



**HAL**  
open science

## **Structure, biosynthesis, and properties of kurstakins, nonribosomal lipopeptides from *Bacillus* spp.**

Max Béchet, Thibault Caradec, Walaa Hussein, Ahmed Abderrahmani,  
Marlène Chollet, Valérie Leclère, Thomas Dubois, Didier Lereclus, Maude  
Pupin, Philippe Jacques

### ► **To cite this version:**

Max Béchet, Thibault Caradec, Walaa Hussein, Ahmed Abderrahmani, Marlène Chollet, et al.. Structure, biosynthesis, and properties of kurstakins, nonribosomal lipopeptides from *Bacillus* spp.. *Applied Microbiology and Biotechnology*, 2012, 95 (3), pp.593-600. 10.1007/s00253-012-4181-2 . hal-00749819

**HAL Id: hal-00749819**

**<https://hal.science/hal-00749819>**

Submitted on 8 Nov 2012

**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 **Structure, Biosynthesis and Properties of Kurstakins, NonRibosomal**

2 **Lipopeptides from *Bacillus* spp.**

3 **Max Béchet<sup>a</sup>, Thibault Caradec<sup>a</sup>, Walaa Hussein<sup>a</sup>, Ahmed Abderrahmani<sup>b</sup>,**

4 **Marlène Chollet<sup>a</sup>, Valérie Leclère<sup>b</sup>, Thomas Dubois<sup>c</sup>, Didier Lereclus<sup>c</sup>,**

5 **Maude Pupin<sup>d</sup> and Philippe Jacques<sup>a\*</sup>**

6 <sup>a</sup>Laboratoire des Procédés Biologiques, Génie Enzymatique et Microbien (ProBioGEM),

7 UPRES-EA 1026, Polytech'Lille/IUT A, Université Lille Nord de France, Avenue Paul

8 Langevin, F-59655 Villeneuve d'Ascq Cedex, France

9 <sup>b</sup>Laboratoire de Biologie Cellulaire et Moléculaire, Faculté des Sciences Biologiques,

10 Université des Sciences et de la Technologie Houari Boumediene, B.P. 32, El Alia, Alger,

11 Algérie

12 <sup>c</sup>INRA, UMR1319 Micalis, La Minière, F-78280 Guyancourt, France

13 <sup>d</sup>LIFL, UMR USTL/CNRS 8022, INRIA Lille-Nord Europe, F-59655 Villeneuve d'Ascq

14 Cedex, France

15

16

17

18 **Running title:** Kurstakins

19

20 **Keywords:** kurstakins, lipopeptides, *Bacillus cereus*, *Bacillus thuringiensis*, NRPS, spreading

21

22 \*Correspondance to: Dr Philippe Jacques, Laboratoire des Procédés Biologiques, Génie

23 Enzymatique et Microbien, UPRES-EA 1026, Polytech'Lille/IUT A, Université Lille Nord de

24 France - Sciences et Technologies, USTL, Avenue Paul Langevin, F-59655 Villeneuve

25 d'Ascq Cedex, France

26 Tel: +33328767440, Fax +33328767356 (e-mail: [philippe.jacques@polytech-lille.fr](mailto:philippe.jacques@polytech-lille.fr))

## 27Abstract

28 A new family of lipopeptides produced by *Bacillus thuringiensis*, the kurstakins, was  
29discovered in 2000 and considered as a biomarker of this species. Kurstakins are  
30lipoheptapeptides displaying antifungal activities against *Stachybotrys charatum*. Recently,  
31the biosynthesis mechanism, the regulation of this biosynthesis and the potential new  
32properties of kurstakins were described in the literature. In addition, kurstakins were also  
33detected in other species belonging to *Bacillus* genus such as *Bacillus cereus*. This mini-  
34review gathers all the information about these promising bioactive molecules.

35

### 36Introduction

37 Between 1949 and 1986, three different families of nonribosomal lipopeptides were  
38identified in *Bacillus* spp.: surfactins, iturins and fengycins (Jacques, 2011). A new family  
39produced by *Bacillus thuringiensis* and named the kurstakins, was discovered in 2000. The  
40recent characterization of the biosynthesis mechanism of these compounds (in 2009 and  
412011), their main properties (in 2011 and 2012) and the regulation of their biosynthesis in  
422012, open new perspectives for these lipopeptidic compounds. This mini-review is the first  
43one dedicated to them.

### 44Discovery, structure and mass spectrometry detection

45 Among the spores from six ATCC *Bacillus thuringiensis* strains, the presence of a  
46lipophilic biomarker, named kurstakin, was detected for the first time from the *B.*  
47*thuringiensis* subsp. *kurstaki* strain HD-1 (Hathout et al. 2000). After a spore washing, the  
48authors identified using LC-MS four  $[M+H]^+$  molecular ions of  $m/z$  879, 893, 893 and 907  
49(with a common fragment ion at  $m/z$  of 609). The molecular masses of these four compounds  
50differed by 14 Da ( $-CH_2-$ ), suggesting that these molecules were homologous lipopeptides.  
51Further acid hydrolysis led to the identification of the corresponding free fatty acids from the  
52four compounds which are 9-methyldecanoic, dodecanoic, 10-methylundecanoic and 11-  
53methyl dodecanoic acids, generating kurstakins  $C_{11}$  (*iso-*),  $C_{12}$  (n- and *iso-*) and  $C_{13}$  (*iso-*),  
54respectively. Amino acid analyses revealed the same residues for the four molecules: Thr,  
55His, Ala, Gly, Ser and Glx (Gln or Glu) with molar ratios of 1:1:1:1:1:2. The ascertainment of  
56the chemical structures of the kurstakins was completed by the determinations of (i) the  
57sequence of the heptapeptide: Thr-Gly-Ala-Ser-His-Gln-Gln; (ii) the presence of an amide  
58bond between the fatty acid chain and the first threonine residue and (iii) the presence of a  
59lactone linkage between the serine at position 4 and the C terminus of glutamine at position 7

