

# Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides

Jan Claesen and Mervyn Bibb<sup>1</sup>

Department of Molecular Microbiology, John Innes Centre, Norwich NR4 7UH, United Kingdom

Edited by Bonnie L. Bassler, Howard Hughes Medical Institute and Princeton University, Princeton, NJ, and approved August 6, 2010 (received for review June 23, 2010)

Posttranslational modification of amino acids confers a range of structural features and activities on ribosomally synthesized peptides, many of which have potent antimicrobial or other biological activities. Cypemycin is an extensively modified linear peptide produced by *Streptomyces* sp. OH-4156 with potent in vitro activity against mouse leukemia cells. Cypemycin does not contain lanthionine bridges but exhibits some of the structural features of lantibiotics, notably dehydrated threonines (dehydrobutyrines) and a C-terminal *S*-[(*Z*)-2-aminovinyl]-*D*-cysteine. Consequently it was classified as a member of the lantibiotic family of posttranslationally modified peptides. Cypemycin also possesses two *L*-*allo*-isoleucine residues and an N-terminal *N,N*-dimethylalanine, both unique amino acid modifications. We identified and heterologously expressed the cypemycin biosynthetic gene cluster and performed a mutational analysis of each individual gene. We show that even the previously described modifications are carried out by unusual enzymes or via a modification pathway unrelated to lantibiotic biosynthesis. Bioinformatic analysis revealed the widespread occurrence of cypemycin-like gene clusters within the bacterial kingdom and in the Archaea. Cypemycin is the founding member of an unusual class of posttranslationally modified ribosomally synthesized peptides, the linaridins.

Actinomycetes | antibiotic | enzymology | natural products | *Streptomyces*

There is currently considerable interest in the identification of novel biologically active modified peptides (1) and in understanding the underlying enzymology that potentially will enable rational drug design (2, 3). Many different posttranslational modifications are introduced into ribosomally synthesized peptides (4), and they often are required for biological activity, contributing structural properties and peptide stability (5). Because of the ribosomal origin of the peptides, variant generation can be accomplished readily by mutagenesis of the corresponding structural gene (6) or by the introduction of tailoring enzymes.

Cypemycin is a modified peptide antibiotic that was identified because of its potent activity against mouse leukemia cells (7). It was classified as a member of the lantibiotic family of peptides because of common structural features, notably the presence of dehydrobutyrine (Dhb) and *S*-[(*Z*)-2-aminovinyl]-*D*-cysteine (AviCys) residues (8, 9). In this study, we identified the cypemycin biosynthetic gene cluster of *Streptomyces* sp. OH-4156 (7) and show that these previously described modifications are introduced by unusual enzymes or by an unusual modification pathway. Six distinct posttranslational modifications participate in cypemycin biosynthesis, altering 9 of the 22 amino acids found in the mature peptide. Four threonine (Thr) residues are dehydrated to form Dhb and one cysteine (Cys) is dethiolated to form dehydroalanine (Dha). The C-terminal Cys is oxidatively decarboxylated to form an enethiolate; addition of this reactive intermediate to the Dha results in the formation of the C-terminal AviCys. The side chains of two isoleucine (Ile) residues are isomerized to form *L*-*allo*-isoleucine (*L*-*allo*-Ile). Finally, the leader peptide is removed, and two subsequent *N*-methylations of the N-terminal alanine (Ala) generate *N,N*-dimethylalanine (Me<sub>2</sub>-Ala).

## Results

**Identification of *cypA*, the Gene Encoding the Cypemycin Precursor Peptide.** All amino acids in cypemycin are in the *L*-configuration (Fig. 1A) (8), presumed to reflect the posttranslational modification of a ribosomal precursor peptide. Initial PCR-based and Southern blotting approaches using degenerate nucleotide sequences derived from the sequence of fully modified cypemycin failed to identify the gene encoding the precursor peptide. Because cypemycin was believed to be a member of the lantibiotic family (9), degenerate primers were designed to amplify a conserved region of lanthionine M (LanM) lantibiotic dehydratases but failed to identify a possible cypemycin dehydratase.

We next carried out Solexa-based genome scanning of *Streptomyces* sp. OH-4156. We generated 15,471 contigs with a median length of 378 bp and an additive length of 8.5 Mb, typical of a *Streptomyces* genome (10). A tBLASTn search of the contig database with cypemycin's predicted propeptide sequence identified a 1,888-bp contig containing the cypemycin prepropeptide gene, *cypA*, confirming that cypemycin is indeed ribosomally synthesized (Fig. 1B).

The 22-aa CypA propeptide sequence revealed the identity of the residues subjected to posttranslational modification. As expected, the four Dhb residues of cypemycin result from dehydration of Thr, Me<sub>2</sub>-Ala is formed by methylation of an Ala residue, and the two *L*-*allo*-Iles are introduced by side-chain isomerization of *L*-Ile. Surprisingly, the AviCys modification is formed from two Cys residues. In contrast, previously described lantibiotic AviCys residues are generated from an internal serine (Ser) and a C-terminal Cys. LanD-mediated oxidative decarboxylation of the propeptide's C-terminal Cys yields an enethiol intermediate that can form AviCys by addition to a Dha created by dehydration of an internal Ser (11).

As in most other posttranslationally modified peptides, the propeptide is preceded by a leader sequence that is removed during or after maturation (12). The CypA propeptide is preceded by a 42-aa leader sequence that does not display similarity to lantibiotic leader peptides. Moreover, although lantibiotic leader sequences are always devoid of Cys (9), the CypA leader peptide contains a Cys at position -6 relative to the cleavage site.

Additional tBLASTn searches of the Solexa data were performed to identify putative cypemycin biosynthetic genes. A few contigs with possible *lanM* homologs were identified, but physical

Author contributions: J.C. and M.B. designed research; J.C. performed research; J.C. and M.B. analyzed data; and J.C. and M.B. wrote the paper.

