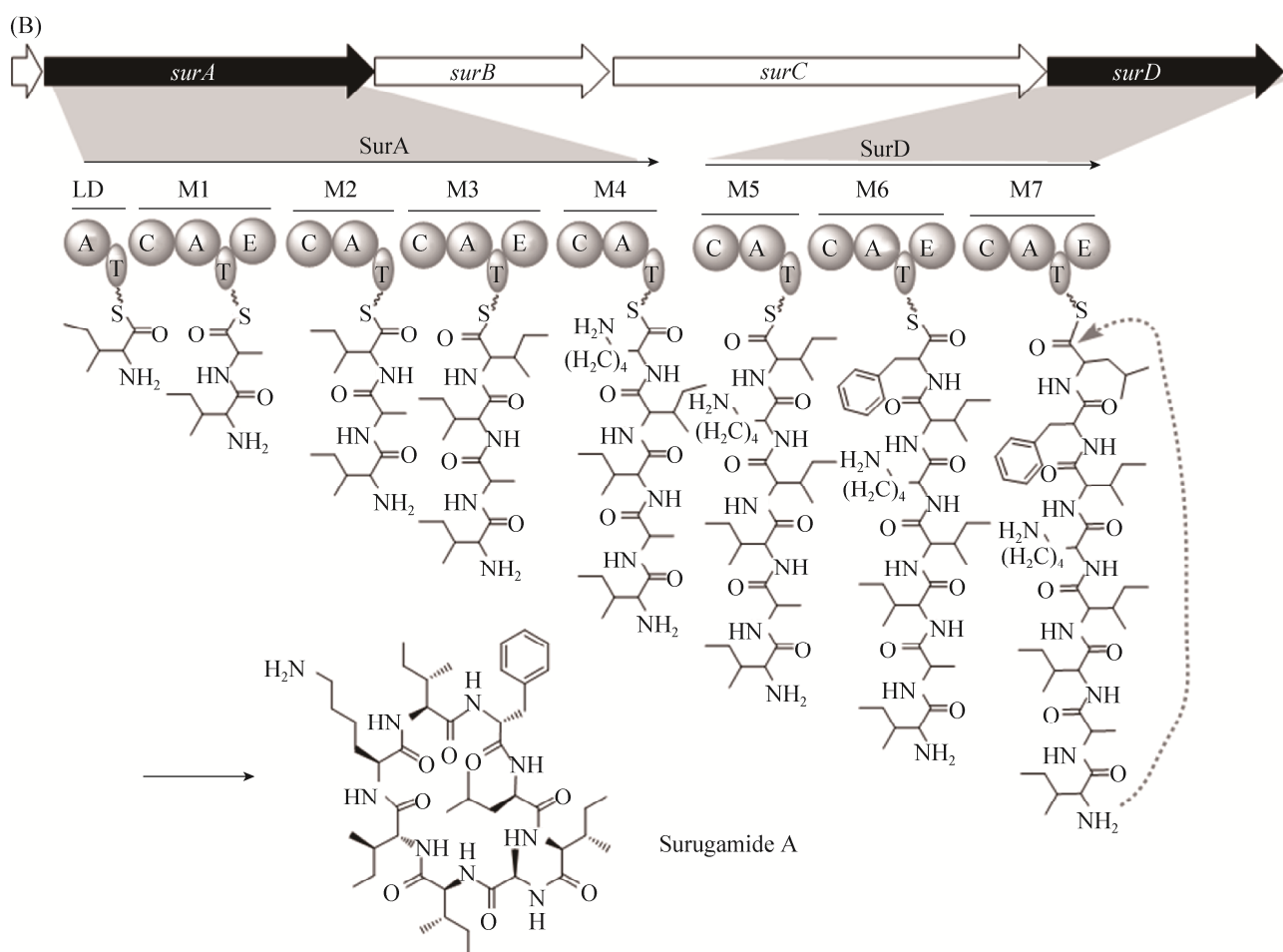


图 2. Surugamide F(A)和 Surugamide A(B)生物合成基因簇及其生物合成途径推导

Figure 2. The gene clusters and proposed biosynthetic pathways for surugamide F (A) and surugamide A (B).



