



HAL
open science

Revised structure of anthelvencin A and characterization of the anthelvencin biosynthetic gene cluster

Céline Aubry, Paolo Clerici, Claude Gerbaud, Laurent Micouin, Jean-Luc Pernodet, Sylvie Lautru

► To cite this version:

Céline Aubry, Paolo Clerici, Claude Gerbaud, Laurent Micouin, Jean-Luc Pernodet, et al.. Revised structure of anthelvencin A and characterization of the anthelvencin biosynthetic gene cluster. ACS Chemical Biology, 2020, 10.1021/acscchembio.9b00960 . hal-02502312

HAL Id: hal-02502312

<https://hal.science/hal-02502312v1>

Submitted on 10 Nov 2020

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 Revised structure of anthelvencin A and characterization
2 of the anthelvencin biosynthetic gene cluster
3
4
5

6 Céline Aubry^{a, b}, Paolo Clerici^c, Claude Gerbaud^a, Laurent Micouin^c, Jean-Luc
7 Pernodet^a, and Sylvie Lautru^{a*}
8
9
10

11 ^a Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université
12 Paris-Saclay, 91198, Gif-sur-Yvette cedex, France
13

14 ^b present address: Unité Molécules de Communication et Adaptation des Microorganismes
15 (MCAM, UMR 7245), Muséum National d'Histoire Naturelle (MNHN), Centre National de la
16 Recherche Scientifique (CNRS), CP 54 57 rue Cuvier, 75005 Paris, France
17

18 ^c Université de Paris, Laboratoire de Chimie et Biochimie pharmacologiques et
19 toxicologiques, LCBPT, UMR 8601 CNRS, F-75006 Paris, France
20

21 * Corresponding author: Sylvie LAUTRU, sylvie.lautru@i2bc.paris-saclay.fr
22
23
24

25 **ABSTRACT**

26

27 Anthelvencins A and B are pyrrolamide metabolites produced by *Streptomyces venezuelae*
28 ATCC 14583 and 14585. Isolated in 1965, they were reported to exhibit anthelmintic and
29 moderate antibacterial activities. In this study, we revise the structure of anthelvencin A and
30 identify a third anthelvencin metabolite, bearing two N-methylated pyrrole groups, which we
31 named anthelvencin C. We sequenced the genome of *S. venezuelae* ATCC 14583 and
32 identified a gene cluster predicted to direct the biosynthesis of anthelvencins. Functional
33 analysis of this gene cluster confirmed its involvement in anthelvencin biosynthesis and
34 allowed us to propose a biosynthetic pathway for anthelvencins. In addition to a non-
35 ribosomal peptide synthetase (NRPS), the assembly of anthelvencins involves an enzyme
36 from the ATP-grasp ligase family, Ant23. We propose that Ant23 uses a PCP-loaded 4-
37 aminopyrrole-2-carboxylate as substrate. As observed for the biosynthesis of the other
38 pyrrolamides congocidine (produced by *Streptomyces ambofaciens* ATCC 25877) and
39 distamycin (produced by *Streptomyces netropsis* DSM 40846), the NRPS assembling
40 anthelvencins is composed of stand-alone domains only. Such NRPSs, sometimes called
41 type II NRPSs, are less studied than the classical multimodular NRPSs. Yet, they constitute
42 an interesting model to study protein-protein interactions in NRPSs and are good candidates
43 for combinatorial biosynthesis approaches.

44

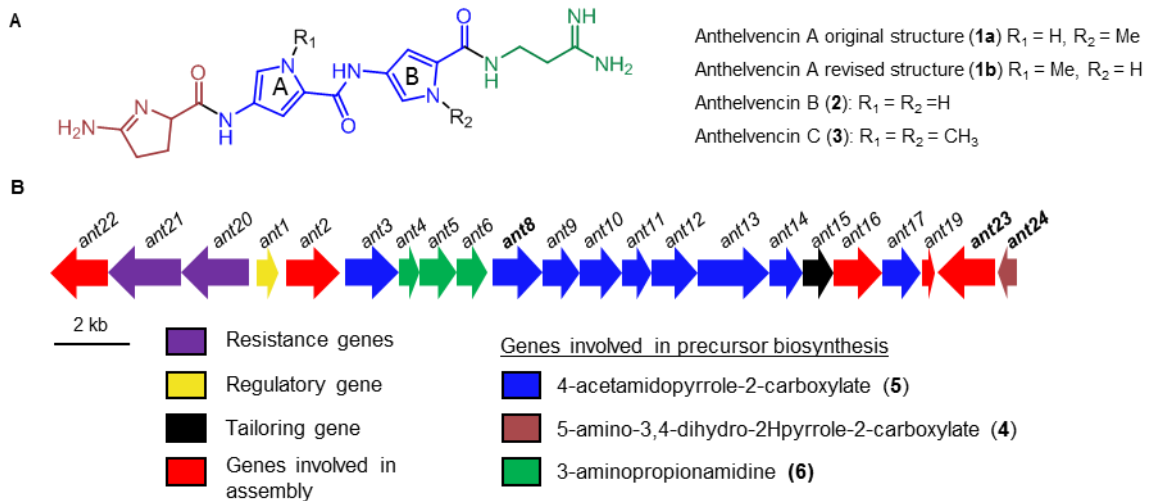
45

46 Anthelvencins A and B (Figure 1A) are specialized metabolites that were isolated in
47 1965 from cultures of *Streptomyces venezuelae* ATCC 14583 and 14585 and exhibit
48 moderate antibacterial and anthelmintic activities¹. They belong to the family of pyrrolamide
49 metabolites, the best-characterized members of which are congocidine (also called
50 netropsin, produced by *Streptomyces ambofaciens* ATCC 25877) and distamycin (produced
51 by *Streptomyces netropsis* DSM 40846) (see Supplementary Figure S1). These metabolites
52 are known to bind to the DNA minor groove with some sequence specificity for regions of
53 four (or more) A or T bases². During the last decade, the biosynthetic gene clusters of
54 congocidine and distamycin have been identified and the biosynthesis of these metabolites
55 has been elucidated³⁻⁷. One remarkable aspect of this biosynthesis is that it involves non-
56 canonical non-ribosomal peptide synthetases (NRPSs), solely constituted of stand-alone
57 modules or domains. Such NRPSs are sometimes called type II NRPSs. They participate in
58 the biosynthesis of unusual precursors of non-ribosomal peptides (NRP)⁸ and are involved in
59 the assembly of some NRP such as phosphinothricin⁹ and rhabdopeptide/xenortide
60 peptides¹⁰.

