Published in final edited form as:
J Am Chem Soc. 2011 November 16; 133(45): 18010–18013. doi:10.1021/ja2040877.

Imaging mass spectrometry and genome mining via short sequence tagging identified the anti-infective agent arylomycin in *Streptomyces roseosporus*

Wei-Ting Liu[1], **Roland D. Kersten**[2], **Yu-Liang Yang**[3], **Bradley S. Moore**[2,3], and **Pieter C. Dorrestein**[1,2,3,*]

[1]Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, California, USA

[2]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California, USA

[3]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, California, USA

Abstract

Here, we described the discovery of anti-infective agent arylomycin and its biosynthetic gene cluster in an industrial daptomycin producing strain *Streptomyces roseosporus*. This was accomplished via the use of MALDI imaging mass spectrometry (IMS) along with peptidogenomic approach in which we have expanded to short sequence tagging (SST) described herein. Using IMS we have observed that prior to the production of daptomycin, a cluster of ions (**1**–**3**) were produced by *S. roseosporus* and correlated well with the decreased staphylococcal cell growth. Further adopted SST peptidogenomics approach, which relies on the generation of sequence tags from tandem mass spectrometric data and query against genomes to identify the biosynthetic genes, we were able to identify these three molecules (**1**–**3**) to arylomycins, a class of broad-spectrum antibiotics that targets type I signal peptidase. The gene cluster responsible for arylomycin production in *S. roseosporus* was then identified. The identification of arylomycins and their biosynthetic gene cluster from intensely studied microorganism highlights the strength of IMS and MS guided genome mining approaches in effectively bridging the gap between phenotypes, chemotypes and genotypes.

Natural products that are made by non-ribosomal peptide synthetases (NRPS) have an unrivaled track record as anti-infective agents in the clinic[1,2]. Penicillin, vancomycin, and daptomycin are examples of antibiotics that are NRPS-derived[3–6] (Figure 1). With the emergence of antibiotic-resistant microbes, there is a great interest in molecules that target drug resistant microbes[7,8]. However, the last broad-spectrum antibiotic introduced on the market was over 50 years ago.

Our laboratory has been interested in the development of mass spectrometric methodologies that interconnects phenotypes, chemotypes, and genotypes. A part of the motivation for these tools is not only to discover new biology but also apply these tools to the discovery of antimicrobials. Here we report the use of imaging mass spectrometry in combination with a short sequence tagging (SST)-based genome mining approach that connects phenotypes and

pdorrest@ucsd.edu.

Supporting Information Available: Materials, detailed experimental procedures and supplementary figures are available free of charge via the Internet at http://pubs.acs.org.

chemotypes with genotypes. We applied this approach to the discovery of the arylomycins (**1**–**3**, Figure 1) and their biosynthetic pathway in *Streptomyces roseosporus*.

To connect phenotypes with chemotypes, our laboratory has recently developed methods to investigate microbial metabolic interactions via imaging mass spectrometry (IMS)[9–11]. One of the goals for the development of IMS approaches to detect metabolic exchange is to enable the discovery of new therapeutic leads. Herein the pathogens *S. aureus*[12] and *S. epidermidis*[13] were co-cultured with *Streptomyces roseosporus* NRRL 15998, whose genome has been sequenced. This actinomycete produces daptomycin, an antibiotic used in the clinic to treat gram-positive bacterial infections[4,6,14–17].

To demonstrate that IMS can be used to observe the molecules responsible for the inhibition of pathogens, we prepared lawns of *S. epidermidis* and *S. aureus* and then spotted *S. roseosporus* in the center (Figure 2, Figure S1). After 36 hours incubation, inhibition zones were observed as expected in both staphylococcal lawns. Surprisingly, even though we determined that the IMS methodology can detect as little as 10 pmole of daptomycin, ions corresponding to daptomycin were not observed. Instead, a cluster of ions at *m/z* 863, 877, and 891, referred to as compounds **1**–**3** in this paper, were observed to localize at the zone of inhibition area. The absence of daptomycin-related signals in the zone-of-inhibition experiment suggested that *S. roseosporus* produced additional antibiotics.

A time course experiment of methanol extracts of *S. roseosporus* starter cultures revealed that compounds **1**–**3** were observed at 36 hours (Figure S2), in agreement with the incubation time in the zone-of-inhibition experiment described above. Not until 48 hours, the production of signals at *m/z* 1634.72, 1648.74, 1662.75, which correspond to daptomycin variants (A21978C1-3, Figure 1) were observed. That daptomycin is not produced until 48 hours is consistent with the absence of daptomycin variants signals in the IMS data. MS-guided purification revealed that the molecules at *m/z* 863, 877, and 891 have monoisotopic masses of 825.439 (**1**), 839.455 (**2**), and 853.471 (**3**) Da, suggesting that the ion cluster observed in IMS exists as the potassium adduct. Compound **2** was purified and shown to exhibit antibiotic activity against *S. epidermidis* with similar efficacy to daptomycin but milder activity towards *S. aureus*, in agreement with the smaller zone of clearing for the *S. aureus* observed in Figure 2 (Figure S1, S3).

