Title

Discovery of the Showdomycin Gene Cluster from *Streptomyces showdoensis* ATCC 15227 Yields Insight into the Biosynthetic Logic of C-Nucleoside Antibiotics

Authors

Kaisa Palmu,[†] Petja Rosenqvist,[‡] Keshav Thapa,[†] Yulia Ilina,[†] Vilja Siitonen,[†] Bikash Baral,[†] Janne Mäkinen,[†] Georgi Belogurov,[†] Pasi Virta,[‡] Jarmo Niemi[†] and Mikko Metsä-Ketelä[*,†]

Affiliations

[†]Departments of Biochemistry and [‡]Chemistry, University of Turku, FIN-20014 Turku, Finland

* Authors for correspondence

Contact information

Mikko Metsä-Ketelä, E-mail: mianme@utu.fi, Tel. +35823336847, Fax. +35823336860.

Abstract

Nucleoside antibiotics are a large class of pharmaceutically relevant chemical entities, which exhibit a broad spectrum of biological activities. Most nucleosides belong to the canonical N-nucleoside family, where the heterocyclic unit is connected to the carbohydrate through a carbon-nitrogen bond. However, atypical C-nucleosides have been isolated from *Streptomyces* bacteria over 50 years ago, but the molecular basis for formation of these metabolites has been unknown. Here we have sequenced the genome of *S. showdoensis* ATCC 15227 and identified the gene cluster responsible for showdomycin production. Key to the detection was the presence of *sdmA,* encoding an enzyme of the pseudouridine monophosphate glycosidase family, which could catalyze formation of the C-glycosidic bond. Sequence analysis revealed an unusual combination of biosynthetic genes, while inactivation and subsequent complementation of *sdmA* confirmed the involvement of the locus in showdomycin formation. The study provides the first steps toward generation of novel C-nucleosides by pathway engineering.

Nucleoside analogues, both synthetic and of natural product origin, have been widely used in medicine, in particular as antiviral or anticancer agents.[1,2] All nucleosides presently in clinical use belong to the classical N-nucleoside family of compounds,[3] where the heterocyclic unit (typically a nucleobase analog) is connected to the carbohydrate part (modified ribose or 2-deoxyribose) through a carbon-nitrogen bond. Notable examples include the antiviral compound vidarabine (**1**) and the anticancer agent gemcitabine (**2**, Figure 1). However, N-nucleosides are typically susceptible to loss of bioactivity by cleavage of the glycosidic linkage, which has raised considerable interest in the development of C-nucleosides that have a non-hydrolysable carbon-carbon bond between the nucleobase and the pentose sugar.[3,4] Two recently reported examples of synthetic C-nucleosides have been shown to be effective against filoviruses [5] and inhibit the hepatitis C virus polymerase.[6]

The isolation and chemical characterization of selected microbial C-nucleosides, namely showdomycin (**3**), oxazinomycin (**4**), pyrazomycin (**5**), and formycin (**6**, Figure 1) from cultures of *Streptomyces* soil bacteria have been described in the 1960s and 1970s.[7-10] Subsequent feeding experiments confirmed the close biosynthetic relationships of these metabolites and established that all of them are synthesized from D-ribose and L-glutamic acid precursors.[11,12] However, the biosynthetic pathways for C-nucleosides and the molecular mechanisms for their formation have been unknown to date.

Recent studies into the biosynthesis of an unrelated polyketide, alnumycin A, led to the identification and biochemical characterization of an atypical enzyme pair AlnA and AlnB, which catalyze the formation of the C-ribosylated pathway intermediate alnumycin C (**7**, Figure 1) in a two-step process.[13-15] In the initial step, AlnA is responsible for the attachment of D-ribose-5-phosphate to the polyketide aglycone via a C_8–$C_{1'}$ bond, while the reaction is

completed by dephosphorylation of the intermediate by AlnB, an enzyme of the haloacid dehalogenase superfamily.[15] The similarity of the C-glycosynthase AlnA to pseudouridine-5′-phosphate glycosidases such as YeiN,[16] which are involved in the degradation of pseudouridine (**8**, Figure 1), implied that homologous enzymes might be found from the biosynthetic pathways of C-nucleoside antibiotics. Here we utilized this strategy for the identification and initial characterization of the gene cluster responsible formation of **1** in *S. showdoensis* ATCC 15227.