The authors declare no conflict of interest.

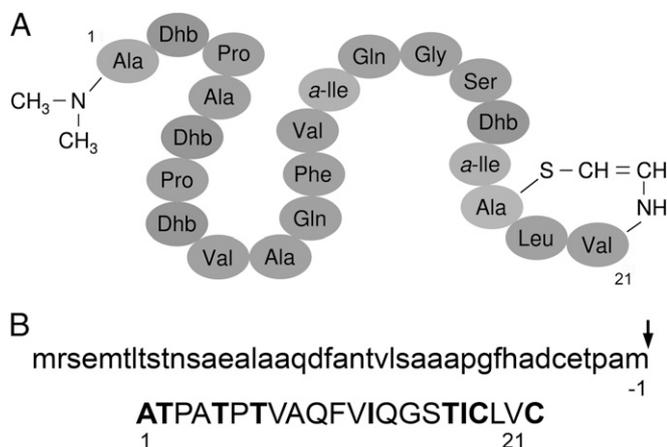
This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. HQ148718).

<sup>1</sup>To whom correspondence should be addressed. E-mail: mervyn.bibb@bbsrc.ac.uk.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1008608107/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1008608107/-DCSupplemental).



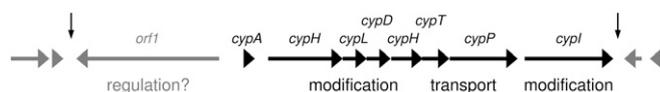
**Fig. 1.** Structure of cypemycin. (A) Schematic representation of fully modified cypemycin and (B) the unmodified CypA preproprotein.

linkage between these and the *cypA* contig could not be established by PCR analysis. Although tBLASTn searches with LanD decarboxylases did not identify a LanD homolog, a contig with the partial sequence of a 4'-phosphopantothenoylcysteine (PPC) decarboxylase was found. PPC decarboxylases belong to the homooligomeric flavin-containing Cys-decarboxylase (HFCD) family that also includes LanD proteins (13).

**Identification of the Cypemycin Biosynthetic Gene Cluster Reveals Unusual Modification Enzymes.** To clone the cypemycin biosynthetic gene cluster, a cosmid library was generated from genomic DNA of *Streptomyces* sp. OH-4156 in SuperCosI. A nylon membrane was spotted with 3,072 library clones and hybridized with a 480-bp <sup>32</sup>P-labeled *cypA* probe to identify 14 putative *cypA*-containing cosmids.

To determine which cosmids contained all of the genes required for cypemycin production, the vector backbones of nine of the hybridizing cosmids were PCR-targeted with a 5.2-kb SspI fragment from pIJ10702 that contained *oriT*, enabling the cosmid to be conjugated into a *Streptomyces* host, and the  $\phi$ C31 *int* gene and phage attachment site (*attP*) for integration of the cosmid into the chromosome of the heterologous host. Stable integration at the  $\phi$ C31 *attB* site obviates the need for continued antibiotic selection, facilitating subsequent bioactivity assays. *Streptomyces venezuelae* was chosen as the initial heterologous host because of the high level of nucleotide sequence identity between its genome and the Solexa data from *Streptomyces* sp. OH-4156. The nine exconjugants and a control strain with the integrated cosmid backbone were assessed for cypemycin production in a *Micrococcus luteus* bioassay and by MALDI-TOF MS. Six of the nine strains produced cypemycin with MALDI-TOF peaks of  $[M+H]^+ = 2,096$  Da,  $[M+Na]^+ = 2,118$  Da, and  $[M+K]^+ = 2,134$  Da, indicating that the integrated cosmids contained all of the genes required for cypemycin production (Fig. S1). Judged by the sizes of inhibition zones, cypemycin production by the *S. venezuelae* exconjugants (Fig. S2) was much lower than that of the natural producer. Because it gave the largest zone of inhibition upon heterologous expression, pIJ12404 was chosen for sequencing.

Analysis of ORFs in the pIJ12404 nucleotide sequence identified a putative biosynthetic cluster of nine genes (Fig. 2). Upstream of *cypA* and divergently transcribed from it is *orf1*, which encodes a putative transcriptional regulator. Seven genes with predicted biosynthetic and transport functions lie directly downstream of *cypA*, the first six of which are likely to be cotranscribed because their coding sequences overlap by several nucleotides (i.e., they appear to be translationally coupled). Immediately downstream of *cypA* is *cypH*, which encodes a product with no significant

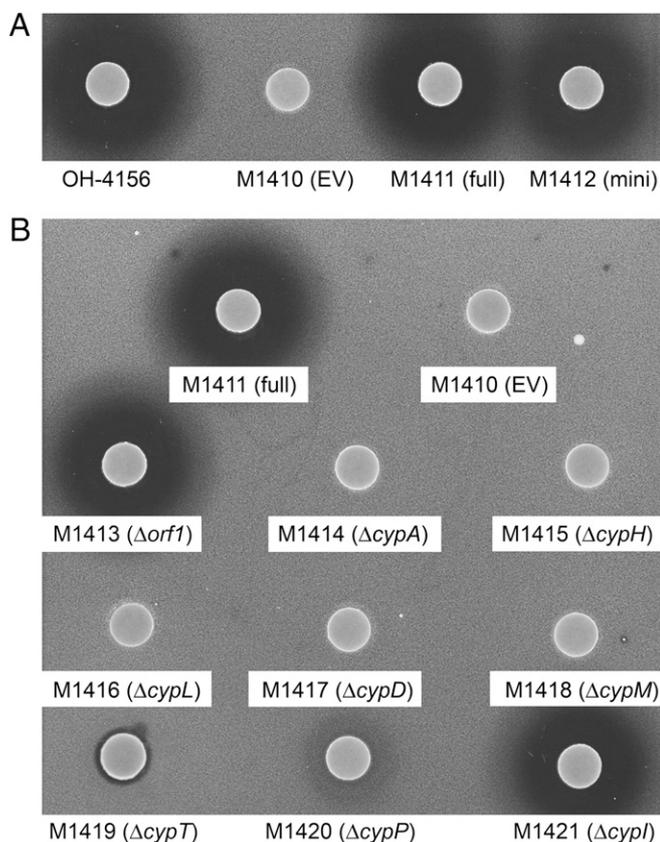


**Fig. 2.** Schematic representation of the cypemycin biosynthetic gene cluster. Flanking genes not expected to be involved in cypemycin biosynthesis are in gray. Vertical arrows delineate the putative *cyp* gene cluster used to generate a minimal gene set.