To link to genotypes, our laboratory has recently developed a peptidogenomic mining approach to the discovery of peptidyl natural products[18]. The approach itself relies on the generation of peptide sequence tags from tandem mass spectrometric data to query genomes and to identify the biosynthetic genes. In turn, in an iterative fashion, the biosynthetic gene cluster supports the identification of a peptide as either a ribosomal or non-ribosomal product and facilitates the prediction of a (partial) structure. For ribosomally-encoded peptides, a 5–6 consecutive amino acid residue sequence tag is often needed to successfully match to its precursor gene because of the larger proteomic search space. In this report, we show that for NRPS-derived peptides, this approach could be expanded to short sequence tagging (SST) with only one or two amino acid residues to identify the candidate biosynthetic gene clusters as we suggested would be possible[18]. SST can be employed to carry out genome mining with molecules that are NRPS-derived. This is possible because the search tags can be more specific due to additional non-proteinogenic amino acids and the much smaller query space because of the small number (often <10) of NRPS gene clusters within a microbial genome. This scenario is similar to matching a peptide to a small database in a proteomics experiment where it also becomes possible to match to the correct peptide with minimal fragmentation data while much more fragmentation information is needed when a large database is used. Therefore, even with a very short sequence tag we can still narrow down to the candidate biosynthetic gene cluster. The identification of candidate gene clusters, in turn, aids in the structural characterization of the molecule.

As proof-of-principle for SST, we first demonstrate how this works with daptomycin. The ions at *m/z* 1634.73, 1648.74, and 1662.75, corresponding to daptomycin variants, were subjected to tandem MS using collision-induced dissociation and resulted in fragment masses at m/z 1051.43, 1166.46, 1280.50 which suggested a sequence tag of Asp-Asn (Figure S4). Such a tag provides a minimal search unit that can be searched against all predicted NRPS biosynthetic pathways found on the *S. roseosporus* genome. All three tandem MS datasets of ions at *m/z* 1634.73, 1648.74, and 1662.75 resulted in an identical sequence tags (Figure S4). The combination of NP.searcher and NRPS predictor[19,20], two programs designed to identify the amino acid specificity of non-ribosomal peptide synthetases were utilized to predict all possible NRPS gene clusters and their amino acid codes in the *S. roseosporus* NRRL 15998 genome. Seven gene clusters that display NRPS features were found by NP.searcher on the *S. roseosporus* genome. Matching the sequence tag obtained from the tandem MS data of daptomycin variants against the NRPS predictor and NP.searcher predicted amino acids identified the daptomycin gene cluster. Therefore the proof-of-principle experiment with daptomycin variants demonstrated that the correct gene cluster could be identified from the genome through the SST approach.

Next we set out to identify the series of ions at *m/z* 825.439, 839.455 and 853.471 (**1–3**). These molecules are separated by 14.013 Da consistent with CH_2 mass shifts. Possible explanations to account for this 14 Da difference may arise from different length of fatty acid chain, amino acid substitution, or methylation. Each scenario is commonly found in NRPS biosynthetic pathways. Therefore SST was employed to match these molecules to one of the remaining six NRPS gene clusters. To *de novo* sequence peptides, it is often challenging to separate the ions belonging to y-ion series from the b-ion series. In this experiment however, all three ions were first subjected to low-resolution tandem MS, and the spectra were aligned (Figure 3A). This revealed a series of ions that displayed 14 Da mass shifts (shifting ions) and a series of ions that did not display the mass shifts (non-shifting ions). We were able to retrieve a sequence tag from the non-shifting ions which displayed mass differences of 57 and 71 Da, suggesting glycine and alanine, respectively. There was only one predicted NRPS out of the remaining six NRPS gene clusters that were identified on the *S. roseosporus* genome that contained this sequence tag. That gene cluster is predicted to encode six amino acids, Ser, Ala/Gly, Gly, Hpg, Ala/Gly and Tyr. Although high-resolution MS spectra could provide more unambiguous sequence tag, we show that SST works with high-resolution as well as low-resolution MS data. We consulted the NORINE database that contains greater than 1000 NRPS-derived molecules and enables users to input specific residue(s) to search for molecules that have specified structural units[21]. Searching the NORINE database for the Ser, Ala/Gly, Gly, Hpg, Ala/Gly and Tyr tag resulted in one group of candidate molecules, the arylomycins[22,23].

Having a candidate molecule in hand, the intact masses of arylomycins were compared and the fragmentation data was re-inspected (Figure 3B, 3C). The intact masses of the observed ions matched to the calculated masses of arylomycins within 0.5–2 ppm[22]. The analysis of the fragmentation of compound **1** revealed that the observed b, y and c and z ions were within 1 ppm, in agreement with arylomycin A2 (Figure S5).

Arylomycins are an exciting set of biologically active molecules. Arylomycins are a class of broad-spectrum antibiotics that targets type I signal peptidases (SPase)[24–28]. SPase is responsible for the cleavage of signal peptides from secreted protein and is an attractive antimicrobial target because it is highly conserved among bacteria, located on the extracellular surface of the cytoplasmic membrane, and is essential for bacterial viability.

