J|A|C|S COMMUNICATIONS

Published on Web 11/23/2009

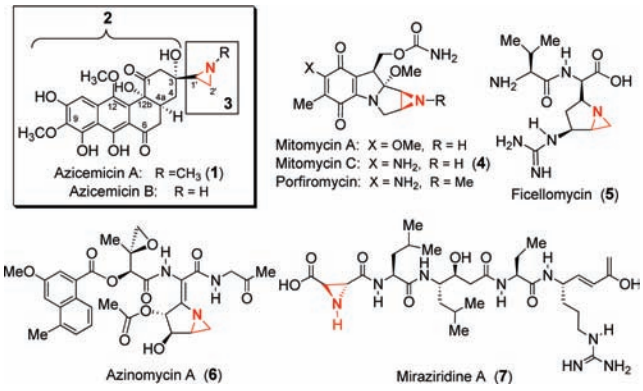
Biosynthetic Studies of Aziridine Formation in Azicemicins

Yasushi Ogasawara and Hung-wen Liu*

Division of Medicinal Chemistry, College of Pharmacy, and Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, Texas 78712

Received August 28, 2009; E-mail: h.w.liu@mail.utexas.edu

Azicemicins (**1**) are aromatic polyketides isolated from *Kibdelosporangium* sp. MJ126-NF4 (formerly known as *Amycolatopsis* sp. MJ126-NF4) that have antimicrobial activity against Gram-positive bacteria, yet show little signs of toxicity to mice.[1−3] Azicemicins consist of an angucycline core structure (**2**) and a unique aziridine moiety (**3**) at the C-3 position. Only a few aziridine-containing natural products are known. These include mitomycin (**4**), ficellomycin (**5**),[4] azinomycin (**6**),[5] and miraziridine (**7**).[6] Studies on the mode of action of mitomycins[7] and azinomycins[8] showed that the aziridine moiety plays an important role in the observed antitumor/antimicrobial activities of these compounds. The biosynthetic gene clusters for mitomycin[9] (**4**) and azinomycin[10] (**6**) have been cloned. Comparison of these two gene clusters revealed that mitomycin and azinomycin employ disparate strategies to generate the aziridine unit. In fact, isotope-tracer experiments indicated that glucosamine is the precursor of the aziridine moiety in the biosynthesis of mitomycin (**4**),[11] whereas either ornithine[12] or a glutamate[13] derivative is the aziridine precursor for azinomycin (**6**). However, the details of the aziridine formation remain obscure in both cases. Clearly, studies of the biosynthesis of aziridine will provide new insight into novel enzyme chemistry as this relatively uncharted research area is explored. In this paper, we report isotope-tracer experiments and cloning of the azicemicin biosynthetic gene cluster to begin the elucidation of the pathway and mechanism of aziridine formation in azicemicin.



First, feeding experiments using [1-^{13}C]- and [1,2-$^{13}C_2$]acetate were carried out. As expected, the labeling pattern revealed that the angucycline skeleton of **1** is constructed in a standard manner involving the assembly of 10 acetate units (Table S1).[14] An intact acetate unit was found to be incorporated into C-1′ to C-3, and the C-2′ is enriched as a singlet in experiment with [1,2-$^{13}C_2$]acetate (Scheme 1). Interestingly, ^{13}C enrichment at C-2′ is not observed when [1-^{13}C]acetate was fed to the producing strain. Thus, the C-2′ of azicemicin A must be derived from C-2 of acetate, and the C−C bond in the corresponding acetate unit is cleaved via decarboxylation. These results suggested that the starter unit of azicemicin is likely an amino acid derived from acetate and may be an aziridine carboxylate (**16**) equivalent.

Since biosynthesis of angucycline-type polyketides is typically catalyzed by type II polyketide synthases (PKSs), the azicemicin biosynthetic gene cluster should contain a set of type II PKS genes. To identify the azicemicin gene cluster, we employed a PCR-based screening method to locate KSα (ketosynthase α subunit) genes, one of the type II PKS genes, in the genome of the producing strain. This approach is based on the fact that KSα is highly conserved in actinomycetes and can be used as a probe for cloning of the biosynthetic gene clusters containing type II PKS genes.[15] To achieve this goal, a cosmid library harboring 2000 unique clones was constructed from the genomic DNA of *Kibdelosporangium*, packaged into phage, and used to transfect *E. coli* XL-1 Blue MRF′ cells. This cosmid library represents ∼10-fold coverage of the genome. Screening of the cosmid library by PCR amplification of the imbedded KSα genes using degenerate primers gave three independent cosmids encoding type II KSα genes.