续图 2.

(R)-methylpropionic acid(AMPA)。该 NRPS 装配线含有的 4 个差向异构酶(epimerase, E)功能域分别位于第一、五、七和九延伸模块,负责转化 L-型氨基酸为 D-型氨基酸。尚未在 sgm 基因簇边界处找到可能负责 AMPA 生物合成的基因。

可能负责 sgm A 生物合成的 NRPS 蛋白 SurA 和 SurD 共加载 8 个氨基酸(图 2-B), SurA 包括一个起始模块和 4 个延伸模块,各模块 A 功能域加载的氨基酸依次为异亮氨酸、丙氨酸、异亮氨酸、异亮氨酸、赖氨酸; SurD 含 3 个延伸模块,其 A 功能域分别加载异亮氨酸、苯丙氨酸和亮氨酸。第一、三、六和七延伸模块中包含的差向异构酶功能域应该负责 sgm A 环肽分子对应氨基酸的 D 型构相的生成。

基于以上对 *S. albidoflavus* LHW3101 来源的非核糖体多肽 Surugamides 生物合成基因簇的生物信息学分析,提示该菌具有生产 sgm A 和 sgm F 的潜力。

### 2.3 化合物 Surugamides 的分离纯化和结构鉴定

对 *S. albidoflavus* LHW3101 和 *S. albus* J1074 的发酵产物乙酸乙酯萃取物进行了高分辨质谱 UPLC-ESI-QTOF 分析对比,发现两者均有化合物 sgm A 和 sgm F 对应的分子离子峰: sgm A,  $[M+H]^+=912.6277$ (计算值 912.6281, 误差值  $-0.4$  mg/L); sgm F,  $[M+H]^+=1056.6486$ (计算值 1056.6452, 误差值 3.2 mg/L)(图 3, 图 4)。为进一步确证 LHW3101 发酵产物中的 sgm A 化合物的结构,对其进行了 36 L 液体发酵。用乙酸乙酯萃

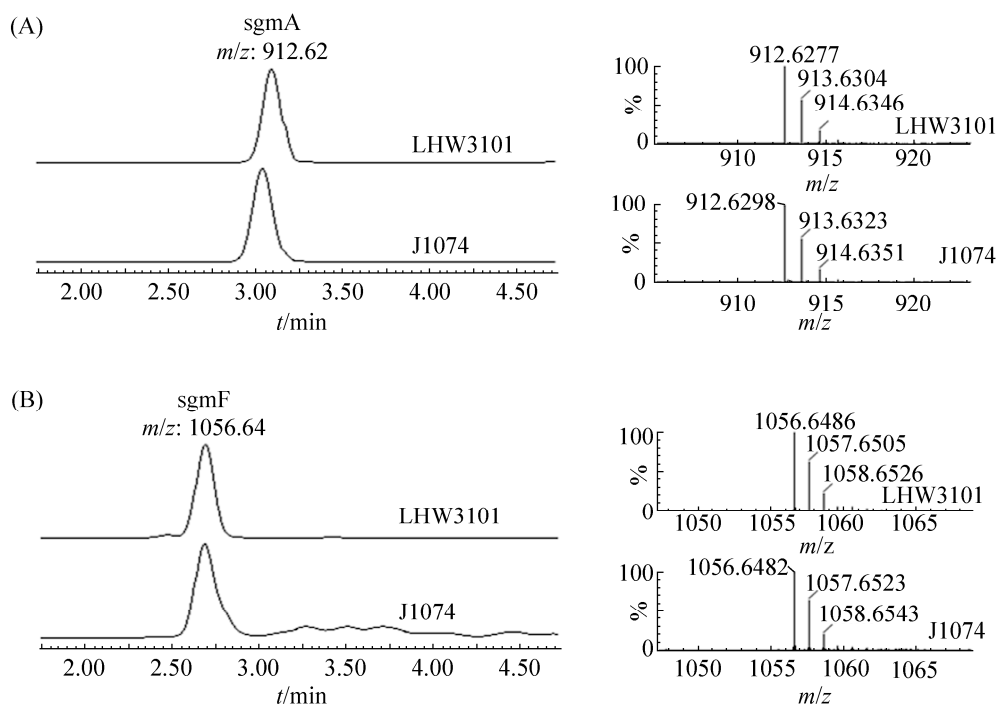


图 3. 超高压液相-四级杆-飞行时间质谱(UPLC-Q-TOF-MS)分析确认链霉菌 LHW3101 和 J1074 发酵提取物中的化合物 sgm A(A)和 sgm F(B)