To aid detection of **3** from culture broths of *S. showdoensis* ATCC 15227, we took advantage of the reactivity of maleimide ring to thiols and developed a colorimetric assay based on glutathione and 5,5'-dithiobis(2-nitrobenzoic acid) to follow the formation of the metabolite from culture extracts.[17,18] Best yields were detected surprisingly early, one day after inoculation, followed by degradation of **3** after prolonged cultivations (Figure 2a). The compound observed was subsequently confirmed as **3** by structure elucidation by NMR (Figure S1 and S2) and comparison to an authentic standard in LC-MS analysis (Figure 2b).

The genome sequence of *S. showdoensis* was acquired by MiSeq technology and assembly of the raw data led to a draft genome of 8,203,801 bp in 174 contigs (N50 value of 101,438 bp). A gene cluster (Figure 3a) encoding proteins homologous to AlnA and AlnB, denoted as SdmA and SdmB, respectively, were detected by genome mining. Phylogenetic analysis of SdmA revealed that the protein clustered in a clade together with AlnA and separated clearly from confirmed pseudouridine glycosidases [19] and unknown sequences found in gene clusters involved in the biosynthesis of the blue pigment indigoidine [20] (Figure 3b). The surrounding area consisted of two apparent operons covering approximately 12.2 kb, which could be responsible for the biosynthesis of **3**. In order to affirm this supposition, next we inactivated

the putative C-glycosynthase *sdmA* by CRISPR/Cas,[21] which led to cessation of the production of **3** in *S. showdoensis ΔsdmA* (Figure 2b). Complementation of the mutation by an intact copy of *sdmA* resulted in the reappearance of the original phenotype in *S. showdoensis ΔsdmA*/pIJE486-sdmA and **3** could again be detected from culture broths (Figure 2b). The complementation also indicated that the loss of production in *ΔsdmA* is not due to polar effects on other biosynthetic genes.

Sequence analysis of the gene cluster (Table 1) revealed an unfamiliar combination of genes, which remained undetected by software such as antiSMASH [22] that have been developed for detection of secondary metabolite biosynthesis gene clusters. Only two genes typically associated with secondary metabolism, *sdmC* and *sdmD,* encoding standalone amino acid adenylation (A) and peptidyl carrier protein (PCP) -domains, respectively, of non-ribosomal peptide synthetases (NRPS), [23] were detected. Genome mining revealed that clusters harboring genes homologous to *sdmABCD* are relatively rare in Actinobacteria, but two such instances were detected in *S. globisporus* C-1027 and *Nocardiopsis alba* ATCC BAA-2165 (Figure 3a).

Previous labeling experiments have established that the maleimide ring of **3** is made from carbons 2-5 and the nitrogen atom of L-glutamate.[12] However, prediction software for NRPS A-domain specificity [24] indicated that SdmC would use L-glycine as the preferred substrate. In order to test the substrate specificity, we produced and purified recombinant SdmC as a histidine tagged protein and evaluated enzyme activity [25] in the presence of the 20 proteinogenic amino acids and ATP. Surprisingly, formation of pyrophosphate could not be observed with any of the amino acids (Figure S3) despite extensive optimization of reaction

conditions, which led us to believe that SdmC does not utilize proteinogenic amino acids as substrates.

For these reasons we propose that the initial step in the biosynthesis of **3** is the cyclization of L-glutamine by either SdmE or SdmG, which harbor sequence similarity to L-ectoine synthases and transglutaminases, respectively. L-ectoine synthases typically utilize N-acetyl-2,4-diaminobutyrate as a substrate, but remarkably cyclization of L-glutamine has been reported.[26] Transglutaminases are typically responsible for intermolecular transamidation reactions, but an intramolecular cyclization of L-glutamate can be envisioned. After cyclization, the intermediate may be activated by aminoadenylation catalyzed by SdmC and transferred to the phosphopantheteine arm of SdmD (Figure 3c).

Although it is challenging to deduce the order of the biosynthetic steps, the next reaction may be deamination of the enamine by SdmM (Figure 3c), an enzyme of the guanine deaminase family. This reaction could be followed by 3,4-desaturation of the five membered ring by SdmF (Figure 3c), which is homologous to FAD-dependent aryl-CoA dehydrogenases. Based on sequence analysis, only SdmM and SdmF are likely to utilize PCP bound substrates and subsequent pathway intermediates may be released from the phosphopantetheine arm of SdmD. However, there are no genes homologous to thioesterase (TE) –domains (Table 1), which are typically responsible for the release or cyclization of protein bound peptides in NRPS systems. [23] The hydrolysis could be catalyzed by the one of the cyclase candidates, possibly SdmG, on the showdomycin pathway.

