



**HAL**  
open science

# Integrated metabolomic, molecular networking, and genome mining analyses uncover novel angucyclines from *Streptomyces* sp. RO-S4 strain isolated from Bejaia Bay, Algeria

Rima Ouchene, Didier Stien, Juliette Segret, Mouloud Kecha, Alice Rodrigues, Carole Veckerlé, Marcelino Suzuki

## ► To cite this version:

Rima Ouchene, Didier Stien, Juliette Segret, Mouloud Kecha, Alice Rodrigues, et al.. Integrated metabolomic, molecular networking, and genome mining analyses uncover novel angucyclines from *Streptomyces* sp. RO-S4 strain isolated from Bejaia Bay, Algeria. 2022. hal-03746686v1

**HAL Id: hal-03746686**

**<https://hal.sorbonne-universite.fr/hal-03746686v1>**

Preprint submitted on 4 Oct 2022 (v1), last revised 5 Aug 2022 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1           **Integrated metabolomic, molecular networking, and genome mining**  
2           **analyses uncover novel angucyclines from *Streptomyces* sp. RO-S4 strain**  
3                           **isolated from Bejaia Bay, Algeria**

4  
5   **Rima Ouchene<sup>1,2</sup>, Didier Stien<sup>2\*</sup>, Juliette Segret<sup>2</sup>, Mouloud Kecha<sup>1</sup>, Alice M. S. Rodrigues<sup>2</sup>,**  
6   **Carole Veckerlé<sup>2</sup>, Marcelino T. Suzuki<sup>2\*</sup>**

7  
8   <sup>1</sup>Laboratoire de Microbiologie Appliquée (LMA), Faculté des Sciences de la Nature et de la  
9   Vie, Université de Bejaia, 06000 Bejaia, Algérie.

10   <sup>2</sup>Sorbonne Université, CNRS, Laboratoire de Biodiversité et Biotechnologies Microbiennes,  
11   LBBM, Observatoire Océanologique, F-66650, Banyuls-sur-mer, France.

12  
13  
14   \*Corresponding authors: Didier Stien ([didier.stien@cnrs.fr](mailto:didier.stien@cnrs.fr)), Marcelino T. Suzuki  
15   ([suzuki@obs-banyuls.fr](mailto:suzuki@obs-banyuls.fr)).

16   **Key words:** Marine *Streptomyces*, Antibacterial activity, MRSA, Metabolomic analysis,  
17   Molecular Networking, Genome mining.

18  
19

20 **Abstract**

21 Multi-omic approaches have recently made big strides towards the effective exploration of  
22 microorganisms and accelerating the discovery of new bioactive compounds. We combined  
23 metabolomic, molecular networking, and genomic-based approaches to investigate the  
24 metabolic potential of the *Streptomyces* sp. RO-S4 strain isolated from the polluted waters of  
25 Bejaia Bay in Algeria. Antagonistic assays against *methicillin-resistant Staphylococcus aureus*  
26 with RO-S4 organic extracts showed an inhibition zone of 20 mm by the agar diffusion method,  
27 and its minimum inhibitory concentration was 16 µg/mL. A molecular network was created  
28 using GNPS and annotated through the comparison of MS/MS spectra against several  
29 databases. The predominant compounds in the RO-S4 extract belonged to the angucyclines  
30 family. Three compounds were annotated as known metabolites, while all the others were  
31 putatively new to Science. Notably, all compounds had fridamycin-like aglycones, and several  
32 of them had a lactonized D ring analogous to that of urdamycin L. The whole genome of  
33 *Streptomyces* RO-S4 was sequenced to identify the biosynthetic gene cluster (BGC) encoding  
34 for these angucyclines, which yielded a draft genome of 7,497,846 bp with 72.4% G+C content.  
35 Subsequently, a genome mining analysis revealed 19 putative biosynthetic gene clusters,  
36 including a grincamycin-like BGC with a high similarity to that of *Streptomyces* sp. CZN-748  
37 previously reported to also produce mostly open fridamycin-like aglycones. As the ring-  
38 opening process leading to these compounds is still not defined, we performed comparative  
39 analysis with other angucycline BGCs and advanced some hypotheses to explain the ring-  
40 opening and lactone formation, possibly linked to the uncoupling between the activity of GcnE  
41 and GcnM homologues in the RO-S4 strain. The combination of metabolomic and genomic  
42 approaches greatly improved the interpretation of the metabolic potential of the RO-S4 strain.

43

## 44 **Introduction**

45 The emergence of novel mechanisms of antimicrobial resistance is increasing and spreading  
46 worldwide, posing a challenge to mankind. The World Health Organization has stated that  
47 antibiotic resistance will be one of the biggest threats to human health in the future<sup>1</sup>. Multidrug-  
48 resistant organisms have become common not only in hospital settings but also in the wide  
49 community settings, suggesting that reservoirs of antibiotic-resistant bacteria are present  
50 outside hospitals (reviewed by Munita and Arias)<sup>2</sup>. This antibiotic resistance crisis has been  
51 attributed to overuse and inappropriate use of these drugs, as well as the lack of antimicrobial  
52 drug development by the pharmaceutical industry due to reduced economic incentives and  
53 difficult regulatory requirements<sup>3,4</sup>. Methicillin-resistant *Staphylococcus aureus* (MRSA) is the  
54 most common cause of nosocomial infections as it is very capable of developing antibiotic  
55 resistance<sup>5,6</sup>. Many challenges are faced by laboratories and clinicians in the diagnosis and  
56 treatment of MRSA infections, some of which were highlighted by Edwards and coworkers<sup>7</sup>. It  
57 is thus clear that the search for new bioactive compounds to combat antimicrobial resistance is  
58 a research priority.

59 Marine environments represent a largely unexplored source for the isolation of new  
60 microorganisms<sup>8</sup>. They display a unique combination of environmental conditions and  
61 organisms with distinct metabolic capabilities to adapt and thrive<sup>9,10</sup>. A large number of  
62 bioactive compounds have been isolated from marine organisms<sup>11,12,13</sup>, particularly  
63 Actinobacteria, which have been a main source of natural products in the past<sup>14,15,16</sup>. Among  
64 the latter, the *Streptomyces* genus is well known for its ability to produce a wide range of  
65 bioactive metabolites as well as antibacterial, anticancer, antifungal, antiparasitic, and  
66 immunosuppressive agents<sup>17,18</sup>, representing the most prolific source of bioactive metabolites  
67 that have been approved for clinical use, notably as antibiotics<sup>19</sup>.

68 Traditionally, activity-guided fractionation of metabolite extracts, followed by purification and  
69 characterization of metabolites, has commonly been used for natural product research, but this  
70 approach often leads to the isolation of already known molecules. More recently, significant  
71 developments in genetics, genomics, and data analysis have greatly changed natural product  
72 research, leading to a new era in the emerging field of systems biology. Consequently, new  
73 avenues were opened for the discovery of novel compounds from actinomycetes  
74 (e.g.<sup>20,13,21,22,23,16</sup>). Interestingly, metabolomics and genomics approaches have proven to be  
75 efficient and promising tools for defining phenotypes in a dynamic context, with the potential  
76 to reduce rediscovery rates<sup>24,25,26</sup>, and several tools have been designed for this purpose, as  
77 reported by Caesar and colleagues<sup>27</sup>. These approaches have been successfully applied to study

78 the chemical diversity of marine bacteria and to uncover novel bioactive molecules<sup>28,29,30</sup>,  
79 despite the challenges encountered due to the complexity of biological matrices<sup>31</sup>. More  
80 recently, molecular networking, a tandem mass spectrometry (MS/MS) data organizational  
81 approach, has been introduced in the field of drug discovery<sup>32</sup>, and the combination of system  
82 analyses involving multi-omics data and genome-scale, metabolic network models has greatly  
83 contributed to exploring bioactive Actinobacteria<sup>33</sup>, and have great potential to accelerate  
84 natural product discovery.

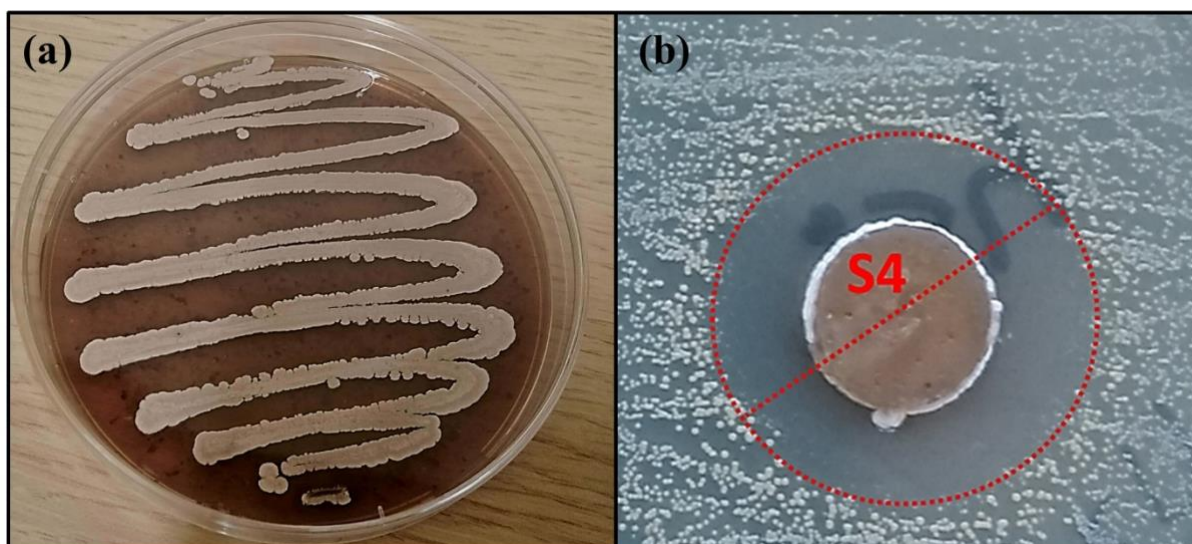
85 Here, we investigated the secreted metabolome of *Streptomyces* sp. RO-S4 in the quest for  
86 novel antimicrobial compounds against antibiotic-multi-resistant *S. aureus* (MRSA). For this  
87 purpose, we used a metabolomic approach based on ultra-performance high resolution tandem  
88 mass spectrometry (UPLC-HRMS/MS) followed by the creation of a molecular network using  
89 the Global Natural Product Social Molecular Networking (GNPS) analysis. These analyses  
90 were combined with a genomic analysis to refine and further annotate the structural hypothesis  
91 generated, and conversely, to understand the biosynthesis of the major angucyclines produced  
92 by this strain.

## 93 **Results**

### 94 **Isolation and antimicrobial assays of the RO-S4 strain**

95 The RO-S4 strain was isolated from the bay of Bejaia City in Algeria. It grew well on the M2  
96 medium, showing substrate growth typical for *Streptomyces* strains, with a brown powdery  
97 aspect and producing a dark-brown pigment (Fig. 1a). Antimicrobial activity was first evaluated  
98 against the MRSA strain by the agar diffusion method. It exhibited antagonistic activity against  
99 this bacterium with an inhibition zone estimated at 20 mm (Fig. 1b). The 16S rRNA gene  
100 sequence indicated that the strain belongs to the *Streptomyces* genus, with 99.79% identity to  
101 *Streptomyces albogriseolus* NRRL B-1305 (T). The MIC of the ethyl acetate extract produced  
102 by the RO-S4 strain was measured by the broth microdilution method on a 96-well plate. A  
103 MIC value of 16 µg/mL was observed against the MRSA ATCC 43300 strain.

104



105

106 **Figure 1. (a):** The morphological appearance of *Streptomyces sp.* RO-S4 strain grown on M2  
107 medium for 12 days at 28 °C. **(b):** MRSA inhibitory potential of the RO-S4 strain evaluated by  
108 the agar diffusion method.