Figure 3. UPLC-Q-TOF-MS analysis of the sgm A (A) and sgm F (B) in the fermentation extract of strain LHW3101 and J1074. The left part is the extracted molecular ion peak and the right one is the mass signal of target molecular ion.

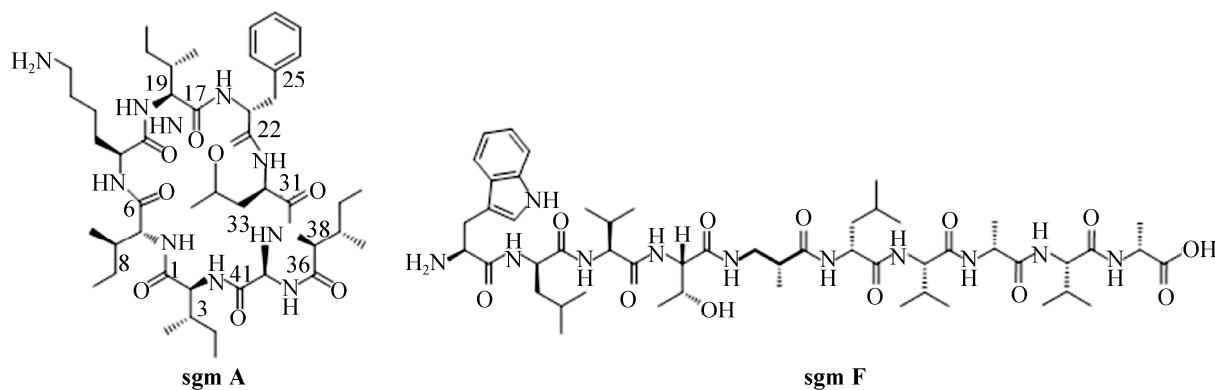


图 4. sgm A 和 sgm F 的化学结构图

Figure 4. The chemical structure of sgm A and sgm F.

取发酵液后获得 16 g 浸膏, 经凝胶色谱柱 Sephadex HP-20 分离得到 A-E 5 个馏分。把含有目标化合物的 B 馏分上样中压液相 ODS 制备柱, 通过 LC-MS 检测获得目标馏分, 最终通过制备型

HPLC 分离获得 sgm A 化合物 9.7 mg。对化合物 sgm A 进行了  $^{13}\text{C}$  NMR 分析, 通过对比文献报道碳谱化学位移确证了其结构的正确性<sup>[12]</sup>(表 2)。由于 sgm F 产量很低而未获得足够的化合物进行



NMR 测试, 但其高分辨质谱数据和预期完全符合。

### 2.4 确认 *sgm A* 和 *sgm F* 的生物合成相关基因

为了确认镶嵌于 *sgm* 基因簇内部的 *surB* 和 *surC* 基因仅负责 *sgm F* 的生物合成而与 *sgm A* 无关, 构建了用于缺失基因 *surB* 和 *surC* 的同源重组双交换载体 pRJ26, 同时在该同源交换臂中间装载了组成型强启动子 *ermEp\**, 使得缺失 *surB* 和 *surC* 的同时在 *surD* 基因阅读框前面引入强启动子 *ermEp\** (图 5-A)。通过双亲接合转移把 pRJ26 导入宿主菌 LHW3101 后筛选同源重组双交换突变株。