The C-glycosylation is most likely catalyzed by SdmA and is likely to occur after release of the substrate from SdmD, since neither AlnA or YeiN utilize substrates bound to prosthetic

groups.[15,27] The reaction could be followed by release of the phosphate group by the haloacid dehalogenase SdmB (Figure 3c) in a manner similar to AlnB.[15] In alnumycin A biosynthesis, the C-glycosylation reaction is not stereospecific, leading to racemization at C-1′ and formation of several products.[28] However, the showdomycin cluster harbours two homologous proteins, SdmH and SdmP, which belong to an enzyme family that contains glutamate racemases and arylmalonate decarboxylases (Table 1). We propose that one of these could be responsible for maintaining the correct stereochemistry at C-1′, while the other may catalyze decarboxylation at C-1 in accordance to the labeling experiments. Finally, SdmN, an enzyme of the thioredoxin fold family of proteins, may be involved in the formation of the maleimide ring in the last step of the biosynthesis.

In conclusion, our work presents for the first time insight into the molecular basis for the formation of C-nucleoside natural products. The gene cluster responsible for biosynthesis of **3** codes for a diverse selection of proteins that are rarely found in microbial secondary metabolism. Examples include similarity to enzymes involved in the modification of nucleosides, and in the biosynthesis of L-ectoine and NRPS. We show that the key carbon-carbon bond between the pentose sugar and the nucleobase is most likely catalyzed by SdmA, an enzyme evolutionarily related to proteins involved in the catabolism of pseudouridine. Investigations into the biosynthesis of other C-nucleosides currently in progress in our laboratory will reveal whether the same biosynthetic logic applies to the whole family of C-nucleoside secondary metabolites.

METHODS

Strains and Culture Conditions. *Escherichia coli* TOP10 (Invitrogen) and *S. lividans* TK24 were utilized as cloning hosts. *E. coli* ET12567 / pUZ8002 [29] was used as a host for conjugation into *Streptomyces*. *E. coli* strains were cultivated at 37 °C in 2xTY (Tryptone-Yeast extract) medium or Luria-Bertani agar plates with appropriate antibiotics for selection. MS, R2YE and ISP4 solid media[29], supplemented with apramycin (50 µg/ml) if necessary, were used for sporulation, protoplast transformation and storage, respectively, of *Streptomyces* strains. For production of **3**, *Streptomyces showdoensis* ATCC 15227 was grown in a medium containing 5 g/l potato starch, 5 g/l glycerol, 5 g/l D-glucose, 5 g/l peptone, 4 g/l potato juice, 3 g/l NaCl and milli-Q water. First, a preculture was prepared by inoculating 30 ml of culture medium and incubating on a reciprocal shaker at 30 °C for 72 h. For production of **3**, 600 µl of seed culture was added to 250 ml Erlenmeyer flasks each filled with 50 ml of culture medium, and the cultures were incubated on a reciprocal shaker at 30 °C.

Analysis of Showdomycin. Showdomycin (**3**) was assayed in fermentations and purification fractions by the reaction of the maleimide group with thiols,[17] in this assay glutathione (GSH). Remaining GSH was measured with Ellman's reagent, 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB) at 412 nm. For the assay, 200 mM glutathione stock was diluted in 1 ml 0.1 M K-phosphate buffer, pH 7.27, 1 mM EDTA, to a final concentration of 3.65 x 10-5 mol/l (corresponding to A412= 0.5 after detection). The sample was added in a volume of 10-100 µl and after 5 minute incubation at room temperature, the remaining GSH was determined by adding 10 µl DTNB stock (3 mg/ml) and measuring A412. The consumption of GSH by **3** was calculated based on a molar extinction coefficient of 14150.[18] For analysis by LC-MS, cultures (40 ml) were cleared from cells by centrifugation, pH was adjusted to 5 and **3** was adsorbed to active charcoal (1g), which was collected by centrifugation. Compounds bound to

the charcoal were extracted with 1:4 water:acetone, 500 µl samples were evaporated to dryness and dissolved in 100 µl water for analysis by LC-MS, which is described in the SI text.