### 109 **Untargeted metabolomic analysis and Molecular Networking**

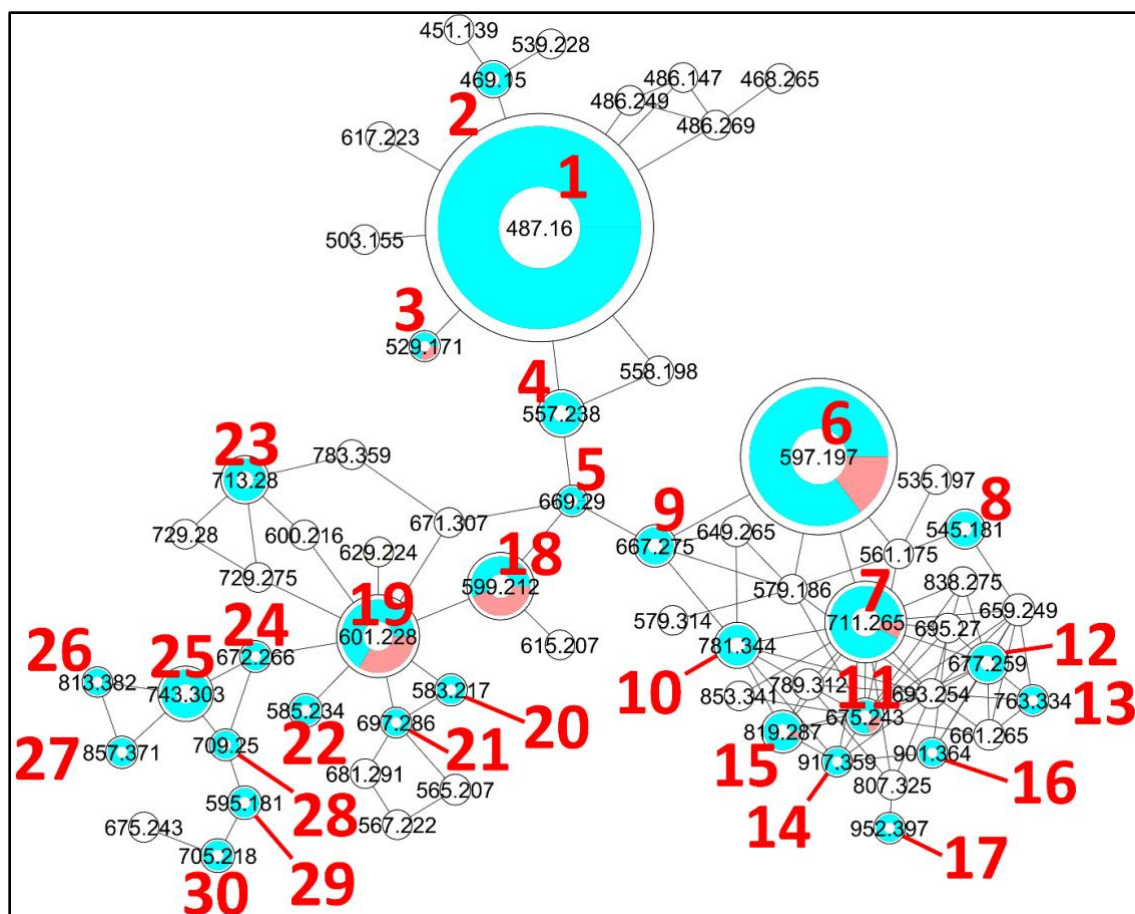
110 The metabolomic profile of the active ethyl acetate (EtOAc) crude extract was investigated  
111 using UPLC-HRMS. An examination of the MS and collision induced MS/MS (MS<sup>2</sup> hereafter)  
112 spectra of the main constituents of the mixture indicated that most metabolite molecular ions  
113 fragmented to yield a product at  $m/z$  487.1600 corresponding to the formula C<sub>25</sub>H<sub>27</sub>O<sub>10</sub><sup>+</sup> (Calcd.  
114 487.1599). The same ion was also detected as a protonated molecular ion corresponding to  
115 compound **1** (Table 1). Compound **1** was annotated either as aquayamycin or as fridamycin A  
116 or B by various dereplication tools. A comparison with experimental spectra from the MoNA  
117 database confirmed that compound **1** was fridamycin A or its diastereomer fridamycin B. The  
118 MS<sup>2</sup> spectra of fridamycin A and aquayamycin are very similar (Fig. S3, Supporting  
119 Information). Nevertheless, two fragment ions are diagnostic. These are the ions at  $m/z$  347.09  
120 and 427.14, the relative intensities of which are very low in fridamycin A when compared to  
121 those of aquayamycin.

122 As mentioned above, the [1+H]<sup>+</sup> ion was also produced as a fragment resulting from in-source  
123 fragmentation of many metabolites in the profile. The MS<sup>2</sup> spectra of all ions leading to a  
124 487.1599 fragment clustered in the same node in the molecular network (MN). The MS<sup>2</sup>  
125 spectrum obtained for the protonated molecular ion of compound **1** was compared to all other  
126 MS<sup>2</sup> spectra of equimassic ions found whenever other more complex metabolites were

127 fragmented in source. The fragmentation patterns were all similar, confirming that many  
128 metabolites in the MN were fridamycin A or B analogues. It was deduced that the strain  
129 biosynthesizes the central fridamycin core and then adds various substituents to generate its  
130 diverse products. Notably, the pentacyclic aquayamycin subunit was not detected in any of the  
131 annotated metabolites. This observation was supported by the presence of a biosynthetic gene  
132 cluster very close to that of *Streptomyces* sp. CNZ-748<sup>34</sup> which also produces a majority of  
133 fridamycin-like compounds.

134 The parameters for MN were set to construct the best representative network containing all  
135 fridamycin analogs (Fig. 2). The MN was constituted of three groups of ions. In the first group,  
136 the annotation was propagated from **1** as follows. Compound **2** is a dehydrofridamycin based  
137 on its molecular formula and MS<sup>2</sup> spectrum. Compound **3** molecular formula was C<sub>27</sub>H<sub>28</sub>O<sub>11</sub>,  
138 which might be annotated as an acetyl-fridamycin A or B, while the position of the acetyl group  
139 could not be inferred from MS<sup>2</sup>. Compound **4** whose molecular formula was C<sub>30</sub>H<sub>36</sub>O<sub>10</sub> (*m/z*  
140 for [M+H]<sup>+</sup> 669.2903, calcd. 669.2905) was a fridamycin bearing a C<sub>5</sub>H<sub>10</sub> substituent. A  
141 fridamycin isopentyl ester was thought to be a reasonable putative structure based on the  
142 biosynthetic considerations below. Compound **5** could not be annotated more precisely than  
143 just with its molecular formula, but its MS<sup>2</sup> spectrum is also one of a fridamycin analog. These  
144 considerations indicated that **5** was new to Science.





145

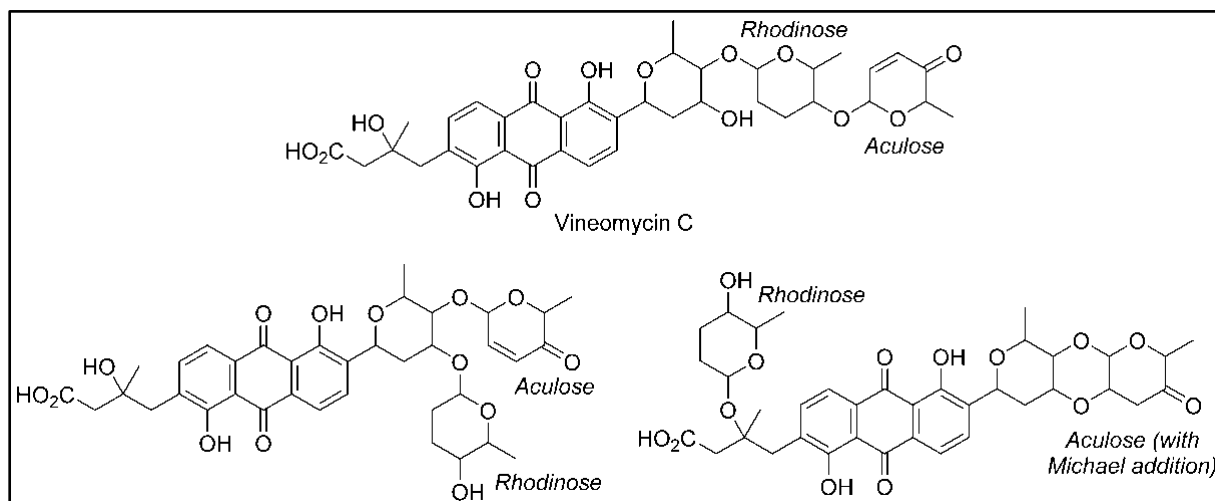
146 **Figure 2.** Molecular network generated with the GNPS Molecular Networking tool. The  
147 diameter of the nodes represents the total extracted ion chromatogram integration of the  
148 corresponding ion peak(s) and the blue/red pie chart represents the proportions of each isomer  
149 in the cluster. The colorless nodes are clusters of MS<sup>2</sup> spectra of ions produced by in-source  
150 fragmentation of diverse compounds and were thus neither annotated nor integrated  
151 (integration was set to 0). Only protonated molecular ions were considered for integration  
152 measurement.

153 In the second part of the fridamycin MN, compound **6a** protonated molecular ion was at *m/z*  
154 597.1973, corresponding to the formula C<sub>31</sub>H<sub>33</sub>O<sub>12</sub><sup>+</sup> (Calcd. *m/z* 597.1967). The formula and  
155 fragmentation pattern were consistent with the annotation of **6a** as fridamycin D<sup>35</sup>. Annotation  
156 as fridamycin D was also supported by Sirius. An isomer of **6a** (**6b**) was also detected in smaller  
157 relative proportions in the strain's metabolomic profile. Another major constituent of the strain  
158 metabolome was compound **7a**, whose protonated molecular ion at *m/z* 711.2645, corresponded  
159 to the formula C<sub>37</sub>H<sub>43</sub>O<sub>14</sub><sup>+</sup> (Calcd. 711.2647). The molecular formula indicated that angucycline  
160 **7a** may be vineomycin C<sup>36</sup>. It was annotated as vineomycin C by Sirius as well, although with  
161 57% confidence. In the collision-induced MS<sup>2</sup> spectrum of **7a** protonated molecular ion, the



162 rhodinose (or its distereoisomer amicetose) and the aculose oxonium ions were present at  $m/z$   
163 115.0756 and 111.0443 (Table S1), respectively, while the presence of the fridamycin A  
164 aglycone was ascertained based on the common fridamycin A fragment ions (Fig. S16,  
165 Supporting Information). Nonetheless, the first fragmentation steps in the MS<sup>2</sup> spectrum  
166 suggested that the sugar sequence might be different to that of vineomycin C. The protonated  
167 molecular ion lost both water ( $m/z$  for C<sub>37</sub>H<sub>41</sub>O<sub>13</sub><sup>+</sup> 693.2258) and a C<sub>6</sub>H<sub>12</sub>O<sub>3</sub> group ( $m/z$  for  
168 C<sub>31</sub>H<sub>31</sub>O<sub>11</sub><sup>+</sup> 597.1938) which could only be assigned to the rhodinose/amicetose moiety. Hence,  
169 the rhodinose could not be placed in between the aglycone and the aculose, as in vineomycin  
170 C, and compound **7a** must be considered new to Science, unless the published structure of  
171 vineomycin C requires revision. Possible annotations are reported in Fig. 3. These alternative  
172 proposals could account for the preferential fragmentation observed in the MS<sup>2</sup> spectrum. The  
173 proximity of **7a** with fridamycin D (**6a**) in the MN suggested that the most probable annotation  
174 might be the one in which aculose underwent a Michael addition, as in **6a**.

175



177

177 **Figure 3.** Published structure of vineomycin C and putative structures for compound **6a**. The  
178 sugar groups annotated as rhodinose are either rhodinose or amicetose (indistinguishable  
179 stereoisomers in MS).

180 A minor isomer of **7b** also appeared in the MN under the same cluster. Based on the same  
181 considerations as for **7a**, **7b** could not be annotated as vineomycin C either. It was therefore  
182 new to Science. Next to both **6** and **7**, compound **8** protonated molecular ions at  $m/z$  545.1808  
183 corresponded to the formula C<sub>31</sub>H<sub>29</sub>O<sub>9</sub><sup>+</sup> (Calcd. 545.1806). The formula suggested that **8** might  
184 be marangucycline B<sup>37</sup>. Sirius also annotated this compound as marangucycline B, although  
185 with 58% confidence only. Marangucycline B is an analog of fridamycin D with a modified  
186 aglycone. The presence of an aculose subunit has been confirmed by MS<sup>2</sup> ( $m/z$  111.0442), but

187 the fragmentation pattern was not attributable to marangucycline B. Hence, while this  
188 compound is probably not marangucycline B it could not be annotated further. Angucycline **9**  
189 was also new to Science. Its protonated molecular ion at  $m/z$  667.2753 corresponded to the  
190 formula  $C_{36}H_{43}O_{12}^+$  (Calcd. 667.2749). In the  $MS^2$  spectrum, the fragmentation from the  
191 protonated molecular ion to product at  $m/z$  579.1851 corresponded to the loss of a neutral  
192 fragment the formula of which was  $C_5H_{12}O$  (Fig. S22, Supporting Information). This group has  
193 been annotated above as an isopentanol, probably linked to the carboxylic acid moiety. A  
194 carbon monoxide loss was also detected from  $m/z$  561.1750 to 533.1812, indicating that the  
195 carboxylic acid side chain of the aglycone should be present. The  $MS^2$  spectrum also showed a  
196 hydrated aculose oxonium ion at  $m/z$  129.0549 and its dehydrated form at  $m/z$  111.0443. For all  
197 these reasons, compound **9** was annotated as shown in Table 1. The protonated molecular ion  
198 of compound **10** was found at  $m/z$  781.3438, a mass that corresponds to the formula  $C_{42}H_{53}O_{14}^+$   
199 (Calcd. 781.3430). An isopentanol, a rhodnose/amicetose, and an aculose were pointed out in  
200 the fragmentation spectrum of the parent ion. Rhodnose and aculose oxonium ions were also  
201 detected at  $m/z$  115.0754 and 111.0442, respectively, showing that **10** should be annotated as  
202 the isopentanol ester of **7a/b**. The formula of compound **11a** protonated molecular ion was  
203 found to be  $C_{42}H_{53}O_{14}^+$  (exp.  $m/z$  675.2438, calcd. 675.2436). This corresponded to the  
204 grincamycin H molecular formula<sup>38</sup>. Nonetheless, the  $MS^2$  spectrum indicated that **11a**  
205 successively lost rhodnose/amicetose and aculose, and therefore could not be annotated as  
206 grincamycin H. The mass of the protonated aglycone after oses fragmentation was  $m/z$   
207 451.1388, corresponding to a doubly dehydrated fridamycin A. We hypothesized that one  
208 hydroxyl group of the olivose side chain might be dehydrated, and that the second  $H_2O$  loss  
209 might be explained by a ring closing of the lactone as in urdamycin L<sup>39</sup>, a possibility supported  
210 by the analysis of the biosynthetic gene cluster discussed below. Compound **11b** was  
211 presumably a minor diastereoisomer of **11a** due to the high similarity of their respective  $MS^2$   
212 spectra. Compound **12** protonated molecular ion at  $m/z$  677.2597 indicated the formula  
213  $C_{37}H_{41}O_{12}^+$  (Calcd. 677.2593). The  $MS^2$  fragmentation pattern showed the successive losses of  
214 rhodnose/amicetose, and aculose. The masses of the fragments generated by cleavage of the  
215 aglycone were all shifted by 2 Da relative to those of compound **11a**, indicating that the two  
216 additional hydrogens were in the center of the aglycone. Thus, it could be deduced that **12** was  
217 likely the hydroquinone form of **11a**. Compound **13** is a minor metabolite and could not be  
218 annotated with reasonable confidence. The molecular formula of compound **14** was  $C_{49}H_{56}O_{17}$   
219 (exp.  $m/z$  917.3601 ( $[M+H]^+$ ), calcd. 917.3590). The  $MS^2$  spectrum indicated a loss of an  
220 aculosyl-rhodnose (or aculosyl-amicetose) moiety. Then the product at  $m/z$  675.2443 lost the

221 neutral group C<sub>6</sub>H<sub>8</sub>O (a dehydro-rhodinose), indicating that the aculosyl-rhodinose moiety was  
222 linked to another rhodinose. Then the aglycone ion at  $m/z$  451.1398 was produced by the loss  
223 of another aculose. The molecular weight of this aglycone ion (fridamycin A -2 H<sub>2</sub>O) along  
224 with the proximity of other lactonic aglycones in the MN spoke in favor of **14** also being a  
225 lactone derivative, as reported in Table 1. The fragmentation spectrum of compound **15** was not  
226 very clear and **15** could not be annotated with enough confidence. At  $m/z$  901.3653, compound  
227 **16** protonated molecular ion indicated the formula C<sub>49</sub>H<sub>57</sub>O<sub>16</sub><sup>+</sup> (calcd. 901.3641). In MS<sup>2</sup>, the  
228 ion **16**+H<sup>+</sup> lost an aculosyl-rhodinose group to give the product at  $m/z$  659.2499, which then  
229 lost dehydro-rhodinose. Further dehydration produced an ion at  $m/z$  545.1790 which again lost  
230 aculose to yield the aglycone ion at  $m/z$  435.1471. This aglycone was doubly dehydrated  
231 compared to fridamycin A, indicating that the C-oliviosyl group in **16** may be dehydrated. The  
232 structure proposed in Table 1 appeared to be a reasonable hypothesis for **16**. The molecular  
233 formula of compound **17** is C<sub>49</sub>H<sub>58</sub>O<sub>18</sub> (exp.  $m/z$  952.3960 ([M+NH<sub>4</sub>]<sup>+</sup>), calcd. 952.3961). This  
234 molecular formula and sodium adduct fragmentation pattern in which two successive aculosyl-  
235 rhodinose losses were recorded were compatible with the annotation of **17** as vineomycin B<sup>40</sup>.