用引物 *surBC-2S* 和 *surBC-2A* 对候选突变株进行 PCR 筛选, 双交换突变株 PCR 目标产物为 1122 bp (图 5-B 左), 野生型的 PCR 目标产物因理论长度为 39 kb 所以无法扩增出产物, 同时用引物 *ermP-sur-A* 和 *ermP-sur-S* 进一步验证 *ermEp\** 已引入 *surD* 基因前, 目标产物为 212 bp (图 5-B 右), 经过测序确认两对引物验证突变株的 PCR 产物, 最终获得突变株 RJ9。LC-MS 分析 RJ9 液体发酵乙酸乙酯萃取物发现化合物 *sgm F* 不再产生, 而 *sgm A* 的产量提高到野生型产量约 2 倍水平(图 5-C)。

表 2. 与文献比对确认 *sgm A* 的 NMR 碳谱归属(在 DMSO-*d*<sub>6</sub> 溶剂中检测)

Table 2. NMR assignments for *sgm A* analyzed in DMSO-*d*<sub>6</sub> by comparing with the published data

Residue	Position	Experiment <sup>[11]</sup>		Residue	Position	Experiment <sup>[11]</sup>	
		$\delta_C$	$\delta_C$			$\delta_C$	$\delta_C$
Ile1	1	ND	172.1	Phe	22	171.3	171.5
	2	57.8	57.8		23	54.8	55.0
	3	34.8	35.3		24	36.4	36.8
	3'	14.7	14.8		25	138.1	138.0
	4	24.5	24.7		26, 30	129.0	129.3
Ile2	5	11.5	11.9	27, 29	128.1	128.4	
	6	171.0	ND	28	126.3	126.6	
	7	57.0	57.0	Leu	31	172.9	172.9
	8	35.9	36.3		32	52.7	52.7
	8'	14.5	14.8		33	ND	40.4
Lys	9	25.7	26.0	34	24.4	24.6	
	10	11.4	11.9	35	21.3	21.7	
	11			35'	23.2	23.5	
	12	51.7	52.1	Ile4	36	169.8	169.8
	13	31.4	31.7		37	57.6	57.7
14	22.1	22.3	38		35.8	36.1	
Ile3	15	29.1	28.2	38'	15.2	15.4	
	16	38.8	39.7	39	23.9	24.2	
	17	171.1	ND	40	10.6	ND	
	18	58.2	58.2	Ala	41	172.6	172.8
	19	35.4	35.9		42	48.1	48.2
	19'	15.0	15.0		43	18.9	19.2
	20	25.0	25.1				
21	11.1	11.4					