Purification of **3** for structure elucidation by NMR is presented in the SI text. NMR measurements were recorded on a 500 Mhz Bruker AVANCE-III NMR-system with a BB/1H –smartprobe. ^{1}H NMR δ_H (500 MHz, D_2O): 6.65 (d, *J* = 1.72 Hz, C=CH), 4.74 (dd, *J* = 1.73 Hz, 4.65 Hz, 1 H, H1´), 4.21 (t, *J* = 4.84 Hz, 1 H H2´), 4.04 (dd, *J* = 6.15 Hz, 5.2 Hz, 1 H, H3´), 3.98 (td, *J* = 6.15 Hz, 5.2 Hz, 1 H, H3´), 3.81 – 3.76 (dd, *J* = 3.11 Hz, 12.57 Hz, 1 H, H5´), 3.67 – 3.62 (dd, *J* = 5.25 Hz, 12.57 Hz, 1 H, H5´´). ^{13}C NMR δ_C (125 MHz, D_2O): 173.1 (C=O), 172.4 (C=O) 147.1 (C-C=C), 129.2 (C=CH), 83.2 (C4´), 77.5 (C1´), 74.4 (C2´), 70.8 (C3´), 61.3 (C5´).

General DNA Techniques and Genome Sequencing. The high molecular weight chromosomal DNA was extracted as reported previously,[30] except that achromopeptidase was not utilized. Concentration of the chromosomal DNA was determined with fluorometric measurement. Sequencing of DNA was completed in Eurofins Genomics by shotgun genome sequencing on MiSeq (2X300bp) with Illumina. CLC Genomics Workbench 7.5 (Qiagen Bioinformatics) was utilized to perform the De novo assembly of the sequencing data. Further, Rapid Annotation using Subsystem Technology (RAST) was used for the annotation of assembled genomic data.[31] PCR conditions and list of oligonucleotides are described in the SI text and Table S1, respectively. Cloning of the CRISPR/Cas knock-out construct in pCRISPomyces2, [21] *sdmA* complementation construct in pIJE486 [32] and expression of recombinant SdmC from pBADΔHis [33] is presented in the SI text.

Enzyme activity assays. Expression and purification of SdmC is presented in the SI text. To evaluate the substrate specificity of SdmC, activity assays were performed as described previously [25, 34] with minor modifications as described in detail in the SI text.

Phylogenetic Analysis. The SEAVIEW v. 4 [35] graphical user interface was used both for the generation of the multiple sequence alignment with MUSCLE [36] and phylogenetic trees using the distance-based neighbour-joining [37] method with 100 rounds in bootstrap analysis. The phylogenetic tree shown in Figure 3 was drawn using the tree-viewing program FIGTREE. Accession numbers for the sequences used in the analysis are presented in the SI.

Accession numbers. The genomes sequence reported in this paper has been deposited into the data bases under the accession number LAQS00000000.

ASSOCIATED CONTENT

Supporting Information

This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

* (M.M.-K.) E-mail: mianme@utu.fi

Notes

The authors declare no competing financial interest.

ACKNOWLEDGEMENTS

The financial support of the Sigrid Jusélius Foundation (to M. Metsä-Ketelä and G. Belogurov) and the Academy of Finland (grant no.s 136060 and 251486) are acknowledged. We thank O. Virtanen and K. Panchal for assistance with some of the cloning steps and A. Salminen for the generous gift of purified PPase.

REFERENCES

(1) Clercq, E. D. (2013) Highlights in antiviral drug research: antivirals at the horizon. *Med. Res. Rev. 33*, 1215–1248.

(2) Périgaud, C. G. J. L., Gosselin, G., and Imbach, J. L. (1992) Nucleoside analogues as chemotherapeutic agents: a review. *Nucleosides & nucleotides 11*, 903–945.

(3) De Clercq, E. (2015) C-Nucleosides to be revisited: Miniperspective. *J. Med. Chem. 59*, 2301–2311.

(4) Stambasky, J., Hocek, M., and Kocovsky, P. (2009) C-nucleosides: synthetic strategies and biological applications. *Chem. Rev. 109*, 6729–6764.

(5) Warren, T. K., Wells, J., Panchal, R. G., Stuthman, K. S., Garza, N. L., Van Tongeren, S. A., Dong, L., Retterer, C. J., Eaton, B. P., Pegoraro, G., and Honnold, S. (2014) Protection against filovirus diseases by a novel broad spectrum nucleoside analogue BCX4430. *Nature 508*, 402–405.

(6) Cho, A., Zhang, L., Xu, J., Lee, R., Butler, T., Metobo, S., Aktoudianakis, V., Lew, W., Ye, H., Clarke, M., and Doerffler, E. (2013) Discovery of the first C-nucleoside HCV polymerase inhibitor (GS-6620) with demonstrated antiviral response in HCV infected patients. *J. Med. Chem. 57*, 1812–1825.

