nature chemical biology

Cyclodipeptide synthases are a family of tRNA-dependent peptide bond–forming enzymes

Muriel Gondry[1,5], Ludovic Sauguet[1,4,5], Pascal Belin[1], Robert Thai[1], Rachel Amouroux[1,2,4], Carine Tellier[1,2], Karine Tuphile[2,4], Mickaël Jacquet[1], Sandrine Braud[1,2,4], Marie Courçon[1], Cédric Masson[1], Steven Dubois[1], Sylvie Lautru[2], Alain Lecoq[1], Shin-ichi Hashimoto[3,4], Roger Genet[1,4] & Jean-Luc Pernodet[2]

Cyclodipeptides and their derivatives belong to the diketopiperazine (DKP) family, which is comprised of a broad array of natural products that exhibit useful biological properties. In the few known DKP biosynthetic pathways, nonribosomal peptide synthetases (NRPSs) are involved in the synthesis of cyclodipeptides that constitute the DKP scaffold, except in the albonoursin (1) pathway. Albonoursin, or cyclo(α,β-dehydroPhe-α,β-dehydroLeu), is an antibacterial DKP produced by *Streptomyces noursei*. In this pathway, the formation of the cyclo(Phe-Leu) (2) intermediate is catalyzed by AlbC, a small protein unrelated to NRPSs. We demonstrated that AlbC uses aminoacyl-tRNAs as substrates to catalyze the formation of the DKP peptide bonds. Moreover, several other bacterial proteins, presenting moderate similarity to AlbC, also use aminoacyl-tRNAs to synthesize various cyclodipeptides. Therefore, AlbC and these related proteins belong to a newly defined family of enzymes that we have named cyclodipeptide synthases (CDPSs).

Cyclodipeptides and their derivatives, the DKPs (**Fig. 1**), constitute a large class of secondary metabolites synthesized by a wide range of organisms, most of which are microorganisms. Many of these natural products exhibit diverse and noteworthy biological activities. For example, cyclo(L-Phe-L-Pro) (cFP, **3**) and cyclo (L-Phe-*trans*-4-OH-L-Pro) (**4**) act as antifungal compounds[1]. Bicyclomycin (**5**) is an antibacterial agent[2]. Ambewelamides A (**6**) and B (**7**) (which are epipolythiodioxopiperazines (ETPs) characterized by a disulfide bridge across the DKP ring) and phenylahistin (**8**) exhibit antitumor properties[3,4]. The ETP gliotoxin (**9**) and sirodesmin PL (**10**) have antibacterial, antiviral and immunosuppressive properties[5].

Although DKPs are noteworthy bioactive molecules, there are few characterized DKP biosynthetic pathways. In almost all known cases, synthesis of the cyclodipeptide constituting the DKP scaffold is catalyzed by NRPSs[6–11]. NRPSs are large modular multifunctional enzymes, and each module is responsible for the incorporation of one amino acid into the final peptide. Each module consists of domains responsible for a particular stage of the synthesis. Adenylation domains select and activate a free amino acid, forming an aminoacyl adenylate in a process requiring ATP. The activated amino acid is transferred to the peptidyl carrier protein domain, to which it is covalently tethered by a 4′-phosphopantethenyl arm. Condensation domains are responsible for peptide bond formation between aminoacyl substrates bound to adjacent modules. Cyclodipeptides can be formed by dedicated NRPSs (as in thaxtomin A (**11**) or gliotoxin biosynthesis[8,10]), which are then expected to be comprised of at least two modules. Alternatively, cyclodipeptides can be formed by NRPSs during the synthesis of longer peptides, as truncated side products—for example, cyclo(D-Phe-L-Pro) (**12**), which is released prematurely during the synthesis of the decapeptide tyrocidine A (refs. 12,13) (**13**), and cyclomarazines A (**14**) and B (**15**), which are released during the synthesis of the heptapeptide cyclomarin D (ref. 9) (**16**).

The albonoursin biosynthetic pathway is thus far the only known DKP biosynthetic pathway that does not involve enzymes related to NRPSs[14]. Albonoursin (**Fig. 1**) is an antibacterial DKP produced by *S. noursei*[15]. The cyclodipeptide albonoursin precursor, cyclo(L-Phe-L-Leu) (cFL) (**Fig. 1**), is formed by AlbC, a 239-residue protein that is unrelated not only to NRPSs but also to all other functionally characterized proteins[14]. We wondered how such a small enzyme could catalyze the formation of cyclodipeptides, the synthesis of which requires a large NRPS enzyme complex with at least two modules in other pathways. In particular, we questioned how amino acid activation is performed, as AlbC contains no predicted ATP binding motif. Thus, AlbC may use a previously unknown mechanism to form the peptide bonds in DKP scaffolds.

[1]Commissariat à l'Energie Atomique, Institut de Biologie et Technologies de Saclay, Service d'Ingénierie Moléculaire des Protéines, Gif-sur-Yvette, France. [2]Université Paris-Sud 11, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 8621, Institut de Génétique et Microbiologie, Orsay, France. [3]Technical Research Laboratories, Kyowa Hakko Bio Co. Ltd., Hofu-city, Yamaguchi, Japan. [4]Present addresses: Department of Biochemistry, Cambridge University, Cambridge, UK (L.S.); Commissariat à l'Energie Atomique, Institut de Radiobiologie Cellulaire et Moléculaire, Service Instabilité Génétique Réparation Recombinaison, Fontenay-aux-Roses, France (R.A.); Université Paris-Sud 11, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 8619, Institut de Biochimie et de Biophysique Moléculaire et Cellulaire, Orsay, France (K.T.); ObeTherapy Biotechnology, Evry, France (S.B.); Technology Development & Research Department, Kyowa Hakko Bio Co. Ltd., Chiyoda-ku, Tokyo, Japan (S.-i.H.); Cemagref, Antony Cedex, France (R.G.). [5]These authors contributed equally to this work. Correspondence should be addressed to M.G. (muriel.gondry@cea.fr) or J.-L.P. (jean-luc.pernodet@igmors.u-psud.fr).

Received 23 January; accepted 12 March; published online 3 May 2009; doi:10.1038/nchembio.175

Here, we demonstrated that AlbC activity can be reconstituted *in vitro* and that purified AlbC uses activated amino acids in the form of aminoacyl-tRNAs (aa-tRNAs) as substrates to catalyze the ATP-independent formation of cyclodipeptides. We also showed that AlbC is not the only enzyme that has this type of activity, as several proteins originating from various bacterial phyla display similar activities and synthesize diverse cyclodipeptides. AlbC and these proteins constitute a family of tRNA-dependent peptide bond–forming enzymes that we have named cyclodipeptide synthases (CDPSs).