Two of these cosmids (Y36J1 and Y18C1) contain angucycline-type PKS genes, and the third contains type II PKS for a longer chain polyketide product, as determined by phylogenetic analysis of the KSα sequences.[16] Subsequent spot sequencing showed that cosmid Y18C1 actually encodes production of a glycosylated angucycline-type compound, since it also harbors genes involved in aminosugar biosynthesis. Thus, cosmid Y36J1 was selected for further studies. To ensure the complete coverage of the azicemicin cluster, a second round of cosmid library screening was carried out using probes derived from the sequenced portions of Y36J1, and cosmid Y41B1 was obtained. Both strands of Y36J1 and Y41B1 were sequenced, covering ∼50 kb of contiguous DNA. The nucleotide sequence has been deposited in GenBank under the accession number GU134622. As shown in Figure 1, a total of 45 ORFs were identified and designated as the *azic* cluster. Among those, a few genes are likely involved in aziridine formation, based on sequence comparison with other known genes in the databank.

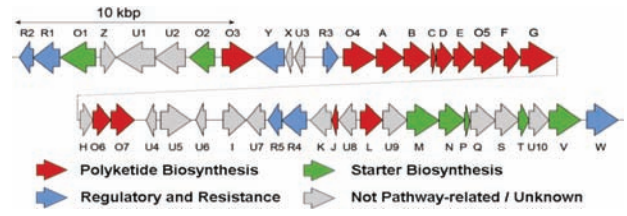


Figure 1. Azicemicin biosynthetic gene cluster of *Kibdelosporangium* sp. MJ126-NF4 showing the location and arrangement of all ORFs.

A BLAST database search indicated that AzicM and AzicV may be involved in the activation of amino acids, because both ORFs show sequence homology to the adenylation domains of nonribosomal peptide synthetases (NRPSs). Given that an amino acid has been implicated as the precursor of the aziridine moiety by the

10.1021/ja907307h

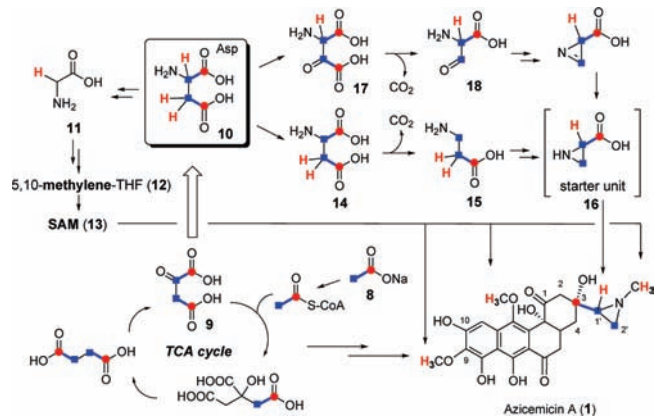
isotope-tracer experiments, one of these two adenylation enzymes is expected to catalyze the adenylation of this amino acid precursor in the biosynthesis of azicemicins. AzicP exhibits modest sequence identity to several carrier proteins from the PKS/NRPS hybrid systems, including LnmP (41% identity)[17] from the leinamycin pathway and VinL (38% identity)[18] from the vicenistatin pathway. AzicP may accept the adenylated amino acid from AzicM or AzicV for aziridine formation. AzicT shows moderate sequence homology to several phosphopantetheinyl transferases, which catalyze activation of carrier proteins by attaching the phosphopantetheine group to a conserved serine residue of the carrier protein. AzicT may play a similar role in the activation of the carrier protein, AzicP, in azicemicin biosynthesis.

AzicN shares significant sequence identity to VinO (38% identity)[18] from the vicenistatin pathway and BtrK (30% identity)[19] of the butirosin pathway and is likely a decarboxylase. BtrK is a PLP-dependent decarboxylase,[20] which removes the α-carboxyl group of a glutamic acid residue tethered via its γ-carboxylate group to BtrI, the acyl carrier protein in the butirosin biosynthetic pathway. AzicN may play a similar role in the construction of the aziridine carboxylate starter unit (**16**), and its substrate is expected to be an AzicP-linked amino acid derivative.

The *azic* cluster contains three oxygenase genes (*azicO1*, *azicO2*, and *azicO6*) in addition to the four oxygenase genes (*azicO3*, *azicO4*, *azicO5*, and *azicO7*) involved in the oxidation/hydroxylation of the angucycline core. AzicO1 is likely a flavin-dependent oxygenase with high sequence identity to StfE (57% identity)[21] from the steffimycin pathway and SpnJ (54% identity)[22] of the spinosyn pathway. AzicO2 shows sequence homology to various cytochrome P450 enzymes, such as NocL (46% identity)[23] from the nocardicin A pathway and LmnA (43% identity)[17] from the leinamycin pathway. The sequence of AzicO6 is similar to the coenzyme F_{420}-dependent N^5,N^{10}-methylene-tetrahydromethanopterin reductases and various flavin-dependent oxidoreductases. These proteins may play a role in the biosynthesis of the aziridine carboxylate starter unit (**16**). Unfortunately, the *Kibdelosporangium* sp. strain is recalcitrant to genetic manipulations, including protoplast transformation, conjugation, and electroporation. We were therefore unable to verify the functions of the genes in azicemicin biosynthesis by common genetic methods, and instead had to rely on further isotope tracing analysis and biochemical experiments to study the biosynthetic pathway.