236 In the third part of the fridamycin MN, compounds **18a** and **18b** were annotated as isomers with  
237 the molecular formula C<sub>31</sub>H<sub>34</sub>O<sub>12</sub> (exp.  $m/z$  599.2124 for [M+H]<sup>+</sup>, calcd. 599.2123). Both  
238 isomers fragmented extensively in the ESI source to produce the fridamycin A protonated  
239 molecular ion [1+H]<sup>+</sup> losing C<sub>6</sub>H<sub>8</sub>O<sub>2</sub>, *i.e.*, a dehydro-cinerulose A moiety. The cinerulose A  
240 oxonium ion was also detected in MS<sup>2</sup>, while the aglycone fragmentation was very similar to  
241 what was recorded for [1+H]<sup>+</sup>. Angucyclins **18a** and **18b** were therefore annotated as  
242 cinerulosyl-fridamycin A or B. The position of the cinerulosyl side chain was not determined  
243 and may not be identical for both isomers. Compounds **19a** and **19b** were annotated as isomers  
244 with the molecular formula C<sub>31</sub>H<sub>36</sub>O<sub>12</sub> (exp.  $m/z$  601.2284 for [M+H]<sup>+</sup>, calcd. 601.2279). Both  
245 isomers also fragmented extensively in the ESI source to produce the fridamycin A protonated  
246 molecular ion [1+H]<sup>+</sup> losing C<sub>6</sub>H<sub>10</sub>O<sub>2</sub>, *i.e.*, dehydro-rhodinose/amicetose subunit, the  
247 corresponding oxonium of which was also present in MS<sup>2</sup>. As mentioned above, the MS<sup>2</sup>  
248 spectrum of the aglycone protonated ion formed by in-source fragmentation was identical to  
249 the one of [1+H]<sup>+</sup>, therefore confirming that **19a** or **b** should not be annotated as grincamycin  
250 L<sup>41</sup>. Instead, **19a** and **19b** were annotated as rhodinოსyl- and/or amicetosyl-fridamycin A/B and  
251 should be considered as new to Science. Angucycline **19a** was one of the major constituents in  
252 the profile of the strain. Compound **20** molecular formula was C<sub>31</sub>H<sub>34</sub>O<sub>11</sub> (exp.  $m/z$  583.2170  
253 for [M+H]<sup>+</sup>, calcd. 583.2174). Its MS<sup>2</sup> spectrum showed a rhodinose/amicetose subunit and a

254 dehydrated protonated aglycone. Therefore, compound **20** was annotated as a dehydro-**19a**.  
255 Compound **21** molecular formula was  $C_{37}H_{45}O_{13}$  (exp.  $m/z$  697.2861 for  $[M+H]^+$ , calcd.  
256 697.2855). Its  $MS^2$  spectrum revealed the successive loss of two rhodinose/amicetose subunits,  
257 yielding a dehydrated protonated aglycone. It was thus annotated as shown in Table 1.  
258 Compound **22** molecular formula was  $C_{31}H_{37}O_{11}$  (exp.  $m/z$  585.2333 for  $[M+H]^+$ , calcd.  
259 585.2330). Its  $MS^2$  spectrum showed the loss of one rhodinose/amicetose subunit, yielding a  
260 protonated deoxy-aglycone. The structure of the aglycone could not be readily inferred from  
261 the  $MS^2$  spectrum and compound **22** could not be annotated further. Compound **23**'s molecular  
262 formula was  $C_{37}H_{44}O_{14}$  (exp.  $m/z$  713.2782 for  $[M+H]^+$ , calcd. 713.2804). A  
263 rhodinose/amicetose was visible in MS, but the  $MS^2$  spectrum was impure and further  
264 annotation was not possible. Compound **24** molecular formula was  $C_{34}H_{41}NO_{13}$  (exp.  $m/z$   
265 672.2655 for  $[M+H]^+$ , calcd. 672.2652). Its  $MS^2$  spectrum showed the loss of 1  
266 rhodinose/amicetose subunit yielding a product at  $m/z$  558.1983 ( $C_{28}H_{32}NO_{11}^+$ ), which then  
267 successively lost water,  $CH_2O_2$  (formic acid or  $H_2O+CO$ ), and  $C_2H_5N$  to generate a protonated  
268 didehydro-fridamycin A at  $m/z$  451.1393. This fragmentation pattern was compatible with **24**  
269 being an alanine amide of **19**, as shown in Table 1. The presence of an alanine subunit was also  
270 supported by the fragment ion at  $m/z$  90.0550, corresponding to an alaninium ion<sup>42</sup>. Compound  
271 **25** molecular formula was  $C_{37}H_{46}N_2O_{14}$  (exp.  $m/z$  743.3026 for  $[M+H]^+$ , calcd. 743.3022). In-  
272 source fragmentation indicated the successive loss of a dehydro-rhodinose/amicetose group and  
273 a  $C_6H_{10}N_2O_2$  subunit. The  $MS^2$  spectrum highlighted the loss of one dehydro-  
274 rhodinose/amicetose and two alanine subunits to generate the protonated didehydro-fridamycin  
275 A at  $m/z$  451.1393 (see Fig. S64, Supporting information). Overall, compound **25** could be  
276 annotated with high confidence as shown in Table 1. Neighboring minor compound **26** in the  
277 MN could not be annotate. Compound **27** was an analog of **25** with one rhodinose/amicetose  
278 subunit more, the presence of which could be ascertained by examination of both in-source  
279 fragmentation scheme and the collision-induced  $MS^2$  spectrum. Compound **28** molecular  
280 formula was  $C_{37}H_{40}O_{14}$  (exp.  $m/z$  709.2490 for  $[M+H]^+$ , calcd. 709.2491). The  $MS^2$  spectrum  
281 clearly indicated that the protonated molecular ion lost both dehydro-rhodinose/amicetose and  
282 a  $C_6H_8O_4$  neutral fragment, therefore confirming that this compound was new and should not  
283 be annotated as saprolmycin B<sup>43</sup>. However, the annotation remained ambiguous and **28** was not  
284 annotated further. Compounds **29** and **30** were analogs of **28**; **29** did not have the  
285 rhodinose/amicetose subunit, while **30** had an aculose in place of the rhodinose/amicetose. All  
286 the annotated compounds'  $MS^2$  spectra were provided in the supporting information (Figs S1-

287 [S74](#), Supporting Information), and a list of sugars potentially linked to the angucyclins is  
 288 provided in Table [S1](#).

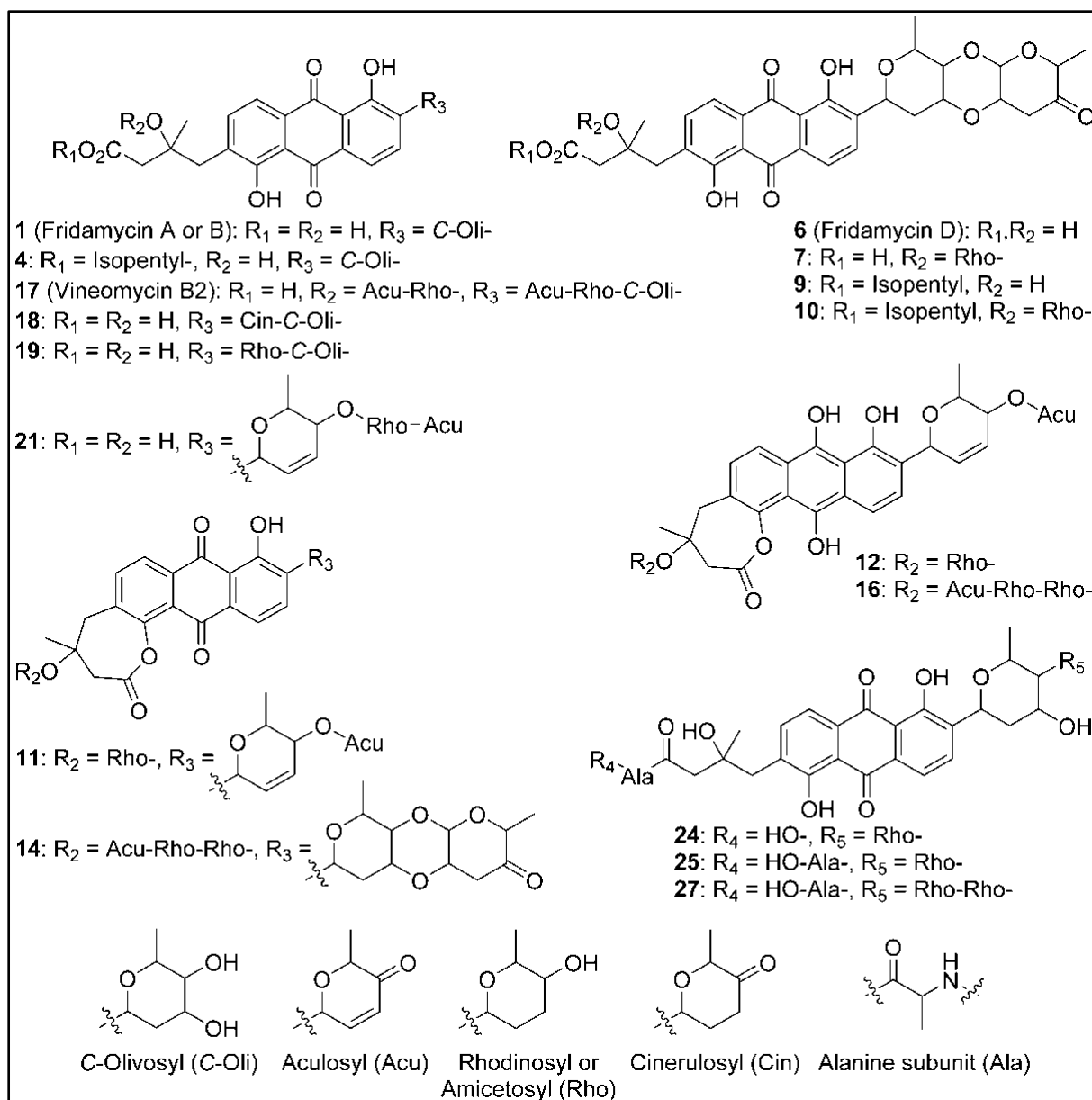
289 **Table 1.** Annotated angucyclines in the metabolomic profile of strain RO-S4