RESULTS

In vitro activity of AlbC

AlbC expressed in *Escherichia coli* synthesizes cFL and cFF (17), which are released into the culture medium[14]. To reconstitute the enzyme activity *in vitro*, we expressed AlbC in *E. coli* and purified the protein to homogeneity by a three-step chromatographic procedure (**Supplementary Fig. 1** online). We incubated the purified enzyme in the presence of tritiated L-phenylalanine and tritiated L-leucine as potential substrates and ATP as the energy donor in various experimental conditions. However, the tritiated amino acids were never consumed, and no tritiated products were detected (**Fig. 2a**). One possible explanation could be that components essential for the enzyme activity had been lost during the purification procedure. To test this hypothesis, we added a soluble extract of *E. coli* cells not expressing AlbC to the incubation mixture. In this case, the tritiated amino acids were incorporated into three products, and this incorporation was dependent on the presence of AlbC (**Fig. 2a**). The two major products were determined to be cFL and cFF, as expected, and the minor one was identified as cLL (18) (see below). These results clearly showed that AlbC is active *in vitro* and is required for the formation of cyclodipeptides. They also demonstrated that one or several components present in the soluble extract are necessary for the enzyme activity but are lost during the purification.

RNA-dependent activity of AlbC

To identify the components required for AlbC activity, we first investigated AlbC-containing extracts of *E. coli* cells, which contained all the components required for AlbC activity. Small molecules with molecular weights below 1,000, including cyclodipeptides synthesized by AlbC during cell culture[14], were removed by desalting the *E. coli* cell extracts. We incubated a desalted AlbC-containing extract with L-phenylalanine, L-leucine and ATP, and we analyzed the resulting incubation mixture for the detection and identification of cyclodipeptides by LC-MS/MS. The three compounds cFL, cFF and cLL were detected in the sample and were absent from the controls (**Fig. 2b–e**). This confirmed that the synthesis of cyclodipeptides is dependent on AlbC, and on the addition of amino acids and ATP. Moreover, it indicated that the components required have a molecular weight over 1,000. To determine which treatment in the purification process was responsible for the loss of the components required for AlbC activity, we submitted the extracts to treatments similar to those used during the purification. Treatment with RNase A resulted in a complete abolition of the enzyme activity (**Fig. 2b,c** and **Supplementary Fig. 2** online), whereas treatment with DNase I had no effect on the formation of cyclodipeptides (**Supplementary Fig. 2**). Thus, ribonucleic components are required for AlbC activity.

Aminoacyl-tRNAs as substrates of AlbC

Having shown that an RNA component is required and knowing that two protein families, aa-tRNA protein transferases and aminoacyl transferases[16,17], use aa-tRNAs as substrates to form peptide bonds, we hypothesized that aa-tRNAs are the substrates used by AlbC to form cyclodipeptides. Thus, under our experimental conditions, the added amino acids would be loaded onto tRNAs, in an ATP-dependent mechanism, by aa-tRNA synthetases present in the extracts, forming the aa-tRNAs used by AlbC. This is consistent with the observed requirements for amino acids, ATP and a ribonucleic component (tRNAs in our hypothesis). Based on this rationale, the abolition of cyclodipeptide formation following RNase A treatment is explained by the degradation of all tRNAs present in the extract, thereby preventing the formation of aa-tRNAs, the substrates of AlbC.

To test our hypothesis, we added *E. coli* tRNAs to AlbC-containing extract treated with RNase A. To prevent the degradation of added tRNAs, RNase A was inactivated with RNasin, an inhibitor of RNase A, before addition of tRNAs. We observed the formation of cFF (**Fig. 2b–e**) after *E. coli* $tRNAs^{Phe}$ were added with L-phenylalanine, L-leucine and ATP to these AlbC-containing extracts; no cyclodipeptide was detected in controls without tRNA addition. If *E. coli* total tRNAs were used instead of $tRNAs^{Phe}$, the synthesis of cFF and also that of cFL and cLL was observed (**Supplementary Fig. 3** online), as expected from previous observations. These results support our hypothesis and indicate that, within AlbC specificity, the nature of the cyclodipeptide synthesized is also dependent on the nature of the aa-tRNAs available. A minor product, identified as cFL, was formed with $tRNA^{Phe}$ use (**Fig. 2c**), which indicates that $tRNAs^{Phe}$ were contaminated by $tRNAs^{Leu}$.

To demonstrate unambiguously that aa-tRNAs are substrates for AlbC, we developed a direct enzymatic assay that uses purified AlbC and purified tritiated Phe-$tRNA^{Phe}$. This *in vitro* reaction led to the formation of tritiated cFF (**Fig. 2f**), showing that aa-tRNAs are the sole substrates required by AlbC to catalyze the formation of cyclodipeptides. In particular, ATP is not necessary, as AlbC uses already-activated amino acids in the form of aa-tRNAs. In light of these results, we propose that AlbC is a member of a new class of tRNA-dependent enzymes catalyzing the formation of cyclodipeptides.

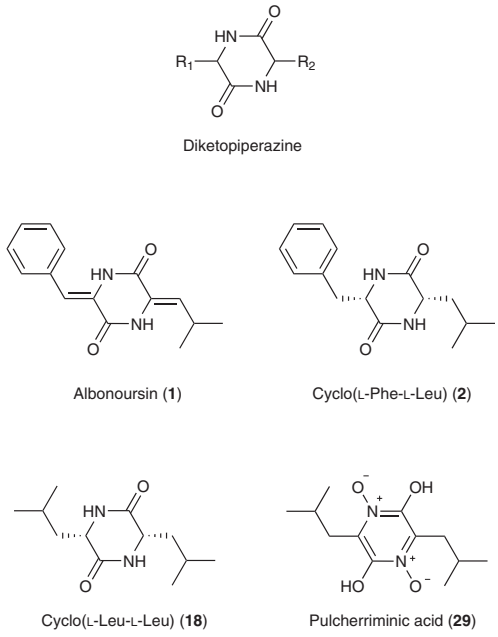

Figure 1 The DKP scaffold and the structures of albonoursin and cyclo(Phe-Leu).