homology to functionally characterized proteins. Although the N terminus of CypH shows homology to a conserved horizontally transferred transmembrane helix domain (14), its C terminus possesses a possible  $\alpha/\beta$  hydrolase fold. Following *cypH* is *cypL*, again with no functionally identified homologs. The next gene, *cypD*, encodes a decarboxylase of the HFCD family (13) and could be responsible for the introduction of AviCys, analogous to LanD enzymes in lantibiotic biosynthesis. The fourth gene downstream of *cypA*, *cypM*, encodes an S-adenosyl methionine (SAM)-dependent methyltransferase presumably required for methylation of the N-terminal Ala of cypemycin. These genes are followed by *cypT*, which encodes an ATP-binding subunit of an ATP-binding cassette (ABC) transporter. The last gene in which the start codon overlaps with the upstream ORF is *cypP*, encoding a 516-aa protein with 12 predicted transmembrane helices and with no functionally annotated homologs in the public databases. Because of its location downstream of *cypT*, we suggest that this protein forms a pore in the cytoplasmic membrane to allow cypemycin export from cell. No leader peptidase domains were found in CypT or CypP. *cypI* is transcribed in the same direction as *cypAHLDMTP*, but its start codon does not overlap with the coding sequence of *cypP*. CypI is a member of the large DUF255 family of conserved proteins.

**Mutational Analysis of Cypemycin Biosynthesis.** Bioinformatic analysis suggested that the cypemycin biosynthetic gene cluster extends from *cypA* (or possibly *orf1*) to *cypI* (Fig. 2). The region upstream of and including *orf1* is syntenous with SCO4966 to SCO4969 in *S. coelicolor* (the latter being the *orf1* ortholog) (10). Genes to the left of *orf1* are predicted to be involved in mycothiol detoxification, and no function in cypemycin biosynthesis is envisaged. Genes downstream of *cypI* encode rodlinins and a chaplin (homologs of SCO2716 to SCO2719) (10) that have been implicated in morphological development in *S. coelicolor* (15) and therefore also are unlikely to be involved in cypemycin biosynthesis.

Because no convenient restriction sites were available to excise the putative minimal set of *cyp* genes and subsequently confirm their identity, PCR targeting was used to introduce unique restriction sites either side of the cluster. Briefly, pIJ12404 was PCR-targeted to the left of *orf1* to introduce a unique XbaI restriction site. The antibiotic resistance cassette was removed by Flippase recombination enzyme (FLP)-mediated recombination, and the resulting cosmid was targeted a second time downstream of *cypI*, introducing a unique SspI restriction site. The 12.2-kb fragment containing the putative minimal gene cluster was excised by digestion with XbaI and SspI and was ligated into XbaI/EcoRV-digested pSET152 to give pIJ12421. *S. venezuelae* was abandoned as a heterologous host because of low levels of cypemycin production. Instead, *S. coelicolor* M1146, from which four antibiotic gene clusters had been deleted and which lacks antibiotic activity, was used. Conjugation resulted in the stable integration of pIJ12421 into the  $\phi$ C31 *attB* site of M1146 to give M1412 that did not require subsequent antibiotic selection to maintain the construct. Cypemycin production was confirmed by both an inhibition assay against *M. luteus* and MALDI-TOF analysis. The halo produced in the bioassay was comparable in size to that produced by M1411 (M1146 harboring pIJ12413), indicating that the putatively assigned minimal gene set was indeed sufficient for cypemycin production in a heterologous host (Fig. 3A).



**Fig. 3.** *M. luteus* bioassays for (A) the minimal gene set and (B) the deletion mutants thereof in *S. coelicolor* M1146.

To investigate the function of each gene within the minimal gene set, individual in-frame “scarred” deletion mutants were generated by PCR targeting of pIJ12404. The backbones of the mutagenized cosmids were targeted subsequently with the 5.2-kb *SspI* fragment from pIJ10702 to allow integration into the *attB* site of M1146. Data obtained from heterologous expression in M1146 (Fig. 3 and Fig. S3) was corroborated by generating apramycin-marked deletions of all genes in the minimal gene set in *Streptomyces* sp. OH-4156 (Fig. S4). Unless otherwise stated, the mutant described is always the heterologously expressed in-frame deletion mutant. We also used quadrupole TOF (Q-TOF) MS to analyze the structure of cypemycin and some of its mutant derivatives. This analysis allowed determination of the amino acid sequences between Pro3 and the AviCys residue (which does not fragment into readily identifiable subfragments) (Fig. S5).