Since an acetate unit (**8**) can be converted to aspartate (**10**) through the TCA cycle, as illustrated in Scheme 1, and alanine, cysteine, and serine can also be derived from acetate,[24] the utilization of any of these amino acids to make the aziridine moiety can explain the labeling patterns observed in the isotope-tracer studies using labeled acetate. To discriminate among these possibilities, feeding experiments using deuterium-labeled amino acids (L-Ala, D,L-Ser, and D,L-Asp) were carried out. When L-[3,3,3-$^2H_3$]alanine was used, no deuterium incorporation was detected. Thus, alanine is clearly not a precursor of the starter unit in the azicemicin pathway. When D,L-[2,3,3-$^2H_3$]serine was used in the feeding experiment, incorporation of deuterium was noted, but only into the *O*- and *N*-methyl groups of azicemicin A (Figure S1). This outcome is not surprising because the deuteria at the β-C of the labeled serine can be relayed through 5,10-methylene tetrahydrofolate (THF, **12**) to the *S*-methyl group of SAM (**13**), which is then used as a methyl donor for the corresponding SAM-dependent methyltransferases in azicemicin biosynthesis. This result also excludes cysteine as a precursor, since serine can be converted to cysteine via acetylserine or cystathionine.[25]

Scheme 1



However when D,L-[2,3,3-$^2H_3$]aspartic acid (**10**) was used, deuterium was found to be incorporated into H-1′, as well as the *O*- and *N*-methyl groups of azicemicin A (Scheme 1). Labeling of the three methyl groups is expected and can be rationalized by the conversion of **10** to glycine (**11**), which should have a deuterium at the α-position. The deuterium in glycine is then transferred to **12** and is further incorporated into the *N*- and *O*-methyl group of **1** via **13**. The finding of deuterium incorporation into C-1′ is most significant, providing strong evidence that aspartate is the precursor of aziridine (**3**) in azicemicins (**1**).

Interestingly, no incorporation of deuterium into C-2′ was observed, revealing that two of the three deuteria in **10** were washed out during the construction of the aziridine moiety. Two possible scenarios can account for such deuterium loss (see Scheme 1). The deuterium at C-1′ of azicemicin could be derived from the β-deuterium of **10**, and the loss of the α-deuterium is due to the interconversion of L- and D-aspartate (racemization) and of L-aspartate and oxaloacetate (**9**) (transamination). In this case, decarboxylation of the α-carboxylate group takes place (**14** → **15**) to generate the aziridine starter unit (**16**). Alternatively, the deuterium at C-1′ of azicemicin could be derived from the α-deuterium of **10**, and the β-deuterium labels are lost due to oxidation at the β-carbon (**10** → **17**). Decarboxylation of the β-carboxylate group (**17** → **18**) is required in this case to form the aziridine starter unit (**16**). Overall, these results have shed light on the biosynthesis of the aziridine moiety in **1**.

A notable feature of the *azic* cluster is the existence of two discrete A-domain homologues, AzicM and AzicV. The function of adenylation enzymes is to convert a cognate amino acid to the corresponding aminoacyl adenylate mixed anhydride at the expense of ATP and then transfer the acyl group onto the associated peptidyl carrier protein.[26] Recent progress in the study of adenylation enzymes enables one to predict which amino acid is activated by comparing the sequence with many known adenylation enzymes.[27] However, alignment of AzicM and AzicV with other known adenylation enzymes failed to predict their preferred amino acid substrate. This suggests that AzicM and AzicV may be used to activate less common amino acids in azicemicin biosynthesis. To investigate the substrate specificity of AzicM and AzicV and to determine whether one or both are involved in azicemicin biosynthesis, the *azicM* and *azicV* genes were each expressed in *E. coli*, and the desired proteins were purified. Enzyme activities were assayed using various amino acid substrates in the presence of magnesium ion and ATP by monitoring pyrophosphate production using the Nichol's method.[28] Low levels (<1 mM) of ATP and orthophosphate contamination do not interfere with the pyrophosphate quantification.