61
62 A structural analysis of anthelvencins shows that these metabolites most likely share
63 two precursors with congocidine and distamycin: 4-acetamidopyrrole-2-carboxylate (**5**) and
64 3-aminopropionamidine (**6**). The remaining precursor is probably 5-amino-3,4-dihydro-2H-
65 pyrrole-2-carboxylate (**4**), a precursor shared with other pyrrolamides such as kikumycins¹¹ or
66 TAN 868A¹² (Figure S1). In fact, a limited number of precursors seems to be assembled in
67 various combinations to give birth to the different members of the pyrrolamide family.
68 Understanding how these precursors are assembled and combined may improve our
69 comprehension of type II NRPS enzymatic mechanisms and help to design functional
70 synthetic NRPSs using synthetic biology. For these reasons, we undertook to identify and
71 characterize the anthelvencin biosynthetic gene cluster of *S. venezuelae* ATCC 14583. In
72 this study, we revise the structure of anthelvencin A based on HR-MS² and NMR data. We
73 show that *S. venezuelae* ATCC 14583 produces, in addition to the already known

74 anthelvencins A and B, a third anthelvencin, methylated on the two pyrrole groups. We
 75 named this new anthelvencin anthelvencin C. We also identify the gene cluster directing the
 76 biosynthesis of anthelvencins in *S. venezuelae* ATCC 14583 genome and we functionally
 77 characterize this gene cluster.

78



79

80 **Figure 1. (A) Chemical structure of anthelvencins A, B and C; (B) Genetic organization**
 81 **of the anthelvencin biosynthetic gene cluster in *S. venezuelae* ATCC 14583.** Genes in
 82 boldface are genes that were used as targets for genetic inactivation. The numbers in
 83 boldface (**1a-6**) are used to refer to the corresponding chemical structures.

84

85 RESULTS AND DISCUSSION

86 ***In silico* identification of a gene cluster putatively involved in anthelvencin**
 87 ***biosynthesis* in *S. venezuelae* ATCC 14583.** To identify the gene cluster directing
 88 anthelvencin biosynthesis, we sequenced the genome of the *S. venezuelae* ATCC 14583
 89 strain by the Illumina technology, using a paired-end genomic library. The 5.45 million reads
 90 of 301 bps were assembled using Velvet v1.2.10, resulting in 63 contigs with a total length of
 91 9.08 Mbps (180-fold coverage).

92 The gene cluster directing the biosynthesis of anthelvencins was identified by mining
 93 the genome of *S. venezuelae* ATCC 14583 for homologs of genes involved in the
 94 biosynthesis of congocidine³. We identified a gene cluster (*ant*, GenBank accession number
 95 MK483114) that spans 26 kb and contains 22 genes (Figure 1B). The deduced products of

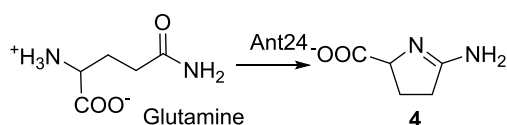
96 twenty of these genes exhibit a high amino acid sequence identity with Cgc proteins involved
97 in congocidine biosynthesis (from 64 to 84 % sequence identity, Table 1) and they most likely
98 have similar

99 **Table 1.** Sequence identities between Ant and Cgc proteins

Protein	Putative protein function	Cgc ortholog (Accession number)	% sequence identity
Ant1	Transcriptional regulator	Cgc1 (AKZ59702.1)	71
Ant2	NRPS, C domain	Cgc2 (AKZ59703.1)	66
Ant3	4-Acetamidopyrrole-2-carboxaldehyde dehydrogenase	Cgc3 (AKZ59704.1)	74
Ant4	Cytosine monophosphate hydrolase	Cgc4 (AKZ59705.1)	83
Ant5	Cytosine reductase	Cgc5 (AKZ59706.1)	77
Ant6	Dihydrocytosine hydrolase	Cgc6 (AKZ59707.1)	78
Ant8	Nucleotidyl N-acetylglucosamine dehydrogenase	Cgc8 (AKZ59709.1)	84
Ant9	Nucleotidyl-2-acetamido-2- deoxyglucopyranuronate decarboxylase	Cgc9 (AKZ59710.1)	84
Ant10	Glycosyltransferase-like enzyme	Cgc10 (AKZ59711.1)	81
Ant11	N-acetylglucosamine-1-phosphate nucleotidyltransferase	Cgc11 (AKZ59712.1)	76
Ant12	Nucleotidyl threo-2-acetamido-2-deoxy- pentopyran-4-ulose aminotransferase	Cgc12 (AKZ59713.1)	79
Ant13	Glycoside hydrolase	Cgc13 (AKZ59714.1)	78
Ant14	4-Acetamidopyrrole-2-carboxylate deacetylase	Cgc14 (AKZ59715.1)	80
Ant15	Methyltransferase	Cgc15 (AKZ59716.1)	84
Ant16	NRPS, C domain	Cgc16 (AKZ59717.1)	68
Ant17	4-Acetamidopyrrole-2-carboxaldehyde dehydrogenase	Cgc17 (AKZ59718.1)	83
Ant19	NRPS, PCP domain	Cgc19 (AKZ59720.1)	64
Ant20	ABC transporter	Cgc20 (AKZ59701.1)	81
Ant21	ABC transporter	Cgc21 (AKZ59700.1)	81
Ant22	Acyl co-A synthetase	Cgc22 (AKZ59699.1)	72
Ant23	ATP-grasp domain-containing protein	None	
Ant24	Ectoine synthase-like protein	None	