60(Fig. 1). Such a peculiar lactone ring between the fourth and seventh amino acids of the  
61peptidic part was recently reported for a new biosurfactant (licheniformin) produced by  
62*Bacillus licheniformis* MS3 (Biria et al. 2010). However, neither the nonribosomal peptide  
63synthesis statute of kurstakins in *B. thuringiensis* subsp. *kurstaki* strain HD-1 nor the possible  
64occurrence of D- forms among the seven amino acids of their peptidic moiety (compared to  
65other known nonribosomal lipopeptides from *Bacillus* spp.) were demonstrated. Three years  
66later, a homologous series of three ions at  $m/z$  892, 906, and 920 similar to those of kurstakins  
67was detected by Madonna et al. (2003) in *Bacillus subtilis* ATCC 6051 but these results were  
68not confirmed by genetic analyses (see below). Another study dealing with numerous *Bacillus*  
69sp. strains isolated worldwide further revealed the presence of kurstakins in 20 from 54 strains  
70tested, using MALDI-TOF-MS fingerprinting of whole bacterial cells (Price et al. 2007).  
71These were typically identified by the molecular ions of  $m/z$  889, 905, 917, and 933 but their  
72primary structures or those of other putative kurstakins of  $m/z$  about 942 and 958 were not  
73elucidated. The authors confirmed that these secondary metabolites were retained by the  
74spores or cells and not secreted because they predominantly found them in the bacterial  
75colonies on agar plates.

76       New information about the structural diversity of the kurstakins was recently reported  
77by MALDI-TOF-MS analyses of the *Bacillus* sp. strain NK2018 (Bumpus et al. 2009) and six  
78*B. thuringiensis* strains (Abderrahmani et al. 2011). From culture supernatants of strain  
79NK2018 grown in an M9 minimal medium, the first authors showed the presence of six  
80kurstakin variants in differing amounts, with molecular masses ranging from 907.4765 to  
81953.5192 Da (formulae from  $C_{40}H_{65}O_{13}N_{11}$  to  $C_{42}H_{71}O_{14}N_{11}$ ), corresponding to kurstakins with  
82a  $\beta$ -hydroxylated fatty acid chain with 12, 13 and 14 atoms of carbon. The occurrence of three  
83molecules differing by exactly 18.0103 Da from three other ones was attributed to the  
84presence or lack of a lactone ring inside the peptidic part of the molecules, suggesting that

85kurstakins might be found currently in culture supernatant as variants with linear peptidic  
86parts. After growth on either AK or LB medium, Abderrahmani et al. (2011) detected the  
87presence of the three C<sub>11</sub>, C<sub>12</sub> and C<sub>13</sub> kurstakin isoforms in six *B. thuringiensis* strains from  
88the 11 tested. Some other molecular ions with m/z of 920, 942 and 958 were also sometimes  
89detected and could correspond to the kurstakin with a C<sub>14</sub> fatty acid chain (Abderrahmani  
902011). To summarize, the kurstakins synthesized by *Bacillus* spp. consist of lipopeptides  
91which (i) are linked to between C<sub>11</sub> and C<sub>14</sub> fatty acids, β-hydroxylated or not, with two  
92isoforms (n-, iso-); (ii) are partially cyclic (lactone bond between Ser/4 and C-terminal Gln/7)  
93and might even be linear; and (iii) are expected to contain two D-configured amino acids. In  
94MALDI-TOF-MS experiments, the values of the [M+H]<sup>+</sup>, [M+Na]<sup>+</sup> and [M+K]<sup>+</sup> molecular  
95ions detected should range from 878.473 (cyclic C<sub>11</sub> isoform [M+H]<sup>+</sup>) to 992.481 Da (linear  
96β-hydroxylated C<sub>14</sub> isoform [M+K]<sup>+</sup>), under conditions allowing these biomarkers to be  
97detected (Figs. 1 and 2).

## 98Biosynthesis

99 The operons potentially encoding kurstakin synthetases in *B. cereus* and *B.*  
100*thuringiensis* were identified by bioinformatics analyses using two new approaches. In the  
101first one, Bumpus et al. (2009) took advantage of the size of the NRPS enzymes and the  
102presence of unique marker ions derived from the common phosphopantetheinyl cofactor to  
103adapt mass spectrometry-based proteomics to detect selectively NRPS and PKS gene clusters  
104in microbial proteomes without requiring genome sequence information. In these conditions,  
105the authors highlighted in strains *Bacillus* sp. NK2018 and *Bacillus cereus* AH1134 the genes  
106involved in the biosynthesis of the kurstakins. The second approach used PCR with  
107degenerate primers based on the intraoperon DNA sequence alignment of adenylation and  
108thiolation domains of all enzymes implicated in the biosynthesis of the lipopeptide family

109(Tapi et al. 2010). Two sets of primers elaborated from first bacillomycin genes, then from  
110kurstakin genes led to the discovery of genes implied in kurstakin biosynthesis (Tapi 2010;  
111Tapi et al. 2010; Abderrahmani et al. 2011). From these two studies the organization of the  
112kurstakin cluster could be predicted. This cluster (Fig. 3) contains three genes (*krsA*, *krsB* and  
113*krsC*) which encode three large multifunctional proteins (KrsA, KrsB and KrsC) constituting  
114the complete synthetase. This latter is organized as follows: KrsA comprises one module  
115(m1), KrsB is constituted of two modules (m2 and m3) and KrsC includes four modules (m4  
116to m7) and for each module a condensation-adenylation-thiolation motif can be found. In  
117addition, m1 and m6 harbour a supplementary epimerization domain. The module 7 includes a  
118final thioesterase domain enabling the unhooking of the neo-formed peptide from the NRPS  
119and its possible cyclization. The combination of the predictions obtained from different  
120bioinformatics tools (Ansari et al. 2004; Raush et al. 2005; Bachmann and Ravel 2009) led to  
121a peptide with the primary structure: D-Thr\_Gly/Ala/Gln/Glu\_Ala\_Ser\_X\_D-Gln\_Gln.