(7) Darnall, K. R., Townsend, L. B., and Robins, R. K. (1967) The structure of showdomycin, a novel carbon-linked nucleoside antibiotic related to uridine. *Proc. Natl. Acad. Sci. USA 57*, 548–553.

(8) Haneishi, T., Okazaki, T., Hata, T., Tamura, C., Nomura, M., Naito, A., Seki, I., and Arai, M. (1971) Oxazinomycin, a new carbon-linked nucleoside antibiotic. *J. Antibiot. 24*, 797–799.

(9) Gutowski, G. E., Chaney, M. O., Jones, N. D., Hamill, R. L., Davis, F. A., and Miller, R. D. (1973) Pyrazomycin B: Isolation and characterization of an α-C-nucleoside antibiotic related to pyrazomycin. *Biochem. Biophys. Res. Commun. 51*, 312–317.

(10) Koyama, G., Maeda, K., Umezawa, H., and Iitaka, Y. (1966) The structural studies of formycin and formycin B. *Tetrahedron Lett. 7*, 597–602.

(11) Buchanan, J. G., Hamblin, M. R., Sood, G. R., and Wightman, R. H. (1980) The biosynthesis of pyrazofurin and formycin *J. Chem. Soc., Chem. Commun.*, 917–918

(12) Buchanan, J. G., Hamblin, M. R., Kumar, A., and Wightman, R. H. (1984) The biosynthesis of showdomycin: studies with stable isotopes and the determination of principal precursors. *J. Chem. Soc. Chem. Commun.*, 1515–1517.

(13) Oja, T., Palmu, K., Lehmussola, H., Leppäranta, O., Hännikäinen, K., Niemi, J., Mäntsälä, P. and Metsä-Ketelä, M. (2008) Characterization of the alnumycin gene cluster reveals unusual gene products for pyran ring formation and dioxan biosynthesis. *Chem. Biol. 15*, 1046–1057.

(14) Oja, T., Klika, K. D., Appassamy, L., Sinkkonen, J., Mäntsälä, P., Niemi, J., and Metsä-Ketelä, M. (2012) Biosynthetic pathway toward carbohydrate-like moieties of alnumycins contains unusual steps for C-C bond formation and cleavage. *Proc. Natl. Acad. Sci. U. S. A. 109*, 6024–6029.

(15) Oja, T., Niiranen, L., Sandalova, T., Klika, K. D., Niemi, J., Mäntsälä, J., Schneider, G., and Metsä-Ketelä, M. (2013) Structural basis for C-ribosylation in the alnumycin A biosynthetic pathway. *Proc. Natl. Acad. Sci. U. S. A. 110*, 1291–1296.

(16) Preumont, A., Snoussi, K., Stroobant, V., Collet, J. F., and van Schaftingen, E. (2008) Molecular identification of pseudouridine-metabolizing enzymes. *J. Biol. Chem. 283*, 25238–25246.

(17) Böttcher, T. and Sieber, S. A. (2010) Showdomycin as a versatile chemical tool for the detection of pathogenesis-associated enzymes in bacteria *J. Am. Chem. Soc. 132*, 6964–6972.

(18) Riddles, P. W., Blakeley, R. L., and Zerner, B. (1983) Reassessment of Ellman's reagent. *Methods Enzymol. 91*, 49–60.

(19) Thapa, K., Oja, T., and Metsä-Ketelä, M. (2014) Molecular evolution of the bacterial pseudouridine-5'-phosphate glycosidase protein family. *FEBS J. 281*, 4439–4449.

(20) Takahashi, H., Kumagai, T., Kitani, K., Mori, M., Matoba, Y., and Sugiyama, M. (2007) Cloning and characterization of a *Streptomyces* single module type non-ribosomal peptide synthetase catalyzing a blue pigment synthesis. *J. Biol. Chem. 282*, 9073–9081.

(21) Cobb, R. E., Wang, Y., and Zhao, H. (2014) High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth. Biol. 4*, 723–728.

(22) Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., Lee, S. Y., Fischbach, M. A., Müller, R., Wohlleben, W., Breitling, R., Takano, E., and Medema, M. H. (2015) antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res. 43*, W237–W243.

(23) Marahiel, M. A., and Essen, L. O. (2009) Nonribosomal peptide synthetases mechanistic and structural aspects of essential domains. *Methods Enzymol. 458*, 337–351.

(24) Röttig, M., Medema, M. H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011) NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res. 39*, W362–W367.