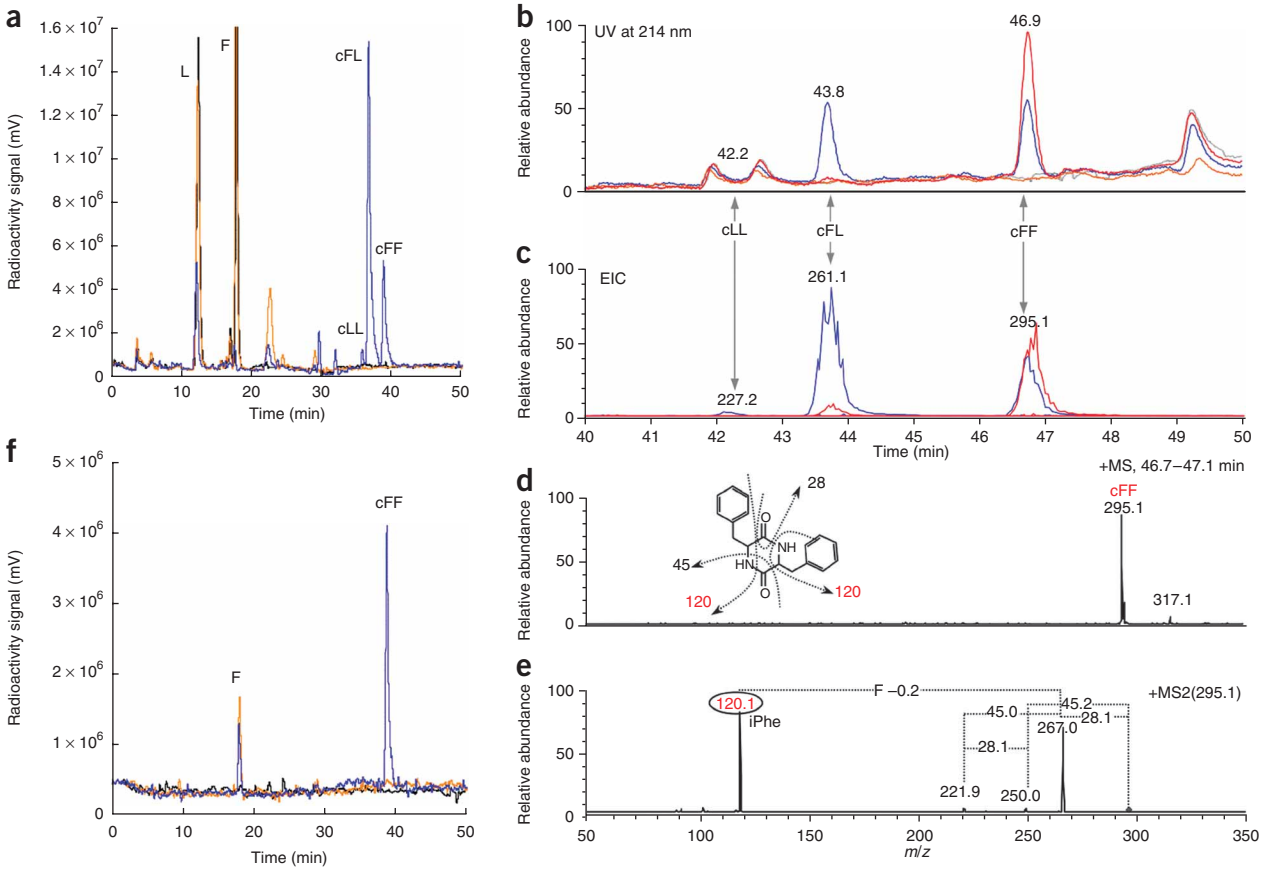

Figure 2 *In vitro* reconstitution of AlbC activity. (**a**) First type of assay (purified AlbC, tritiated L-phenylalanine and L-leucine, ATP, $MgCl_2$), analyzed by radio-HPLC. Radio-chromatograms of an enzyme assay (black), an enzyme assay supplemented with a soluble extract of *E. coli* cells not expressing AlbC (blue), and a control without AlbC but with *E. coli* soluble extract (orange). The compounds obtained in the supplemented enzyme assay were identified as cLL, cFL and cFF by LC-MS/MS analyses (see Methods). (**b**–**e**) Second type of assay (desalted AlbC-containing extracts, L-phenylalanine, L-leucine, ATP, $MgCl_2$), analyzed by LC-MS/MS. (**b**) UV chromatograms of an enzyme assay (blue), a control (extract not containing AlbC, L-phenylalanine, L-leucine, ATP, $MgCl_2$) (orange), an assay in which the AlbC-containing extract was treated sequentially with RNase A and RNasin (gray) and an assay in which *E. coli* $tRNAs^{Phe}$ were added to an extract treated sequentially with RNase A and RNasin (red). (**c**) Corresponding EIC chromatograms with the same color codes. The compounds were identified as cLL, cFL and cFF by LC-MS/MS analyses. (**d**,**e**) Identification of cFF for the assay performed with an extract treated with RNase A and RNasin and supplemented with $tRNAs^{Phe}$. (**d**) MS spectrum of the compound. (**e**) MS/MS spectrum of the compound. The fragmentation pattern is characteristic of a cyclodipeptide, and the encircled peak matches the immonium ion of phenylalanine (iPhe). (**f**) Third type of assay (tritiated Phe-$tRNA^{Phe}$, purified AlbC), analyzed by radio-HPLC. Radio-chromatograms correspond to an enzyme assay (AlbC, tritiated Phe-$tRNA^{Phe}$) (blue) and two controls—an assay without incubation (black) and an assay without purified AlbC (orange).

Cyclodipeptide biosynthesis by AlbC-related proteins

At this stage of the study, AlbC was the only known representative of this newly defined CDPS family. However, we identified by an *in silico* search seven proteins presenting some sequence similarity to AlbC. All were annotated as hypothetical and originated from various bacterial phyla: Firmicutes (*Bacillus*, *Staphylococcus*), Actinobacteria (*Corynebacterium*, *Mycobacterium*) and γ-Proteobacteria (*Photorhabdus*). These seven proteins are similar in size, but display only 19–27% identity and 37–45% similarity with AlbC (**Table 1**)—numbers that were not sufficient for predicting protein function[18]. We aligned sequences for the seven identified proteins and AlbC; conserved residues were found in only 13 positions, half of which are clustered into two small regions, from residues 31 to 37 and 178 to 184 (AlbC numbering) (**Fig. 3**). The two consensus sequences, Hx[LVI][LVI]G[LVI]S and Y[LVI]xxExP, corresponding to these two conserved regions were sufficient to define a specific sequence signature for AlbC-related proteins. Indeed, a search in Uniprot for proteins containing these two consensus sequences separated by 120–160 amino acids using the PATTINPROT software tool[19] detected only AlbC and the seven related proteins. However, no conserved protein domains or secondary structure elements could be identified, even taking into account the protein sequence alignment.

To investigate the biological relevance of this sequence similarity, we tested the ability of these proteins to catalyze the formation of cyclodipeptides, without making assumptions on the nature of the cyclized amino acids. AlbC expressed in *E. coli* synthesizes cyclodipeptides, which are then released into the medium[14]. Thus, each protein was produced in *E. coli*, and the supernatants of these *E. coli* cultures were analyzed for cyclodipeptides by LC-MS/MS (**Supplementary Figs. 4–11** online). Examples of extracted ion current (EIC) chromatograms obtained are given for the culture supernatant of *E. coli* expressing YvmC-Bsub from *Bacillus subtilis* (YvmC-Bsub) and Rv2275 from *Mycobacterium tuberculosis* (**Fig. 4a**,**b**). For each analysis, we compared the chromatograms from the sample and the control—that is, the culture supernatant of *E. coli* cells containing the empty cloning vector. The expression of YvmC-Bsub led to

Table 1 Characteristics of the bacterial proteins related to AlbC

Protein name	Organism	Residue number	Identity/similarity with AlbC (%)	Accession number
AlbC	*Streptomyces noursei* ATCC 11455	239	100/100	AAN07909
Rv2275[a]	*Mycobacterium tuberculosis* H37Rv	289	26/41	CAB00960
YvmC-Bsub[b]	*Bacillus subtilis subtilis* str 168	248	27/44	CAB15512
YvmC-Blic[b,c]	*Bacillus licheniformis* ATCC 14580	249	27/43	AAU25020
YvmC-Bthu[b]	*Bacillus thuringiensis* serovar israelensis ATCC 35646	238	24/43	EAO57133
Plu0297	*Photorhabdus luminescens laumondii* TTO1	234	27/45	CAE12592
pSHaeC06	*Staphylococcus haemolyticus* JCSC1435	234	19/40	BAE05998
Jk0923	*Corynebacterium jeikeium* K411	216	20/37	CAI37087