The  $\Delta cypA$  strain did not produce cypemycin (confirmed by MALDI-TOF), and no halo was produced in the bioassay with *M. luteus* (Fig. 3B). Deletion of the putative regulatory gene *orf1* had no effect on cypemycin biosynthesis (confirmed by bioassay and MALDI-TOF analysis), consistent with production of cypemycin by the  $\Delta orf1$  mutant of *Streptomyces* sp. OH-4156. Individual deletion of *cypH* and *cypL* abolished cypemycin production. The  $\Delta cypD$  knockout strain produced a peptide with a mass corresponding to nondecarboxylated cypemycin ( $[M+H]^+ = 2,142$  Da,  $[M+Na]^+ = 2,164$  Da, and  $[M+K]^+ = 2,180$  Da) that did not inhibit growth of *M. luteus*. We hypothesized that Cys19 in this compound would be modified to Dha and that Cys22 would still be intact because of the lack of decarboxylation. Alkylation with iodoacetamide thus was carried out to identify any thiol groups in the peptide; an increase in mass of 57 Da indicated the presence of only one such group (Fig. S6A). To confirm that this thiol group was derived from the C-terminal Cys, we verified

the presence of the Dha residue at the position of Cys19 by Q-TOF analysis (Fig. S6B). To demonstrate unambiguously that CypD was responsible for decarboxylation, a 6-His-tagged maltose-binding protein (MBP)-CypD fusion protein was incubated with the cypemycin prepropeptide (*SI Materials and Methods*), and decarboxylation was confirmed by MALDI-TOF MS (Fig. S7A and B).

M1146 ( $\Delta cypM$ ) also produced a smaller version of cypemycin detectable by MALDI-TOF ( $[M+H]^+ = 2,068$  Da,  $[M+Na]^+ = 2,090$  Da, and  $[M+K]^+ = 2,106$  Da) that failed to produce a zone of inhibition against *M. luteus*. The mass corresponds to non-methylated cypemycin, the presence of which was confirmed by Q-TOF analysis (Fig. S6C). To confirm that CypM was indeed responsible for dimethylation of the cypemycin propeptide, the peptide product of the  $\Delta cypM$  mutant was incubated in vitro with a 6-His-tagged MBP-CypM fusion protein (*SI Materials and Methods*). Dimethylation was confirmed by MALDI-TOF MS (Fig. S7C and D).

Deletion of *cypT* or *cypP* resulted in reduced halo sizes, but MALDI-TOF analysis confirmed cypemycin production. The *Streptomyces* sp. OH-4156  $\Delta cypT$  and  $\Delta cypP$  mutants displayed similar phenotypes when analyzed by bioassay and MALDI-TOF. It is possible that an alternative transporter could export cypemycin out of the cell or that, given the hydrophobic nature of the peptide, some level of diffusion through the membrane can occur. No putative peptidase domains were identified in either CypT or CypP, and their respective mutants produced lower levels of cypemycin, indicating that they are not essential for modification or cleavage of the peptide and presumably serve solely to export cypemycin.

Deletion of *cypI* had no effect on the production or activity of cypemycin. By a process of elimination, CypI is a possible candidate for isomerization of the Ile side chains to form L-*allo*-Ile, a modification that cannot be studied by MALDI-TOF because no change in mass occurs. Analysis of the *Streptomyces* sp. OH-4156  $\Delta cypI$  strain showed an identical phenotype. If *cypI* is required for introduction of the *allo*-Ile residues in cypemycin, they are not required for antibacterial activity.

The five M1146 derivatives that contained a mutagenized cosmid with a biosynthetic phenotype [M1414 ( $\Delta cypA$ ), M1415 ( $\Delta cypH$ ), M1416 ( $\Delta cypL$ ), M1417 ( $\Delta cypD$ ), and M1418 ( $\Delta cypM$ )] were complemented with wild-type copies of the respective genes to confirm that the mutant phenotypes were indeed caused by the targeted mutations. The coding sequence of each gene was cloned downstream of the constitutive *ermE\** promoter and EF-Tu ribosome binding site of pIJ10257. Upon conjugation, this vector integrates into the  $\phi$ BT1 phage attachment site of the *S. coelicolor* M1146 chromosome; again, antibiotic selection is not required to maintain the construct. Complementation of the  $\Delta cypA$  through  $\Delta cypM$  mutants was verified by MALDI-TOF analysis. Interestingly, in the complemented *cypL* mutant, nondecarboxylated cypemycin also was produced (Fig. S8).

**Cypemycin Is a Member of a Large Family of Posttranslationally Modified Peptides, the Linaridins.** A search of the National Center for Biotechnology Information sequence database identified 10 *cypL* homologs. Their occurrence always coincided with the presence, in close proximity, of genes with homology to *cypH* (sometimes only to the 5' half of *cypH*) (Fig. 4). Interestingly, in *S. viridochromogenes*, the *cypL* homolog is present in a cluster with two other genes, one of which is homologous to the 5' half of *cypH* and the other to the 3' half. All these clusters, with the exception of that in *Solibacter usitatus*, contain a short ORF that could encode a prepropeptide (Table 1). Both *cypL* and *cypH* were essential for cypemycin biosynthesis, and one or other or both were required for introduction of dehydrobutyrine (dehydrated threonine) residues into the mature peptide. Given the rarity of the *cypH* and *cypL* homologs and their invariant co-occurrence, we propose that the