(25) McQuade, T. J., Shallop, A. D., Sheoran, A., Delproposto, J. E., Tsodikov, O.V., and Garneau-Tsodikova, S. (2009) A nonradioactive high-throughput assay for screening and characterization of adenylation domains for nonribosomal peptide combinatorial biosynthesis. *Anal Biochem. 386*, 244–250.

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

(27) Huang, S., Mahanta, N., Begley, T. P., and Ealick, S. E. (2012) Pseudouridine monophosphate glycosidase: a new glycosidase mechanism. *Biochemistry 51*, 9245–9255.

(28) Oja, T., Tähtinen, P., Dreiack, N., Mäntsälä, P., Niemi, J., Metsä-Ketelä, M., and Klika, K. D. (2012) Alnumycins A2 and A3: new inverse-epimeric pairs stereoisomeric to alnumycin A1. *Tetrahedron: Asymmetry 23*, 670–682.

(29) Kieser, T., Bibb M. J., Buttner, M. J., Chater, K. F. and Hopwood D.A. (2000) In: Practical *Streptomyces* genetics. The John Innes Foundation, John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7 UH, England. pp 613.

(30) Nikodinovic, J., Barrow, K. D., and Chuck, J.-A. (2003) High yield preparation of genomic DNA from Streptomyces. *Biotechniques 35*, 932–936.

(31) Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., and Meyer, F. (2008) The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics 9*, 75.

(32) Ylihonko, K., Tuikkanen, J., Jussila, S., Cong, L., and Mäntsälä, P. (1996) A gene cluster involved in nogalamycin biosynthesis from *Streptomyces nogalater*: sequence analysis and complementation of early-block mutations in the anthracycline pathway. *Mol. Gen. Genet. 251*, 113– 120.

(33) Kallio, P., Sultana, A., Niemi, J., Mäntsälä, P., and Schneider, G. (2006) Crystal structure of the polyketide cyclase AknH with bound substrate and product analogue: implications for catalytic mechanism and product stereoselectivity. *J. Mol. Biol. 357*, 210–220.

(34) Salminen, T., Käpylä, J., Heikinheimo, P., Kankare, J., Goldman, A., Heinonen, J., Baykov, A. A., Cooperman, B. S., and Lahti, R. (1995) Structure and function analysis of *Escherichia coli* inorganic pyrophosphatase: is a hydroxide ion the key to catalysis? Biochemistry 34, 782–791.

(35) Gouy, M., Guindon, S., and Gascuel, O. (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol. 27*, 221–224.

(36) Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res. 32*, 1792–1797.

(37) Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol. 4*, 406–425.

Table 1. Deduced functions of the gene products found in the *sdm* gene cluster.

Gene	Size	Putative function	Sequence Similarity / Protein, Origin	Identities / Positives / Gaps	Accession no
sdmP	257	Decarboxylase	Arylmalonate decarboxylase, *Bordetella bronchiseptica*	27 / 42 / 9	3DTV_A
sdmO	462	Transporter	Puromycin resistance protein Pur8, *Streptomyces alboniger*	37 / 52 / 9	P42670
sdmB	246	Phosphatase	Alnumycin P Phosphatase AlnB, *Streptomyces* sp. CM020	43 / 56 / 0	4EX6_A
sdmA	317	C-glycosynthase	Prealnumycin C-glycosynthase AlnA, *Streptomyces* sp. CM020	48 / 64 / 1	4EX8_A
sdmN	239	Oxidoreductase	Novel Disulfide Oxidoreductase, *Deinococcus Radiodurans*	32 / 47 / 14	5CNW_A
sdmM	162	Deaminase	Guanine deaminase, *Bacillus subtilis*	34 / 50 / 2	1WKQ_A
sdmL	105	Transporter	Quaternary ammonium compound-resistance protein SugE, *Photorhabdus luminescens subsp. laumondii*	46 / 67 / 0	Q7MZY0
sdmK	137	Transporter	Quaternary ammonium compound-resistance protein SugE, *Photorhabdus luminescens subsp. laumondii*	42 / 64 / 0	Q7MZY0
sdmJ	151	Transcriptional Regulator	HTH -type transcriptional regulator YkoM, *Bacillus subtilis* (strain 168)	32 / 58 / 2	O34949
sdmC	485	Amino Acyl AMP Transferase	D-alanyl carrier protein ligase DltA, *Bacillus cereus*	25 / 41 / 5	3DHV_A
sdmD	89	Aminoacyl carrier protein	Kosinostatin Biosynthesis Protein Kstb-pcp, *Micromonospora* sp. TP-A0468	39 / 73 / 0	2MY5_A
sdmE	124	Cyclase	L-ectoine synthase EctC, *Sphingopyxis alaskensis* RB2256	36 / 49 / 5	5BXX_A
sdmF	383	Oxidase	Prolyl Acyl Carrier Protein Oxidase Anab, *Oscillatoria* sp. PCC 6506	36 / 57 / 0	4IRN_A
sdmG	239	Transglutaminase	Transglutaminase-like superfamily protein,	75 / 83 / 1	WP_044372119