122        While the precise mechanism of biosynthesis was not yet experimentally analysed, it  
123could easily be deduced from the bioinformatics analysis and the high quantity of available  
124information on the biosynthesis of other lipopeptides from *Bacillus* or *Pseudomonas* species  
125(Sieber and Marahiel, 2005; Raaijmakers et al. 2010; Roongsawang et al. 2010) (Fig. 3). The  
126first synthetase, KrsA, contains one module with a starting condensation domain (Rausch et  
127al. 2007; Kraas et al. 2010) which catalyses the link between a fatty acid chain of different  
128length and isomery and a threonine residue activated by the adenylation domain and  
129covalently fixed to the thiolation domain of this first module. The last domain of this first  
130module is an epimerization domain that will transform the L-Thr to D-Thr. The second  
131synthetase, KrsB, contains two modules responsible for the activation and the incorporation of  
132two other amino acid residues, Gly and Ala. In the third synthetase, KrsC, four modules are  
133involved in the incorporation of the four other amino acid residues: Ser, His, Gln and Gln.

134The presence of an epimerization domain in the third module probably modifies the  
135incorporated L-Gln to form the D-configuration. The last module contains a thioesterase  
136which will catalyse the liberation of the peptide and probably its partial cyclization (Kopp and  
137Marahiel 2007). A gene encoding a phosphopantetheinyl transferase was identified  
138downstream from the NRPS complex. This enzyme could be involved in the transformation of  
139the apo-form of the enzymes to the holo-form by the addition of a phosphopantetheinyl group  
140from the coenzymeA to the different thiolation domains (Mofid et al. 2004). A second  
141thioesterase (TEII) is encoded by the next gene (*krsD*). The results obtained by Schwarzer et  
142al. (2002) on the biochemical characterization of two second similar thioesterases named TEII  
143and involved in the surfactin and bacitracin peptide antibiotics biosynthesis showed that these  
144enzymes play a role in the regeneration of the misacylated Peptidyl Carrier Proteins  
145(Thiolation domains). The presence of linear peptide in strain NK2018 could result from the  
146action of this second thioesterase or an insufficient expression or efficacy of the first one.

147 A sixth gene (*krsE*) situated upstream from the *krsA-C* genes and belonging to the  
148kurstakin cluster was identified by Dubois et al. (2012) in *Bacillus thuringiensis* Bt407. The  
149product of this gene, the protein KrsE, is a presumed efflux protein and could be involved in  
150the secretion of the lipopeptide.

### 151**Overview on kurstakin potentially producing microorganisms**

152 In order to identify the genetic potential for kurstakin production among  
153microorganisms, we performed a BLASTp method using KrsA, KrsB, KrsC, KrsD, KrsE and  
154Ppant sequences from *B. thuringiensis* serovar pondicheriensis BGSC 4BA1 as queries. The  
155same cluster leading to kurstakin synthesis was retrieved in the genome of 32 strains for  
156which genomes are sequenced, assembled and either finished or still as drafts (V. Leclère, M.  
157Pupin, W. Hussein, P. Jacques, unpublished data). Without exception, all the strains pointed



158out belong to the *Bacillus* genus, more especially to the *B. cereus* species group, indicating  
159that kurstakin production could be considered as the marker for this group. However, no  
160sequences of kurstakin synthetases were present in *Bacillus anthracis* and *Bacillus*  
161*cytotoxicus* although they belong to the same *B. cereus* group. However, only one genome is  
162available for *B. cytotoxicus* and five genomes are completely sequenced and assembled for  
163*B. anthracis*. So the question of lack of kurstakin synthetase in the *B. anthracis* species  
164remains open and should be related later to the high virulence of the strains. In addition,  
165kurstakins were also detected in *B. subtilis* ATCC 6051 strain (Madonna et al. 2003). This  
166strain was an ancestor of the reference strain 168 (Zeigler et al. 2008), the genome of which  
167was completely sequenced and, surprisingly, no traces of kurstakin genes were found in this  
168genome. The kurstakin cluster is present in most of the genomes of *B. cereus* and *B.*  
169*thuringiensis* for which genomic data are available (Table 1). Kurstakin genes were also  
170detected in the partially sequenced genome of *B. cereus* BDRD-Cer4 and *B. cereus* AH1134.  
171When the sequences are present in the genomes, they are highly conserved and the  
172organization of the cluster (KrsE-KrsA-KrsB-KrsC-Ppant-TE) is also conserved. However,  
173the strains *B. cereus* AH603, BDRD-ST196, *B. mycooides* DSM 2048 and *B.*  
174*weihenstephanensis* KBAB4 might produce a variant form with a Glu or Asp instead of Gln at  
175the last position as predicted by NRPSpredictor2 (Röttig et al. 2011). As no amino acid  
176residue can be predicted for module 6, the strain *B. cereus* Rock4-2 can be supposed to  
177produce another member of the kurstakin family varying by the residue at this position..

### 178Regulation

179 The regulation system of kurstakin production has been partially described in *Bacillus*  
180*thuringiensis* Bt407 (Fig. 4). A transcriptomic analysis indicates that the four genes *krsEABC*  
181form a cluster whose transcription is activated by the NprR-NprX quorum-sensing system  
182during late stationary phase (Dubois et al. 2012). NprR is a quorum sensor activated by its

183cognate signaling peptide, NprX. NprR-NprX functions as a typical Gram-positive quorum-  
184sensing system: the pro-signaling peptide NprX is exported from the bacterial cell and after  
185being processed to its active form (presumably a heptapeptide), the peptide is reimported into  
186the bacteria, where it binds to NprR allowing the recognition of its DNA target (Perchat et al.  
1872011).

188       The NrpX-NrpR system regulates 41 different genes, divided into four different  
189groups. The first group is composed of genes coding for stress resistance proteins, including  
190cytochrome P450, cysteine dioxygenase and several metabolite exporters. The second group  
191is composed of four genes encoding the Opp permease system, involved in the import of  
192small peptides into the cells. The third group is composed of the NRPS *krs* genes. The last  
193group codes for degradative enzymes and proteins able to bind organic material (Dubois et al.  
1942012).