	<i>m/z</i>	Ion type	Molecular formula	$\#$	<i>t<sub>R</sub></i> (min)	Annotation <sup>b</sup>
<b>1</b>	487.1600	[M+H] <sup>+</sup>	C <sub>25</sub> H <sub>26</sub> O <sub>10</sub>	1	8.59	Fridamycin A or B (Fig. 4)
<b>2<sup>c</sup></b>	469.1492	[M+H] <sup>+</sup>	C <sub>25</sub> H <sub>24</sub> O <sub>9</sub>	1	8.48	[1-H <sub>2</sub> O]
<b>3<sup>c</sup></b>	529.1708	[M+H] <sup>+</sup>	C <sub>27</sub> H <sub>29</sub> O <sub>11</sub>	1	8.58	Acetyl-fridamycin A or B
<b>4<sup>c</sup></b>	557.2383	[M+H] <sup>+</sup>	C <sub>30</sub> H <sub>36</sub> O <sub>10</sub>	1	11.69	See Fig. 4
<b>5<sup>c</sup></b>	669.2903	[M+H] <sup>+</sup>	C <sub>36</sub> H <sub>44</sub> O <sub>12</sub>	1	13.08	n.a.
<b>6</b>	a: 597.1973	[M+H] <sup>+</sup>	C <sub>31</sub> H <sub>32</sub> O <sub>12</sub>	2	a: 11.01	a: Fridamycin D (Fig. 4)
	b: 597.1969				b: 12.57	
<b>7<sup>c</sup></b>	a: 711.2650	[M+H] <sup>+</sup>	C <sub>37</sub> H <sub>42</sub> O <sub>14</sub>	2	a: 11.87	See Fig. 4
	b: 711.2656				b: 11.17	
<b>8<sup>c</sup></b>	545.1808	[M+H] <sup>+</sup>	C <sub>31</sub> H <sub>28</sub> O <sub>9</sub>	1	14.21	n.a.
<b>9<sup>c</sup></b>	667.2751	[M+H] <sup>+</sup>	C <sub>46</sub> H <sub>42</sub> O <sub>12</sub>	1	13.66	See Fig. 4
<b>10<sup>c</sup></b>	781.3436	[M+H] <sup>+</sup>	C <sub>42</sub> H <sub>52</sub> O <sub>14</sub>	1	14.31	See Fig. 4
<b>11<sup>c</sup></b>	a: 675.2438	[M+H] <sup>+</sup>	C <sub>37</sub> H <sub>38</sub> O <sub>12</sub>	2	a: 13.74	See Fig. 4
	b: 675.2436				b: 13.10	
<b>12<sup>c</sup></b>	677.2597	[M+H] <sup>+</sup>	C <sub>37</sub> H <sub>40</sub> O <sub>12</sub>	1	11.02	See Fig. 4
<b>13<sup>c</sup></b>	763.3328	[M+H] <sup>+</sup>	C <sub>42</sub> H <sub>50</sub> O <sub>13</sub>	1	12.88	n.a.
<b>14<sup>c</sup></b>	917.3601	[M+H] <sup>+</sup>	C <sub>49</sub> H <sub>56</sub> O <sub>17</sub>	1	12.03	See Fig. 4
<b>15<sup>c</sup></b>	a: 819.2863	[M+H] <sup>+</sup>	C <sub>43</sub> H <sub>46</sub> O <sub>16</sub>	2	a: 12.64	n.a.
	b: 819.2869				b: 12.95	
<b>16<sup>c</sup></b>	901.3647	[M+H] <sup>+</sup>	C <sub>49</sub> H <sub>56</sub> O <sub>16</sub>	1	12.94	See Fig. 4
<b>17</b>	952.3957	[M+NH <sub>4</sub> ] <sup>+</sup>	C <sub>49</sub> H <sub>58</sub> O <sub>18</sub>	1	12.27	Vineomycin B2 (Fig. 4)
<b>18<sup>c</sup></b>	a: 599.2124	[M+H] <sup>+</sup>	C <sub>31</sub> H <sub>34</sub> O <sub>12</sub>	2	a: 10.11	See Fig. 4
	b: 599.2125				b: 10.53	
<b>19<sup>c</sup></b>	a: 601.2284	[M+H] <sup>+</sup>	C <sub>31</sub> H <sub>36</sub> O <sub>12</sub>	2	a: 10.06	See Fig. 4
	b: 601.2288				b: 10.95	

<b>20<sup>c</sup></b>	583.2170	[M+H] <sup>+</sup>	C <sub>31</sub> H <sub>34</sub> O <sub>11</sub>	1	8.62	[ <b>19a</b> -H <sub>2</sub> O]
<b>21<sup>c</sup></b>	697.2861	[M+H] <sup>+</sup>	C <sub>37</sub> H <sub>44</sub> O <sub>13</sub>	1	9.32	See Fig. 4
<b>22<sup>c</sup></b>	585.2333	[M+H] <sup>+</sup>	C <sub>31</sub> H <sub>36</sub> O <sub>11</sub>	1	10.87	n.a.
<b>23</b>	713.2782	[M+H] <sup>+</sup>	C <sub>37</sub> H <sub>44</sub> O <sub>14</sub>	1	11.88	n.a.
<b>24<sup>c</sup></b>	672.2655	[M+H] <sup>+</sup>	C <sub>34</sub> H <sub>41</sub> NO <sub>13</sub>	1	9.70	See Fig. 4
<b>25<sup>c</sup></b>	743.3031	[M+H] <sup>+</sup>	C <sub>37</sub> H <sub>46</sub> N <sub>2</sub> O <sub>14</sub>	1	9.44	See Fig. 4
<b>26<sup>c</sup></b>	813.3816	[M+H] <sup>+</sup>	C <sub>42</sub> H <sub>56</sub> N <sub>2</sub> O <sub>14</sub>	1	11.55	n.a.
<b>27<sup>c</sup></b>	857.3709	[M+H] <sup>+</sup>	C <sub>42</sub> H <sub>56</sub> N <sub>2</sub> O <sub>14</sub>	1	10.24	See Fig. 4
<b>28<sup>c</sup></b>	709.2490	[M+H] <sup>+</sup>	C <sub>37</sub> H <sub>40</sub> O <sub>14</sub>	1	10.95	n.a.
<b>29<sup>c</sup></b>	595.1808	[M+H] <sup>+</sup>	C <sub>31</sub> H <sub>30</sub> O <sub>12</sub>	1	9.47	n.a.
<b>30<sup>c</sup></b>	705.2182	[M+H] <sup>+</sup>	C <sub>37</sub> H <sub>36</sub> O <sub>14</sub>	1	11.83	n.a.

290 <sup>a</sup>Number of isomers detected in the MN cluster. <sup>b</sup>Proposed structures are annotations based on  
291 literature data and automatic and manual analysis of MS<sup>2</sup> spectra. n.a.: not annotated. <sup>c</sup>Novel  
292 compound.





293

294 **Figure 4.** Annotated metabolites from Strain RO-S4. Stereocenters are intentionally drawn as  
 295 undefined. Only the raw formula of the substituents can be inferred from the mass spectra. Their  
 296 developed formulas and relative positions are putative.

### 297 Whole genome sequencing

298 The whole genome of *Streptomyces* sp. RO-S4 was sequenced using the Illumina Novaseq  
 299 technology. The complete genome consisted of 7,497,846 bp with 72.4% G+C content. The  
 300 closest genome to *Streptomyces* sp. RO-S4 sequenced from a type strain was that of *S.*  
 301 *althoticus* JCM 4344 (assembly GCA\_014649355; 90.73% ANI), and that of *S. tendae* strain  
 302 139 (CP04395; 95.84%) for a non-type strain. The genome of *Streptomyces albogriseolus*

303 NRRL B-1305, the closest strain based on 16S rRNA gene similarity was not available at the  
304 time of analysis.

### 305 **Secondary metabolite biosynthetic gene clusters of *Streptomyces* sp. RO-S4**

306 The genome of the RO-S4 strain was analyzed by the antibiotics & Secondary Metabolite shell  
307 (antiSMASH) to determine its putative biosynthetic capabilities. A total of 19 putative  
308 biosynthetic gene clusters were annotated (Table S2, supporting information), including three  
309 types of polyketide synthases BGCs [Type 1 (T1PKS), Type 2 (T2PKS), and type 2 (T3PKS)]  
310 polyketide synthases, class I Lanthipeptides, a Lasso peptide, a Ribosomally Synthesized and  
311 Post-Translationally Modified Peptide (RiPP), an Ectoine, a Terpene, Phenazines, and a  
312 Butyrolactone BGC. In addition, four hybrid clusters were recovered that were composed of 1)  
313 T2PKS, Oligosaccharide Phenazine, Siderophore, 2) RiPP-like, betalactam, Terpene, 3) Two  
314 hybrid Non-ribosomal Peptide Synthase (NRPS), and T1PKS. The analysis showed that 17 out  
315 of the 19 identified BGCs showed high content similarity with known BGCs, five of which (3,  
316 5, 7, 13, and 17) showed 100% content similarity with known BGCs. Two clusters (Cluster 10  
317 and 16) were annotated as orphan BGCs for which no homologous gene clusters could be  
318 identified, suggesting that they could be responsible for the biosynthesis of novel natural  
319 products or natural products with no characterized BGCs. Many of these clusters are known to  
320 encode genes linked to the production of biologically active natural compounds, such as  
321 antibiotics. Notably, we have recovered a T2PKS BGC very similar to those linked to the  
322 biosynthesis of angucycline compounds, consistent with the metabolomic analysis.

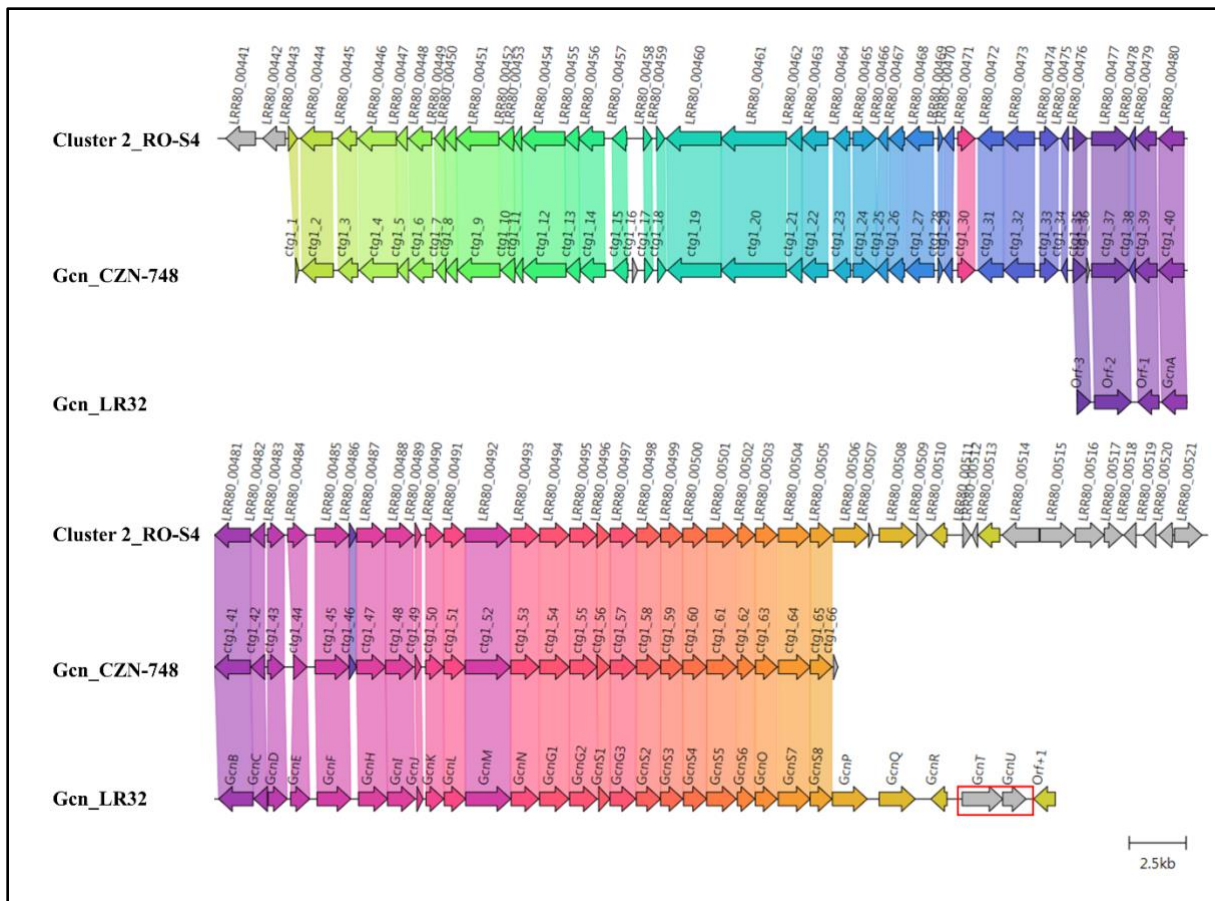
### 323 **Description of the Angucycline Biosynthetic Gene Cluster**

324 AntiSMASH analysis using an unannotated genomic DNA sequence revealed that Cluster 2 in  
325 Contig ROS4\_2 of the assembly displayed a high synteny (> 65% common genes) to that of  
326 BGCs linked to several angucycline compounds, such as grincamycin (97%), saprolmycin E  
327 (83%), saquayamycin A (75%), landomycin A (71%), and saquayamycin Z (67%). All these  
328 compounds share a common tetracyclic angular benz[a]anthraquinone aglycone. Due to the  
329 predominance of tricyclic aglycones ("open" aglycone(s) hereafter) among the major  
330 metabolites of RO-S4, we performed an in-depth analysis of Cluster 2.

331 Cluster 2 contains genes putatively involved in the biosynthesis and modification of the  
332 aglycone core. A typical set of genes responsible for an angucycline core assembly named  
333 "minimal PKS" has been identified supporting the synthesis of angucycline like molecules by  
334 his cluster. This included three genes: a ketoacyl synthase  $\alpha$  (LRR80\_00487), a ketoacyl  
335 synthase  $\beta$ /chain length factor CLF (LRR80\_00488) and an acyl carrier protein (ACP)

336 (LRR80\_004989). Two possible cyclase genes (LRR80\_00486 and LRR80\_00491) were also  
337 annotated, which are likely responsible for the polyketide chain cyclization into the  
338 benz[a]anthracene structure. In addition, this cluster harbors two genes encoding oxygenase  
339 enzymes (LRR80\_00485 and LRR80\_00492) probably involved in the modification of the  
340 aglycone and possibly in the lactonization and opening of the angular aglycone cycle (see  
341 discussion below). (Keto)reductase-coding genes, including LRR80\_00490 and  
342 LRR80\_00470, were annotated to exhibit a high degree of sequence similarity to known  
343 enzymes involved in the modification of aromatic polyketides. Three genes (LRR80\_00495,  
344 LRR80\_00496, and LRR80\_00498) likely associated with the glycosylation steps showed high  
345 similarity to genes coding glycosyltransferases (GTs) in other angucyclines. All the annotated  
346 genes involved in the BGC of Cluster 2 and their homologs are listed in Table S3.

347 The closest BGCs to RO-S4 Cluster-2 are those of the grincamycin-producing *Streptomyces*  
348 *lusitanus* SCSIO LR32 (Gcn LR32)<sup>44</sup>, and *Streptomyces* sp. CZN-748 (Gcn CZN-748)<sup>34</sup>  
349 graciously provided by the authors). We performed a synteny analysis comparing the three  
350 BGCs and notably, when PROKKA<sup>45</sup>-annotations were used for RO-S4 and CZN-748, the  
351 *gcnM* ORF was absent in these strains, as previously described by Shang and co-workers<sup>34</sup>. In  
352 contrast, the BGC annotated by AntiSMASH from genomic sequences identified these ORFs.  
353 This difference could be explained by a possible tRNA<sub>ala</sub> in the region coding the *gcnM*  
354 orthologs (Fig. S75, Supporting Information). In addition, the comparison between the BGC of  
355 LR32 identified two missing genes (*gcnU* and *gcnT*) in the BGC of ROS4\_2, whereas the region  
356 downstream of *gcnS8* was not present in the available BGC of strain CZN-748 (Fig. 5). Cluster  
357 2 of RO-S4 showed near complete synteny and a higher average amino acid identity (99.8 %  
358 for 28 common ORFs) to the grincamycin BGC of CZN-748, which fits the observation that  
359 both strains produce a majority of "open" aglycone angucyclines, whereas LR32 (93.2 %  
360 average amino acid identity for 28 common ORFs) produces primarily tetracyclic angucyclines,  
361 as previously noted by Shang and co-workers<sup>34</sup>.