100 function to their Cgc homologs. Thus, the gene numbers attributed to the *ant* genes were
 101 chosen to follow the *cgc* nomenclature whenever possible. The genetic organization of the
 102 *ant* cluster is remarkably similar to the one of the *cgc* cluster³ from *S. ambofaciens*
 103 ATCC23877 (Figure S2). Two *cgc* genes (*cgc7* and *cgc18*) have no homologs in the *ant*
 104 gene cluster. These two genes are involved in the biosynthesis of the guanidinoacetate
 105 precursor (absent in anthelvencins) and its assembly. Based on the high level of sequence
 106 identity between Ant and Cgc proteins and on the previously published biosynthetic
 107 pathways for congocidine and distamycin^{3,4,6,7}, we propose the following roles for the Ant
 108 proteins. Ant8 to Ant13, Ant3 and Ant17 are involved in the biosynthesis of 4-
 109 acetamidopyrrole-2-carboxylate (**5**) and Ant14 is involved in its deacetylation. Ant4 to Ant6
 110 are involved in the biosynthesis of 3-aminopropionamidine (**6**). Ant2, Ant16, Ant19, and
 111 Ant22 are involved in the assembly of anthelvencins. In addition to these, proteins are
 112 necessary for the biosynthesis and assembly of 5-amino-3,4-dihydro-2H-pyrrole-2-
 113 carboxylate (**4**). The *ant* cluster contains two genes, *ant24* and *ant23*, which have no
 114 homolog in the *cgc* cluster. We hypothesized that they might be involved in the biosynthesis
 115 of (**4**) and its assembly. Indeed, a protein blast¹³ and a conserved domain searches¹⁴ on the
 116 Ant24 sequence suggested that Ant24 belongs to the L-ectoine synthase (EC 4.2.1.108)
 117 family of enzymes. L-ectoine synthases catalyze the ring closure of N γ -acetyl-L-2,4-
 118 diaminobutyric acid, yielding the osmolyte ectoine, a metabolite structurally related to (**4**). In
 119 2011, Witt and collaborators reported that the ectoine synthase from *Halomonas elongata*
 120 can catalyze the intramolecular condensation of glutamine to form (**4**) as a side reaction¹⁵.
 121 This enzyme has 34% sequence identity and 51% sequence similarity with Ant24. Thus, it
 122 appears likely that Ant24 catalyzes the same reaction (Scheme 1).

123



124

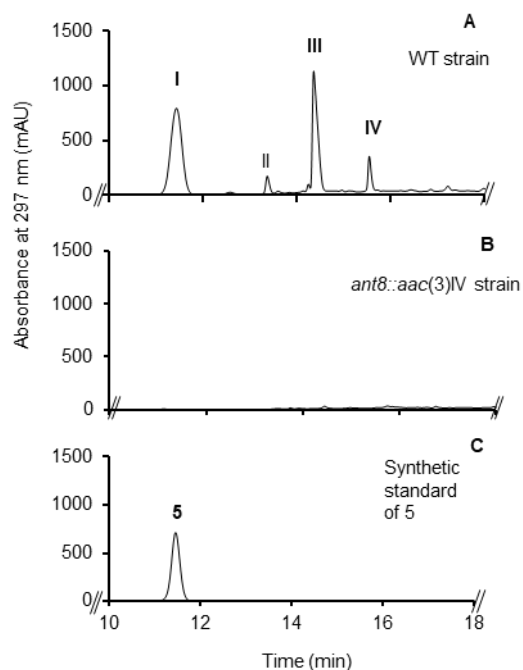
125 **Scheme 1:** Proposed biosynthesis of 5-amino-3,4-dihydro-2H-pyrrole-2-carboxylate (**4**) by
126 Ant24

127

128 Ant23 contains an ATP-grasp domain. ATP-grasp enzymes usually catalyze the ATP-
129 dependent ligation of a carboxylate-containing molecule to an amino or thiol group-containing
130 molecule¹⁶. Some of these ATP-grasp enzymes play a role in the biosynthesis of specialized
131 metabolites¹⁷. They can function as an alternative to or in combination with non-ribosomal
132 peptide synthetases (NRPS), to elongate a peptide chain^{17,18}. Thus, it appears plausible that
133 Ant23 catalyzes the amide bond formation between (**4**) and a PCP (Ant19)-bound 4-
134 aminopyrrole-2-carboxylate.

135

136 ***Abolition of the production of four metabolites in a S. venezuelae ATCC 14583 mutant***
137 ***deleted for ant8 (coding for a putative nucleotidyl N-acetylglucosamine***
138 ***dehydrogenase).*** To verify that the *ant* gene cluster is involved in the biosynthesis of
139 anthelvencins, we inactivated *ant8*. This gene is the ortholog of *cgc8*, the product of which is
140 involved in the biosynthesis of the 4-acetamidopyrrole-2-carboxylate (**5**). The *ant8* gene was
141 replaced by an *aac(3)IV* resistance cassette by homologous recombination using the
142 pANT007 suicide plasmid, yielding the *S. venezuelae* ANT007 strain. This strain and the wild
143 type *S. venezuelae* strain were cultivated for three days in MP5 liquid medium. The culture
144 supernatants were then filtered and analyzed by HPLC. The chromatograms show that four
145 metabolites present in the wild type strain supernatant (peaks I to IV) are absent in the
146 supernatant of the ANT007 mutant strain (Figure 2). The first metabolite (peak I, retention
147 time of 11.5 min) corresponds to 4-aminopyrrole-2-carboxylate (**5**), identified by its UV
148 spectrum and by comparison with an authentic standard (Figure 2 and reference 7). The
149 three peaks II (retention time of 13.3 min), III (retention time of 14.3 min) and IV (retention
150 time of 15.5 min) have UV absorption spectra typical of pyrrolamides (Figure S3 and
151 reference 4).



152

153 **Figure 2: HPLC analysis of culture supernatants** of A) *S. venezuelae* ATCC 14583 wild
 154 type and B) ANT007 (*S. venezuelae* ATCC 14583 *ant8::aac(3)IV*). C) Standard of 4-
 155 acetamidopyrrole-2-carboxylate (**5**). Peaks I to IV are compounds the biosynthesis of which
 156 is linked to the *ant* gene cluster. Peak I corresponds to (**5**).

157

158

159 **Chemical nature of metabolites II, III and IV.** To determine the chemical nature of the
 160 metabolites II, III and IV, we partially purified them. For that purpose, we used ANT012, a
 161 strain that expressed a second copy of the genes *ant23* and *ant24* under the control of the
 162 promoter *rpsL(TP)*¹⁹, as this strain produces compounds III and IV in slightly higher titers
 163 (data not shown). The ANT012 culture supernatant was recovered after three days of culture
 164 in MP5 medium and the compounds of interest were partially purified on a XAD16 resin. The
 165 elution fraction was concentrated to dryness solution, resuspended in water and analyzed by
 166 LC-HR-MS².