[a]Homologs of Rv2275 are encoded in all genome sequences available for strains of the *M. tuberculosis* complex (proteins are identical except in *Mycobacterium bovis* BCG strain Pasteur, in which one amino acid is substituted: E261A). [b]The proteins from *B. subtilis*, *B. licheniformis* and *B. thuringiensis* are called YvmC, YvmC and RBTH_07362, respectively, in GenBank. [c]An identical protein is encoded in the genome of *B. licheniformis* DSM 13. Accession numbers are from GenBank.

the synthesis of one major compound (**Fig. 4a**), identified as cLL (**Supplementary Fig. 6**), and expression of Rv2275 also led to the synthesis of one major compound (**Fig. 4b**), identified as cYY (19) (**Supplementary Fig. 5**). The expression of other AlbC-related proteins predominantly led to the formation of cLL (**Supplementary Figs. 7–11**). Thus, all AlbC-related proteins catalyze the formation of cyclodipeptides.

Aminoacyl-tRNAs as substrates of AlbC-related proteins

We then investigated whether AlbC-related proteins use aa-tRNAs as substrates to synthesize cyclodipeptides. However, preparing their potential substrates (Tyr-tRNA^{Tyr} for Rv2275 and Leu-tRNA^{Leu} for the other proteins) was hampered by an absence of easily available tRNA or corresponding aa-tRNA synthetases, which precluded the direct *in vitro* assay. Instead, we used extracts of *E. coli* cells expressing the enzyme studied, as described above with AlbC-containing extracts. We studied YvmC-Bsub from *B. subtilis* as a representative of cLL-forming enzymes and Rv2275, which forms cYY. We prepared extracts of *E. coli* expressing these two enzymes, cleared them of small molecules and incubated them with (i) ATP and L-leucine for extracts containing YvmC-Bsub, or (ii) ATP and L-tyrosine for those containing Rv2275. Reaction mixtures were analyzed by LC-MS/MS. For each analysis, we compared chromatograms from the sample and the control—that is, cleared *E. coli* cell extracts containing the empty cloning vector. The EIC chromatograms showed that, as expected, YvmC-Bsub catalyzes the formation of cLL (**Fig. 4c**) and Rv2275 catalyzes the formation of cYY (**Fig. 4d**). Sequential treatment of extracts with RNase A and RNasin abolished cyclodipeptide formation in both cases. Addition of *E. coli* total tRNAs restored cLL formation in YvmC-Bsub–containing extracts and cYY formation in Rv2275-containing extracts (**Fig. 4c,d**). Thus, YvmC-Bsub and Rv2275 also use aa-tRNAs as substrates to catalyze the formation of cyclodipeptides, as observed for AlbC. YvmC-Bsub can be considered the representative of the five other enzymes that also synthesize cLL. These biochemical activities and sequence similarities strongly suggest that all AlbC-related enzymes use aa-tRNAs as substrates to form cyclodipeptides and are therefore CDPSs.

Substrate specificity of CDPSs

As previously mentioned, AlbC mostly synthesizes cFL, Rv2275 mostly synthesizes cYY and the other CDPSs mostly synthesize cLL. The amount of the major cyclodipeptide produced by each CDPS was quantified and found to be between 16.6 and 117.5 mg l^{-1} of culture supernatant, except for Jk0923 from *Corynebacterium jeikeium* (0.5 mg l^{-1}) (**Fig. 4e**).

LC-MS/MS analyses also revealed that each CDPS produces other cyclodipeptides in smaller amounts (**Supplementary Figs. 4–11**). We identified and quantified all cyclodipeptides synthesized by the eight CDPSs, thus obtaining information on their substrate specificities. AlbC produced ten other cyclodipeptides in addition to the previously identified products, cFL and cFF (ref. 14) (**Fig. 4e**). Thus, AlbC can incorporate various nonpolar residues, such as L-phenylalanine, L-leucine, L-tyrosine and L-methionine, and to a much lesser extent L-alanine and L-valine, into cyclodipeptides. Indeed, the ten possible cyclodipeptides composed of L-phenylalanine, L-leucine, L-tyrosine and L-methionine are all synthesized to detectable amounts by AlbC. Almost all of the compounds produced by other CDPSs are combinations of the same four amino acids, with the restriction that cyclodipeptides synthesized by Rv2275 always contain L-tyrosine, and those synthesized by the other CDPSs almost always contain L-leucine (**Fig. 4e**).

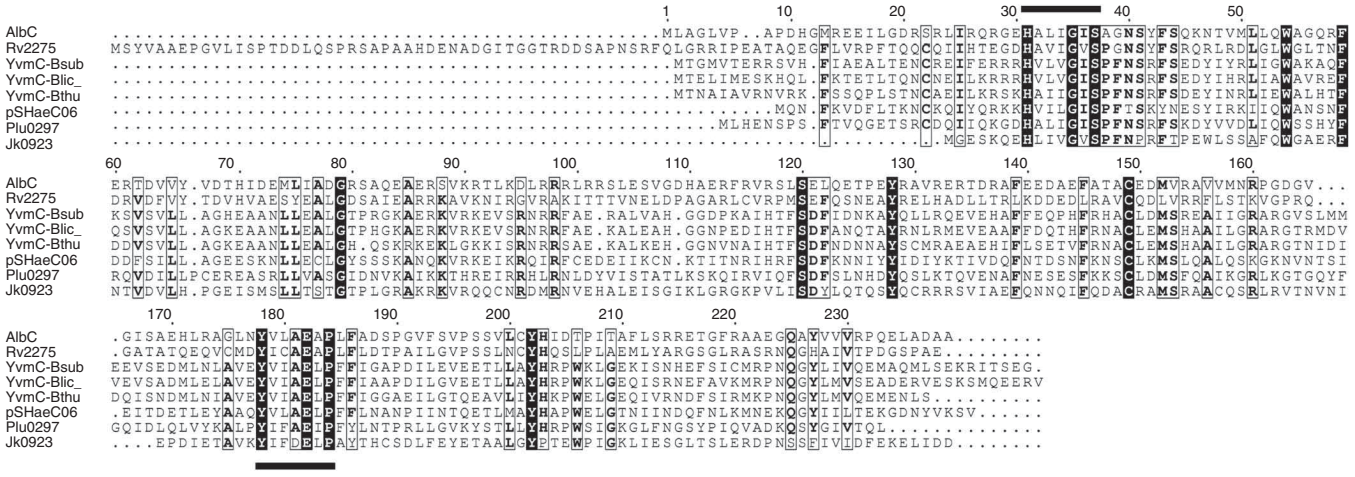

Figure 3 Protein sequence alignments of proteins similar to AlbC. Positions with identical residues in all proteins are indicated by a black background, and positions with identical residues in at least six proteins are boxed. Lines above or below the alignment indicate the position of the two regions with conserved residues, Hx[LVI][LVI]G[LVI]S and Y[LVI]xxExP.