362  
 363 **Figure 5.** Comparison between Cluster 2 of *Streptomyces* sp. RO-S4, the Grincamycin Gene  
 364 Cluster of *Streptomyces lusitanus* SCSIO LR32, and the BGC of Grincamycin-producing  
 365 *Streptomyces* sp. CNZ-748. Gene neighborhoods representative of the compared BGCs are  
 366 shown aligned with an arbitrary color scheme using clinker to highlight the conserved genes.  
 367 Missing genes in RO-S4 compared to LR-32 were highlighted in red.

### 368 Discussion

369 Here we report the use of a combined genomic-metabolomic approach to investigate the  
 370 antagonistic potential of the *Streptomyces* sp. RO-S4 strain isolated from a polluted marine  
 371 environment. Based on 16S rRNA gene sequencing and genomic analysis, the strain belongs to  
 372 the genus *Streptomyces*, but it was not possible to assign it to a species. RO-S4 extracts show  
 373 inhibitory activity against MRSA with a MIC of 16 µg/mL.

374 Metabolomic analyses of the crude extract produced by the RO-S4 strain using mass-  
 375 spectrometry-based molecular networking revealed diverse angucycline derivatives as  
 376 dominant products, which have mostly (but not exclusively) been linked to the *Streptomyces*  
 377 genus<sup>46,47</sup>. Angucyclines represent the largest group of type 2 PKS natural products produced  
 378 by actinobacteria, and they show diverse pharmacological activities including cytotoxicity,  
 379 antitumor, antibacterial, and antiviral properties<sup>48,49</sup>.

380 We have demonstrated that many of the compounds identified by our untargeted metabolomic  
381 analysis are novel to Science, and this high diversity of novel molecules can be explained by  
382 the ability of HRMS to highlight minor compounds even though some of the major metabolites  
383 are also new to Science. Our annotation of the RO-S4 angucycline-like compounds was further  
384 supported by genomic analysis.

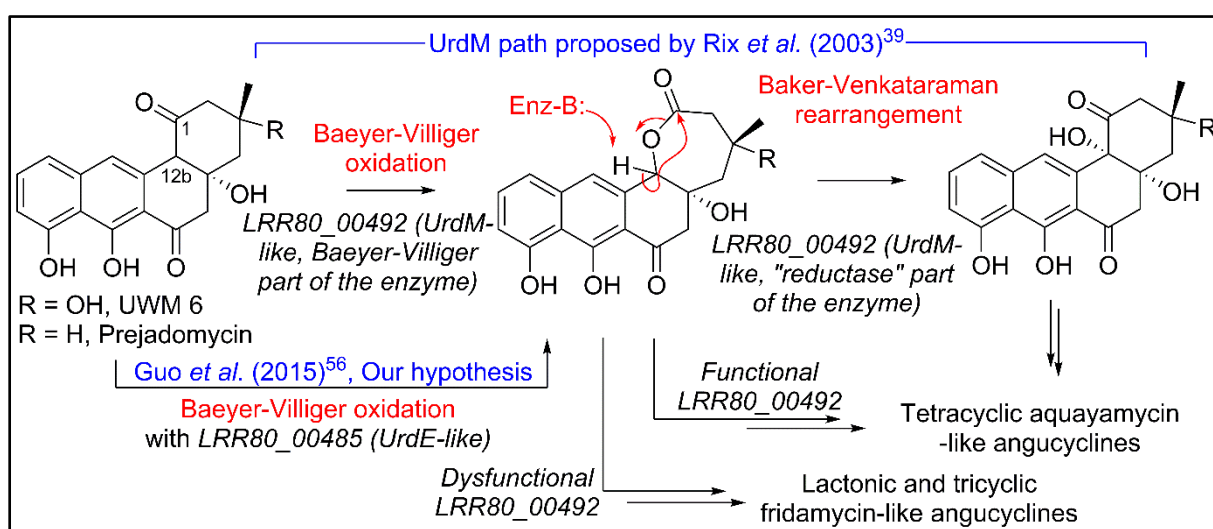
385 The combination of metabolomic and genome analysis provides some interesting insights into  
386 the biosynthesis of angucyclines. Cluster 2 shows high content, synteny, and sequence  
387 similarity to previously described BGCs of grincamycin-like products and is certainly  
388 responsible for producing the structures predicted by the MN analysis. Since the RO-S4 strain  
389 and *Streptomyces* CNZ-748 produce primarily tricyclic glycosylated structures, while LR32  
390 and many other strains produce tetracyclic angucyclines, we were interested in possible  
391 enzymatic processes that could lead to the biosynthesis of these tricyclic aglycons<sup>34</sup>. Several  
392 studies have indicated that oxygenase complexes are required for cyclic C-C bond cleavage, in  
393 particular Baeyer-Villiger type oxygenases<sup>50</sup>. In the case of RO-S4 products, we hypothesized  
394 that this reaction would take place by an oxidation of the C12b–C1 single bond of the aglycone  
395 (an UWM6-like molecule), prior or after glycosylation, and a subsequent hydrolysis of the  
396 lactone (Figure 6). We have therefore focused the analysis on putative Baeyer-Villiger mono-  
397 oxidases (BVMOs) in Cluster 2 of the RO-S4 strain.

398 The observation that – as in the case of the grincamycin BGC of strain CZN-748 – annotation  
399 with PROKKA<sup>45</sup> failed to identify an ORF downstream of the T2PKS synthases and the cyclase  
400 putatively involved in the generation of the angular cycle (the region coding for GcnM in *S.*  
401 *lusitanus* LR32<sup>44</sup>), suggested that this ORF could be involved in the ring opening process. Since  
402 AntiSMASH identified an ORF both in RO-S4 (LRR80\_00492) and CNZ-748 (CTG1-52)  
403 when a genomic sequence was used as the query, we determined that PROKKA failed to  
404 annotate the ORF since its Aragorn<sup>51</sup> step identified a putative tRNA<sub>ala</sub> (Fig. S75, Supporting  
405 Information) in its complementary strand. However, the facts that the putative tRNA is in the  
406 reverse strand, that it contains mismatches in the side harpins, and that 4 more canonical  
407 tRNA<sub>ala</sub> are coded in the genome, suggest that this tRNA<sub>ala</sub> could have been misidentified. We  
408 performed an RNA fold analysis<sup>52</sup> that identified a highly probable and low-entropy hairpin  
409 loop that could potentially affect the translation of this ORF (Fig. S76, Supporting Information)  
410 and possibly decreases the production of the coded protein.

411 The LRR80\_00492 ORF codes a hybrid FAD-dependent oxidase-reductase (GcnM in the  
412 grincamycin BGC) homologous to UrdM, that has been linked to C12b hydroxylation in the



413 biosynthesis of urdamycins by *S. fradiae* TÛ 2717<sup>53,39</sup>. Furthermore, a mutant with an in-frame  
 414 deletion of the reductase domain of UrdM produced small amounts of urdamycin-L, a product  
 415 containing an oxygen between C12b and C1, leading to the hypothesis that UrdM is involved  
 416 in the C12b–C1 bond oxidation and subsequent lactone Baker-Venkataraman rearrangement  
 417 leading to the tetracyclic skeleton of aquayamycin-like angucyclins (Fig. 6)<sup>39</sup>. However, in this  
 418 model, low levels, or absence of LRR80\_00492 due to the secondary structure described above  
 419 would not lead to fridamycin-like aglycones as those in RO-S4 (Table 1), and another BVMO  
 420 should be responsible for the oxidation of the C12b–C1 bond in UWM 6 (or other  
 421 intermediates) leading to compounds 11, 12 and, 16 and fridamycin-like aglycones.



422

423 **Figure 6.** Hypotheses for conversion of UWM 6 or prejadomycin into angucyclines.

424 The BGC of RO-S4 and of all grincamycin-producing strains contains a second FAD-dependent  
 425 putative BVMO product of LRR80\_00485 (GcnE in the grincamycin BGC) that is homologous  
 426 to FAD-dependent monooxygenases involved in angucycline modifications (e.g. UrdE, PgaE,  
 427 BexE, CabE; Fig S77 and S78, Supporting Information). In earlier studies, UrdE was  
 428 hypothesized to directly hydroxylate different positions of the aglycon (C6, C12, C12b)<sup>53,54,55</sup>  
 429 in urdamycin biosynthesis, but more recent evidence have suggested that its homologue PgaE  
 430 might also oxidize the C12b–C1 bond of prejadomycin leading to the tricyclic aglycons of  
 431 gaudimycins D and E<sup>56</sup>. *In vitro* assays using enzymes heterologously expressed in *E. coli* also  
 432 showed that PgaE/CabE oxidizes UWM6 and is dependent on the PgaM<sub>red</sub> homologue CabV to  
 433 complete the hydroxylation of UWM6 at C12b<sup>57</sup>.

434 Since in the RO-S4 BGC there are two possible FAD-dependent mono-oxidases that could be  
 435 involved in oxidation of the C12b–C1 bond and subsequent ring opening of RO-S4 and CNZ-



436 748, we attempted to compare the sequences of both ORFs to different enzymes oxidizing  
437 analogous cyclic compounds, including MtmOIV, shown to perform a Baeyer-Villiger  
438 oxidation and ring opening of premithramycin B to mithramycin. *Blastp* analyses indicated that  
439 the LRR80\_00492 (UrdM like)/mtmOIV alignment was shorter and had a lower overall score,  
440 but a higher number of identical positions, whereas the LRR80\_00485 (UrdE-like)/mtmOIV  
441 alignment was longer and had a higher total score, but with fewer identical positions.  
442 Phylogenetic analyses including UrdE homologues and the oxidase portion of UrdM  
443 homologues separated these oxidases into two groups, with maximum likelihood and distance  
444 methods showing that LRR80\_00485 was slightly closer to MtmOIV than LRR80\_00492, and  
445 in a subclade including PgaE (Figs. [S77](#) and [S78](#), Supporting Information).

446 In aggregate, these results lead to different possibilities that could be tested in the future using  
447 genetic modifications of the different ORFs in the angucyline BGC of the RO-S4 strain or  
448 CNZ-748. Hypothesis 1): the product of LRR80\_00492 would be solely responsible for  
449 oxidation of the C12b–C1 bond. This hypothesis is supported by the prediction that this enzyme  
450 has several AAs unique to RO-S4 and CNZ-748. On the other hand, since the reductase portion  
451 of the enzyme is present, one would expect that tetracyclic angucyclines with a hydroxylated  
452 C12b would be produced. Hypothesis 2): the most likely hypothesis based on our work on the  
453 RO-S4 strain is that the LRR80\_00492 ORF is inactive due to the presence of a tRNA<sub>Ala</sub> in the  
454 coding region or that its translation is affected by secondary structure, in which case  
455 LRR80\_00485 would generate the lactone via a Baeyer-Villiger oxidation, possibly allowing  
456 for a later opening of the ring. Other alternative hypotheses that could be related to the ring  
457 opening are: Hypothesis 3): that the LRR80\_00492 ORF would be partly transcribed due to the  
458 mRNA secondary structure, which would allow its oxidase portion to be transcribed but not the  
459 reductase, much like the case with the *urdM* partial knockout mutant that produces urdamycin  
460 L<sup>39</sup> and is responsible for C12b–C1 single bond oxidation, and Hypothesis 4): another BVMO  
461 enzyme not in the BGC could be responsible for ring opening. The genome of RO-S4 codes for  
462 another enzyme with a slightly higher *blastp* score when queried with MtmOIV. That ORF is  
463 present in a hybrid NRPS-T1PKS hybrid BGC similar to that of polyoxypeptin A BGC<sup>58</sup>.  
464 However, the function of that homologue (ORF4) in the polyoxypeptin A BGC has not yet been  
465 described.

466 In addition to these oxygenases, three genes are presumed to code for glycosyltransferase (GT)  
467 enzymes, which show similarities to known GTs involved in angucycline biosynthesis. The  
468 first, RO-LRR80\_00495, shows high homology to GcnG1 (96.98%<sup>44</sup>), sqnG1 (88.14%<sup>59</sup>),

469 sprGT1 (86.74%<sup>60</sup>), and SchS10 (81.63%<sup>61</sup>). The second (LRR80\_00496) is closely related to  
470 GcnG2 (92.46%<sup>44</sup>), sprGT2 (83.92%<sup>60</sup>), sqnG2 (81.91%<sup>59</sup>), and schS9 (74.74%<sup>61</sup>). The third  
471 GT (LRR80\_00498) is most similar to enzymes involved in the glycosylation of D-olivose at  
472 the C9 position of several angucycline-like molecules [ex. SchS7 in the Sch-47554 biosynthetic  
473 gene cluster<sup>61</sup>; SprGT3 in saprolmycin biosynthesis<sup>60</sup>; UrdGT2 in urdamycin biosynthesis in  
474 *Streptomyces fradiae* Tü 2717<sup>62</sup>, among others], supporting well the predicted structures in  
475 Table 1, all of which are predicted to have been glycosylated with an olivose at the C9-position.  
476 SchS9 and SchS10 are thought to be *O*-glycosyltransferases involved in the biosynthesis of  
477 Sch-47554<sup>61</sup>. Genetic studies using heterologous expression and targeted gene disruption have  
478 shown that SchS7 attaches D-amicetose at C-9 and SchS9 further extends the saccharide chain,  
479 while SchS10 attaches L-aculose at the C-3 position<sup>63</sup>. The SqnGT1-G3 are glycotransferases  
480 involved in the biosynthesis of saquayamycin A in *Streptomyces sp.* KY40-1. According to  
481 genetic experimentation, sqnG2 was identified as catalyzing both *O*- and *C*-glycosylations<sup>59</sup>.