167

168 The exact mass and fragmentation pattern of compound II (Figure S4) are consistent
 169 with II being anthelvencin B (2) ($[M+H]^+$ $m/z = 414.1998$; calculated 414.1997). The exact

170 mass of compound III (Figure S5) is consistent with III being anthelvencin A ($[M+H]^+$ $m/z =$
171 428.2151; calculated 428.2153). The fragmentation pattern however (Figure S5, fragments 4
172 and 5), indicates that the position of the methyl group is not on the B pyrrole ring, as
173 previously proposed (but never experimentally established¹) but rather on the A pyrrole ring
174 (Figure 1). To confirm the structure of anthelvencin A, we purified compound III and carried
175 out NMR experiments. ¹H NMR of anthelvencin A hydrochloride proved to be very similar to
176 the one reported for the authentic sample synthesized by Lee and coworkers¹⁷ assuming the
177 position of the methyl group on the B pyrrole ring. However, a combination of two-
178 dimensional NMR experiments confirmed the position of this methyl on the A ring. We
179 therefore propose a revised structure for anthelvencin A (Figure S6).

180

181 The exact mass and fragmentation pattern of compound IV (Figure S7) are consistent
182 with IV being an anthelvencin metabolite methylated on both pyrrole groups ($[M+H]^+$ $m/z =$
183 442.2311; calculated 442.2310), a metabolite that we named anthelvencin C (Figure 1A, 3).
184 We tried to purify anthelvencin C to confirm its chemical structure with NMR analyses but this
185 metabolite turned out to be highly unstable, as already observed by Lee and coworkers²⁰.

186

187

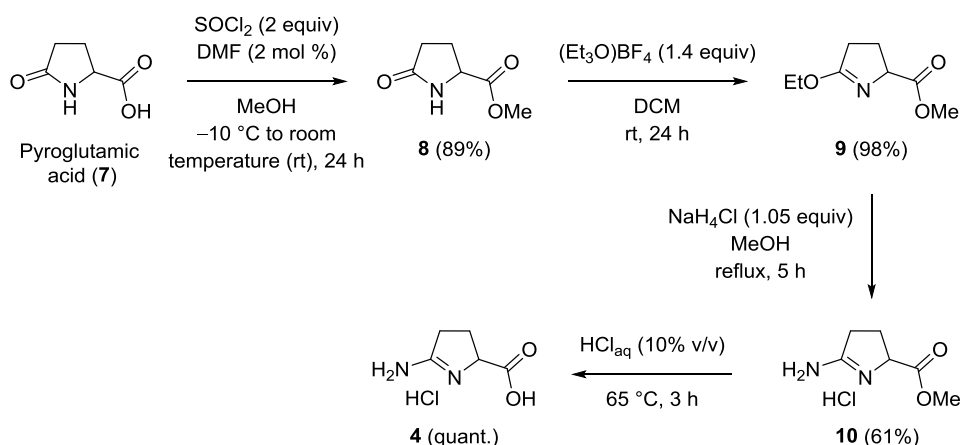
188 ***Involvement of ant24 in 5-amino-3,4-dihydro-2H-pyrrole-2-carboxylate (4) biosynthesis.***

189 To verify that *ant24* is involved in the biosynthesis of anthelvencins, we replaced it by an
190 *aac(3)IV* resistance cassette by homologous recombination, following the same procedure as
191 described above. The culture supernatant of the resulting mutant strain, called ANT009, was
192 analyzed by HPLC. No production of anthelvencins was observed, confirming that *ant24* is
193 necessary for the production of these metabolites (Figure 3A). To ensure that the observed
194 phenotype was due to the replacement of *ant24* by the *aac(3)IV* cassette, we genetically
195 complemented the ANT009 strain using a plasmid expressing *ant23* and *ant24* under a
196 constitutive promoter. The production of anthelvencins was restored in the complemented

197 strain, named ANT014 (Figure S8A), thus confirming that *ant24* is involved in anthelvincin
198 biosynthesis.

199 To examine Ant24 putative function in the biosynthesis of (4), we undertook the chemical
200 complementation of ANT009 by adding to the culture medium (4), the putative product of
201 Ant24-catalyzed reaction. Thus, (4) was synthesized according to a previously described
202 synthetic procedure²¹ (Scheme 2).

203



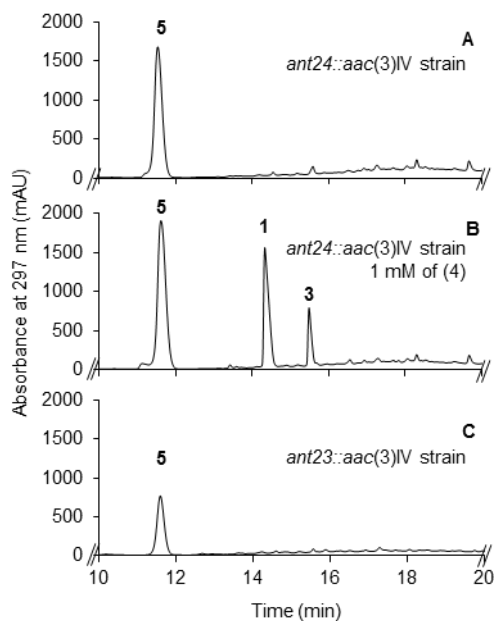
204

205 **Scheme 2:** Synthesis of 5-amino-3,4-dihydro-2Hpyrrole-2-carboxylate (4)

206

207 We next fed the ANT009 strain with (4). As shown in Figure 3B, this resulted in the
208 restoration of the production of anthelvincins A and C, hence confirming the involvement of
209 *ant24* in the biosynthesis of the anthelvincin precursor (4).

210



211

212 **Figure 3:** HPLC analysis of culture supernatants of (A) ANT009 (*S. venezuelae* ATCC 14583
 213 *ant24::aac(3)IV*), (B) ANT009 (*S. venezuelae* ATCC 14583 *ant24::aac(3)IV*) cultivated in
 214 presence of 1mM of (4), and (C) ANT008 (*S. venezuelae* ATCC 14583 *ant23::aac(3)IV*).
 215 Numbers above peaks correspond to the metabolite numbers in the text.