### 195**Properties**

196       Kurstakins are not recovered in the culture supernatant of producing strains but are  
197found in association with the bacterial cells and particularly on spores (Hathout et al. 2000;  
198Price et al. 2007; Abderrahmani et al. 2011). However, co-infection experiments carried out  
199with a producing strain and a non-producing one in the insect larvae *Galleria mellonella*,  
200suggest that this molecule is secreted (Dubois et al. 2012). This apparent discrepancy between  
201these results suggests that kurstakin is a secreted molecule with a high affinity for membranes.  
202This high affinity could be due to the presence of the basic amino acid histidine which confers  
203a cationic charge to the lipopeptide and thus facilitates its electrostatic interaction with  
204phospholipids of the cell membrane.

205       Purified kurstakins displayed an antifungal activity against *Stachybotrys charatum*,  
206showing a halo of inhibition identical to the one obtained with polymyxin B used as positive

207control (Hathout et al. 2000). Nevertheless, Abderrahmani et al. (2011) showed that no  
208correlation exists between the antifungal activities of the strains and the presence of  
209kurstakins. Indeed, some producing strains did not show any antifungal activity whereas some  
210other ones did not produce kurstakin and showed antifungal properties. However, their  
211evaluation was made with the fungi *Mucor rouxii* DSM 1191, *Rhizopus orizae* DSM 907,  
212*Penicillium roqueforti* DSM 1080, *Aspergillus niger* DSM 737, and *Fusarium oxysporum*  
213DSM 62297, different from those used by Hathout et al. These data indicate that kurstakin  
214might be a pore-forming molecule with a limited spectrum of activity.

215       The fact that significant colonization of solid media was detected neither for non-  
216producing kurstakin strains nor for a kurstakin-deficient mutant indicates that kurstakins are  
217responsible for the invasive growth (Abderrahmani et al. 2011). Moreover, a strain where the  
218genes *krsA*, *krsB* and *krsC* ( $\Delta krsABC$ ) were deleted was unable both to swarm and to form a  
219biofilm at the air/liquid interface (Dubois et al. 2012). A very interesting property of kurstakin  
220is its ability to enhance the survival of *B. thuringiensis* in the insect cadaver (Dubois et al.  
2212012). In view of these various properties, kurstakin might allow *B. thuringiensis* to spread  
222across the cadaver, thereby facilitating access to new substrates and increasing its ability to  
223disseminate in the environment.

#### 224**Perspectives**

225       The research on this fourth family of lipopeptides produced by *Bacillus* spp. has yet to  
226be developed, and several perspectives are worth considering. The precise structure of the  
227different variants should be confirmed by chemical analysis: the presence of the D-amino acid  
228residues should be validated by analysis of amino acid residues after acid hydrolysis and  
229derivatization, e.g., by GC using chirasyl-L-Val column. Confirmation of the presence of  
230linear structure or C14 fatty acid chain should be done by LC-MS-MS analysis and NMR.  
231The predicted biosynthetic pathway proposed in this review should be confirmed by

232biochemical analysis of the different domains of the synthetase. Particular attention will have  
233to be paid to the thioesterase domains and their role in the concomitant presence of cyclized  
234and linear forms of the lipopeptides.

235 Hathout et al. (2000) have evaluated the amount of kurstakin produced at about 15-20  
236µg/mg of spore. Overproducing mutant cells could be constructed using similar strategies  
237developed by Leclère et al. (2005), Fickers et al. (2009) and Coutte et al. (2010), for the  
238overproduction of other families of lipopeptides from *B. subtilis*. Purification techniques need  
239to be developed to extract the lipopeptides or collect them in the supernatant.

240 Lipopeptides from *Bacillus* spp. are well known for their potential applications in  
241several fields (Jacques, 2011) including biocontrol of plant pathogens (Ongena and Jacques,  
2422008). Purified compounds could be thus used in different physico-chemical or biological  
243tests in order to characterize their physico-chemical properties and biological activities and  
244their potential applications.

#### 245**Acknowledgements**

246 This work received financial support from the Université Lille 1, Sciences et  
247Technologies and ARCIR funds from Nord-Pas-de-Calais and the European Funds of  
248INTERREG IV PhytoBio Project. We acknowledge William Everett for his kind English  
249proof reading.

250

#### 251**References**

252Abderrahmani A, Tapi A, Nateche F, Chollet M, Leclère V, Wathelet B, Hacene H, Jacques P  
253(2011) Bioinformatics and molecular approaches to detect NRPS genes involved in the  
254biosynthesis of kurstakin from *Bacillus thuringiensis*. Appl Microbiol Biotechnol 92:571-581

255

256Abderrahmani A (2011) Identification du mécanisme de synthèse non ribosomique d'un  
257nouveau lipopeptide, la kurstakine et étude de son influence sur le phénotype de souches de  
258*Bacillus thuringiensis* isolées en Algérie. Thèse de Doctorat d'Etat, Université des Sciences et  
259de la Technologie Houari Boumediene, Alger, Algérie

260

261Ansari MZ, Yadav G, Gokhale, RS, Mohanty D (2004) NRPS-PKS: a knowledge-based  
262resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res* 32:405-413

263

264Bachmann BO, Ravel J (2009) Methods for *in silico* prediction of microbial polyketide and  
265nonribosomal peptide biosynthetic pathways from DNA sequence data. *Methods Enzymol*  
266458:181-217

267

268Birria D, Maghsoudi E, Roostaazad R, Dadafarin H, Sahebghadam Lotfi A, Amoozegar MA  
269(2010) Purification and characterization of a novel biosurfactant produced by *Bacillus*  
270*licheniformis* MS3. *World J Microbiol Biotechnol* 26:871-878

271

272Bumpus SB, Evans BS, Thomas PM, Ntai I, Kelleher NL (2009) A proteomics approach to  
273discovering natural products and their biosynthetic pathways. *Nat Biotechnol* 27:951-956

274

275Coutte F, Leclère V, Béchet M, Guez JS, Lecouturier D, Chollet-Imbert M, Dhulster P,  
276Jacques P (2010) Effect of *pps* disruption and constitutive expression of *surfA* on surfactin  
277productivity, spreading and antagonistic properties of *Bacillus subtilis* 168 derivatives. *J Appl*  
278*Microbiol* 109:480-491