## Bacteria

### Actinobacteria

*Streptomyces* sp. OH-4156

*Streptomyces griseus* NBRC 13350

*Streptomyces scabiei* 87.22

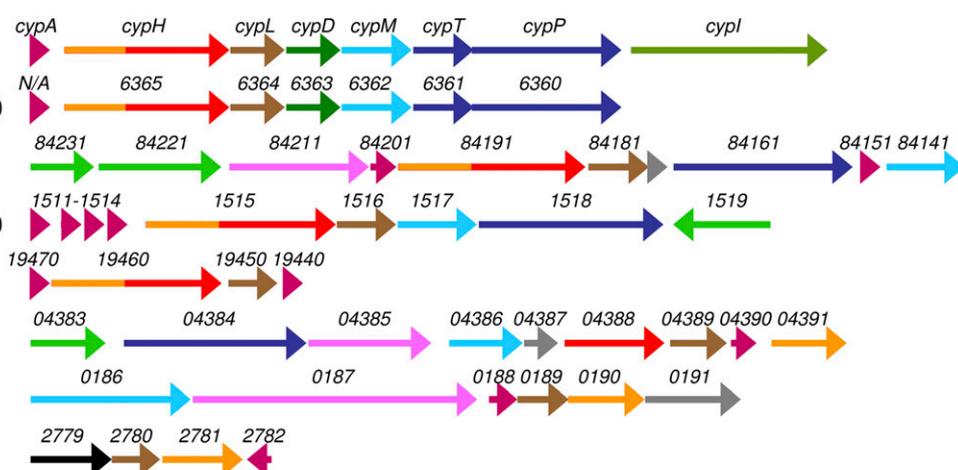
*Streptomyces griseus* NBRC 13350

*Leifsonia xyli* CTCB07

*Streptomyces viridochromogenes*

*Frankia* sp. EUN1f

*Mycobacterium vanbaalenii* PYR-1

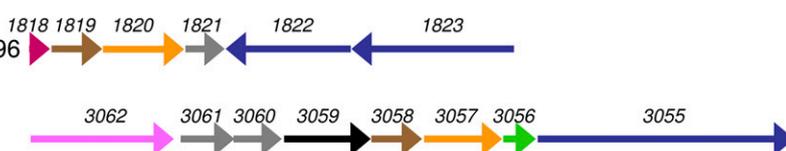


### Proteobacteria

*Nitrosospira multiformis* ATCC 25196

### Acidobacteria

*Solibacter usitatus* Ellin6076



## Archaea

### Euryarchaeota

*Haloterrigena turkmenica* DSM 5511



**Fig. 4.** The biosynthetic gene clusters of the linaridin family of peptides. Genes are color-coded by the predicted function of their products: structural genes (purple), *cypH* homologs (orange for the N terminus, red for the C terminus), *cypL* homologs (brown), decarboxylases (dark green), SAM-dependent methyltransferases (light blue), transporters (dark blue), regulation (light green), catalytic (pink), unknown (gray), and unknown but specific to the linaridin gene clusters (black).

gene clusters in Fig. 4 represent a previously undescribed family of posttranslationally modified peptides. Because these peptides are predicted to be linear (lin) and noncyclized, and to contain dehydrated (arid) amino acids, we propose the name “linaridins.” Although most of the linaridin gene clusters identified thus far occur in Actinomycetes (Gram-positive bacteria known for their ability to produce a wide range of secondary metabolites), they also occur in other bacterial phyla and even in Archaea (Fig. 4).

So far, the occurrence of a decarboxylase (the *cypD* homolog) is restricted to cypemycin and the orthologous cluster from *S.*

*griseus*. SAM-dependent methyltransferases are more prevalent in linaridin gene clusters, indicating that a subset of compounds could be modified by methylation of their N-terminal amino acid. Examples of genes other than those in the cypemycin cluster that could play a potential role in posttranslational modification are oxidoreductase-dehydrogenases (SCAB84211 and SSQG\_04385), a metallophosphoesterase (FrEUN1DRAFT\_0187), and a multi-copper oxidase (Acid\_3062). Interestingly, homologous genes encoding a protein of unknown function are present in or immediately adjacent to linaridin gene clusters in *Mycobacterium*

**Table 1.** Putative structural peptides

CypA	MRSEMTLSTNSAEALAAQDFANTVLSAAAPGFHAD <sub>C</sub> ETPAMATPATPTVAQFVIQGSTICLV <sub>C</sub>
SGR_N/A	MRLDSIATQETATALPESMATQDFANSVLAGAVPGFHSDAETPAMATPAVAQFVIQGSTICLV <sub>C</sub>
SCAB_84201	MSSIENALNSVEIPVEGVVYVAARPTLGTPRIARIGRIAQAEEGIGIAAAAATAGVGVAAEANNLAAEANAQNAALAAVGGAPAS
SCAB_84151	MNDFLLIPSVVALGIVGFLIATRAVSTPAVIGVALVVLWGAC <sub>S</sub> QARSTVHPRHTSSKKRRQHP
SGR_1511	MPEFRQPGWTRGVAPLDESAGGQVFGGASPVAAATPAVVATAGAVVFAAGVAARHLANGGNVLP
SGR_1512	MASPHRPPTSEEKSLNTASIPAVLSETGALSES <sub>D</sub> HGRALLDTPVVASVFTMTA <sub>C</sub> VEVS <sub>V</sub> CLTGSV <sub>I</sub> QLPQ
SGR_1513	MNKSSAPAVLTATGALTESDAGVTLPTLAPVAATPVAIAATMGVAVFVAGYAAGRAATGNVLP
SGR_1514	MSMSPTPAALRGAGLSESDPGRALSSLAPVTATPGVVAGVALGVALVNAFAAGYNHCCGGNVLP
Lxx19470	MSLQRTAQALSDLVATELNTDVTAGSSVPLHEVTC <sub>L</sub> GVIAVVAATAMDIAAYDVASGAAALAVTTL <sub>S</sub> V
Lxx19440	MSLQKTAQTLNDRVTEIDPDATASTSVLHRENT <sub>C</sub> IGVV <sub>T</sub> IIVGAAADVAADFVATGASALAATALAI
SvirD4_22614	MPRESAPSRRTAGELALRITDRKRTTMSVADFANTE <sub>L</sub> ADVT <sub>P</sub> GRIGNDATPTMLT <sub>P</sub> LAALATPEGVAVTAAATAYALNEVTNDLAG
FrEUN1fDRAFT_0188	MSMRSEPGSLRLSQLARIDALITEAQS <sub>R</sub> GFGLSD <sub>R</sub> FRIHITEEQA <sub>A</sub> ATPD <sub>A</sub> HHP <sub>L</sub> FDLSEHDREILNQI <sub>E</sub> L <sub>T</sub> GQLEHTTSIGELVEMRAQV <sub>V</sub> Q
Mvan_2782	MRVIVIVT <sub>T</sub> AERFRAAAAHPAATVSIAGV <sub>R</sub> WPTYK <sub>V</sub> SVLLVGLGVFVAVATTA <sub>A</sub> APAVLSGAGVATLVWLGLGLYRTSRR
Nmul_A1818	MMTKLAEALAGLD <sub>A</sub> VEARRM <sub>T</sub> SD <sub>T</sub> SD <sub>K</sub> YEQ <sub>V</sub> IPIGNAGTAVAGDNAGLSGEGELSLGVL <sub>M</sub> KLRENALS
Htur_3018	MSSQ <sub>T</sub> TFGWSL <sub>F</sub> TSGIV <sub>T</sub> L <sub>V</sub> LKALPG <sub>D</sub> SLW <sub>W</sub> GLMLLAVGLVLLYYR