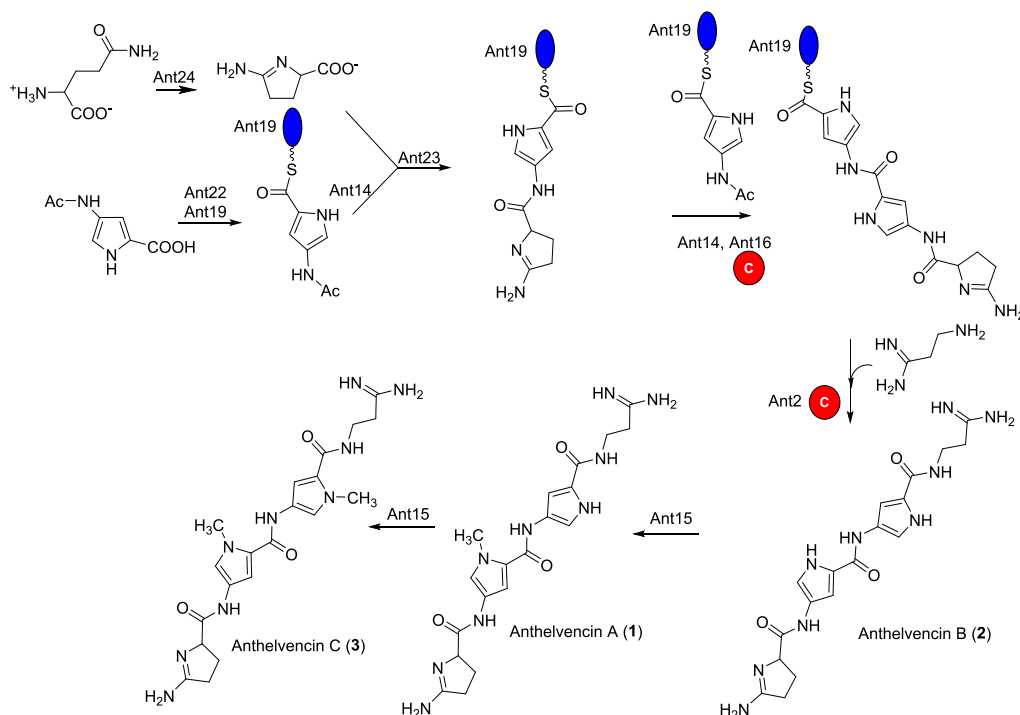
216

217 **Involvement of ant23 in the biosynthesis of anthelvencins.** To confirm that Ant23 is
 218 involved in anthelvencin biosynthesis, we replaced *ant23* by the *aac(3)IV* resistance cassette
 219 following the previously described protocol. The resulting mutant strain was called ANT008. It
 220 was cultivated for three days in MP5 medium at 28°C and the culture supernatant was
 221 analyzed by HPLC. Figure 3C shows that no anthelvencin is produced by the ANT008
 222 mutant. To ensure that the observed phenotype was due to the replacement of *ant23* by the
 223 *aac(3)IV* cassette, we genetically complemented the ANT008 strain using the plasmid
 224 expressing *ant23* and *ant24* under a constitutive promoter. The production of anthelvencins
 225 was restored in the complemented strain, named ANT013 (Figure S8B), thus confirming that
 226 *ant23* is involved in anthelvencin biosynthesis.

227

228 **Proposed biosynthetic pathway for anthelvencin biosynthesis.** Based on the results
 229 presented above and on previous characterizations of pyrrolamide biosyntheses^{3,4,6}, we

230 proposed that anthelvencins are assembled from 3-amidinopropionamidine, 4-aminopyrrole-
231 2-carboxylate and 5-amino-3,4-dihydro-2H-pyrrole-2-carboxylate following the biosynthetic
232 pathway presented in Figure 4. As already observed for the biosynthesis of other
233 pyrrolamides (congocidine, distamycin), the non-ribosomal peptide synthetase involved in
234 anthelvencins is constituted solely of stand-alone domains (C domains Ant2 and Ant16 and
235 PCP domain Ant19). No adenylation domain is involved in the activation of the carboxylate
236 groups of the precursors. Instead, activation of the carboxylate group of the pyrrole precursor
237 **(5)** and the covalent attachment of the activated precursor to the PCP domain Ant19 is
238 catalyzed by Ant22, which belongs to the family of acyl-CoA synthetases. The formation of
239 the first amide bond between **(4)** and Ant19-bound **(5)** is likely catalyzed by Ant23, an
240 enzyme from the ATP-grasp ligase family, which form acylphosphate intermediates. Two
241 stand-alone condensation domains, Ant16 and Ant2, catalyze the formation of the other
242 amide bonds, adding respectively a second pyrrole precursor and the 3-
243 aminopropionamidine to Ant19-bound intermediates. Thus, during anthelvencin assembly,
244 the stand-alone PCP domain, Ant19, interacts with two different families of enzymes
245 (condensation domain and ATP-grasp ligase).



246

247 **Figure 4:** Proposed biosynthetic pathway for anthelvencins A, B and C
 248 The blue ellipses represent PCP domains, and the red circles represent C domains
 249

250 In conclusion, we have identified and characterized the gene cluster directing the
 251 biosynthesis of anthelvencins in *Streptomyces venezuelae* ATCC 14583. We showed that
 252 this cluster directs the biosynthesis of two known metabolites, anthelvencin A, for which we
 253 propose a revised structure, anthelvencin B, and new anthelvencin that we named
 254 anthelvencin C. As already observed for the assembly of other pyrrolamides (congoicidine,
 255 distamycin), the assembly of anthelvencins involves a non-canonical NRPS constituted of
 256 stand-alone domains only. Such structural organization, rarely encountered in NRPS,
 257 appears to be characteristic of pyrrolamide NRPSs. Characterizing biosynthetic pathways
 258 involving type II NRPSs is of interest for the production of novel specialized metabolites by
 259 synthetic biology as the architecture of these NRPSs is particularly suited for combinatorial
 260 biosynthesis approaches.

261

262

263 **MATERIALS AND METHODS**

264 **Bacterial strains, plasmids and growth conditions.** Strains and plasmids used in this
265 study are listed in Table S1 and S2. *Escherichia coli* strains were grown at 37 °C in LB or
266 SOB complemented with MgSO₄ (20 mM final), supplemented with appropriate antibiotics as
267 needed²². The Soya Flour Mannitol (SFM) medium²³ was used for genetic manipulations of
268 *Streptomyces* strains and spore stocks preparations at 28°C. *Streptomyces* strains were
269 grown for three days at 28°C in MP5²⁴ for anthelvencins (1-3) production.