279

280Dubois T, Faegri K, Perchat S, Lemy C, Buisson C, Nielsen-LeRoux C, Gohar M, Jacques P,  
281Ramarao N, Kolsto AB, Lereclus D (2012) Necrotrophism is a quorum-regulated lifestyle in  
282*Bacillus thuringiensis*. PloS Pathogens 8:e1002629

283

284Fickers P, Guez JS, Damblon C, Leclère V, Béchet M, Jacques P, Joris B (2009) High-level  
285biosynthesis of the anteiso-C(17) isoform of the antibiotic mycosubtilin in *Bacillus subtilis*  
286and characterization of its candidacidal activity. Appl Environ Microbiol 75:4636-4640

287

288Hathout Y, Ho YP, Ryzhov V, Demirev P, Fenselau C (2000) Kurstakins: a new class of  
289lipopeptides isolated from *Bacillus thuringiensis*. J Nat Prod 63:1492-1496

290

291Jacques P (2011) Surfactin and other lipopeptides from *Bacillus spp.* In: Soberon-Chavez G  
292(ed) Biosurfactants microbiology monographs, Volume 20, Springer-Verlag, Berlin  
293Heidelberg, pp 57-91

294

295Kopp F, Marahiel MA (2007) Macrocyclization strategies in polyketide and nonribosomal  
296peptide biosynthesis. Nat Prod Rep 24:735-749

297

298Kraas FI, Helmetag V, Wittman M, Strieker M, Marahiel MA (2010) Functional dissection of  
299surfactin synthetase initiation module reveals insights into the mechanism of lipoinitiation.  
300Chem Biol 17:872-880

301

302Leclère V, Béchet M, Adam A, Guez JS, Wathelet B, Ongena M, Thonart P, Gancel F,  
303Chollet-Imbert M, Jacques P (2005) Mycosubtilin overproduction by *Bacillus subtilis*

304BBG100 enhances the organism's antagonistic and biocontrol activities. Appl Environ  
305Microbiol 71:4577-4584

306

307Madonna AJ, Voorhees KJ, Taranenko NI, Laiko VV, Doroshenko VM (2003) Detection of  
308cyclic lipopeptide biomarkers from *Bacillus* species using atmospheric pressure matrix-  
309assisted laser desorption/ionization mass spectrometry. Anal Chem 75:1628-1637

310

311Mofid MR, Finking R, Essen L, Marahiel MA (2004) Structure-based mutational analysis of  
312the 4'-phosphopantetheinyl transferases Sfp from *Bacillus subtilis*: carrier protein recognition  
313and reaction mechanism. Biochemistry 43:4128-4136

314

315Ongena M, Jacques P (2008) *Bacillus* lipopeptides: versatile weapons for plant diseases  
316biocontrol. Trends Microbiol 16:115-125

317

318Perchat S, Dubois T, Zouhir S, Gominet M, Poncet S, Lemy C, Aumont-Nicaise M,  
319Deutscher J, Gohar M, Nessler S, Lereclus D (2011) A cell-cell communication system  
320regulates protease production during sporulation in bacteria of the *Bacillus cereus* group. Mol  
321Microbiol 82:619-633

322

323Price NP, Rooney AP, Swezey JL, Perry E, Cohan FM (2007) Mass spectrometric analyses of  
324lipopeptides from *Bacillus* strains isolated from diverse geographical locations. FEMS  
325Microbiol Lett 271:83-89

326

327Raaijmakers JM, De Bruijn I, Nybroe O, Ongena M (2010) Natural functions of lipopeptides  
328from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics. FEMS Microbiol Rev  
32934:1037-1062

330

331

332Rausch C, Weber T, Kohlbacher O, Wohlleben W, Huson DH (2005) Specificity prediction of  
333adenylation domains in nonribosomal peptide synthetases (NRPS) using Transductive Support  
334Vector Machines (TSVM). Nucl Acids Res 33:5799-5808

335

336Rausch C, Hoof I, Weber T, Wohlleben W, Huson DH (2007) Phylogenetic analysis of  
337condensation domains in NRPS sheds light on their functional evolution. BMC Evol Biol 7:78

338

339Roongsawang N, Washio K, Morikawa M (2010) Diversity of nonribosomal peptide  
340synthetases involved in the biosynthesis of lipopeptide biosurfactants. Int J Mol Sci 12:141-

341172

342

343Röttig M, Medema MH, Blin K, Weber T, Rausch C, Kohlbacher O (2011) NRSPredictor2-a  
344web server for predicting NRPS adenylation domain specificity. Nucl Acids Res 39 (Web  
345Server issue):W362-367

346

347Schwarzer D, Mootz HD, Linne U, Marahiel MA (2002) [Regeneration of misprimed](#)  
348[nonribosomal peptide synthetases by type II thioesterases](#). Proc Natl Acad Sci USA 99:14083-

34914088

350



351Sieber SA, Marahiel MA (2005) Molecular mechanisms underlying nonribosomal peptide  
352synthesis: approaches to new antibiotics. *Chem Rev* 105:715-738

353

354Tapi A (2010) Stratégie moléculaire de mise en évidence de peptides actifs d'origine non  
355ribosomiale chez *Bacillus* sp. et *Lactobacillus* sp. PhD thesis, Université Lille1 Sciences et  
356Technologies, France