482 Hence, based on similarities between known GTs and the sugars annotated by metabolomic  
483 analysis, we hypothesize that a D-olivose is added to C9 by LRR80\_00498 and further *O*-  
484 glycosylated with rhodinose/amicetose by one of the three glycosylases, and in some cases (ex.  
485 compound **9**), experiencing a Michael addition, or further glycosylation as in compounds **19**  
486 and **27** (and in previously described angucyclines). Interestingly, all compounds glycosylated  
487 at position C3 in RO-S4 appear to have a rhodinose/amicetose as the sugar moiety, as do all  
488 grincamycins. It is worth noting that the composition and varying lengths of the oligosaccharide  
489 chains in angucyclines have a considerable impact on their biological potential, as previously  
490 reported by Elshahawi *et al.*<sup>64</sup>.

491 In addition to the rare lactonized D cycle, we also detected other modifications that, to the best  
492 of our knowledge, have so far not been shown for fridamycin-like molecules, including alanyl  
493 amidation(s) (compounds **25-28**). Amide modifications have been previously shown in  
494 fridamycin G<sup>65</sup>, and fridamycin I produced by *Actinokineospora spheciospongiae*<sup>66</sup>.  
495 Fridamycin G contains an ethanolamine moiety and was produced heterologously, and the  
496 amidation was attributed to a process linked to the host, since the BGC source (*S. cyanogenus*  
497 S136) does not produce it. Fridamycin I contains a benzylamine moiety, the biosynthetic origin  
498 of which was not discussed. As compounds **25-28** contain multiple alanine moieties, we  
499 speculate that fridamycin-like molecules could undergo peptide-like elongation with  
500 aminoacids (alanine or *L-p*-hydroxyphenylglycine in the case of fridamycin I), perhaps via an  
501 NRPS in a manner analogous to what has been described in the biosynthesis of actinomycin-

502 D<sup>67</sup>. The presence of an *hpgT* homologue – the enzyme linked to L-*p*-hydroxyphenylglycine  
503 production in actinobacteria<sup>68</sup> – in the *Actinokineospora spheciospongiae*, supports this  
504 hypothesis.

## 505 **Conclusions**

506 Our combination of untargeted UPLC-HRMS/MS metabolomic, molecular networking, and  
507 genomic analyses generated many structural and biosynthetic hypotheses for targeted structural  
508 determination and genetic manipulation, possibly using recently developed gene editing  
509 approaches (e.g.<sup>69,70</sup>). This approach does not substitute traditional isolation-NMR structural  
510 analyses and genetic manipulations, which ultimately will be needed to confirm these  
511 hypotheses. However, as it yields structural information of many minor compounds linked to  
512 BGCs in a relatively fast manner, it can streamline and accelerate pipelines of discovery of new  
513 drugs, biosynthetic pathways, and enzymes, and hopefully inspire the discovery to novel  
514 antibiotics effective against multi-resistant microorganisms.

## 515 **Methods**

### 516 **Isolation of the RO-S4 strain**

517 The RO-S4 strain was isolated from polluted seawater that was collected from the coastline of  
518 Bejaia City (36°43'55.2"N5°04'37.9"E), Algeria on August 2017. It was isolated after filtration  
519 onto a 0.22 µm pore size membrane filter as described by<sup>71</sup> and laid onto solid M2 medium  
520 prepared according to Jensen *et al.*<sup>72</sup> with small modifications , which consisted of starch (10  
521 g), casein bovine milk (1 g), microbiological agar (18 g), and natural 100% of seawater (1 L),  
522 and 1 mL (per liter of final medium) of trace salt solution that was prepared according to  
523 Shirling and Gottlieb<sup>73</sup>.

### 524 **Molecular identification of the RO-S4 isolate**

525 The initial molecular identification of the RO-S4 strain was based on the 16S rRNA gene  
526 sequence. Genomic DNA was extracted from the grown strain using the Wizard® Genomic  
527 DNA Purification Kit (Promega, USA) according to the manufacturer's instructions. PCR and  
528 sequencing were realized as previously described<sup>74</sup>, utilizing universal primers recommended  
529 for bacteria 27F mod: 5'AGRGTTCGATCMTGGCTCAG 3' and 1492R mod:  
530 5'TACGGYTACCTTGTTAYGACTT 3'. The PCR product was purified with a purification  
531 kit (Promega, USA), and then sequenced by the dideoxy termination reaction using an AB3130  
532 DNAXl sequencer. The obtained 16S rRNA sequence was identified by comparison to the

533 EZBiocloud database (<https://www.ezbiocloud.net/>) recommended by Yoon and co-workers<sup>75</sup>.  
534 The RO-S4 16S rRNA gene sequence was deposited in the GenBank database under the  
535 accession number (MW448345)

### 536 **Antimicrobial Assays**

537 The antibacterial potential of RO-S4 isolate was evaluated against *methicillin-resistant*  
538 *Staphylococcus aureus* (MRSA) ATCC 43300 by the agar diffusion method<sup>76</sup>. 8 mm diameter  
539 agar cylinders of the RO-S4 strain (M2 medium, incubation for 14 days at 28 °C) were inserted  
540 into Muller Hinton plates previously seeded with the targeted bacterium at 10<sup>7</sup> UFC/mL. The  
541 plates were placed for 2 h at 4 °C and antibacterial activity was estimated by measuring the  
542 inhibition zone around the agar disc after incubation of the plates for 24 h at 37 °C.

### 543 **Culture strain and the production of raw extract**

544 The production of bioactive compounds by the selected strain was carried out by agar surface  
545 fermentation (ASF), according to Nkanga and Hagedorn<sup>77</sup>. Briefly, the RO-S4 isolate was  
546 initially grown on M2 agar plates. After 14 days, the mycelium layers were peeled off and  
547 extracted overnight in ethyl acetate (EtOAc), covering the entire surface, and the ethyl acetate  
548 extract was drawn. Subsequently, the organic extract was concentrated under vacuum with a  
549 rotary evaporator at 40 °C and then stored at -80 °C until further analysis. A control,  
550 uninoculated medium, was extracted with the same protocol.

### 551 **Minimum Inhibitory Concentration (MIC) of the RO-S4 crude extract**

552 The minimal inhibitory concentration (MIC) of the RO-S4 EtOAc extract was evaluated against  
553 the MRSA ATCC 43300 strain by the broth microdilution method as recommended by the  
554 Clinical and Laboratory Standards Institute<sup>78</sup>. The assays were performed in serial dilutions (in  
555 triplicate) in 96-well plates. Briefly, the EtOAc extract was diluted in DMSO and tested at  
556 different concentrations ranging from 256 to 0.5 µg/mL. The targeted bacterium culture was  
557 prepared in Muller Hinton broth at 2•10<sup>5</sup> UFC/mL. Afterward, 10 µl of the test bacterial culture  
558 was pipetted into each well. The last column (column 12) with no inoculum served as a sterility  
559 control, while wells that were not treated with the crude extract served as a negative control  
560 (column 11). The final volume of each well was adjusted to 100 µL. The microplate was shaken  
561 gently, then incubated for 24 h at 37 °C. Inhibition was evaluated as well, where the growth  
562 medium appeared clear, indicating that the test extract prevented the growth or killed the  
563 bacteria.

## 564 **UHPLC-HRMS Profiling**

565 The protocol for high-resolution Full MS data dependent MS<sup>2</sup> analyses was adapted from  
566 previous reports<sup>79,80</sup>. Here, crude bacterial and culture medium (M2) extracts were dissolved in  
567 MeOH at a concentration of 1.5 mg/mL. Pure methanol injections were used as blanks for  
568 metabolomics. In HPLC, the solvent system was a mixture of water (solution A) with increasing  
569 proportions of acetonitrile (solution B), both solvents modified with 0.1% formic acid. Here,  
570 the gradient was as follows: 5% B 5 min before injection, then from 1 to 12 min, a linear  
571 increase of B up to 100%, followed by 100% B for 8 min.

## 572 **MS/MS Molecular Networking analysis and Spectra Annotation**

573 The molecular network was constructed using the Global Natural Product Social networking  
574 (GNPS) platform available at: (<https://gnps.ucsd.edu>) as recommended by Wang and  
575 collaborators<sup>32</sup>, using the molecular networking (MN) tool. The MS<sup>2</sup> data of the crude extract,  
576 solvent (blank) and culture media were converted from RAW to mzXML files using the  
577 Proteowizard MSConvert tool version (3.0.20104), then uploaded to GNPS. For MN  
578 construction, the precursor ion mass tolerance was set at 0.005 Da and the MS<sup>2</sup> fragment ion  
579 tolerance was set at 0.01 Da. A network was created where edges were filtered to have a cosine  
580 score above 0.77 and 11 or more matched peaks. The maximum size of a molecular family was  
581 set at 85. The MS<sup>2</sup> spectra in the network were searched against the 'GNPS spectra library'. All  
582 matches between network and library spectra were required to have a score above 0.7 and at  
583 least 6 matched peaks. Visualization of the molecular network was performed in Cytoscape  
584 (3.8.0) which allowed its visualization as a network of nodes and edges<sup>81</sup>. Redundancies and  
585 adducts were cleared manually. In our Fig. 2, node numbers are consensus parent masses, node  
586 size is linked to the relative molecular ion intensity based on peak area measured from the  
587 extracted ion chromatogram. Peak areas were measured automatically with the FreeStyle  
588 Genesis algorithm, sometimes modified manually if found unfitting, and then pasted manually  
589 into the Cytoscape table. This information was also used to create pie charts in which each  
590 portion represents the relative peak area of different isomers included in the same node (each  
591 GNPS node is a cluster of MS<sup>2</sup> spectra that may come from different isomeric protonated  
592 molecular ions). For nodes gathering protonated molecular ions and in-source fragments of  
593 higher molecular weight compounds, only protonated molecular ion integrations were included  
594 for peak area information. Any ion present in the network solely due to in-source fragmentation  
595 was given the arbitrary extracted ion intensity of 0 and was white in the network. Furthermore,  
596 spectra of interest were manually annotated using different databases and tools, including

597 Sirius<sup>82</sup>, Metfrag<sup>83</sup> available at (<https://msbi.ipb-halle.de/MetFrag/>), Pubchem, Sci-Finder, and  
598 Mass Bank of North America (MoNa, <https://mona.fiehnlab.ucdavis.edu/>). Detailed spectral  
599 data are provided in supporting information (Figs. S1-S74, Supporting Information), and the  
600 raw spectral files are available (See Data Availability).

### 601 **Whole Genome Sequencing and Assembly of *Streptomyces* sp. RO-S4 strain**

602 Genomic DNA was isolated from 50 mL of RO-S4 grown in M2 broth medium for 12 days at  
603 28 °C with shaking (150 rpm/min). The DNA was extracted using the Bacteria Genomic DNA  
604 Extraction Kit (Promega, United States) according to the manufacturer's instructions. Illumina  
605 whole genome sequencing was performed by the Genotoul facility in Toulouse, France. Briefly,  
606 libraries (Truseq nano HT, Illumina) were constructed using 200 ng of purified DNA and  
607 sequenced on a Novaseq 6000 sequencer (Illumina), generating 93 million paired 150 base pair  
608 (bp) reads. The entire dataset was assembled using SPAdes (v3.14.0)<sup>84</sup> with the option  
609 "careful". The assembly was manually curated to remove contigs with low (<500) coverage and  
610 low (<55%) G+C content. Since a gene cluster of interest was truncated in a contig, we  
611 manually extended it using blast searches against raw reads and re-assembly of reads and of a  
612 downstream contig using the *gap4* tool of the Staden package (<http://staden.sourceforge.net>).  
613 This final assembly was auto-annotated using PROKKA v. 1.14.6<sup>45</sup> using *Streptomyces* sp.  
614 Vc74B-19 protein descriptions and the annotation was manually curated prior to final  
615 submission to 1) add the ORF corresponding to GcnM in *Streptomyces* sp. LR32, 2) remove  
616 partial rRNA genes and possible adapters in contig ends. The genomic sequence was compared  
617 to that of close strains based on the 16S rRNA gene analysis described above, relatives (type  
618 strains) and genomes in the NCBI GenBank, based on top *blastp* hits using the five  
619 housekeeping genes described by Antony-Babu and coworkers<sup>85</sup>. ANI values were calculated  
620 using the OrthoANI tool available through the EZbiocloud server  
621 (<https://www.ezbiocloud.net/tools/orthoani>).

### 622 **Prediction of secondary metabolite biosynthetic gene clusters (BGCs) in the RO-S4** 623 **genome**

624 The sequenced genome (DNA sequence as input) of the RO-S4 strain was analyzed for the  
625 prediction of secondary metabolites and biosynthetic gene clusters (BGCs) using the genome  
626 mining tool antiSMASH<sup>86</sup>, version 6.0.1. available through  
627 (<https://antismash.secondarymetabolites.org>) using the "relaxed" option. Synteny plots of the  
628 RO-S4 angucycline BGC was performed using Clinker v 0.0.21<sup>87</sup>. We also submitted the



629 PROKKA annotated genbank formatted files using different annotation parameters. All results  
630 as well as descriptions of annotation pipelines are available (see data availability below).

### 631 **Phylogenetic analysis of UrdE and UrdM homologues**

632 Phylogenetic analysis of UrdE and UrdM homologues in RO-S4 and in the BGCs of other  
633 angucyclines and the ring-opening Baeyer-Villiger monooxygenase MtmOIV in the  
634 mithramycin BGC was performed using SeaView Version 4.6 and Mega11<sup>88</sup>. Briefly, amino-  
635 acid sequences were aligned using muscle and a mask created using the gblocks options "allow  
636 smaller blocks positions" and "allow gap positions" followed by manual curation of the mask  
637 and a tree were generated by maximum likelihood with the GT+F model and gamma  
638 distribution using MEGA11 based on 359 homologous. A neighbor-joining tree was  
639 constructed using the JTT substitution model. The robustness of both trees was evaluated by  
640 bootstrap analysis using 100 replicates. The raw alignments, the regions used for reconstruction  
641 and results of model testing are available (see Data Availability).