			Streptomyces griseus		
sdmH	269	Decarboxylase	Maleate Isomerase, *Pseudomonas putida* S16	25 / 44 / 10	4FQ7_A
sdmI	81	hypothetical protein	hypothetical protein, *Streptomyces* sp. WM6378	66 / 74 / 0	WP_053724005

Figures and Legends to the Figures

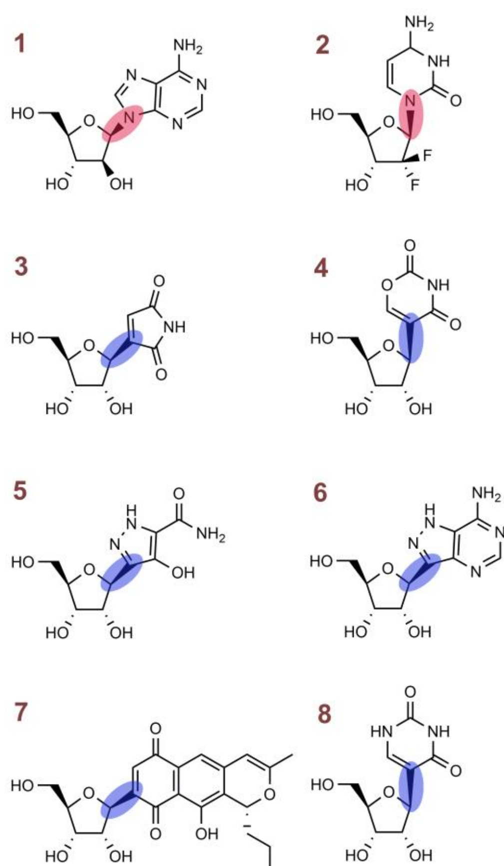

Figure 1. Chemical structures of the N-nucleosides vidarabine (**1**) and gemcitabine (**2**), and the C-nucleoside metabolites showdomycin (**3**), oxazinomycin (**4**), pyrazomycin (**5**), formycin (**6**), the aromatic polyketide alnumycin C (**7**) and pseudouridine (**8**). The key N-$C_{1'}$ and C-$C_{1'}$ bonds are shown in red and blue, respectively.

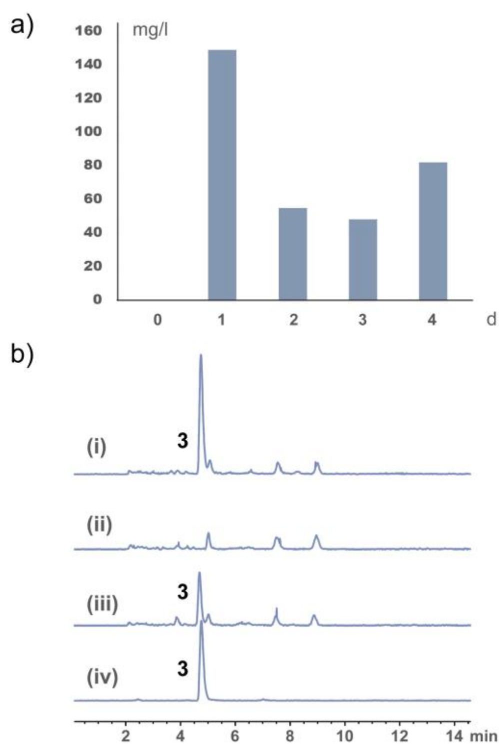

Figure 2. Analysis of showdomycin production by *S. showdoensis* ATCC 15227. a) Time course for production of **3** by *S. showdoensis*, b) Extract ion chromatogram traces (positive mode, **3**: *m/z* $[M + H]^+$ $C_9H_{12}NO_6$, found 230.07)) from culture samples of (i) wild type *S. showdoensis*, (ii) *S. showdoensis ΔsdmA*, (iii) *S. showdoensis ΔsdmA*/pIJE486-sdmA and (iv) authentic commercial reference of **3**.