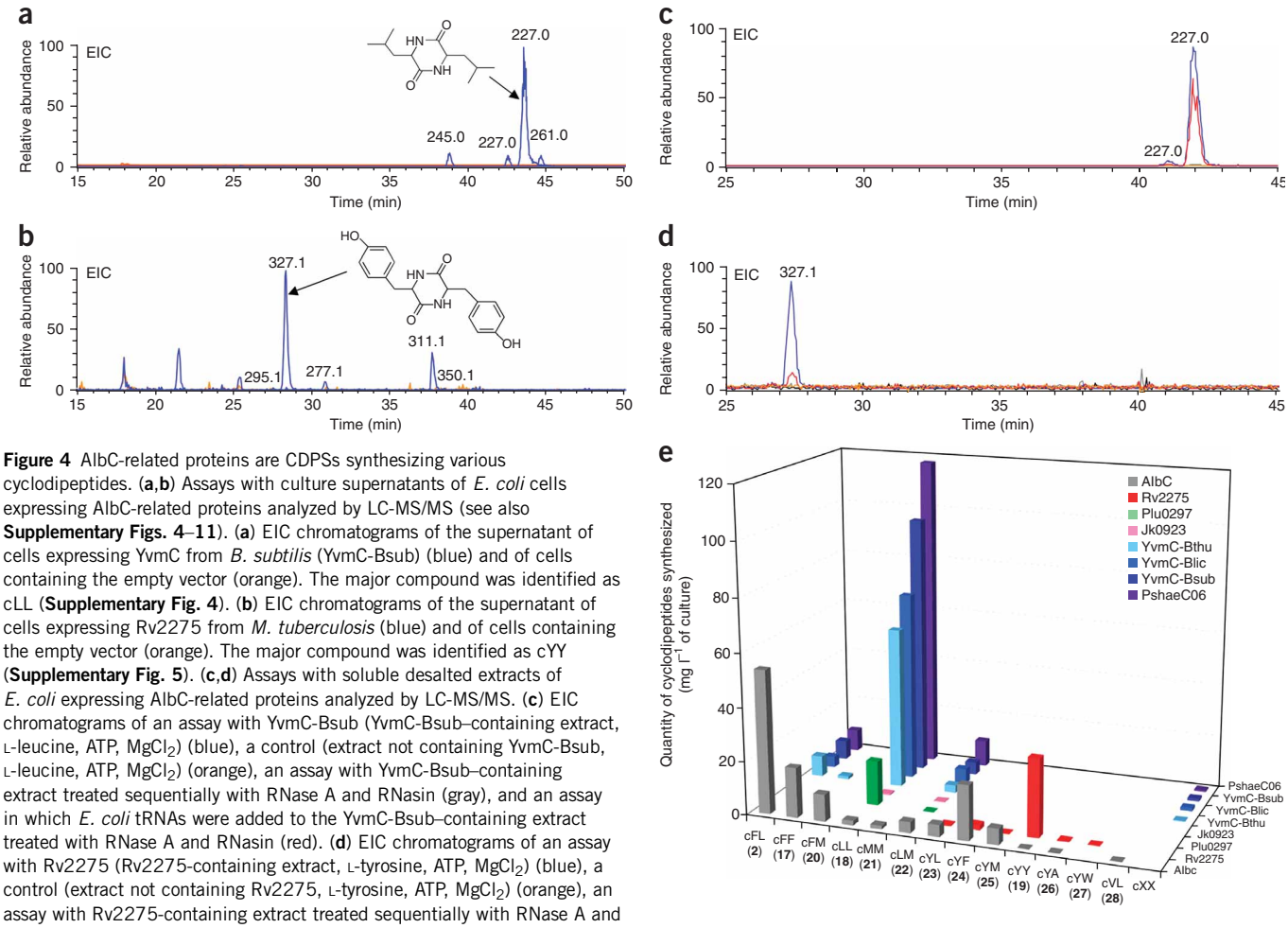

Figure 4 AlbC-related proteins are CDPSs synthesizing various cyclodipeptides. (**a**,**b**) Assays with culture supernatants of *E. coli* cells expressing AlbC-related proteins analyzed by LC-MS/MS (see also **Supplementary Figs. 4–11**). (**a**) EIC chromatograms of the supernatant of cells expressing YvmC from *B. subtilis* (YvmC-Bsub) (blue) and of cells containing the empty vector (orange). The major compound was identified as cLL (**Supplementary Fig. 4**). (**b**) EIC chromatograms of the supernatant of cells expressing Rv2275 from *M. tuberculosis* (blue) and of cells containing the empty vector (orange). The major compound was identified as cYY (**Supplementary Fig. 5**). (**c**,**d**) Assays with soluble desalted extracts of *E. coli* expressing AlbC-related proteins analyzed by LC-MS/MS. (**c**) EIC chromatograms of an assay with YvmC-Bsub (YvmC-Bsub–containing extract, L-leucine, ATP, $MgCl_2$) (blue), a control (extract not containing YvmC-Bsub, L-leucine, ATP, $MgCl_2$) (orange), an assay with YvmC-Bsub–containing extract treated sequentially with RNase A and RNasin (gray), and an assay in which *E. coli* tRNAs were added to the YvmC-Bsub–containing extract treated with RNase A and RNasin (red). (**d**) EIC chromatograms of an assay with Rv2275 (Rv2275-containing extract, L-tyrosine, ATP, $MgCl_2$) (blue), a control (extract not containing Rv2275, L-tyrosine, ATP, $MgCl_2$) (orange), an assay with Rv2275-containing extract treated sequentially with RNase A and RNasin (gray) and an assay in which *E. coli* tRNAs were added to the Rv2275-containing extract treated with RNase A and RNasin (red). (**e**) Histogram of the amounts of the various cyclodipeptides synthesized. cXX corresponds to an unidentified cyclodipeptide of the same molecular mass as cLL.

Modification of cyclodipeptides synthesized by CDPSs

Cyclodipeptide formation is often the first step in the synthesis of more complex DKPs that are obtained after tailoring reactions[11,20]. This also appears to be the case for the cyclodipeptides synthesized by CDPSs, which are probably not the final products, but rather intermediates in DKP biosynthetic pathways. In *S. noursei*, the cyclodipeptide cFL synthesized by AlbC is converted into albonoursin by cyclodipeptide oxidase (CDO), a tailoring enzyme catalyzing the α,β-dehydrogenation of the two aminoacyl side chains[21]. In *S. noursei*, the genes encoding the CDPS AlbC and the tailoring enzyme CDO are clustered. In the bacterial genomes harboring the other CDPS genes, no genes encoding CDO homologs were found. However, all CDPS genes, except for *albC*, are located immediately upstream from genes encoding proteins that are highly similar to those of the cytochrome P450 family (**Supplementary Fig. 12** online). Thus, cyclodipeptides synthesized by these CDPSs may be substrates for these putative cytochrome P450 proteins. Data supporting this hypothesis are available in two cases. In *M. tuberculosis*, *cyp121* is the gene downstream from *Rv2275*. Cytochrome CYP121 was previously characterized and shown to be essential for the viability of *M. tuberculosis*[22], but its natural substrate remained unknown. Our findings showed that Rv2275 synthesizes cYY (**Fig. 4b** and **Supplementary Fig. 5**), and we recently established that CYP121 uses cYY as a substrate to form a new DKP product[23]. In *B. subtilis*, the gene downstream from *yvmC-Bsub* is called *cypX*. In a study combining transcriptional profiling and bioinformatics, *yvmC-Bsub* and *cypX* were suggested to be involved in the biosynthesis of pulcherriminic acid[24] (**29**). Our data established that YvmC-Bsub synthesizes cLL, and we suggested that CypX is responsible for the redox reaction leading to pulcherriminic acid. This is in agreement with the pathway proposed previously on the basis of feeding experiments, in which the relevant enzymes were not identified[25]. Therefore, we suggest that cyclodipeptides synthesized by the other CDPSs are also modified by cytochrome P450 proteins encoded by the downstream genes.