270
271 **DNA preparation and manipulations.** All oligonucleotides used in this study were
272 purchased from Eurofins and are listed in Table S3. The High fidelity DNA polymerase
273 Phusion (Thermo Fisher Scientific) was used to amplify the DNA fragments for the
274 construction of the suicide plasmids. DreamTaq polymerase (Thermo Fisher Scientific) was
275 used for PCR verification of plasmids and of the replacement of the targeted genes by the
276 resistance cassette *aac(3)IV*. DNA fragments were purified from agarose gels using the
277 Nucleospin Gel and PCR clean-up kit from Macherey-Nagel. *E. coli* transformations and *E.*
278 *coli*/*Streptomyces* conjugations were performed according to standard procedures^{25,23}.

279
280 ***S. venezuelae* ATCC 14583 sequencing and assembly.** Total DNA of *S. venezuelae*
281 ATCC 14583 was extracted following the salting out procedure²³ except that the mycelium
282 was ground in liquid nitrogen. A paired-end library of the whole genome was constructed and
283 sequenced at the high throughput sequencing core facility of I2BC with a MiSeq M01342
284 instrument (Illumina), generating 5.45 million 301 bp reads that were assembled using Velvet
285 v1.2.10. The GenBank accession number of the anthelvencin gene cluster is MK483114.

286
287 **Construction of the replacement mutants.** The suicide plasmid pANT007 was constructed
288 to replace the *ant8* gene by an *aac(3)IV* resistance cassette in *S. venezuelae*. This vector
289 was constructed by assembling the three following inserts in the backbone pOSV400
290 (bearing a hygromycin resistance gene)²⁶: a 1.8 kb fragment upstream of *ant8*, the resistance
291 cassette *aac(3)IV* and a 2.0 kb DNA fragment downstream of *ant8*. The 1.8 kb and 2.0 kb

292 DNA fragments from *S. venezuelae* ATCC 14583 were amplified by PCR with the primers
293 CEA001/CEA002 and CEA003/CEA004 respectively. The PCR products were purified and
294 ligated into pCR® Blunt, yielding pANT001 (containing the upstream DNA fragment) and
295 pANT002 (containing the downstream DNA fragment). Both plasmids were verified by
296 sequencing (Genewiz). The *aac(3)IV* resistance cassette was obtained by digestion of
297 pW60²⁷ by *Hind*III. The 1.8 kb *Hind*III-*Xho*I fragment from pANT001, the 1.0 kb *Hind*III
298 *aac(3)IV* fragment, and the 2.0 kb *Hind*III-*Spe*I fragment from pANT002 were then ligated
299 into the *Xho*I-*Spe*I-digested pOSV400, yielding pANT007. The pANT007 plasmid was
300 verified by restriction digestion and was introduced into *S. venezuelae* ATCC 14583 by
301 intergeneric conjugation from the *E. coli* ET12567/pUZ8002/pANT007 strain. Double-
302 recombinant mutants were selected on SFM plates with 50 µg/mL apramycin and screened
303 for hygromycin sensitivity. DNA from the resulting strain, ANT007, was isolated and analyzed
304 by PCR using the primers A5, A6, and CEA013-CEA016 (See Figure S8). The same protocol
305 was used for the construction of the ANT008 mutant strain (replacement of *ant23*), using the
306 plasmid pANT008, and the ANT009 mutant strain (replacement of *ant24*), using the plasmid
307 pANT009 (see Tables S2 and S3 for plasmid names and for primer sequences).

308

309 **Construction of the ANT012 strain overexpressing Ant23 and Ant24.** The DNA region
310 containing *ant23-ant24* was amplified by PCR from *S. venezuelae* ATCC 14583 genomic
311 DNA using the primers CEA034/CEA035. The PCR product was purified and cloned into
312 pCR® Blunt, yielding pANT011. The sequence of the insert was confirmed by sequencing.
313 The 2.0 kb *Nhe*I/*Afl*III fragment from pANT011 was ligated into the *Spe*I/*Afl*III-digested
314 pCEA005²⁸. The obtained plasmid, containing *ant23* and *ant24* under the *rpsL*(TP) promoter,
315 was named pANT012 and confirmed by restriction digestion. This plasmid was introduced
316 into *S. venezuelae* ATCC 14583 by intergeneric conjugation. The correct integration of
317 pANT012 was verified on the extracted DNA by PCR using the primers CEA_vec_seq14 and
318 CEA_vec_seq15 and the strain was named ANT012.

319

320 **Genetic complementation of ANT008.** As the ANT008 strain bears the *aac(3)IV* resistance
321 marker, the previously constructed pANT012 plasmid, bearing an *aac(3)IV* resistance gene
322 cassette, could not be used for the genetic complementation of the strain. Thus, the 2.4 kb
323 *NsiI/AflII* DNA fragment of pANT012 containing *ant23* and *ant24* under the control of the
324 *rpsL*(TP) promoter was ligated into the *NsiI/AflII*-digested pOSV806²⁸. The resulting plasmid
325 was named pANT013 and was introduced into ANT008 by intergeneric conjugation. The
326 strain obtained was named ANT013.

327
328 **Chemical synthesis of 5-amino-3,4-dihydro-2H-pyrrole-2-carboxylate (4).** Compound (4)
329 was prepared according to a previously described synthetic procedure²⁹ (Scheme 2).
330 Commercially available DL-pyroglutamic acid (7) was converted into the corresponding
331 methyl ester (8) by treatment with thionyl chloride (2 equiv.), and N,N-dimethylformamide
332 (DMF) (2 mol %) in methanol. Derivative (8) was then submitted to a reaction with
333 triethylxonium tetrafluoroborate (Meerwein's salt, 1.4 equiv.) in dichloromethane (DCM) to
334 form carboximidate (9) in quantitative yield. This compound subsequently reacted with
335 ammonium chloride (1.05 equiv.) in refluxing methanol to provide product (10) in 61% yield.
336 Hydrolysis of the ester moiety of compound (10) finally afforded the desired acid (4) in a
337 quantitative yield. A detailed synthesis protocol is available in the Supporting Information.