Therefore, it has been suggested that arylomycins as promising therapeutic leads[29,30]. This promise has led to a 2010 start-up company, RQx Pharmaceuticals, that is aiming to develop arylomycin and its analogs into the clinic. Arylomycins were first discovered from *Streptomyces* sp. Tü 6075 isolated in the tropical rain forest at Cape Coast, Ghana[22,23]. Independently, using assays screened for novel signal peptidase I inhibitors, scientists at Eli Lilly & Company reported a similar group of compounds that shares the same skeleton to arylomycins but with a glycosylation on the hydroxyphenylglycine residue[28]. Although natural resistance has been reported[31], arylomycins and its glycosylated congeners are effective against gram-negative bacteria, such as *Helicobacter pylori*, *Yersinia pestis*, and gram-positive bacteria *Streptococcus pneumonia*, *Streptococcus pyrogens*, *Staphylococcus epidermidis* and *Staphylococcus haemolyticus* with MICs of 4–16 μg/ml[22,23,28,31–34].

To verify that *S. roseosporus* NRRL 15998 produces arylomycins, the candidate arylomycin gene cluster was annotated (Scheme 1, Supplementary information, Table S1) and analyzed for biosynthetic consistency with the proposed arylomycin product. As the candidate gene cluster in the NRRL 15998 strain contained a frameshift and sequencing gaps, we based our analysis on the complete, almost identical gene cluster sequence of *S. roseosporus* strain NRRL 11379, which also produces the same set of molecules (Figure S6). Arylomycins contains an N-acyl chain and six amino acids, serine (Ser), alanine (Ala), glycine (Gly), hydroxyphenyl glycine (Hpg), alanine (Ala) and tyrosine (Tyr). Ser1 and Hpg4 are *N*-methylated, Ser1 and Ala2 are in D-configuration and finally Tyr6 and Hpg4 are cross-linked by a biaryl carbon-carbon linkage reminiscent of vancomycin (Figure 1). The gene cluster comprises 10 genes and is consistent with the arylomycins core structure. The assembly-line NRPS contains 6 modules on 3 genes (*aryABD*) where all A domains have the predicted substrate specificity of the observed amino acids, Ser, Ala/Gly, Gly, Hpg, Ala/Gly and Tyr respectively[35,36]. The loading module has a C domain that clades with starter C domains based on a phylogenetic analysis (Figure S7). This C domain incorporates the *N*-acyl group into arylomycins as starter C domains are known to catalyze initial *N*-acylation in NRP biosynthesis[37]. The two *N*-methyl groups at positions 1 (Ser) and 4 (Hpg) in arylomycins are in agreement with the methyltransferase domains in the corresponding NRPS modules. Furthermore, the two D-amino acid residues at positions 1 (D-Ser) and 2 (D-Ala) are consistent with the epimerization domains in corresponding NRPS modules, too. Finally the gene cluster contains a cytochrome P450 enzyme with 49% similarity to the vancomycin OxyC protein that is predicted to form the bis-aryl carbon linkages in vancomycin[38] (Figure S8). Thus, AryC is likely to be responsible for the biaryl bond formation in arylomycins. It should be noted that we did not observe the nitrosated congeners that corresponding to the arylomycin B series described in previous reports[22,23] in *S. roseosporus*, which is in agreement with the absence of nitrosating enzymes in the gene cluster[39]. Therefore, based on the annotation of the MS data as well as the annotation of the gene cluster, the data suggest that SST enabled the discovery of the promising anti-infective agent arylomycins from *S. roseosporus*.

To finally confirm the production of the arylomycins lipopeptides from the daptomycin-producing *S. roseosporus* strains, a larger scale fermentation, extraction, and purification using reversed-phase HPLC was undertaken. Compounds **1** and **2** were subjected to NMR analysis. The resulting ^{1}H-NMR spectra revealed the same peptide core as arylomycin by comparing with the NMR spectra described in the original arylomycin report[22,33] (Figure S9). The ^{1}H-NMR spectra of **1** and **2** matched to arylomycin A2 and A4, respectively[33].