DISCUSSION

We showed that AlbC and seven proteins presenting a moderate similarity to AlbC use aa-tRNAs as substrates to synthesize various cyclodipeptides. They form a family of peptide bond–forming enzymes that we have named CDPSs. The fact that CDPSs do not activate the amino acids, but instead use already activated amino acids in the form of aa-tRNAs, explains how relatively small proteins can catalyze the formation of the two peptide bonds of the DKP scaffold. By contrast, NRPSs, the other enzyme family synthesizing cyclodipeptides, are much larger proteins, as they usually have an adenylation domain and a thiolation domain for the activation of each amino acid incorporated.

In addition to CDPSs, a few other enzymes also use aa-tRNAs as substrates to form peptide bonds. These enzymes are aa-tRNA protein transferases, which are involved in N-terminal protein modification, and aminoacyl transferases, which are involved in peptidoglycan biosynthesis[16,17,26]. However, these enzymes differ from CDPSs in that they form only one peptide bond between the amino acid moiety of the aa-tRNA donor and the N-terminal amino group of a peptide or protein acceptor, whereas CDPSs only use aa-tRNAs to form the two peptide bonds of a cyclodipeptide. CDPSs provide a new example of the versatility of aa-tRNAs in donating activated amino acids to different processes in the cell[26].

Different specificity determinants in aa-tRNAs are recognized by peptide bond–forming enzymes. The aa-tRNA protein transferases recognize the aa-tRNA in an RNA sequence–independent manner, the double-stranded acceptor stem being fully dispensable[27]. For aminoacyl transferases, specificity determinants lay not only in the aminoacyl group, but also in the tRNA moieties. A double-stranded acceptor stem is required; the sequence of the stem is also important[28]. Specificity determinants are unknown for AlbC and the other CDPSs. The CDPSs tested originate from different bacteria, but all are able to efficiently use *E. coli* aa-tRNAs to synthesize large amounts of cyclodipeptides. This could be explained by the similarity between tRNAs from diverse bacteria. Comparing isoacceptor tRNA sequences from *E. coli* and from the bacteria in which the CDPSs are found did not provide much information on the specificity determinants. For instance, the unique $tRNA^{Tyr}$ molecule from *M. tuberculosis* and the three $tRNA^{Tyr}$ molecules from *E. coli* (sharing an identical acceptor arm sequence) share four of the seven base pairs of the stem. Thus, we cannot exclude the possibility that the double-stranded acceptor stem is required. Further studies will be required to identify these determinants.

Owing to their mode of synthesis, all cyclodipeptides produced by CDPSs most likely have an L-L configuration. This is in agreement with the observed incorporation of L-amino acids into cyclodipeptides. This is also in agreement with previous observations on cFL synthesized by AlbC (ref. 29) and on cLL biosynthesis in *B. subtilis*[25].

The biological role of the DKP final products in organisms that produce them is unknown, but albonoursin has antibacterial activity[15], and pulcherriminic acid is able to chelate iron[25]. As aa-tRNAs are ubiquitous, CDPSs are offered a supply of substrates in any host in which they are expressed, but biosynthetic pathways involving CDPSs might be considered parasitic to the translation pathway. The use of aa-tRNAs as substrates also implies that DKPs are produced mostly during the exponential growth phase, in which the substrates are abundant; this is indeed what we observed for albonoursin biosynthesis in *S. noursei*[21].

In conclusion, we identified and characterized a new family of enzymes using a combination of genome mining and chemical biology experiments. These enzymes, CDPSs, are unrelated to NRPSs and catalyze the formation of various cyclodipeptides using aa-tRNAs as substrates. They are associated with cyclodipeptide-tailoring enzymes in DKP biosynthetic pathways. The characterization of new and versatile cyclodipeptide-synthesizing and cyclodipeptide-tailoring enzymes opens the way to pathway engineering and combinatorial approaches to further increase the natural diversity of DKPs, a family of compounds with diverse biological properties.

METHODS

Expression of *S. noursei* AlbC in *E. coli* and purification. Recombinant *S. noursei* AlbC was expressed in *E. coli* M15 [pREP4] cells (Invitrogen) transformed with the plasmid pQE60-AlbC (**Supplementary Methods** and **Supplementary Table 1** online). The expressed His_6-tagged AlbC protein was submitted to a three-step chromatographic procedure (**Supplementary Methods**). The purity of AlbC was >90%, as estimated by SDS-PAGE analyses, and about 20 mg of AlbC was obtained per liter of culture.

***In vitro* reconstitution of AlbC activity.** The first type of enzyme assay was carried out with radiolabeled amino acids and purified AlbC. They were performed in a total volume of 100 µl, with 35 µM AlbC, 5 µM [^{3}H] L-phenylalanine (1.85×10^3 Bq µl^{-1}), 5 µM [^{3}H]L-leucine (1.85×10^3 Bq µl^{-1}), 5 mM ATP, 10 mM $MgCl_2$ and 100 mM Tris-HCl pH 8.0. Some assays contained 60 µl of soluble extracts of *E. coli* cells not expressing AlbC (**Supplementary Methods**). After 17 h at 30 °C, reactions were stopped with the addition of 2% (v/v) trichloroacetic acid (TCA) and were analyzed by radio-HPLC.

The second type of enzyme assay was carried out with nonlabeled amino acids and AlbC-containing extracts. The extracts were prepared and desalted, thus removing small molecules, including cyclodipeptides synthesized during the cell culture, as described in **Supplementary Methods**. Reaction assays (200 µl) contained 180 µl of desalted extracts, 1 mM L-phenylalanine, 1 mM L-leucine, 5 mM ATP and 10 mM $MgCl_2$. These extracts were subjected to various treatments (DNase I, RNase A, RNase A followed by RNasin), as indicated in the Results. DNase I treatment consisted of digestion for 15 min at 30 °C with 1 µg ml^{-1} of enzyme. RNase A treatment consisted of digestion for 15 min at 30 °C with 0.1 µg ml^{-1} of enzyme. In some cases, the RNase A–treated extracts were further incubated for 5 min at 30 °C with 250 units ml^{-1} RNasin (Promega). Extracts treated with RNase A and RNasin were then supplemented with 0.125 A_{260} units ml^{-1} of *E. coli* $tRNA^{Phe}$ or 5 A_{260} units ml^{-1} of *E. coli* total tRNAs (Sigma). All the extracts were used in reaction assays as described above and were incubated for 1 h at 30 °C. Reactions were stopped by the addition of 2% (v/v) TCA and were analyzed by online LC-MS/MS.