357

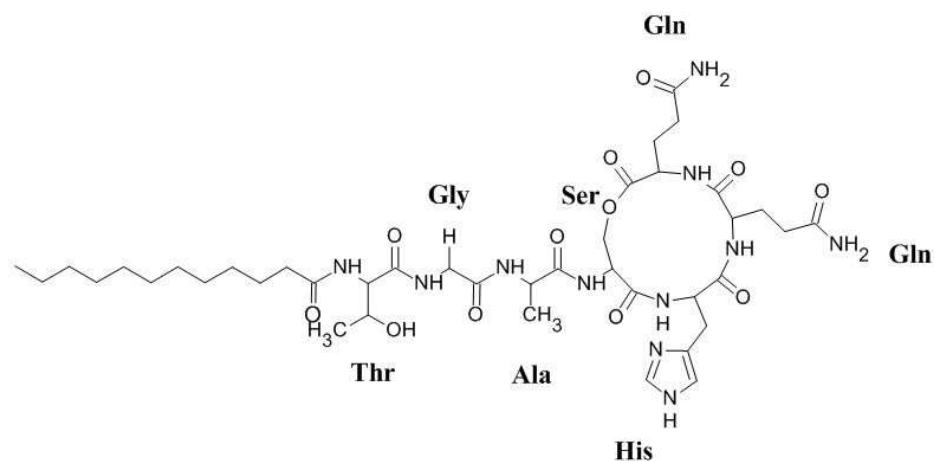
358Tapi A, Chollet-Imbert M, Scherens B, Jacques P (2010) New approach for the detection of  
359non ribosomal peptide synthetase genes in *Bacillus* strains by polymerase chain reaction. *Appl*  
360*Microbiol Biotechnol* 85:1521–1531

361

362Zeigler DR, Pragai Z, Rodriguez S, Chevreux B, Muffler A, Albert T, Bai R, Wyss M,  
363Perkins JB (2008) The origins of 168, W23, and other *Bacillus subtilis* legacy strains. *J*  
364*Bacteriol* 190:6983-6995

| Table 1 Presence of kurstakin genes in sequenced genomes of <i>B. cereus</i> and <i>B. thuringiensis</i>   |                                      |                    |             |             |             |            |             |
|--|--------------------------------------|--------------------|-------------|-------------|-------------|------------|-------------|
| Strains  | Sequenced genome status <sup>1</sup> | Genes <sup>2</sup> |             |             |             |            |             |
|  |                                      | <i>krsE</i>        | <i>krsA</i> | <i>krsB</i> | <i>krsC</i> | <i>sfp</i> | <i>krsD</i> |
| Reference strain: <i>Bacillus thuringiensis</i> serovar pondicheriensis BGSC 4BA1  | C                                    | +                  | +           | +           | +           | +          | +           |
| <i>B. thuringiensis</i>  |                                      |                    |             |             |             |            |             |
| BMB171; sv chinensis CT-43   | C                                    | +                  | +           | +           | +           | +          | +           |
| sv finitimus; sv konkukian str. 97-27; str.A1Hakam   | C                                    | -                  | -           | -           | -           | -          | -           |
| Bt407; IBL 200; sv berliner ATCC 10792; sv huazhongensis BGSC 4BD1; sv kurstaki str. T03a001; sv thuringiensis str. T01001; sv pakistani str. T13001; sv pulsiensis BGSC 4CC1; sv sotto str. T04001              | U                                    | +                  | +           | +           | +           | +          | +           |
| IBL 4222   | U                                    | +                  | +           | +           | TCT ?       | +          | +           |
| <i>B. cereus</i>   |                                      |                    |             |             |             |            |             |
| ATCC 14579; B4264; G9842   | C                                    | +                  | +           | +           | +           | +          | +           |
| 03BB102; AH187; AH820; ATCC10987; E33L; Q1; bv anthracis str.CI  | C                                    | -                  | -           | -           | -           | -          | -           |
| 172560W; ATCC 10876; BDRD-ST24; BGSC 6E1; F65185; m1550; Rock1-15; Rock1-3   | U                                    | +                  | +           | +           | +           | +          | +           |
| BDRD-Cer4  | P                                    | +                  | +           | +           | +           | +          | +           |
| AH1134   | P                                    | +                  | +           | NF          | +           | NF         | NF          |
| AH603; BDRD-ST196  | U                                    | +                  | +           | +           | 7-D/E       | +          | +           |
| AH621  | U                                    | +                  | NF          | NF          | TCT ?       | +          | NF          |
| AH676  | U                                    | +                  | +           | +           | +           | +          | NF          |
| Rock4-2  | U                                    | +                  | +           | +           | 6-X         | +          | +           |
| Others species   |                                      |                    |             |             |             |            |             |
| <i>B. mycoides</i> DSM 2048  | C                                    | +                  | +           | +           | 7-D/E       | +          | +           |
| <i>B. weihenstephanensis</i> KBAB4   | C                                    | +                  | +           | +           | 7-D/E       | +          | +           |
| <sup>1</sup> Status of genome sequencing (GS): C complete, U: unfinished, P: partial   |                                      |                    |             |             |             |            |             |
| <sup>2</sup> NF: Not Found, TCT?: truncated ?, 7-D/E: prediction of amino acid residue incorporated by module 7 is D/E instead of Q, 6-X: no possible prediction for amino acid residue incorporated by module 6 |                                      |                    |             |             |             |            |             |
| + : protein present with at least 90% identity with the reference one, - : not present   |                                      |                    |             |             |             |            |             |

365



366I.

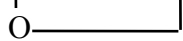
367

368II.

369FA\_Thr-Gly-Ala-Ser-His-Gln-Gln

FA\_Thr-Gly-Ala-Ser-His-Gln-Gln

370



371iC<sub>11</sub>

877.465

n.r.

372C<sub>12</sub>

891.481

n.r.

373C<sub>12</sub>[-OH(3)]

907.476

925.486

374iC<sub>12</sub>

891.481

n.r.

375iC<sub>13</sub>

905.496

n.r.