Ser and Thr are represented in bold face. Cys are underlined. SGR\_N/A = gene not annotated.

*vanbaalenii* (Mvan\_2779), *Solibacter usitatus* (Acid\_3059), and *Haloterrigena turkmenica* (two copies: Htur\_3023 and Htur\_3024; shown in black in Fig. 4). The presence of this homolog only in the linaridin gene clusters suggests a role that is dedicated to the production of modified peptides.

## Discussion

We have identified, by genome scanning, the gene cluster for cypemycin biosynthesis. We showed, contrary to previous classification, that cypemycin is not a member of the lantibiotic family (9) but instead is a member of an unusual family of posttranslationally modified peptides. A putative minimal gene cluster was constructed, and nine genes contained in an 8.3-kb region were shown to be required for cypemycin production in a heterologous host. Individual in-frame deletions of these genes allowed identification of their functions in posttranslational modification and transport.

Cypemycin contains two modifications that are unique to ribosomally synthesized peptides, an N-terminal Me<sub>2</sub>-Ala and two L-*allo*-Iles (8). Recently, D-*allo*-Ile residues also were identified in the aerucyclamides, a family of cyclic cyanobacterial peptides (16). We were not able to identify unambiguously a gene responsible for the isomerization of the side chains of Ile13 and Ile18 to form the L-*allo*-Ile residues present in cypemycin. CypI is a candidate isomerase and belongs to the DUF255 family of proteins that contain a thioredoxin domain and show similarity to the N-acetyl-D-glucosamine epimerase superfamily. As yet, we have been unable to obtain enough pure peptide from the  $\Delta$ cypI mutants to verify whether the L-*allo*-Ile modifications are absent.

N,N-dimethylation of the N-terminal Ala of cypemycin is another unusual peptide modification (4). We unambiguously identified CypM, a SAM-dependent methyltransferase, as the enzyme that introduces these two methyl groups. Nonmethylated cypemycin was not active in a *M. luteus* bioassay.

In addition to these two modifications, cypemycin contains structural motifs, Dhb and AviCys, found in many lantibiotics. However, much to our surprise, these motifs are introduced through enzyme activities and/or via a pathway distinct from their lantibiotic counterparts. No conventional dehydratase was identified in the cluster, but both *cypH* and *cypL* were required for cypemycin biosynthesis. Our mutational analysis dictates, by a process of elimination, that CypH and/or CypL must be responsible for dehydration of the Thr residues of cypemycin to Dhb. It is possible that this modification is an early step in biosynthesis and is required for the subsequent modifications, hence the absence of other modified forms of the propeptide. Alternatively, the lack of the Dhb residues may lead to rapid protease-mediated degradation. The mature cypemycin molecule contains an unmodified Ser; interestingly, lantibiotic Ser residues also are more likely than Thr to escape dehydration (17). We have been unable to demonstrate CypH or CypL activity in vitro, and, perhaps significantly, similar difficulties have been encountered with LanB dehydratases.

The occurrence of Cys at position 19 in the propeptide, rather than the expected Ser (9, 11), initially led us to think that formation of AviCys in cypemycin occurred via an unprecedented

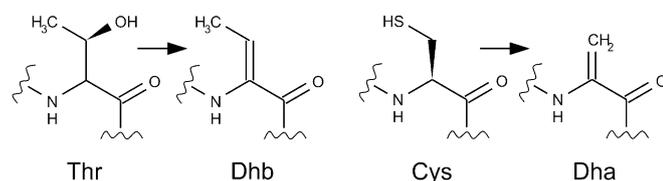


Fig. 5. CypH and/or CypL are involved in the dehydration of Thr residues to form Dhb and catalyze a similar dethiolation reaction that converts Cys to Dha.

mechanism. Analysis of the  $\Delta$ cypD mutant and an in vitro enzyme assay revealed that CypD is solely involved in decarboxylation of the C-terminal Cys22. Alkylation of the peptide produced by the *cypD* mutant with iodoacetamide and subsequent Q-TOF analysis confirmed that Cys19 had been converted to Dha. Because the chemistry behind this reaction is similar to the dehydration of Thr to Dhb, we propose that Cys dethiolation also is catalyzed by CypH and/or CypL (Fig. 5).

Although they share little sequence similarity, CypD decarboxylates the C-terminal Cys in a manner similar to LanD enzymes. The presence of an AviCys motif at the C terminus of members of both the lantibiotic and linaridin family of modified peptides might reflect a general role for AviCys in peptide stability, protecting the modified peptide from degradation by carboxypeptidases. AviCys, like the Me<sub>2</sub>-Ala residue, is essential for activity against *M. luteus*. A  $\Delta$ cypDM double mutant produced a truncated peptide lacking both modifications. As expected, this compound did not display bioactivity.