### 642 **References**

- 643 1. Prestinaci, F., Pezzotti, P. & Pantosti, A. Antimicrobial Resistance: A Global Multifaceted  
644 Phenomenon. *Pathog Glob Health* **109**, 309–318;  
645 <https://doi.org/10.1179/2047773215Y.0000000030> (2015).
- 646 2. Munita, M. & Arias, C.A. Mechanisms of Antibiotic Resistance. Edited by Indira T. Kudva  
647 and Qijing Zhang. *Microbiol Spectr* **4**, 1-37; [https://doi.org/10.1128/microbiolspec.VMBF-](https://doi.org/10.1128/microbiolspec.VMBF-0016-2015)  
648 [0016-2015](https://doi.org/10.1128/microbiolspec.VMBF-0016-2015) (2016).
- 649 3. Ventola, C.L. The Antibiotic Resistance Crisis. *P & T* **40**, 277-283 (2015).
- 650 4. Chokshi, A., Sifri, Z., Cennimo, D. & Helen Horng. Global Contributors to Antibiotic  
651 Resistance. *J Global Infect Dis* **11**, 36-42; [https://doi.org/10.4103/jgid.jgid\\_110\\_18](https://doi.org/10.4103/jgid.jgid_110_18) (2019).
- 652 5. Chambers, H. F. & DeLeo, F. R. Waves of resistance: *Staphylococcus aureus* in the  
653 antibiotic era. *Nat Rev Microbiol* **7**, 629–641 (2009).
- 654 6. Barbosa, F., Pinto, E., Kijjoa, A., Pinto, M. & Sousa, E. Targeting Antimicrobial Drug  
655 Resistance with Marine Natural Products. *Int J Antimicrob* **56**, 1-29;  
656 <https://doi.org/10.1016/j.ijantimicag.2020.106005> (2020).
- 657 7. Edwards, B et al. Treatment Options for Methicillin-Resistant *Staphylococcus Aureus*  
658 (MRSA) Infection: Where Are We Now?. *J. Glob. Antimicrob Resist* **2**, 133–140;  
659 <https://doi.org/10.1016/j.jgar.2014.03.009> (2014).
- 660 8. Bhatnagar, I. & Kim, S.K. Immense Essence of Excellence: Marine Microbial Bioactive  
661 Compounds. *Mar Drugs* **8**, 2673–2701; <https://doi.org/10.3390/md8102673> (2010).

- 662 9. Bose, U. et al. LC-MS-Based Metabolomics Study of Marine Bacterial Secondary  
663 Metabolite and Antibiotic Production in *Salinispora arenicola*. *Mar Drugs* **13**, 249–266;  
664 <https://doi.org/10.3390/md13010249> (2015).
- 665 10. Stien, D. Marine Microbial Diversity as a Source of Bioactive Natural Products. *Mar Drugs*  
666 **18**, 1-5; <https://doi.org/10.3390/md18040215> (2020).
- 667 11. Lane, A.L. & Moore, B.S. A Sea of Biosynthesis: Marine Natural Products Meet the  
668 Molecular Age. *Nat Prod Rep.* **28**, 411–428; <https://doi.org/10.1039/C0NP90032J> (2011).
- 669 12. Blunt, J.W., Copp, B.R., Keyzers, R.A., Munro, M.H.G. & Prinsep, M.R. Marine Natural  
670 Products. *Nat. Prod. Rep.* **30**, 144-222; <https://doi.org/10.1039/C2NP20112G> (2012).
- 671 13. Jose, P.A. & Jha, B. New Dimensions of Research on Actinomycetes: Quest for Next  
672 Generation Antibiotics. *Front Microbiol* **7**, 1295;  
673 <https://doi.org/10.3389/fmicb.2016.01295> (2016).
- 674 14. Hassan, S.S. & Shaikh, A.L. Marine Actinobacteria as a Drug Treasure House.  
675 *Biomed Pharmacother* **87**, 46–57; <https://doi.org/10.1016/j.biopha.2016.12.086> (2017).
- 676 15. Lewis, K. The Science of Antibiotic Discovery. *Cell* **181**, 29-45;  
677 <https://doi.org/10.1016/j.cell.2020.02.056> (2020).
- 678 16. Jose, P.A., Maharshi, A. & Jha. B. Actinobacteria in Natural Products Research: Progress  
679 and Prospects. *Microbiol Res* **246**, 126708; <https://doi.org/10.1016/j.micres.2021.126708>  
680 (2021).
- 681 17. Baltz, R. Renaissance in Antibacterial Discovery from Actinomycetes. *Curr Opin*  
682 *Pharmacol* **8**, 557–563; <https://doi.org/10.1016/j.coph.2008.04.008> (2008).
- 683 18. Olano, C., Méndez, C. & Salas. A. Antitumor Compounds from Marine Actinomycetes.  
684 *Mar Drugs* **7**, 210–248; <https://doi.org/10.3390/md7020210> (2009).
- 685 19. Bérdy, J. Bioactive Microbial Metabolites: A Personal View. *J Antibiot* **58**, 1–26;  
686 <https://doi.org/10.1038/ja.2005.1> (2005).
- 687 20. Dettmer, K., Aronov, P.A. & Hammock, B.D. Mass Spectrometry-Based Metabolomics.  
688 *Mass Spectrom Rev* **26**,51–78; <https://doi.org/10.1002/mas.20108> (2007).
- 689 21. Gomez-Escribano, J.P., Alt, S. & Bibb, M. Next Generation Sequencing of  
690 Actinobacteria for the Discovery of Novel Natural Products. *Mar Drugs* **14**, 78;  
691 <https://doi.org/10.3390/md14040078> (2016).
- 692 22. Cornell, R.C., Marasini, D. & Mohamed F.K. Molecular Characterization of Plasmids  
693 Harbored by Actinomycetes Isolated From the Great Salt Plains of Oklahoma Using PFGE

- 694 and Next Generation Whole Genome Sequencing. *Front Microbiol* **9**, 1-13;  
695 <https://doi.org/10.3389/fmicb.2018.02282> (2018).
- 696 23. Hu, D. et al. Exploring the Potential of Antibiotic Production From Rare Actinobacteria by  
697 Whole-Genome Sequencing and Guided MS/MS Analysis. *Front Microbiol* **11**, 1-12;  
698 <https://doi.org/10.3389/fmicb.2020.01540> (2020).
- 699 24. Khoomrung, S. et al. Metabolomics and Integrative Omics for the Development of Thai  
700 Traditional Medicine. *Front Pharmacol* **8**, 474; <https://doi.org/10.3389/fphar.2017.00474>  
701 (2017).
- 702 25. Pye, C.R., Bertin, M.J., Lokey, R.S., Gerwick, W.H. & Linington, R.G. Retrospective  
703 Analysis of Natural Products Provides Insights for Future Discovery Trends. *PNAS* **114**,  
704 5601–5606; <https://doi.org/10.1073/pnas.1614680114> (2017).
- 705 26. Van der Hooft, J.J.J et al. Linking Genomics and Metabolomics to Chart Specialized  
706 Metabolic Diversity. *Chem. Soc. Rev.* **49**, 3297–3314;  
707 <https://doi.org/10.1039/D0CS00162G> (2020).
- 708 27. Caesar, L.K., Montaser, R., Keller, N.P. & Kelleher, N.L. Metabolomics and Genomics in  
709 Natural Products Research: Complementary Tools for Targeting New Chemical Entities.  
710 *Nat Prod Rep* **38**, 2041-2065; <https://doi.org/10.1039/D1NP00036E> (2021).
- 711 28. Macintyre, L. et al. Metabolomic Tools for Secondary Metabolite Discovery from Marine  
712 Microbial Symbionts. *Mar Drugs* **12**, 3416–3448; <https://doi.org/10.3390/md12063416>  
713 (2014).
- 714 29. Schneider, O. et al. Genome Mining of *Streptomyces* Sp. YIM 130001 Isolated From  
715 Lichen Affords New Thiopeptide Antibiotic. *Front Microbiol* **9**, 1-12;  
716 <https://doi.org/10.3389/fmicb.2018.03139> (2018).
- 717 30. Ye, Y. et al. Discovery of Three 22-Membered Macrolides by Deciphering the Streamlined  
718 Genome of Mangrove-Derived *Streptomyces* Sp. HM190. *Front Microbiol* **11**, 1-13;  
719 <https://doi.org/10.3389/fmicb.2020.01464> (2020).
- 720 31. Liu, X. & Locasale, J.W. Metabolomics: A Primer. *Trends Biochem Sci* **42**, 274-284;  
721 <https://doi.org/10.1016/j.tibs.2017.01.004> (2017).
- 722 32. Wang, M. et al. Sharing and Community Curation of Mass Spectrometry Data with Global  
723 Natural Products Social Molecular Networking. *Nat. Biotechnol* **34**, 828–837;  
724 <https://doi.org/10.1038/nbt.3597> (2016).
- 725 33. Palazzotto, E., Tong, Y., Lee, S.Y. & Weber, T. Synthetic Biology and Metabolic  
726 Engineering of Actinomycetes for Natural Product Discovery. *Biotechnol Adv* **37**, 1-15;  
727 <https://doi.org/10.1016/j.biotechadv.2019.03.005> (2019).

- 728 34. Shang, Z. et al. Grincamycins P–T: Rearranged Angucyclines from the Marine Sediment-  
729 Derived *Streptomyces* Sp. CNZ-748 Inhibit Cell Lines of the Rare Cancer Pseudomyxoma  
730 Peritonei. *J Nat Prod* **84**, 1638–1648; <https://doi.org/10.1021/acs.jnatprod.1c00179> (2021).
- 731 35. Maskey, R.P., Helmke, E. & Laatsch, H. Himalomycin A and B: isolation and structure  
732 elucidation of new fridamycin type antibiotics from a marine *Streptomyces* isolate. *J*  
733 *Antibiot* **56**, 942-949; <https://doi: 10.7164/antibiotics.56.942>. PMID: 14763560 (2003).
- 734 36. Alvi, A.A., Baker, D.D., Stienecker, V., Hosken, M. & Nair, B.G. Identification of  
735 Inhibitors of Inducible Nitric Oxide Synthase from Microbial Extracts. *J Antibiot* **53**, 496–  
736 501; <https://doi.org/10.7164/antibiotics.53.496> (2000).
- 737 37. Song, Y. et al. Cytotoxic and Antibacterial Angucycline- and Prodigiosin- Analogues from  
738 the Deep-Sea Derived *Streptomyces* Sp. SCSIO 11594. *Mar Drugs* **13**, 1304–1316;  
739 <https://doi.org/10.3390/md13031304> (2015).
- 740 38. Zhu, X. et al. Cytotoxic Rearranged Angucycline Glycosides from Deep Sea-Derived  
741 *Streptomyces Lusitanus* SCSIO LR32. *J Antibiot* **70**, 819–822;  
742 <https://doi.org/10.1038/ja.2017.17> (2017).
- 743 39. Rix, U., Rensing, L.L, Hoffmeister, D., Bechthold, A. & Rohr, J. Urdamycin L: A Novel  
744 Metabolic Shunt Product That Provides Evidence for the Role of the UrdM Gene in the  
745 Urdamycin A Biosynthetic Pathway of *Streptomyces Fradiae* Tü 2717. *ChemBioChem* **4**,  
746 109–911; <https://doi.org/10.1002/cbic.200390002> (2003).
- 747 40. Imamura, N., Kakinuma, K., Ikekawa, N., Tanaka, H. & Omura S. Biosynthesis of  
748 vineomycins A1 and B2. *J Antibiot* **35**, 602-608; <https://doi: 10.7164/antibiotics.35.602>  
749 (1982).
- 750 41. Bao, J et al. Cytotoxic Antibiotic Angucyclines and Actinomycins from the *Streptomyces*  
751 Sp. XZHG99T. *J Antibiot* **71**, 1018–1024; <https://doi.org/10.1038/s41429-018-0096-1>  
752 (2018).
- 753 42. Tan, B. et al. Identification of Endogenous Acyl Amino Acids Based on a Targeted  
754 Lipidomics Approach. *J Lipid Res* **51**, 112–219; [https://doi.org/10.1194/jlr.M900198-  
755 JLR200](https://doi.org/10.1194/jlr.M900198-JLR200) (2010).
- 756 43. Nakagawa, K., Hara, C., Tokuyama, S. Takada, K. & Imamura, I. Saprolymycins A–E, New  
757 Angucycline Antibiotics Active against *Saprolegnia Parasitica*. *J Antibiot* **65**, 599–607;  
758 <https://doi.org/10.1038/ja.2012.86> (2012).