The identification of arylomycins and their biosynthetic gene cluster from intensely studied microorganism of commercial importance highlights the strength of MS-guided genome mining approaches and IMS effectively bridging the gap between phenotypes, chemotypes and genotypes. The SST approach described herein enables matching of molecules identified through imaging mass spectrometry to NRPS biosynthetic machinery using only a minimal sequence tag. We anticipate that SST will also prove capable of identifying the biosynthetic machinery for molecules that contain non-standard amino acids, which are often incorporated in NRPs. According to the NCBI genome database, there are now ~1700 fully sequenced bacterial genomes as assessed in July 2011 opposed to ~1000 in October 2009[40] which represents a more than 70% increase in less than 2 years (and there are ~5000 bacterial genome sequencing projects in progress). Since the full repertoire of genome sequence continues to expand at a rapid pace, there is a need to increase our effectiveness in genome mining to identify new natural products; genome mining of one molecule at a time will not be able to keep up the pace by which genomes are sequenced. Current high-throughput approaches (e.g. metabolomic or proteomic) do not efficiently identify natural products and therefore there is a need to develop genome mining approaches that enables us to rapidly connect chemotypes to genotypes and ultimately phenotypes. Imaging mass spectrometry and/or peptidogenomics, especially with its expansion to smaller peptides and SSTs, are approaches that expedite genome mining for amino acid containing natural products and their connection to phenotypes. Furthermore our findings will enable the discovery of the arylomycin biosynthetic gene cluster thereby enabling future bioengineering to produce novel arylomycin analogs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support was provided by the National Institute of Health (GM085770 to B.S.M. and GM086283 and AI095125 to P.C.D.) and the Beckman Foundation. W.-T.L. was supported, in part, by a study abroad grant SAS-98116-2-US-108 from Taiwan.

References

1. Newman DJ, Cragg GM. J. Nat. Prod. 2007; 70:461–477. [PubMed: 17309302]

2. Li JW, Vederas JC. Science. 2009; 325:161–165. [PubMed: 19589993]

3. Aharonowitz Y, Cohen G, Martin JF. Annu. Rev. Microbiol. 1992; 46:461–495. [PubMed: 1444264]

4. Miao V, Coeffet-Legal MF, Brian P, Brost R, Penn J, Whiting A, Martin S, Ford R, Parr I, Bouchard M, Silva CJ, Wrigley SK, Baltz RH. Microbiology. 2005; 151:1507–1523. [PubMed: 15870461]

5. van Wageningen AM, Kirkpatrick PN, Williams DH, Harris BR, Kershaw JK, Lennard NJ, Jones M, Jones SJ, Solenberg PJ. Chem. Biol. 1998; 5:155–162. [PubMed: 9545426]

6. Robbel L, Marahiel MA. J. Biol. Chem. 2010; 285:27501–27508. [PubMed: 20522545]

7. Payne DJ, Gwynn MN, Holmes DJ, Pompliano DL. Nat. Rev. Drug. Discov. 2007; 6:29–40. [PubMed: 17159923]

8. Fischbach MA, Walsh CT. Science. 2009; 325:1089–1093. [PubMed: 19713519]

9. Yang YL, Xu Y, Straight P, Dorrestein PC. Nat. Chem. Biol. 2009; 5:885–887. [PubMed: 19915536]

10. Liu WT, Yang YL, Xu Y, Lamsa A, Haste NM, Yang JY, Ng J, Gonzalez D, Ellermeier CD, Straight PD, Pevzner PA, Pogliano J, Nizet V, Pogliano K, Dorrestein PC. Proc. Natl. Acad. Sci. U. S. A. 2010; 107:16286–16290. [PubMed: 20805502]

11. Yang YL, Xu Y, Kersten RD, Liu WT, Meehan MJ, Moore BS, Bandeira N, Dorrestein PC. Angew. Chem. Int. Ed. Engl. 2011; 50:5839–5842. [PubMed: 21574228]

12. Chambers HF, Deleo FR. Nat. Rev. Microbiol. 2009; 7:629–641. [PubMed: 19680247]

13. Otto M. Nat. Rev. Microbiol. 2009; 7:555–566. [PubMed: 19609257]

14. Gu JQ, Nguyen KT, Gandhi C, Rajgarhia V, Baltz RH, Brian P, Chu M. J. Nat. Prod. 2007; 70:233–240. [PubMed: 17284073]

15. Baltz RH. J. Antibiot. 2010; 63:506–511. [PubMed: 20648020]

16. Nguyen KT, Ritz D, Gu JQ, Alexander D, Chu M, Miao V, Brian P, Baltz RH. Proc. Natl. Acad. Sci. U. S. A. 2006; 103:17462–17467. [PubMed: 17090667]

17. Baltz RH, Miao V, Wrigley SK. Nat. Prod. Rep. 2005; 22:717–741. [PubMed: 16311632]

18. Kersten RD, Yang YL, Xu Y, Sang-Jip Nam SJ, Fenical W, Cimermancic P, Fischbach M, Moore BS, Dorrestein PC. Nat. Chem. Biol. 2011 Embargo date Oct 9 2011.

19. Rausch C, Weber T, Kohlbacher O, Wohlleben W, Huson DH. Nucleic Acids Res. 2005; 33:5799–5808. [PubMed: 16221976]

20. Li MH, Ung PM, Zajkowski J, Garneau-Tsodikova S, Sherman DH. BMC Bioinformatics. 2009; 10:185–194. [PubMed: 19531248]

21. Caboche S, Pupin M, Leclere V, Fontaine A, Jacques P, Kucherov G. Nucleic Acids Res. 2008; 36:326–331.

22. Holtzel A, Schmid DG, Nicholson GJ, Stevanovic S, Schimana J, Gebhardt K, Fiedler HP, Jung G. J. Antibiot. 2002; 55:571–577. [PubMed: 12195963]