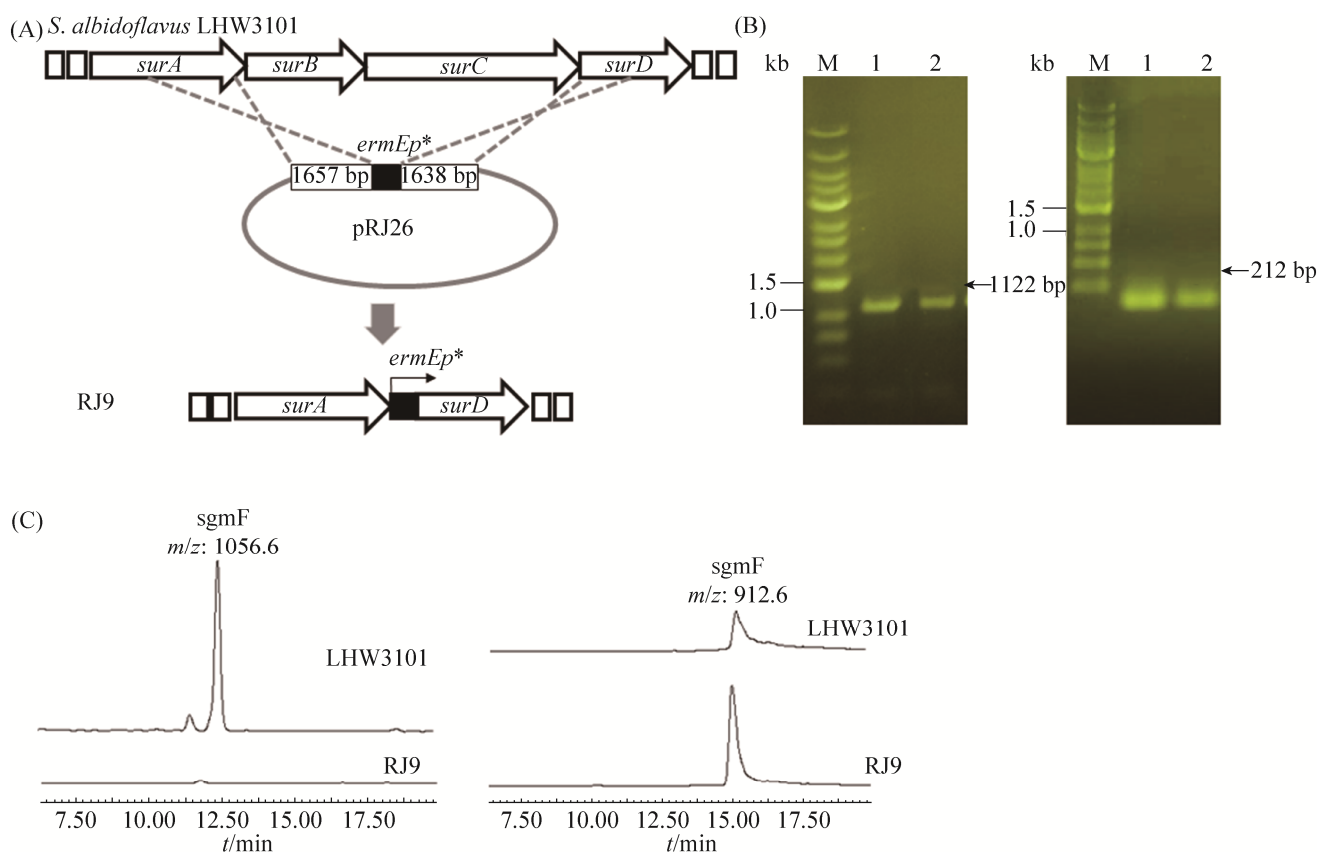


图 5. *surB-C* 被替换突变株 RJ9 的构建(A, B)以及发酵产物 LC-MS 检测(C)

Figure 5. Construction and LC-MS analysis of the *surB-C* replacement mutant RJ9. A: Replacement of *surB-surC* genes fragment with a constitutive strong promoter *ermEp\** just before the open reading frame of *surD* by homologous recombination; B: PCR confirmation of mutant RJ9. The PCR product of either positive control (PC) or RJ9 is 1122 bp for primer *surBC-2S* and *surBC-2A* (left gel) and 212 bp for primer *ermP-sur-S* and *ermP-sur-A* (right gel); C: LC-MS analysis of the fermentation extract of RJ9 and LHW3101. The ion peaks corresponding to *sgmF* and *sgm A* are labeled.

### 3 讨论

Surugamides(*sgm*)生物合成基因簇最先发现于海洋链霉菌 *Streptomyces* sp. JAMM992 中<sup>[7]</sup>, 但其 DNA 序列一直未公布。2017 年许飞等再次从 *S. albus* J1074 中发现了该基因簇, 并对 *surA* 和 *surB* 基因分别进行了基因中断突变, 发现 *surA* 是化合物 Surugamide A (*sgm A*)生物合成必需, 而 *surB* 为化合物 Surugamide F (*sgm F*)生物合成必需, 结合对 NRPS 大蛋白功能域构成的分析, 推

测 NRPS 基因 *surA* 和 *surD* 负责 *sgmA* 生物合成, 而 *surB* 和 *surC* 则专一负责 *sgmF* 的生物合成<sup>[6]</sup>。我们从中国西沙胃甲海绵(*Dactylospongia elegans*)中分离到了链霉菌 *S. albidoflavus* LHW3101, 在对其基因组天然产物生物合成基因簇分析中发现了 *sgm* 基因簇, 确认了 LHW3101 代谢产物中的化合物 *sgm A* 和 *sgm F*。前期报道对基因 *surC* 和 *surD* 功能归属尚无实验证据, 我们通过基因缺失突变确认了 *surB* 和 *surC* 基因与 *sgm A* 产物无关, 在 *surD* 基因编码框前引入强启动子后显著提高了

sgm A 的产量, 从而进一步确证了 *surA* 和 *surD* 基因负责 sgm A 生物合成的推论。此外, *surB* 和 *surC* 基因编码的 NRPS 功能域构成与化合物 sgm F 链状十肽的结构对应, 同时缺失这两个基因后 sgm F 不再产生, 提示 *surB* 和 *surC* 可能负责 sgm F 的生物合成。