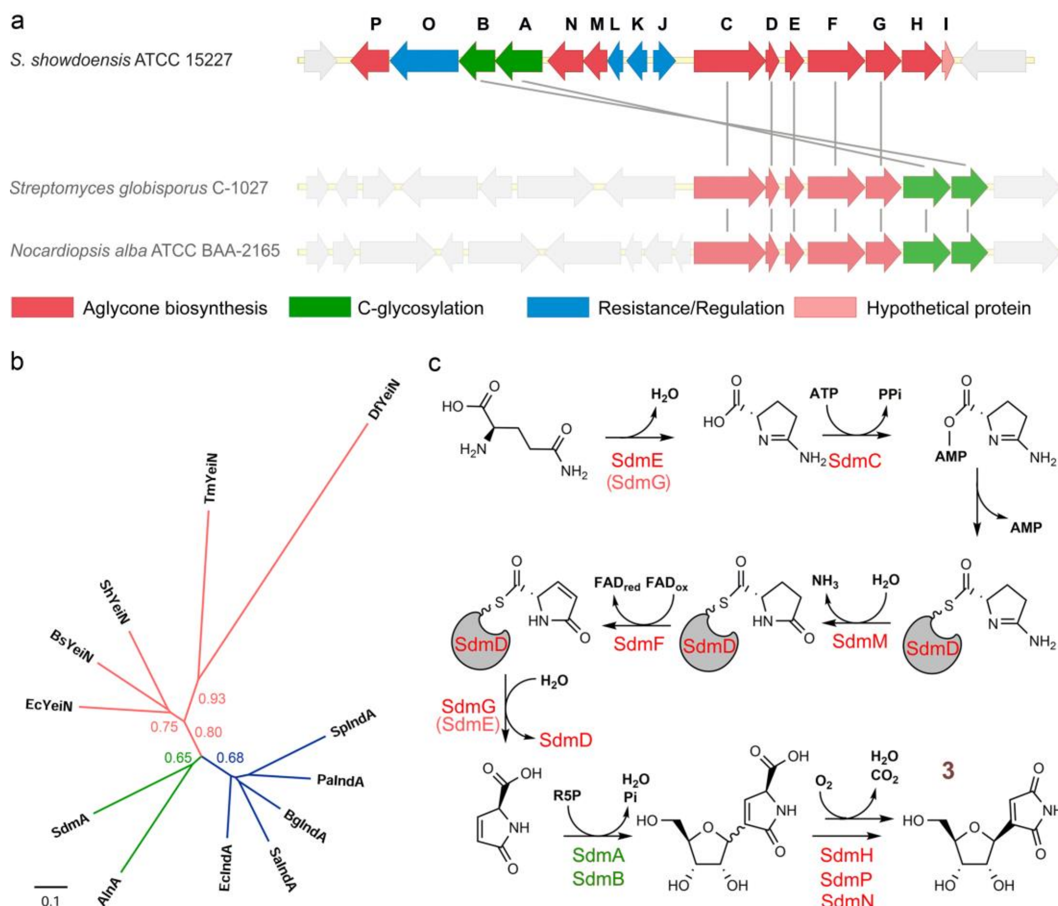

Figure 3. Discovery of the showdomycin gene cluster and model for its biosynthesis. a) Graphical representation of the showdomycin gene cluster from *S. showdoensis ATCC 15227* and comparison to two related clusters of unknown function from *S. globisporus* C-1027 and *Nocardiopsis alba* ATCC BAA-2165. Grey lines connect genes that show sequence similarity, b) phylogenetic analysis of the pseudouridine glycosidase protein family. Confirmed monophosphate glycosidases[19] are marked in red, sequences found in gene clusters for indigoidine pigments in blue and C-glycosynthases involved in antibiotic biosynthesis in green. Legend: DfYeiN, *Dorea formicigenerans*; TmYeiN, *Thermotoga maritima*; ShYeiN, *S. himastatinicus*; EcYeiN, *Escherichia coli*; EcIndA*, Erwinia chrysanthemi*; SaIndA, *Streptomyces albus*; BgIndA, *Burkholderia gladioli*; PaIndA, *Photorhabdus asymbiotica*; SpIndA, *Serratia proteamaculans*. Scale bar represents 10% dissimilarity, c) proposal for the

biosynthetic steps in the formation of **3**. Legend: PPi, pyrophosphate; R5P, D-ribose-5-phosphate; FAD, flavin adenine dinucleotide.