338
339 **Chemical complementation of ANT009.** *S. venezuelae* ANT009 strain was cultivated in
340 50 mL of MP5. After 24 h, the cultures were separated in two 25 mL cultures, and 1 mM of
341 (4) (final concentration) was added to one of the cultures. After a total of 72 h of cultivation,
342 culture supernatants were analyzed by HPLC as described below.

343
344 **HPLC analysis of culture supernatants.** *S. venezuelae* ATCC 14583 and its derivatives
345 were cultivated in MP5 medium for three days at 28°C. The supernatants were filtered using
346 Mini-UniPrep syringeless filter devices (0.2 µm, Whatman). The samples were analyzed on
347 an Atlantis C₁₈ T3 column (250 mm x 4.6 mm, 5 µm, column temperature 28°C) using an

348 Agilent 1200 HPLC instrument with a quaternary pump. Samples were eluted in isocratic
349 conditions with 0.1% v/v HCOOH in H₂O (solvent A)/ 0.1% v/v HCOOH in CH₃CN (solvent B)
350 (95:5) at 1 ml.min⁻¹ for 7 min, followed by a gradient to 40:60 A/B over 23 min. Anthelvencins
351 were detected by monitoring absorbance at 297 nm.

352

353 **LC-HR-MS-MS analyses.** The resuspended elution fraction obtained above was analyzed by
354 LC-HR-MS². The analysis was performed using a Dionex Ultimate 3000 HPLC system
355 coupled with a Maxis II™ QTOF mass spectrometer (Bruker, MA, USA) fitted with an
356 electrospray ionization (ESI) source.

357

358 Chromatographic analysis was performed using a C₁₈ Acclaim™ RSLC
359 PolarAdvantage II (2.1 x 100 mm, 2.2 μm pore size) column (Thermo Scientific, MA, USA).
360 Column temperature was set at 40 °C and 2 μL of each sample was injected via an
361 autosampler cooled to 4 °C. A flow rate of 0.3 mL/min was used and the eluent was
362 introduced directly into the MS for ion detection. Elution was conducted with a mobile phase
363 consisting of 0.1% v/v HCOOH in H₂O (solvent A) and 0.1% v/v HCOOH in CH₃CN (solvent
364 B) following the gradient elution profile: 0 min, 5% solvent B; 2 min, 5% solvent B; 9 min,
365 50% solvent B; 15 min 90% solvent B; 17 min 90% solvent B; 19 min 5% solvent B; 21 min
366 5% solvent B. In the first half minute of each run, a sodium formate solution was injected
367 directly as an internal reference for calibration. The acquisition parameters of the ESI source
368 were set up as follows: electrospray voltage for the ESI source: 3500V, nebulising gas (N₂)
369 pressure: 35 psi, drying gas (N₂) flow: 8 L/min, and drying temperature: 200°C. Mass spectra
370 were recorded over the m/z range 100-1300 at a frequency of 2 Hz, in positive ion mode. For
371 MS/MS analysis, the cycle time was of 3 sec. Mass spectra were recorded over the m/z
372 range 100-1300 at a frequency of 2 Hz, in positive ion mode. Selected parent ion at m/z
373 442.23 was fragmented at a fixed collision energy value of 40 eV and an isolation window of
374 0.5 amu.

375

376 **Accession codes.** The GenBank accession number of the anthelvencin gene cluster is
377 MK483114. Its MiBig³⁰ accession number is BGC0002042.

378
379 *Supporting Information Available:* The protocol for the synthesis of 5-amino-
380 3,4-dihydro-2H-pyrrole-2-carboxylate (4), the strains, plasmids and oligonucleotides used in
381 this study and seven supplementary figures are available in the Supporting Information. This
382 material is available free of charge via the internet at <http://pubs.acs.org>

383

384 **Acknowledgements**

385 We acknowledge the High-throughput sequencing facility of I2BC for its sequencing
386 and bioinformatics expertise (Centre de Recherche de Gif – <http://www.i2bc.paris-saclay.fr/>).
387 We thank Z. Hong, Y. Li, S. Prado and D. Touboul for their help with the LC-HRMS² analysis.
388 The research received funding from ANR-14-CE16-0003-01. The funders had no role in
389 study design, data collection and interpretation, or the decision to submit the work for
390 publication.

391

392

393 **References**

394

- 395 (1) Probst, G. W., Hoehn, M. M., and Woods, B. L. (1965) Anthelvencins, new antibiotics with
396 anthelmintic properties. *Antimicrob. Agents Chemother.* *5*, 789–795.
397 (2) Neidle, S. (2001) DNA minor-groove recognition by small molecules. *Nat. Prod. Rep.* *18*,
398 291–309.
399 (3) Juguet, M., Lautru, S., Francou, F.-X., Nezbedová, S., Leblond, P., Gondry, M., and
400 Pernodet, J.-L. (2009) An iterative nonribosomal peptide synthetase assembles the pyrrole-
401 amide antibiotic congocidine in *Streptomyces ambofaciens*. *Chem. Biol.* *16*, 421–431.
402 (4) Vingadassalon, A., Lorieux, F., Juguet, M., Le Goff, G., Gerbaud, C., Pernodet, J.-L., and
403 Lautru, S. (2015) Natural combinatorial biosynthesis involving two clusters for the synthesis
404 of three pyrrolamides in *Streptomyces netropsis*. *ACS Chem. Biol.* *10*, 601–610.
405 (5) Hao, C., Huang, S., Deng, Z., Zhao, C., and Yu, Y. (2014) Mining of the pyrrolamide
406 antibiotics analogs in *Streptomyces netropsis* reveals the amidohydrolase-dependent
407 “iterative strategy” underlying the pyrrole polymerization. *PLoS One* *9*, e99077.
408 (6) Al-Mestarihi, A. H., Garzan, A., Kim, J. M., and Garneau-Tsodikova, S. (2015) Enzymatic
409 Evidence for a revised congocidine biosynthetic pathway. *Chembiochem Eur. J. Chem. Biol.*
410 *16*, 1307–1313.

411 (7) Lautru, S., Song, L., Demange, L., Lombès, T., Galons, H., Challis, G. L., and Pernodet,
412 J.-L. (2012) A sweet origin for the key congocidine precursor 4-acetamidopyrrole-2-
413 carboxylate. *Angew. Chem. Int. Ed Engl.* 51, 7454–7458.