Deletion of *orfI* had no effect on cypemycin production, and the cluster contains no apparent regulatory genes. Interestingly, *cypH*, the first gene in the likely biosynthetic operon, contains a rare TTA codon, suggesting that cypemycin production might be developmentally regulated and controlled by the tRNA-encoding *blaD* (18). The cypemycin operon appears to be transcribed from a promoter upstream of *cypA*. A predicted transcriptional attenuator (a stem-loop structure with a calculated  $\Delta G$  of  $-31.2$  kcal) lies between *cypA* and the other biosynthetic genes and may ensure appropriately different levels of production of the modification enzymes and their peptide substrate.

Bioinformatic analysis identified 10 additional gene clusters that contain a *cypL* homolog together with a gene with full or partial homology to *cypH* and, with one exception, a gene encoding a putative prepropeptide. These clusters are expected to be capable of producing linear (noncyclized) dehydrated peptides (linaridins). The gene clusters are phylogenetically widespread, occurring in different phyla of bacteria and Archaea, suggesting that they play an important adaptive role in microbial physiology. Although cypemycin has antibiotic activity against *M. luteus*, it is inactive against many other bacteria and fungi, and the primary role of cypemycin, and potentially the linaridin family of peptides, may be to function as extracellular signaling molecules rather than as antimicrobial agents.

Interestingly, some of the putative linaridins from gene clusters with a full-length *cypH* homolog contain Cys residues, whereas peptides from clusters with genes homologous to only the 5' end of *cypH* are devoid of Cys (Fig. 4 and Table 1). It is tempting to speculate that the C-terminal domain of CypH is involved in the dethiolation of Cys. The occurrence of other potential biosynthetic genes (shown in pink in Fig. 4) suggests that other members of the linaridin family contain novel posttranslational modifications.

## Materials and Methods

**General Methods.** Plasmids and bacterial strains are summarized in Table S1. Media compositions, culture conditions, antibiotic concentrations, and general *Escherichia coli* and *Streptomyces* manipulations are described in ref. 19. Unless otherwise stated, restriction enzymes were purchased from Roche Diagnostics and chemicals from Sigma-Aldrich. Oligonucleotide primers and probes (Table S1) were purchased from Sigma-Aldrich. *Streptomyces* genomic DNA isolation, conjugative plasmid transfer, and standard PCR techniques and general cloning methods are described in ref. 19.

**Construction of a *Streptomyces* sp. OH-4156 Cosmid Library.** High molecular weight genomic DNA was isolated from a stationary-phase culture and partially cleaved with *Sau3A*I. The DNA was size-fractionated by pulsed-field gel electrophoresis, and fragments between 35 and 50 kb were excised and gel purified using the Qiaex II gel extraction kit (Qiagen). The DNA was ligated with *Bam*HI-cleaved SuperCosII cosmid vector and encapsulated in phage particles [using the Gigapack III Gold in vitro packaging system (Stratagene)] which then were used to infect *E. coli* XL1-Blue. Then 3,072

library clones were transferred to 384-well microtiter plates and arrayed on nylon membranes. The filters were baked, and bacterial debris was rinsed off before hybridization.

**Isolation of Cosmids Containing *cypA*.** A 480-bp PCR product corresponding to *cypA* and flanking regions was generated using primers *cypA* T1 and *cypA* T2 and genomic DNA of *Streptomyces* sp. OH-4156 as template. The fragment was labeled with  $\alpha$ -<sup>32</sup>P-dCTP following the protocol of the Prime-It Random Primer Labeling Kit (Stratagene) and was used to identify cosmids containing *cypA*.

**PCR Targeting.** PCR targeting was performed as described in ref. 20 and *SI Materials and Methods*.

**Generation of Complementation Constructs.** The ORFs of genes *cypA* to *cypM* were amplified by PCR using primer couples *cypA* 10257 F/*cypA* 10257 R to *cypM* 10257 F/*cypM* 10257 R (Table S1) and cosmid pIJ12404 as template. The PCR products were purified using the Qiagen PCR purification kit and were digested with NdeI and HindIII, whose recognition sequences were incorporated into the primers. The resulting fragments were cloned in NdeI and HindIII-digested pIJ10257 and verified by sequencing using primers 10257 seq F or 10257 seq R. Constructs were introduced into ET12567 pUZ8002 and then into the corresponding *Streptomyces* mutants by conjugation, whereupon they integrated stably at the  $\phi$ BT1 *attB* site.

**Cypemycin Bioassays.** Seed cultures were grown in SOC medium [tryptone 20 g/L, yeast extract 5 g/L, NaCl 0.5 g/L, KCl 0.19 g/L, MgCl<sub>2</sub> 0.95 g/L, glucose 3.6 g/L (21)] for 3 d, used to inoculate production medium (7), and grown for a further 3 d. Equal volumes of culture medium were taken from comparably grown cultures and extracted with CHCl<sub>3</sub> unless otherwise stated. The solvent was evaporated, and the resulting pellet was dissolved in methanol (at 1/10 of the original CHCl<sub>3</sub> volume). Aliquots (20  $\mu$ L) of these samples were spotted onto 6-mm Whatman paper discs. After all methanol had evaporated, the discs were placed on top of soft nutrient agar containing the indicator strain *M. luteus*. Typically, 1 vol of a *M. luteus* culture grown in L medium at 37 °C to an OD<sub>600</sub> of 0.4–0.5 was used to inoculate 8 vol of soft nutrient agar. Bioassay plates were incubated over night at 30 °C and zones of inhibition were recorded after 18–24 h.