Sgm 生物合成途径还存在一些未解之谜值得深入探索。两个 NRPS 生物合成途径均缺乏负责链释放或环化功能的硫酯酶功能域, 推测位于 sgm 基因簇左边界的内酰胺酶(lactamase)基因或基因簇外部的硫酯酶基因负责完成这一功能, 类似的情况如南昌霉素(nanchangmycin)聚酮生物合成末端释放由单独存在硫酯酶通过反式作用机制负责完成<sup>[13]</sup>。此外 sgm F 肽链构成中含有的  $\beta$ -氨基酸 AMPA 需要相关的基因负责合成, 该前体也发现于链霉菌来源化合物 Vicenistatin 的聚酮生物合成途径中, 是由谷氨酸经谷氨酸变位酶和脱羧酶负责合成的<sup>[14]</sup>, 用这 2 个蛋白的序列分别对 *S. albidoflavus* LHW3101 基因组进行 BLAST 比对搜索却并未找到同源基因。

## 参 考 文 献

- [1] Kries H. Biosynthetic engineering of nonribosomal peptide synthetases. *Journal of Peptide Science*, 2016, 22(9): 564–570.
- [2] Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, Mueller A, Schäberle TF, Hughes DE, Epstein S, Jones M, Lazarides L, Steadman VA, Cohen DR, Felix CR, Fetterman KA, Millett WP, Nitti AG, Zullo AM, Chen C, Lewis K. A new antibiotic kills pathogens without detectable resistance. *Nature*, 2015, 17(7535): 455–459.
- [3] Walsh CT. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Natural Product Reports*, 2016, 33(2): 127–135.
- [4] Wang SY. Advances in the study of the mechanism and application of nonribosomal peptide synthetases. *Acta Microbiologica Sinica*, 2007, 47(4): 734–737 (in Chinese). 王世媛. 非核糖体肽合成酶(NRPSs)作用机理与应用的研究进展. *微生物学报*, 2007, 47(4): 734–737.
- [5] Horsman ME, Hari TPA, Boddy CN. Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: logic gate or a victim of fate? *Natural Product Reports*, 2016, 33(2): 183–202.
- [6] Xu F, Nazari B, Moon K, Bushin LB, Seyedsayamdost MR. Discovery of a cryptic antifungal compound from *Streptomyces albus* J1074 using high-throughput elicitor screens. *Journal of the American Chemical Society*, 2017, 139(27): 9203–9212.
- [7] Ninomiya A, Katsuyama Y, Kuranaga T, Miyazaki M, Nogi Y, Okada S, Wakimoto T, Ohnishi Y, Matsunaga S, Takada K. Biosynthetic gene cluster for surugamide A encompasses an unrelated decapeptide, surugamide F. *Chembiochem*, 2016, 17(18): 1709–1712.
- [8] Sun YH, He XY, Liang JD, Zhou XF, Deng ZX. Analysis of functions in plasmid pHZ1358 influencing its genetic and structural stability in *Streptomyces lividans* 1326. *Applied Microbiology and Biotechnology*, 2009, 82(2): 303–310.
- [9] Wilkinson CJ, Hughes-Thomas ZA, Martin CJ, Böhm I, Mironenko T, Deacon M, Wheatcroft M, Wirtz G, Staunton J, Leadlay PF. Increasing the efficiency of heterologous promoters in actinomycetes. *Journal of Molecular Microbiology and Biotechnology*, 2002, 4(4): 417–426.
- [10] Gibson DG, Young L, Chuang RY, Venter JC, Hutchison III CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 2009, 6(5): 343–345.
- [11] Hain T, Ward-Rainey N, Kroppenstedt RM, Stackebrandt E, Rainey FA. Discrimination of *Streptomyces albidoflavus* strains based on the size and number of 16S-23S ribosomal DNA intergenic spacers. *International Journal of Systematic Bacteriology*, 1997, 47(1): 202–206.
- [12] Takada K, Ninomiya A, Naruse M, Sun Y, Miyazaki M, Nogi Y, Okada S, Matsunaga S. 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 distinction of D- and L-amino acid residues in the sequence. *Journal of Organic Chemistry*, 2013, 78(13): 6746–6750.
- [13] Liu TG, You DL, Valenzano C, Sun YH, Li JL, Yu Q, Zhou XF, Cane DE, Deng ZX. Identification of NanE as the thioesterase for polyether chain release in nanchangmycin biosynthesis. *Chemistry & Biology*, 2006, 13(9): 945–955.
- [14] Miyanaga A, Iwasawa S, Shinohara Y, Kudo F, Eguchi T. Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier protein in vicenistatin biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 2016, 113(7): 1802–1807.