23. Schimana J, Gebhardt K, Hoeltzel A, Schmid DG, Suessmuth R, Mueller J, Pukall R, Fiedler HP. J. Antibiot. 2002; 55:565–570. [PubMed: 12195962]

24. Paetzel M, Goodall JJ, Kania M, Dalbey RE, Page MG. J. Biol. Chem. 2004; 279:30781–30790. [PubMed: 15136583]

25. Bockstael K, Geukens N, Van Mellaert L, Herdewijn P, Anne J, Van Aerschot A. Microbiology. 2009; 155:3719–3729. [PubMed: 19696105]

26. Powers ME, Smith PA, Roberts TC, Fowler BJ, King CC, Trauger SA, Siuzdak G, Romesberg FE. J. Bacteriol. 2011; 193:340–348. [PubMed: 21075926]

27. Luo C, Roussel P, Dreier J, Page MG, Paetzel M. Biochemistry. 2009; 48:8976–8984. [PubMed: 19655811]

28. Kulanthaivel P, Kreuzman AJ, Strege MA, Belvo MD, Smitka TA, Clemens M, Swartling JR, Minton KL, Zheng F, Angleton EL. J. Biol. Chem. 2004; 279:36250–36258. [PubMed: 15173160]

29. Clardy J, Fischbach MA, Walsh CT. Nat. Biotechnol. 2006; 24:1541–1550. [PubMed: 17160060]

30. Butler MS, Buss AD. Biochem. Pharmacol. 2006; 71:919–929. [PubMed: 16289393]

31. Smith PA, Roberts TC, Romesberg FE. Chem. Biol. 2010; 17:1223–1231. [PubMed: 21095572]

32. Smith PA, Powers ME, Roberts TC, Romesberg FE. Antimicrob. Agents Chemother. 2011; 55:1130–1134. [PubMed: 21189343]

33. Roberts TC, Smith PA, Cirz RT, Romesberg FE. J. Am. Chem. Soc. 2007; 129:15830–15838. [PubMed: 18052061]

34. Dufour J, Neuville L, Zhu J. Chemistry. 2010; 16:10523–10534. [PubMed: 20658499]

35. Challis GL, Ravel J, Townsend CA. Chem. Biol. 2000; 7:211–214. [PubMed: 10712928]

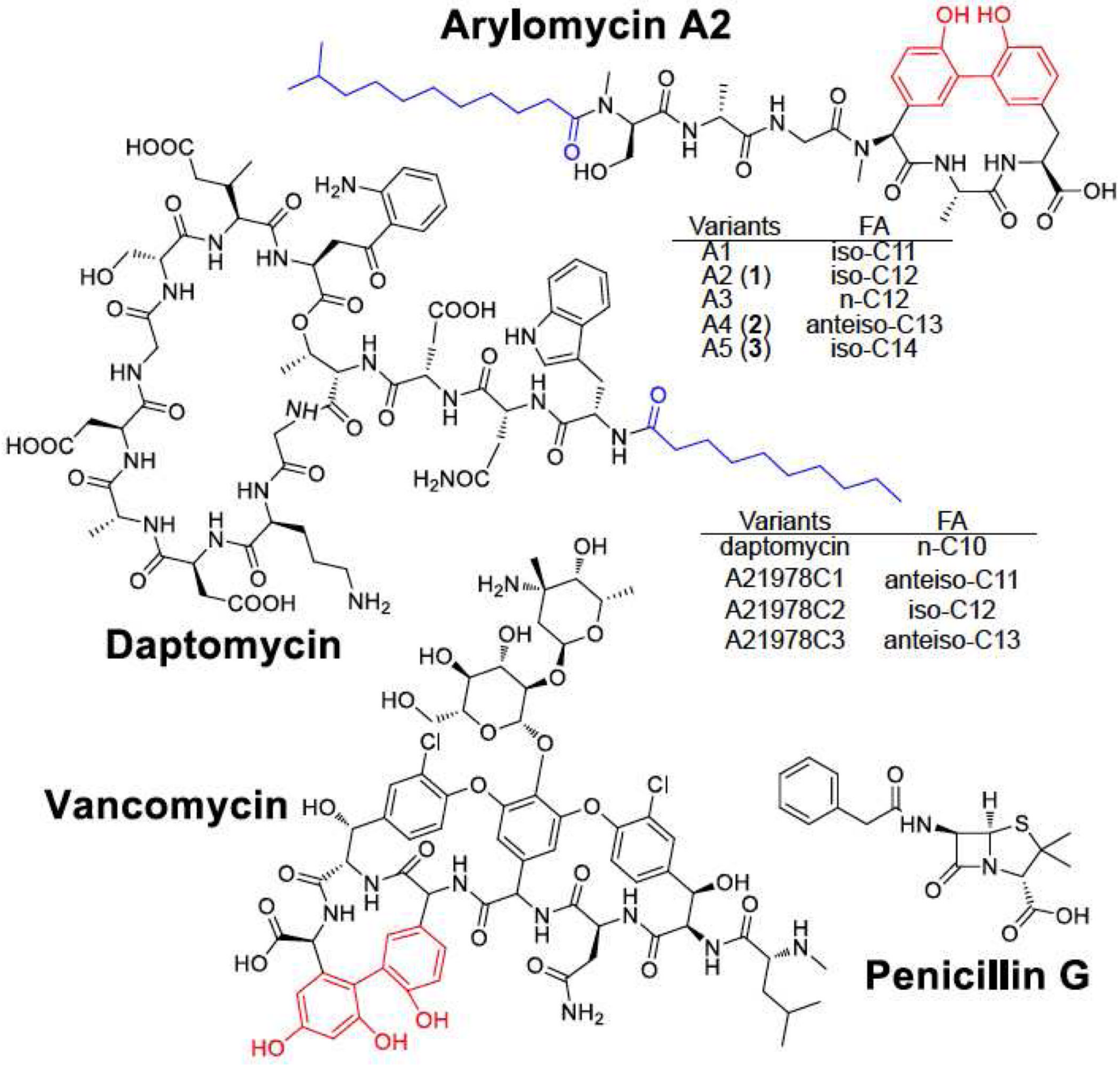
36. Stachelhaus T, Mootz HD, Marahiel MA. Chem. Biol. 1999; 6:493–505. [PubMed: 10421756]