**Detection of Cypemycin by MALDI-TOF and Q-TOF MS Analyses.** CHCl<sub>3</sub> was evaporated from 1 mL of production culture extract, and the resulting pellet was dissolved in 50  $\mu$ L 5% formic acid. Samples (~0.8  $\mu$ L) were spotted onto a Prespotted AnchorChip MALDI target plate (Bruker Daltonics) and washed briefly with 8  $\mu$ L 5% formic acid. After drying, the samples were analyzed by MALDI-ToF MS on a Bruker Ultraflex TOF/TOF. The instrument was calibrated using prespotted standards (~200 laser shots). Samples were analyzed using a laser power of ~25%, and spectra were summed from ~20  $\times$ 20 laser shots. For Q-TOF MS analysis, the peptide was infused directly into a QToF II (Waters) and analyzed with MassLynx 4.1 (Waters UK, Elstree, UK). The sample was diluted into 30% methanol/30% acetonitrile/1% acetic acid and applied with a GlassTip (New Objective) by nano-electrospray. Full MS scan analysis was performed with standard settings, and fragmentation was achieved by increasing the collision energy to 40. Alkylation of the free thiol group in the cypemycin intermediate produced by M1517 ( $\Delta$ *cypD*) was adapted from ref. 22. Briefly, culture medium from a 3-d-old production culture was incubated at 60 °C for 30 min under reducing conditions in the presence of 5 mM Tris(2-carboxyethyl) phosphine (TCEP). The mixture was allowed to cool to 25 °C. Iodoacetamide was added to a final concentration of 10 mM, and the mixture incubated in the dark at 25 °C for 30 min. CHCl<sub>3</sub> extraction was performed as described previously, and the sample was analyzed by MALDI-TOF and Q-TOF MS.

**Fusion Proteins and Enzyme Assays.** Production and purification of the 6-His-tagged MBP fusions to CypD, CypM, and the CypA preproprotein and the conditions used in the in vitro enzyme assays are described in *SI Materials and Methods*.

**ACKNOWLEDGMENTS.** We thank Gerhard Saalbach and Mike Naldrett for many mass spectroscopy analyses, David Studholme for assembly of the Solexa sequence data, Govind Chandra for help with bioinformatics, Stephen Bornemann for many useful discussions, and Satoshi Omura (Kitasato Institute, Tokyo, Japan), Juan Pablo Gomez-Escribano (John Innes Centre, Norwich, UK), and Novacta Biosystems Ltd. for providing bacterial strains. J.C. was supported by Marie Curie Actions Early Stage Training Programme Grant MEST-CT-2005-019727 to the John Innes Centre, and M.B. was supported by a grant to the John Innes Centre from the Biotechnology and Biological Sciences Research Council ([www.bbsrc.ac.uk](http://www.bbsrc.ac.uk)).

- Piper C, Cotter PD, Ross RP, Hill C (2009) Discovery of medically significant lantibiotics. *Curr Drug Discov Technol* 6:1–18.
- Pag U, Sahl HG (2002) Multiple activities in lantibiotics—models for the design of novel antibiotics? *Curr Pharm Des* 8:815–833.
- Cortés J, Appleyard AN, Dawson MJ (2009) Chapter 22. Whole-cell generation of lantibiotic variants. *Methods Enzymol* 458:559–574.
- McIntosh JA, Donia MS, Schmidt EW (2009) Ribosomal peptide natural products: Bridging the ribosomal and nonribosomal worlds. *Nat Prod Rep* 26:537–559.
- Rink R, et al. (2010) To protect peptide pharmaceuticals against peptidases. *J Pharmacol Toxicol Methods* 61:210–218.
- Boakes S, Cortés J, Appleyard AN, Rudd BA, Dawson MJ (2009) Organization of the genes encoding the biosynthesis of actagardine and engineering of a variant generation system. *Mol Microbiol* 72:1126–1136.
- Komiyama K, et al. (1993) A new antibiotic, cypemycin. Taxonomy, fermentation, isolation and biological characteristics. *J Antibiot (Tokyo)* 46:1666–1671.
- Minami Y, et al. (1994) Structure of cypemycin, a new peptide antibiotic. *Tetrahedron Lett* 35:8001–8004.
- Chatterjee C, Paul M, Xie L, van der Donk WA (2005) Biosynthesis and mode of action of lantibiotics. *Chem Rev* 105:633–684.
- Bentley SD, et al. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417:141–147.
- Blaesse M, Kupke T, Huber R, Steinbacher S (2000) Crystal structure of the peptidyl-cysteine decarboxylase EpiD complexed with a pentapeptide substrate. *EMBO J* 19: 6299–6310.
- Oman TJ, van der Donk WA (2010) Follow the leader: The use of leader peptides to guide natural product biosynthesis. *Nat Chem Biol* 6:9–18.
- Kupke T, et al. (2000) Molecular characterization of lantibiotic-synthesizing enzyme EpiD reveals a function for bacterial Dfp proteins in coenzyme A biosynthesis. *J Biol Chem* 275:31838–31846.
- Schultz J (2004) HTTM, a horizontally transferred transmembrane domain. *Trends Biochem Sci* 29:4–7.
- Claessen D, et al. (2004) The formation of the rodlet layer of streptomycetes is the result of the interplay between rodlin and chaplin. *Mol Microbiol* 53:433–443.
- Portmann C, Blom JF, Gademann K, Jüttner F (2008) Aerucyclamides A and B: isolation and synthesis of toxic ribosomal heterocyclic peptides from the cyanobacterium *Microcystis aeruginosa* PCC 7806. *J Nat Prod* 71:1193–1196.
- Rink R, et al. (2005) Lantibiotic structures as guidelines for the design of peptides that can be modified by lantibiotic enzymes. *Biochemistry* 44:8873–8882.
- Leskiw BK, Lawlor EJ, Fernandez-Abalos JM, Chater KF (1991) TTA codons in some genes prevent their expression in a class of developmental, antibiotic-negative, *Streptomyces* mutants. *Proc Natl Acad Sci USA* 88:2461–2465.
- Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA (2000) *Practical Streptomyces Genetics* (The John Innes Foundation, Norfolk, United Kingdom).
- Gust B, et al. (2004) Lambda red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. *Adv Appl Microbiol* 54:107–128.
- Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166:557–580.
- Sechi S, Chait BT (1998) Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Anal Chem* 70:5150–5158.