- 759 44. Zhang, Y. et al. Identification of the Grincamycin Gene Cluster Unveils Divergent Roles  
760 for GcnQ in Different Hosts, Tailoring the L -Rhodinose Moiety. *Org Lett* **15**, 3254–3257;  
761 <https://doi.org/10.1021/ol401253p> (2013).
- 762 45. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinform* **30**, 2068–2069;  
763 <https://doi.org/10.1093/bioinformatics/btu153> (2014).
- 764 46. Peng, A. et al. Angucycline Glycosides from an Intertidal Sediments Strain *Streptomyces*  
765 sp. and Their Cytotoxic Activity against Hepatoma Carcinoma Cells. *Mar Drugs* **16**, 1-10;  
766 <https://doi.org/10.3390/md16120470> (2018).
- 767 47. Hu, Z. et al. Angucycline Antibiotics and Its Derivatives from Marine-Derived  
768 Actinomycete *Streptomyces* Sp. A6H. *Nat Prod Res* **30**, 2551–2558;  
769 <https://doi.org/10.1080/14786419.2015.1120730> (2016).
- 770 48. Rohr, J. & Thiericke, R. Angucycline Group Antibiotics. *Nat Prod Rep* **9**, 103-137;  
771 <https://doi.org/10.1039/np9920900103> (1992).
- 772 49. Kharel, M.k. et al. Angucyclines: Biosynthesis, Mode-of-Action, New Natural Products,  
773 and Synthesis. *Nat Prod Rep* **29**, 264–325; <https://doi.org/10.1039/C1NP00068C> (2012).
- 774 50. Fürst, M.J. L. J., Gran-Scheuch, A., Aalbers, F.S & Fraaije. M.W. Baeyer–Villiger  
775 Monooxygenases: Tunable Oxidative Biocatalysts. *ACS Catal* **9**, 11207–11241.  
776 <https://doi.org/10.1021/acscatal.9b03396> (2019).
- 777 51. Laslett, D. & Canback, B. ARAGORN, a Program to Detect TRNA Genes and TmRNA  
778 Genes in Nucleotide Sequences. *Nucleic Acids Res* **32**, 11–16;  
779 <https://doi.org/10.1093/nar/gkh152> (2004).
- 780 52. Gruber, A. R., Lorenz, R., Bernhart, S., Neubock, R. & Hofacker, I.L. The Vienna RNA  
781 Websuite. *Nucleic Acids Res* **36**, 70–74; <https://doi.org/10.1093/nar/gkn188> (2008).
- 782 53. Faust, B et al. Two New Tailoring Enzymes, a Glycosyltransferase and an Oxygenase,  
783 Involved in Biosynthesis of the Angucycline Antibiotic Urdamycin A in *Streptomyces*  
784 *Fradiæ* Tu\$ 2717. *Microbiology* **146**, 147-154; <https://doi: 10.1099/00221287-146-1-147>  
785 (2000).
- 786 54. Decker, H. & Haag, S. Cloning and Characterization of a Polyketide Synthase Gene from  
787 *Streptomyces Fradiæ* Tü2717, Which Carries the Genes for Biosynthesis of the  
788 Angucycline Antibiotic Urdamycin A and a Gene Probably Involved in Its Oxygenation. *J*  
789 *Bacteriol* **177**, 6126–6136; <https://doi.org/10.1128/jb.177.21.6126-6136.1995> (1995).
- 790 55. Rix, U., Fischer, C., Remsing, L.L. & Rohr, J. Modification of Post-PKS Tailoring Steps  
791 through Combinatorial Biosynthesis. *Nat Prod Rep* **19**, 542–580;  
792 <https://doi.org/10.1039/b103920m> (2002).



- 793 56. Guo, F. *et al.* Targeted activation of silent natural product biosynthesis pathways by  
794 reporter-guided mutant selection. *Metab Eng* **28**, 134–142;  
795 <http://dx.doi.org/10.1016/j.ymben.2014.12.006> (2015).
- 796 57. Kallio, P. *et al.* Flavoprotein Hydroxylase PgaE Catalyzes Two Consecutive Oxygen-  
797 Dependent Tailoring Reactions in Angucycline Biosynthesis. *Biochemistry* **50**, 5535–5543;  
798 <http://dx.doi.org/10.1021/bi200600k> (2011).
- 799 58. Du, Y. *et al.* Identification and Characterization of the Biosynthetic Gene Cluster of  
800 Polyoxypeptin A, a Potent Apoptosis Inducer. *BMC Microbiol* **14**, 30;  
801 <https://doi.org/10.1186/1471-2180-14-30> (2014)
- 802 59. Salem, S. M., Weidenbach, S. & Rohr, J. Two Cooperative Glycosyltransferases Are  
803 Responsible for the Sugar Diversity of Saquayamycins Isolated from *Streptomyces* sp. KY  
804 40-1. *ACS Chem. Biol.* **12**, 2529–2534 (2017).
- 805 60. Kawasaki, T., Moriyama, A., Nakagawa, K. & Imamura, N. Cloning and Identification of  
806 Saprolmycin Biosynthetic Gene Cluster from *Streptomyces* Sp. TK08046.  
807 *Biosci Biotechnol Biochem* **80**, 2144–2150;  
808 <https://doi.org/10.1080/09168451.2016.1196574> (2016).
- 809 61. Basnet, D.B. *et al.* Angucyclines Sch 47554 and Sch 47555 from *Streptomyces* Sp. SCC-  
810 2136: Cloning, Sequencing, and Characterization. *Mol Cells* **9**, 154–162 (2006).
- 811 62. Mittler, M., Bechthold, A. & Schulz, G.E. Structure and Action of the C–C Bond-Forming  
812 Glycosyltransferase UrdGT2 Involved in the Biosynthesis of the Antibiotic Urdamycin. *J*  
813 *Mol Biol* **372**, 67–76; <https://doi.org/10.1016/j.jmb.2007.06.005> (2007).
- 814 63. Fidan, O. *et al.* New Insights into the Glycosylation Steps in the Biosynthesis of Sch47554  
815 and Sch47555. *ChemBioChem* **19**, 1424–1432; <https://doi.org/10.1002/cbic.201800105>  
816 (2018).
- 817 64. Elshahawi, S.I., Shaaban, K.K., Kharel, M.K. & Thorson, J.S. A Comprehensive Review  
818 of Glycosylated Bacterial Natural Products. *Chem Soc Rev* **44**, 7591–7697;  
819 <https://doi.org/10.1039/C4CS00426D> (2015).
- 820 65. Myronovskyi, M. *et al.* Generation of New Compounds through Unbalanced Transcription  
821 of Landomycin A Cluster. *Appl Microbiol Biotechnol* **100**, 9175–9186;  
822 <https://doi.org/10.1007/s00253-016-7721-3> (2016).
- 823 66. Tawfike, A. *et al.* New Bioactive Metabolites from the Elicited Marine Sponge-Derived  
824 Bacterium *Actinokineospora Spheciospongiae* sp. Nov. *AMB Expr* **9**, 1–9;  
825 <https://doi.org/10.1186/s13568-018-0730-0> (2019).



- 826 67. Keller, U, Lang, M., Crnovcic, I., Pfennig, F., & Schauwecker, F. The Actinomycin  
827 Biosynthetic Gene Cluster of *Streptomyces Chrysomallus*: A Genetic Hall of Mirrors for  
828 Synthesis of a Molecule with Mirror Symmetry. *J Bacteriol* **192**, 2583–2595;  
829 <https://doi.org/10.1128/JB.01526-09> (2010).
- 830 68. Hubbard, B.K., Thomas, M.G. & Walsh, C.T. Biosynthesis of L-p-Hydroxyphenylglycine,  
831 a Non-Proteinogenic Amino Acid Constituent of Peptide Antibiotics. *Chem Biol* **7**, 931-  
832 941; [https://doi: 10.1016/s1074-5521\(00\)00043-0](https://doi.org/10.1016/s1074-5521(00)00043-0) (2000).
- 833 69. Cobb, R. E., Wang, Y. & Zhao, H. High-Efficiency Multiplex Genome Editing of  
834 *Streptomyces* Species Using an Engineered CRISPR/Cas System. *ACS Synth. Biol.* **4**, 723–  
835 728; [https://doi: 10.1021/sb500351f](https://doi.org/10.1021/sb500351f) (2015).
- 836 70. Tong, Y. *et al.* CRISPR–Cas9, CRISPRi and CRISPR-BEST-mediated genetic  
837 manipulation in streptomycetes. *Nat Protoc* **15**, 2470–2502 (2020).
- 838 71. Hirsch, C F. & Christensen, D.L. Novel Method for Selective Isolation of Actinomycetes.  
839 *Appl Environ Microbiol* **46**, 925–929; <https://doi.org/10.1128/aem.46.4.925-929.1983>  
840 (1983).
- 841 72. Jensen, P.R., Dwight, R. & Fenical, W. Distribution of Actinomycetes in Near-Shore  
842 Tropical Marine Sediments. *Appl Environ Microbiol* **57**, 1102–1108;  
843 <https://doi.org/10.1128/AEM.57.4.1102-1108.1991> (1991).
- 844 73. Shirling, E. B. & Gottlieb, D. Methods for Characterization of *Streptomyces* Species. *Int J*  
845 *Syst. Bacteriol* **16**, 313–340; <https://doi.org/10.1099/00207713-16-3-313> (1966).
- 846 74. Fagervold, S. K. *et al.* *Pleionea Mediterranea* Gen. Nov., Sp. Nov., a  
847 Gammaproteobacterium Isolated from Coastal Seawater. *Int. J. Syst. Evol* **63**, 2700–2705;  
848 <https://doi.org/10.1099/ijss.0.045575-0> (2013).
- 849 75. Yoon, S. H *et al.* Introducing EzBioCloud: A Taxonomically United Database of 16S  
850 rRNA Gene Sequences and Whole-Genome Assemblies. *Int J Syst Evol.* **67**, 1613–1617;  
851 <https://doi.org/10.1099/ijsem.0.001755> (2017).
- 852 76. Shomura, T. *et al.* Studies on Actinomycetales producing antibiotics only on agar culture.  
853 I. Screening, taxonomy and morphology-productivity relationship of *Streptomyces*  
854 *halstedii*, strain SF-1993. *J Antibiot* **32**, 427-435; [https://doi: 10.7164/antibiotics.32.427](https://doi.org/10.7164/antibiotics.32.427)  
855 (1979).
- 856 77. Nkanga, E.J. & Hagedorn, C. Detection of Antibiotic-Producing *Streptomyces* Inhabiting  
857 Forest Soils. *Antimicrob. Agents Chemother* **14**, 51-59;  
858 <https://doi.org/10.1128/AAC.14.1.51> (1978).

- 859 78. CLSI: Clinical and Laboratory Standards Institute: Performance standards for antimicrobial  
860 susceptibility testing; Twenty-sixth informational supplement. CLSI document M100-S26.  
861 Wayne, PA: Clinical and Laboratory Standards Institute (2016).
- 862 79. Stien, D. et al. Metabolomics Reveal That Octocrylene Accumulates in *Pocillopora*  
863 *Damicornis* Tissues as Fatty Acid Conjugates and Triggers Coral Cell Mitochondrial  
864 Dysfunction. *Anal Chem* **91**, 990–995; <https://doi.org/10.1021/acs.analchem.8b04187>  
865 (2019).
- 866 80. Stien, D. A Unique Approach to Monitor Stress in Coral Exposed to Emerging Pollutants.  
867 *Sci Rep* **10**, 1-11; <https://doi.org/10.1038/s41598-020-66117-3> (2020).
- 868 81. Otasek, D., Morris, J.H., Bouças, Pico, A.R. & Demchak, D. Cytoscape Automation:  
869 Empowering Workflow-Based Network Analysis. *Genome Biol* **20**, 1-15;  
870 <https://doi.org/10.1186/s13059-019-1758-4> (2019).
- 871 82. Dührkop, K. SIRIUS 4: A Rapid Tool for Turning Tandem Mass Spectra into Metabolite  
872 Structure Information. *Nat Methods* **16**, 299–302; [https://doi.org/10.1038/s41592-019-](https://doi.org/10.1038/s41592-019-0344-8)  
873 [0344-8](https://doi.org/10.1038/s41592-019-0344-8) (2019).
- 874 83. Ruttkies, C., Schymanski, E.L., Wolf, W., Hollender, J. & Neumann, S. MetFrag  
875 Relaunch: Incorporating Strategies beyond in Silico Fragmentation. *J Cheminform* **8**, 1-  
876 16; <https://doi.org/10.1186/s13321-016-0115-9> (2016).
- 877 84. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using SPAdes  
878 De Novo Assembler. *Curr Protoc Bioinform* **70**, 1-29; <https://doi.org/10.1002/cpbi.102>  
879 (2020).
- 880 85. Antony-Babu. S et al. Multiple *Streptomyces* Species with Distinct Secondary  
881 Metabolomes Have Identical 16S rRNA Gene Sequences. *Sci Rep* **7**, 11089;  
882 <https://doi.org/10.1038/s41598-017-11363-1> (2017).
- 883 86. Blin, K. et al. AntiSMASH 6.0: Improving Cluster Detection and Comparison Capabilities.  
884 *Nucleic Acids Res* **49**, 29–35; <https://doi.org/10.1093/nar/gkab335> (2021).
- 885 87. Gilchrist, C.L.M. & Chooi, Y.H. clinker & clustermap.js: automatic generation of gene  
886 cluster comparison figures. *Bioinformatics* **37**, 2473–2475; [https://doi:](https://doi.org/10.1093/bioinformatics/btab007)  
887 [10.1093/bioinformatics/btab007](https://doi.org/10.1093/bioinformatics/btab007) (2021).
- 888 88. Gouy, M., Guindon, S. & Gascuel, O. SeaView Version 4: A Multiplatform Graphical User  
889 Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol Biol Evol* **27**, 221–  
890 224; <https://doi.org/10.1093/molbev/msp259> (2010).

891

892

893 **Acknowledgements**

894 We thank the Bio2Mar platform (<http://bio2mar.obs-banyuls.fr>) for providing technical support  
895 and access to instrumentation. This work benefited from access to the Observatoire  
896 Océanologique de Banyuls-sur-Mer, an EMBRC-France and EMBRC-ERIC site.

897 **Author Contributions:** R.O., D.S, and M.S. conceived and designed the experiments. R.O.  
898 and K.M. collected biological material. R.O., A.S.R., C.V. performed the experiments. D.S.  
899 and J.S. performed the metabolomics analysis. R.O and M.T.S the genomic analysis. R.O., D.S.,  
900 and M.T.S. wrote the original draft and prepared tables and figures. All authors read and  
901 approved the final version.

902 **Additional Information**

903 **Data availability:** The draft genome sequence was deposited in the GenBank database under  
904 accession number JAJQKZ000000000. Raw Illumina reads have been deposited in the SRA  
905 under accession SRR17084181. The sequence analysis pipeline plus discussions regarding  
906 angucycline biosynthesis, AntiSMASH results and the files used for the phylogenetic analysis  
907 are available through are available through github ([github.com/suzumar/ROS4\\_manus](https://github.com/suzumar/ROS4_manus)).

908  
909 **Funding:** This work was funded by recurrent funds of the CNRS and Sorbonne University  
910 attributed to the LBBM laboratory.

911 **Competing interests:** The authors declare no conflicts of interest.

912

913

914