37. Imker HJ, Krahn D, Clerc J, Kaiser M, Walsh CT. Chem. Biol. 2010; 17:1077–1083. [PubMed: 21035730]

38. Pylypenko O, Vitali F, Zerbe K, Robinson JA, Schlichting I. J Biol Chem. 2003; 278:46727–46733. [PubMed: 12888556]

39. Kersten RD, Dorrestein PC. Nat. Chem. Biol. 2010; 6:636–637. [PubMed: 20720546]

40. Walsh CT, Fischbach MA. J. Am. Chem. Soc. 2010; 132:2469–2493. [PubMed: 20121095]



Variants	FA
A1	iso-C11
A2 (**1**)	iso-C12
A3	n-C12
A4 (**2**)	anteiso-C13
A5 (**3**)	iso-C14

Variants	FA
daptomycin	n-C10
A21978C1	anteiso-C11
A21978C2	iso-C12
A21978C3	anteiso-C13

Figure 1. Structures of NRPS-derived compounds

Arylomycins and daptomycin have various components with different fatty acid (FA) (labeled in blue) chain lengths. Bis-aryl bridges are highlighted in red.

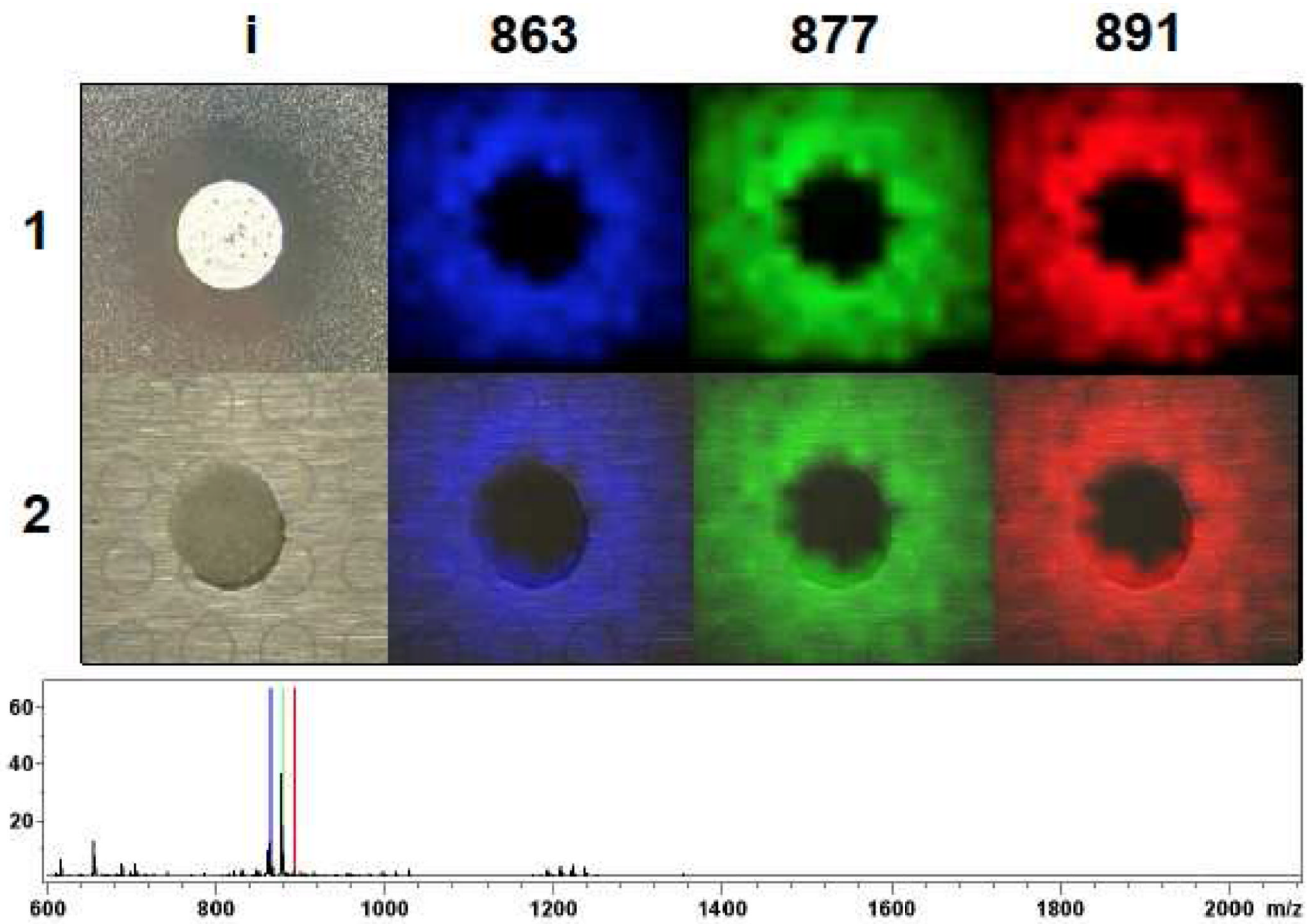
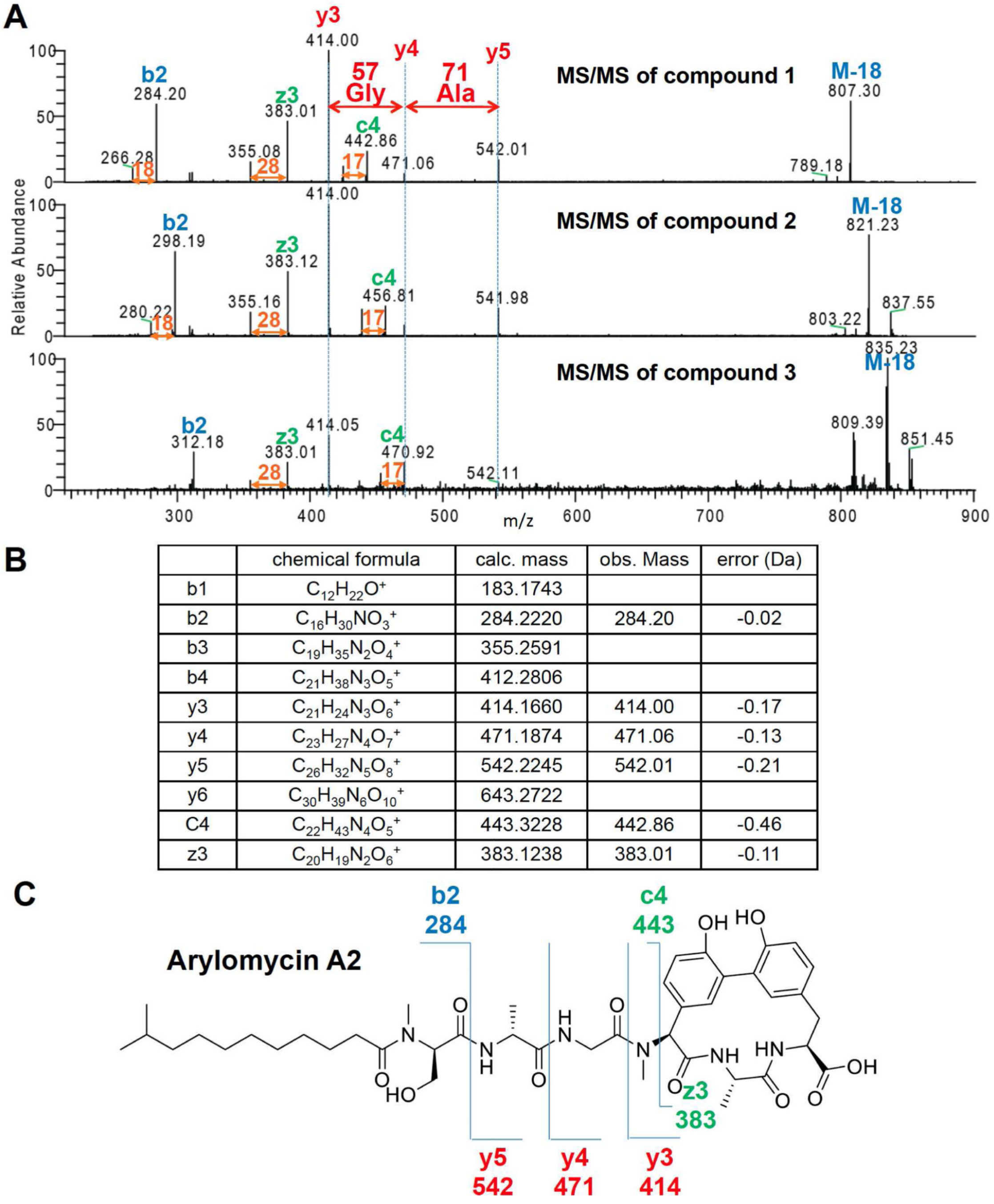