The third type of enzyme assay was carried out with purified [^{3}H]Phe-$tRNAs^{Phe}$ and purified AlbC. [^{3}H]Phe-$tRNAs^{Phe}$ were synthesized *in vitro* with *E. coli* $tRNA^{Phe}$ (Sigma) and *E. coli* Phe-tRNA synthetase, according to published methods[30] with slight modifications (**Supplementary Methods**). Aminoacylation of $tRNAs^{Phe}$, measured by scintillation counting, was estimated to be about 70% (**Supplementary Methods**). The enzymatic assays were carried out in 40 µl 100 mM Tris-HCl pH 8.0 with 2.3 µM [3H]Phe-$tRNA^{Phe}$ (330×10^6 Bq µl^{-1}), 5 nM purified AlbC and 150 mM NaCl. After 2 min at 30 °C, reactions were stopped with the addition of 2% (v/v) TCA and were analyzed by radio-HPLC.

Analysis of enzyme assays by radio-HPLC. Analysis was performed using a Perkin-Elmer 200 series HPLC system with an online radio-flow scintillation detector and a C18 column (4.6 × 250 mm, 5 µm, 300 Å, Vydac) developed with a linear gradient from 0 to 50% (v/v) acetonitrile in 0.2% (v/v) trifluoroacetic acid (TFA) for 55 min (flow rate, 1 ml min^{-1}).

Analysis of enzyme assays by LC-MS/MS. Analysis was performed using an Agilent 1100 HPLC coupled via a split system to an Esquire HCT ion trap mass spectrometer (Bruker Daltonik GmbH). Samples were loaded onto a C18 column (4.6 × 150 mm, 3.5 µm, 100 Å, Atlantis) developed with a linear gradient from 0 to 50% (v/v) acetonitrile in 0.1% (v/v) formic acid for 50 min (flow rate, 0.6 ml min^{-1}). Positive electrospray ionization and mass analysis were optimized for the detection of compounds in the range of natural cyclodipeptides. For MS/MS, an isolation width of 1.0 *m/z* was set for isolating the parent ion, and a fragmentation energy ramp was used for optimizing the MS/MS fragmentation process. All data were acquired and processed using software from the manufacturer (Bruker Daltonik GmbH).

Cyclodipeptides were detected and identified by both their *m/z* value (MS) and their daughter ion spectrums (MS/MS), as a result of their common fragmentation patterns: neutral losses due to cleavages of the DKP ring on either side of peptide bonds[31,32], associated with both the neutral losses corresponding to the departure of amino acid residues and the appearance of *m/z* peaks of the so-called immonium ions and/or their related ions[33–36]. The nature of each detected cyclodipeptide was unambiguously confirmed by comparison with authentic standards.

Cyclodipeptides were quantified on the basis of their peak area at 214 nm using calibration curves obtained with authentic standards. Cyclodipeptide

standards were purchased from Bachem (cFL, cFF, cMM, cYW) and NeoMPS (cLL, cYY), or were synthesized according to previously described methods[37] (cFM, cLM, cYL, cYF, cYM, cYA, cVL) and spectrally characterized (**Supplementary Methods**). The concentrations of the standard solutions were determined by amino acid composition analyses, except for that of cFF, which was determined by UV absorbance.

***In silico* analyses.** BLAST searches[38] were performed to identify AlbC-related proteins. Multiple sequence alignments were performed with Multalin[39]. The PATTINPROT software tool[19] was used to search for proteins presenting a specific sequence signature.

***In vivo* assays for AlbC and its related proteins.** Genes encoding proteins similar to AlbC were PCR-amplified and inserted into expression vectors (**Supplementary Methods** and **Supplementary Table 1**). Recombinant AlbC, pSHaeC06, Jk0923, Rv2275, Plu0297, YvmC-Bsub, YvmC-Blic and YvmC-Bthu were expressed in *E. coli* BL21AI cells (Invitrogen) harboring pREP4 (Qiagen) transformed by the corresponding expression vectors (**Supplementary Methods**). Each culture supernatant (**Supplementary Methods**) was analyzed by LC-MS/MS, as previously described.

Identification of aa-tRNAs as substrates for Rv2275 and YvmC-Bsub. Assays for Rv2275 and YvmC-Bsub were performed with desalted extracts, as described for AlbC. In these cases, each pretreated extract was incubated with total tRNA from *E. coli*, ATP and L-tyrosine (for extracts containing Rv2275) or L-leucine (for extracts containing YvmC-Bsub). The incubation mixtures were analyzed by LC-MS/MS.

Accession codes. GenBank: AAN07909, CAB00960, CAB15512, AAU25020, EAO57133, CAE12592, BAE05998 and CAI37087 (see **Table 1** for details).

Note: Supplementary information and chemical compound information is available on the Nature Chemical Biology website.

ACKNOWLEDGMENTS

We thank C. Olaizola and M. Bahut for technical assistance, S. Zinn-Justin and J. Baillon for helpful discussion, and M. Moutiez for the critical reading of the manuscript. We are indebted to M. Springer (Institut de Biologie Physico-Chimique) for the plasmid pBI and to F. Doucet-Populaire (Université Paris Descartes), T. Msadek (Pasteur Institute), A. Bolotin (Institut National de la Recherche Agronomique), T. Baba (Juntendo University) and E. Krin (Pasteur Institute) for the gifts of bacterial strains and bacterial genomic DNAs. This study was supported by Commissariat à l'Energie Atomique, Centre National de la Recherche Scientifique, Université Paris-Sud 11, Pôle de Recherche et d'Enseignement Supérieur UniverSud Paris and Kyowa Hakko Bio Co. Ltd. L.S. is a recipient of a doctoral fellowship from Commissariat à l'Energie Atomique.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturechemicalbiology/.