376iC<sub>13</sub>[-OH(3)]

921.491

939.502

377C<sub>14</sub>[-OH(3)]

935.507

953.518

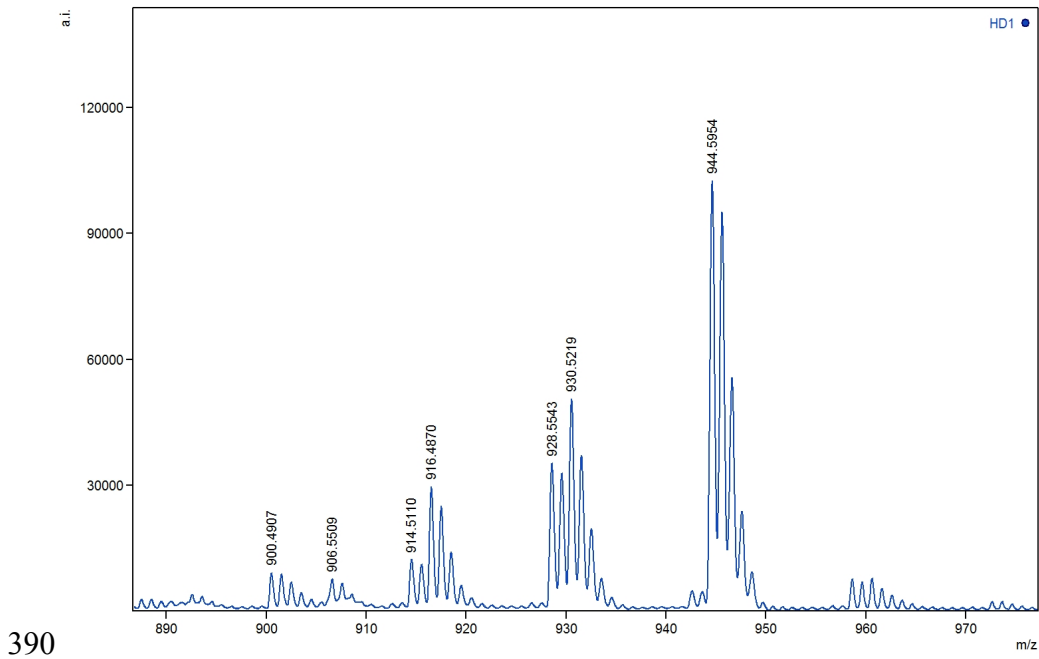
378

**A**

**B**

379**Fig. 1 I.** Chemical structure of cyclic kurstakin with a C12 fatty acid chain. **II.** Structures and  
 380calculated molecular masses (Da) of the different isoforms of kurstakins characterized until  
 381now by MALDI-TOF-MS. **A.** Partially cyclic molecules (Hathout et al. 2000; Bumpus et al.  
 3822009); the square brackets [-OH(3)] in bold correspond to the presence of a  $\beta$ -hydroxylated  
 383fatty acid; **B.** Linear molecules (Bumpus et al. 2009). The amino acids Thr and Gln (in  
 384position 6) are expected to be under the D- form (Abderrahmani et al. 2011) (It is worthy of

385note that L- and D- forms have never been chemically determined). The first kurstakin  
386homologues, chemically characterized by Hathout et al. (2000), did not contain a  $\beta$ -  
387hydroxylated fatty acid, and comprised the sole  $iC_{11}$ -,  $iC_{12}$ - and  $iC_{13}$  isoforms isolated so far.  
388**FA** = fatty acid; n.r. = not reported to date.



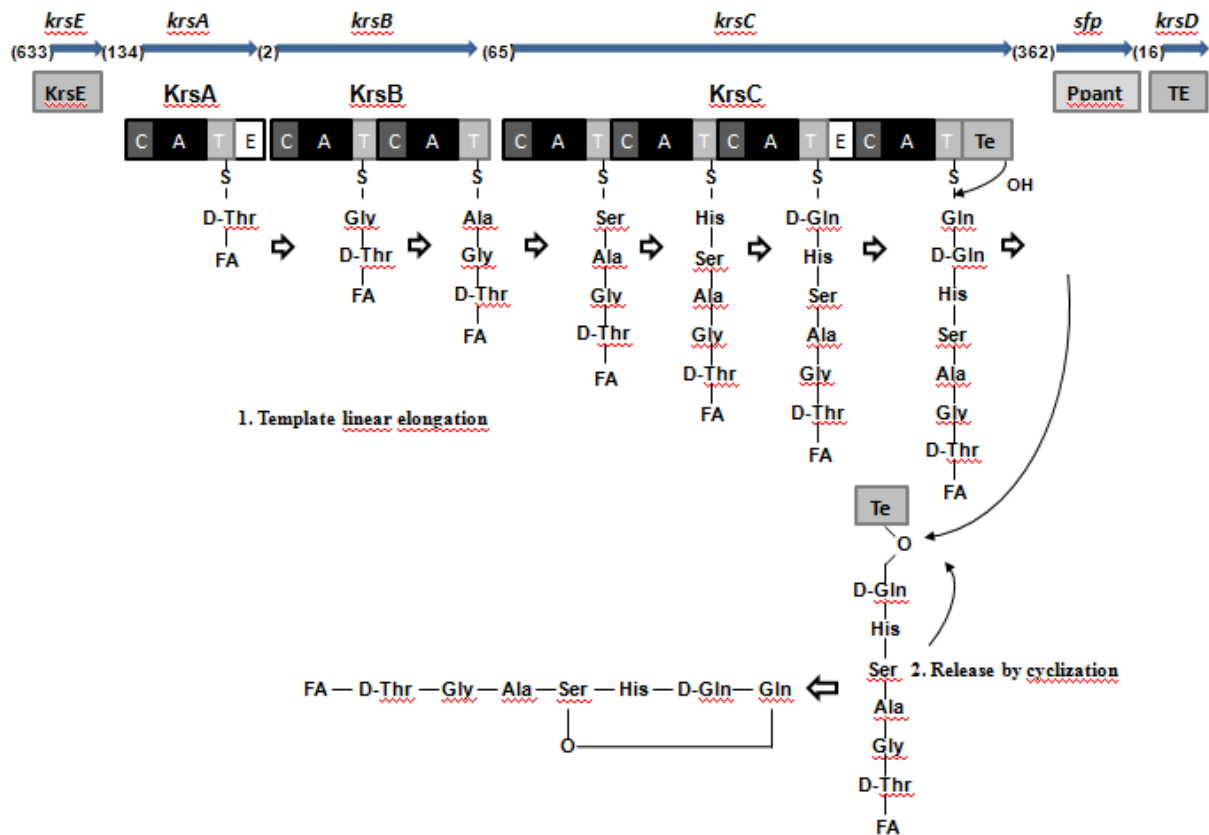
391 **Fig. 2** Typical pattern of whole cell MALDI-TOF MS analysis of the kurstakin producing  
 392 strain *B. thuringiensis* subsp. *kurstaki* HD-1 (Caradec et al., unpublished data).