# Elucidation of an intercalation structure of the gene cluster for nonribosomal peptide Surugamides biosynthesis

Chunmin Xu<sup>1</sup>, Xiao Lin<sup>3</sup>, Lei Li<sup>2</sup>, Shuping Wang<sup>2</sup>, Fan Yang<sup>2</sup>, Shihai Xu<sup>3</sup>, Hai Deng<sup>2</sup>, Yongjun Zhou<sup>2</sup>, Houwen Lin<sup>1,2\*</sup>

<sup>1</sup> Jiangxi University of Traditional Chinese Medicine, Nanchang 330004, Jiangxi Province, China

<sup>2</sup> Department of Pharmacy, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, China

<sup>3</sup> College of Pharmacy, Jinan University, Guangzhou 510632, Guangdong Province, China

**Abstract: [Objective]** The Surugamides (sgm) biosynthetic gene cluster has been reported to contain four NRPS genes *surA–D* which give two unrelated NRPS pathways. It was experimentally confirmed that *surA* and *surB* were essential for Surugamide A and Surugamide F production respectively. However, the functions of *surC* and *surD* genes are not verified with experimental evidence. This work is designed to confirm if *surA* and *surD* genes are responsible for Surugamide A biosynthesis so as to pave the way to either genetically engineering of Surugamides biosynthetic pathways or study the recognition mechanism of their NRPS proteins. **[Methods]** The Actinomycetes isolated from marine sponge were identified by analyzing their 16S rRNA gene sequences. The natural product biosynthetic gene clusters were analyzed by submitting genomic sequences to the online database antiSMASH. The chemical structures of isolated compounds were elucidated by UPLC-Q-TOF-MS and <sup>13</sup>C NMR. To generate the mutation of gene deletion and replacement, a plasmid was constructed with two fragments used for target homologous recombination. The plasmid was then transformed into the target strain for screening of double-crossover mutants. **[Results]** We discovered Surugamides gene cluster from the genome of *Streptomyces albidoflavus* LHW3101 isolated from marine sponge *Dactylospongia elegans*. The compound Surugamide A and Surugamide F were then identified from the fermentation extract of the strain. The *surB* and *surC* gene in Surugamides gene cluster were replaced with a constitutive and strong promoter *ermEp*<sup>\*</sup>, which was located just before the transcription frame of *surD* gene. The resulted mutant RJ9 lost the production of Surugamide F but kept on producing Surugamide A with around two-fold yield of the wide type strain. **[Conclusion]** Gene *surB* and *surC* are verified to be not related to sgm A production. It is reported that *surA* is essential for Surugamide A production. And a remarkable yield improvement for Surugamide A was achieved after introducing a strong promoter before the open reading frame of *surD* in this work. Therefore we concluded that *surA* and *surD* take charge of the biosynthesis of Surugamide A.

**Keywords:** microbial natural products, nonribosomal peptide synthase, biosynthetic gene cluster

(本文责编: 张晓丽)

Supported by the National Natural Science Foundation of China (31670096, 31628001, 31600014)

\*Corresponding author. Tel/Fax: +86-21-68383339; E-mail: franklin67@126.com

Received: 23 January 2018; Revised: 1 April 2018; Published online: 20 April 2018