Published online at http://www.nature.com/naturechemicalbiology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Ström, K., Sjogren, J., Broberg, A. & Schnurer, J. *Lactobacillus plantarum* MiLAB 393 produces the antifungal cyclic dipeptides cyclo(L-Phe-L-Pro) and cyclo(L-Phe-trans-4-OH-L-Pro) and 3-phenyllactic acid. *Appl. Environ. Microbiol.* **68**, 4322–4327 (2002).
2. Kohn, H. & Widger, W. The molecular basis for the mode of action of bicyclomycin. *Curr. Drug Targets Infect. Disord.* **5**, 273–295 (2005).
3. Kanoh, K. *et al.* Antitumor activity of phenylahistin *in vitro* and *in vivo*. *Biosci. Biotechnol. Biochem.* **63**, 1130–1133 (1999).
4. Williams, D.E. *et al.* Ambewelamides A and B, antineoplastic epidithiapiperazinediones isolated from the lichen *Usnea* sp. *Tetrahedr. Lett.* **39**, 9579–9582 (1998).
5. Waring, P. & Beaver, J. Gliotoxin and related epipolythiodioxopiperazines. *Gen. Pharmacol.* **27**, 1311–1316 (1996).
6. Maiya, S., Grundmann, A., Li, S.M. & Turner, G. The fumitremorgin gene cluster of *Aspergillus fumigatus*: identification of a gene encoding brevianamide F synthetase. *ChemBioChem* **7**, 1062–1069 (2006).
7. Walzel, B., Riederer, B. & Keller, U. Mechanism of alkaloid cyclopeptide synthesis in the ergot fungus *Claviceps purpurea*. *Chem. Biol.* **4**, 223–230 (1997).
8. Healy, F.G., Wach, M., Krasnoff, S.B., Gibson, D.M. & Loria, R. The *txtAB* genes of the plant pathogen *Streptomyces acidiscabies* encode a peptide synthetase required for phytotoxin thaxtomin A production and pathogenicity. *Mol. Microbiol.* **38**, 794–804 (2000).
9. Schultz, A.W. *et al.* Biosynthesis and structures of cyclomarins and cyclomarazines, prenylated cyclic peptides of marine actinobacterial origin. *J. Am. Chem. Soc.* **130**, 4507–4516 (2008).
10. Balibar, C.J. & Walsh, C.T. GliP, a multimodular nonribosomal peptide synthetase in *Aspergillus fumigatus*, makes the diketopiperazine scaffold of gliotoxin. *Biochemistry* **45**, 15029–15038 (2006).
11. Gardiner, D.M., Cozijnsen, A.J., Wilson, L.M., Pedras, M.S. & Howlett, B.J. The sirodesmin biosynthetic gene cluster of the plant pathogenic fungus *Leptosphaeria maculans*. *Mol. Microbiol.* **53**, 1307–1318 (2004).
12. Schwarzer, D., Mootz, H.D. & Marahiel, M.A. Exploring the impact of different thioesterase domains for the design of hybrid peptide synthetases. *Chem. Biol.* **8**, 997–1010 (2001).
13. Gruenewald, S., Mootz, H.D., Stehmeier, P. & Stachelhaus, T. *In vivo* production of artificial nonribosomal peptide products in the heterologous host *Escherichia coli*. *Appl. Environ. Microbiol.* **70**, 3282–3291 (2004).
14. Lautru, S., Gondry, M., Genet, R. & Pernodet, J.L. The albonoursin gene cluster of *S. noursei*: biosynthesis of diketopiperazine metabolites independent of nonribosomal peptide synthetases. *Chem. Biol.* **9**, 1355–1364 (2002).
15. Fukushima, K., Yazawa, K. & Arai, T. Biological activities of albonoursin. *J. Antibiot. (Tokyo)* **26**, 175–176 (1973).
16. Watanabe, K. *et al.* Protein-based peptide-bond formation by aminoacyl-tRNA protein transferase. *Nature* **449**, 867–871 (2007).
17. Mainardi, J.L., Villet, R., Bugg, T.D., Mayer, C. & Arthur, M. Evolution of peptidoglycan biosynthesis under the selective pressure of antibiotics in Gram-positive bacteria. *FEMS Microbiol. Rev.* **32**, 386–408 (2008).
18. Rost, B. Twilight zone of protein sequence alignments. *Protein Eng.* **12**, 85–94 (1999).
19. Combet, C., Blanchet, C., Geourjon, C. & Deleage, G. NPS@: network protein sequence analysis. *Trends Biochem. Sci.* **25**, 147–150 (2000).
20. Loria, R. *et al.* Thaxtomin biosynthesis: the path to plant pathogenicity in the genus *Streptomyces*. *Antonie Van Leeuwenhoek* **94**, 3–10 (2008).
21. Gondry, M. *et al.* Cyclic dipeptide oxidase from *Streptomyces noursei*. Isolation, purification and partial characterization of a novel, amino acyl alpha,beta-dehydrogenase. *Eur. J. Biochem.* **268**, 1712–1721 (2001).
22. McLean, K.J. *et al.* Characterization of active site structure in CYP121: a cytochrome P450 essential for viability of *Mycobacterium tuberculosis* H37Rv. *J. Biol. Chem.* **283**, 33406–33416 (2008).
23. Belin, P. *et al.* Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* (in the press).
24. Tang, M.R., Sternberg, D., Behr, R.K., Sloma, A. & Berka, R.M. Use of transcriptional profiling & bioinformatics to solve production problems: eliminating red pigment production in a *Bacillus subtilis* strain producing hyaluronic acid. *Ind. Biotechnol.* **2**, 66–74 (2006).
25. Uffen, R.L. & Canale-Parola, E. Synthesis of pulcherriminic acid by *Bacillus subtilis*. *J. Bacteriol.* **111**, 86–93 (1972).
26. RajBhandary, U.L. & Soll, D. Aminoacyl-tRNAs, the bacterial cell envelope, and antibiotics. *Proc. Natl. Acad. Sci. USA* **105**, 5285–5286 (2008).
27. Abramochkin, G. & Shrader, T.E. Aminoacyl-tRNA recognition by the leucyl/phenylalanyl-tRNA-protein transferase. *J. Biol. Chem.* **271**, 22901–22907 (1996).
28. Villet, R. *et al.* Idiosyncratic features in tRNAs participating in bacterial cell wall synthesis. *Nucleic Acids Res.* **35**, 6870–6883 (2007).
29. Kanzaki, H., Imura, D., Nitoda, T. & Kawazu, K. Enzymatic dehygrogenation of cyclo(L-Leu-L-Phe) to a bioactive derivative, albonoursin. *J. Mol. Catal. B Enzym.* **6**, 265–270 (1999).
30. Abramochkin, G. & Shrader, T.E. The leucyl/phenylalanyl-tRNA-protein transferase. Overexpression and characterization of substrate recognition, domain structure, and secondary structure. *J. Biol. Chem.* **270**, 20621–20628 (1995).
31. Chen, Y.-H., Liou, S.-E. & Chen, C.-C. Two-step mass spectrometric approach for the identification of diketopiperazines in chicken essence. *Eur. Food Res. Technol.* **218**, 589–597 (2004).
32. Stark, T. & Hofmann, T. Structures, sensory activity, and dose/response functions of 2,5-diketopiperazines in roasted cocoa nibs (*Theobroma cacao*). *J. Agric. Food Chem.* **53**, 7222–7231 (2005).
33. Falick, A.M., Hines, W.M., Medzihradszky, K.F., Baldwin, M.A. & Gibson, B.W. Low-mass ions produced from peptides by high-energy collision-induced dissociation in tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **4**, 882–893 (1993).
34. Papayannopoulos, I.A. The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrom. Rev.* **14**, 49–73 (1995).
35. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **11**, 601 (1984).
36. Johnson, R.S., Martin, S.A., Biemann, K., Stults, J.T. & Watson, J.T. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal. Chem.* **59**, 2621–2625 (1987).
37. Jeedigunta, S., Krenisky, J.M. & Kerr, R.G. Diketopiperazines as advanced intermediates in the biosynthesis of Ecteinascidins. *Tetrahedron* **56**, 3303–3307 (2000).
38. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
39. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881–10890 (1988).