414 (8) Jaremko, M. J., Davis, T. D., Corpuz, J. C., and Burkart, M. D. (2020) Type II non-
415 ribosomal peptide synthetase proteins: structure, mechanism, and protein-protein
416 interactions. *Nat. Prod. Rep.* <https://doi.org/10.1039/C9NP00047J>

417 (9) Schinko, E., Schad, K., Eys, S., Keller, U., and Wohlleben, W. (2009) Phosphinothricin-
418 tripeptide biosynthesis: an original version of bacterial secondary metabolism?
419 *Phytochemistry* 70, 1787–1800.

420 (10) Cai, X., Nowak, S., Wesche, F., Bischoff, I., Kaiser, M., Fürst, R., and Bode, H. B. (2017)
421 Entomopathogenic bacteria use multiple mechanisms for bioactive peptide library design.
422 *Nat. Chem.* 9, 379–386.

423 (11) Takaishi, T., Sugawara, Y., and Suzuki, M. (1972) Structure of Kikumycin A and B.
424 *Tetrahedron Lett.* 13, 1873–1876.

425 (12) Takizawa, M., Tsubotani, S., Tanida, S., Harada, S., and Hasegawa, T. (1987) A new
426 pyrrole-amidine antibiotic TAN-868 A. *J. Antibiot. (Tokyo)* 40, 1220–1230.

427 (13) Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local
428 alignment search tool. *J. Mol. Biol.* 215, 403–410.

429 (14) Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F.,
430 Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lu, F., Marchler,
431 G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., Geer, L. Y.,
432 and Bryant, S. H. (2017) CDD/SPARCLE: functional classification of proteins via subfamily
433 domain architectures. *Nucleic Acids Res.* 45, D200–D203.

434 (15) Witt, E. M. H. J., Davies, N. W., and Galinski, E. A. (2011) Unexpected property of
435 ectoine synthase and its application for synthesis of the engineered compatible solute ADPC.
436 *Appl. Microbiol. Biotechnol.* 91, 113–122.

437 (16) Galperin, M. Y., and Koonin, E. V. (1997) A diverse superfamily of enzymes with ATP-
438 dependent carboxylate-amine/thiol ligase activity. *Protein Sci. Publ. Protein Soc.* 6, 2639–
439 2643.

440 (17) Goswami, A., and Van Lanen, S. G. (2015) Enzymatic strategies and biocatalysts for
441 amide bond formation: tricks of the trade outside of the ribosome. *Mol. Biosyst.* 11, 338–353.

442 (18) Hollenhorst, M. A., Clardy, J., and Walsh, C. T. (2009) The ATP-dependent amide
443 ligases DdaG and DdaF assemble the fumaramoyl-dipeptide scaffold of the dapdiamide
444 antibiotics. *Biochemistry* 48, 10467–10472.

445 (19) Shao, Z., Rao, G., Li, C., Abil, Z., Luo, Y., and Zhao, H. (2013) Refactoring the Silent
446 Spectinabilin Gene Cluster Using a Plug-and-Play Scaffold. *ACS Synth. Biol.* 2, 662–669.

447 (20) Lee, M., Coulter, D. M., and Lown, J. W. (1988) Total synthesis and absolute
448 configuration of the antibiotic oligopeptide (4S)-(+)-anthelvencin A and its 4R(-) enantiomer.
449 *J. Org. Chem.* 53, 1855–1859.

450 (21) Lee, M., and Lown, J. W. (1987) Synthesis of (4S)- and (4R)-methyl 2-amino-1-
451 pyrroline-5-carboxylate and their application to the preparation of (4S)-(+)- and (4R)-(-)-
452 dihydrokikumycin B. *J. Org. Chem.* 52, 5717–5721.

453 (22) Sambrook, J., and Russell, D. (2000) Molecular Cloning: A Laboratory Manual 3rd
454 Revised edition. Cold Spring Harbor Laboratory Press, U.S.

455 (23) Kieser, T., Bibb, M., Buttner, M., and Hopwood, D. A. (2000) Practical *Streptomyces*
456 Genetics, John Innes Foundation, Norwich NR47UH, UK.

457 (24) Pernodet, J. L., Alegre, M. T., Blondelet-Rouault, M. H., and Guérineau, M. (1993)
458 Resistance to spiramycin in *Streptomyces ambofaciens*, the producer organism, involves at
459 least two different mechanisms. *J. Gen. Microbiol.* 139, 1003–1011.

460 (25) Sambrook, J., and Russell, D. W. (2001) Molecular cloning: A Laboratory Manual, Third
461 edition. CSHL Press, Cold Spring Harbor, NY.

462 (26) Boubakri, H., Seghezzi, N., Duchateau, M., Gominet, M., Kofroňová, O., Benada, O.,
463 Mazodier, P., and Pernodet, J.-L. (2015) The Absence of pupylation (prokaryotic ubiquitin-
464 like protein modification) affects morphological and physiological differentiation in
465 *Streptomyces coelicolor*. *J. Bacteriol.* 197, 3388–3399.

466 (27) Corre, C., Song, L., O'Rourke, S., Chater, K. F., and Challis, G. L. (2008) 2-Alkyl-4-
467 hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by
468 *Streptomyces coelicolor* genome mining. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 17510–17515.
469 (28) Aubry, C., Pernodet, J.-L., and Lautru, S. (2019) Modular and Integrative Vectors for
470 Synthetic Biology Applications in *Streptomyces* spp. *Appl. Environ. Microbiol.* *85*.
471 (29) Lee, M., and Lown, J. W. (1987) Synthesis of (4S)- and (4R)-methyl 2-amino-1-
472 pyrroline-5-carboxylate and their application to the preparation of (4S)-(+)- and (4R)-(-)-
473 dihydrokikumycin B. *J. Org. Chem.* *52*, 5717–5721.
474 (30) Kautsar, S. A., Blin, K., Shaw, S., Navarro-Muñoz, J. C., Terlouw, B. R., van der Hooft,
475 J. J. J., van Santen, J. A., Tracanna, V., Suarez Duran, H. G., Pascal Andreu, V., Selem-
476 Mojica, N., Alanjary, M., Robinson, S. L., Lund, G., Epstein, S. C., Sisto, A. C., Charkoudian,
477 L. K., Collemare, J., Linington, R. G., Weber, T., and Medema, M. H. (2020) MIBiG 2.0: a
478 repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* *48*, D454–
479 D458.
480