Figure 2. IMS of *S. roseosporus* spotted on top of a *S. epidermidis* lawn

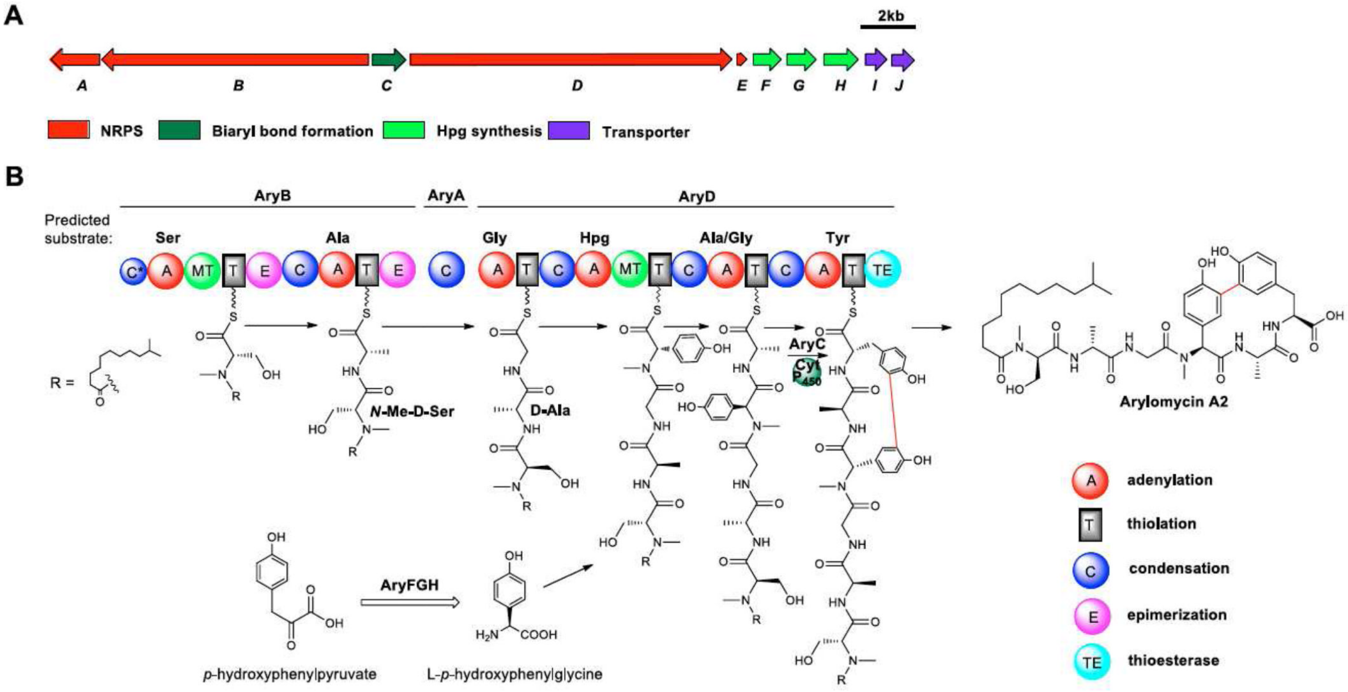
(1) ion distribution of compound **1**–**3** (863, 877, 891) observed in IMS. 1i is a photograph showing *S. roseosporus* inhibit *Staphylococci* growth. (2) Superimposition of the photograph with IMS data on top of MALDI target plate. Average mass spectrum of each IMS experiment was shown below IMS images with signals correlated to compound **1**–**3** labeled with corresponding color as displayed in images.



	chemical formula	calc. mass	obs. Mass	error (Da)
b1	$C_{12}H_{22}O^+$	183.1743		
b2	$C_{16}H_{30}NO_3^+$	284.2220	284.20	-0.02
b3	$C_{19}H_{35}N_2O_4^+$	355.2591		
b4	$C_{21}H_{38}N_3O_5^+$	412.2806		
y3	$C_{21}H_{24}N_3O_6^+$	414.1660	414.00	-0.17
y4	$C_{23}H_{27}N_4O_7^+$	471.1874	471.06	-0.13
y5	$C_{26}H_{32}N_5O_8^+$	542.2245	542.01	-0.21
y6	$C_{30}H_{39}N_6O_{10}^+$	643.2722		
C4	$C_{22}H_{43}N_4O_5^+$	443.3228	442.86	-0.46
z3	$C_{20}H_{19}N_2O_6^+$	383.1238	383.01	-0.11

Figure 3. Correlate compound 1–3 to arylomycins

(A) Alignment of IT MS/MS of compounds **1**–**3** revealed sequence tag Gly-Ala. (B) Annotated ion table corresponds to the IT MS/MS of compound **1**. (C) Ion map showing the fragmentation pattern of compound **1** correlates to arylomycin A2.



Scheme 1. Arylomycin biosynthetic gene cluster and proposed biosynthetic pathway