393 Cyclic isoform with C<sub>11</sub> fatty acid chain: 900.4907 [M+Na]<sup>+</sup>; 916.4870 [M+K]<sup>+</sup>

394 Cyclic isoform with C<sub>12</sub> fatty acid chain: 914.5110 [M+Na]<sup>+</sup>; 930.5219 [M+K]<sup>+</sup>

395 Cyclic isoform with C<sub>13</sub> fatty acid chain: 906.5509 [M+H]<sup>+</sup>; 928.5543 [M+Na]<sup>+</sup>; 944.5944

396 [M+K]<sup>+</sup>



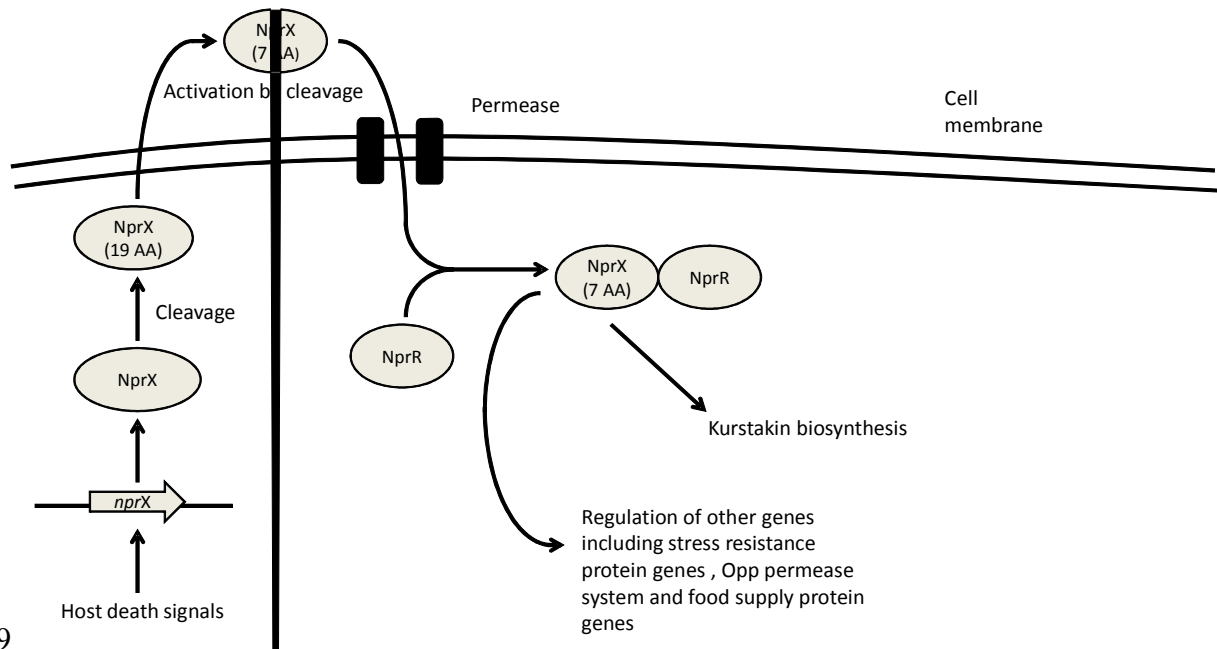
399**Fig. 3** Hypothetical kurstakin biosynthetic assembly line based on bioinformatics analysis

400 Situated downstream from the putative *krsE* gene which potentially encodes for an efflux  
 401 protein which could be involved in the kurstakin secretion (bthur0010\_59510), the  
 402 biosynthetic complex consists of three kurstakin synthetases: KrsA, KrsB and KrsC (encoded  
 403 by three genes: bthur0010\_59520, bthur0010\_59530 bthur0010\_59540 in *Bacillus*  
 404 *thuringiensis* serovar pondicheriensis BSCG 4BA1), divided into seven distinct modules.  
 405 Each module is responsible for recognition, activation and loading of a single amino acid  
 406 substrate. In the first module, a starter condensation domain links a fatty acid chain to the  
 407 amino acid residue (Thr) activated and fixed in this module. Two epimerization domains are  
 408 found in modules 1 and 6, converting the corresponding amino acid in the D-stereoisomer.  
 409 The cyclization and release of the final heptapeptide is catalyzed by the first adjacent Te

410domain. Immediately downstream are situated two genes, coding for a phosphopantetheinyl  
411transferase (bthur0010\_59550) and a closely adjacent type II thioesterase (bthur0010\_59560),  
412which are expected to belong to the kurstakin cluster. The numbers in brackets correspond to  
413the spaces (in nucleotides) between these different open reading frames borne by strain BSCG  
4144BA1. C: Condensation domain; A: Adenylation domain; T: Thiolation domain; E:  
415Epimerization domain; Te: Thioesterase domain; Ppant: Phosphopantetheinyl transferase  
416(Sfp); TE: Type II thioesterase; FA: Fatty acid

417

418



419

420**Fig. 4** Regulation of kurstakin biosynthesis

421The NprX peptide is produced as an inactive form and is then activated by two successive  
 422cleavages. The first cleavage happens in the cell cytoplasm, and leads to a 19 amino acid  
 423peptide formation. After this first cleavage, the peptide is exported out of the cell, and is  
 424cleaved a second time in a 7 amino acids peptide, leading to its active form NprX. This  
 425peptide is then reimported within the bacterial cell through the Opp permease system, and is  
 426bound with the NprR regulator. NprR forms a complex with the heptapeptide NprX whose  
 427production is regulated by host-death signals. The NprX-NprR complex activates the  
 428kurstakin production, by binding to the -35 box of the promoter of the *krs* genes (Perchat et al.  
 4292011).

430