As shown in Figure 2, AzicM demonstrates a clear preference for both L- and D-aspartate (**10**), indicating that both enantiomers are substrates of AzicM. Neither isomer of aziridine carboxylate (**16**) is a substrate of AzicM. These findings, in conjunction with the results of the isotope-tracer experiments, strongly indicate that AzicM loads (D and/or L)-aspartate onto AzicP prior to decarboxylation. Because the sequence of AzicM does not align well with those of the regular aspartate-activating enzymes, which catalyze the adenylation of the α-carboxylate group of aspartate, AzicM may activate the β-carboxylate group of either D- or L-aspartate as an adenylate. An analogous mode of activation is also required for microcystin biosynthesis, in which the γ-carboxyl group of the L-glutamate precursor and the β-carboxyl group of the 3-methyl-aspartate precursor are proposed to be activated and then transferred to the carrier protein.[29] Examination of Figure 2 also shows that L-cysteine is the preferred substrate of AzicV. However, according to the feeding experiments, cysteine is not a precursor of azicemicin. Thus, AzicV is unlikely involved in the biosynthesis of azicemicin.

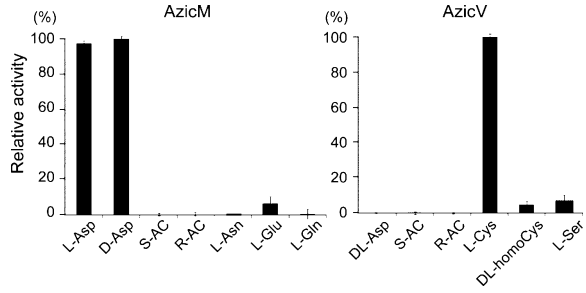
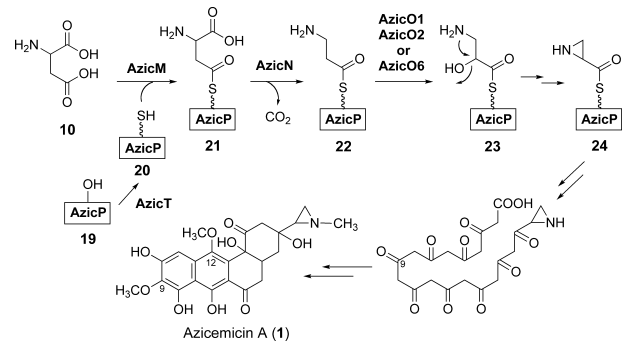


Figure 2. Substrate specificity of AzicM and AzicV from the azicemicin biosynthetic gene cluster (S-AC, (*S*)-aziridine-2-carboxylic acid; R-AC, (*R*)-aziridine-2-carboxylic acid).

On the basis of the isotope labeling patterns, the genetic analysis, and the enzymatic studies, a reasonable biosynthetic pathway for the formation of the aziridine carboxylate starter unit can now be formulated (Scheme 2). The reaction is initiated by AzicM-catalyzed activation of the β-carboxylate group of either D- or L-aspartate (**10**) as an adenylate, which then serves as the aminoacyl donor for the carrier protein AzicP. The AzicP-linked aspartate (**21**), with a free α-carboxylate group, is the substrate for the PLP-dependent decarboxylase, AzicN. After decarboxylation, the resulting AzicP-tethered β-alanine (**22**) is oxidized by a P450 (AzicO2) or flavin-dependent (AzicO1 or O6) monooxygenase to the AzicP-tethered isoserine (**23**). Cyclization of the resulting isoserine, which has a β-aminoalcohol moiety, may occur through an intramolecular S_N2 reaction leading to the displacement of the hydroxyl group by the amino group to generate the aziridine carboxyl starter unit (**24**). However, the identity of the enzyme catalyzing the last cyclization step is not immediately apparent. An alternative route (see Scheme 1) involving the activation of the α-carboxylate of aspartate is less likely, as AzicN would be required to catalyze the decarboxylation of a β-carboxylate.

Overall, we have cloned and sequenced the biosynthetic gene cluster of azicemicin from *Kibdelosporangium* sp. MJ126-NF4 and have clearly demonstrated that aspartate is the precursor of the aziridine moiety in azicemicin by isotope-tracer experiments. This conclusion is supported by the fact that one of the two adenylyl-transferases, AzicM, was characterized as an aspartate-activating enzyme. Although the enzyme(s) responsible for aziridine ring formation remains obscure, this work has laid the foundation for further biochemical and mechanistic studies of aziridine formation in azicemicin biosynthesis. Since the aziridine moiety holds promise to serve as a versatile functional entity in the design and construction of bioactive agents, an understanding of the mechanism of its formation may facilitate biosynthetic pathway re-engineering to prepare new aziridine-containing secondary metabolites.

Scheme 2



Acknowledgment. The azicemicin producing strain was kindly provided by Dr. Yuzuru Akamatsu (Microbial Chemistry Research Center, Tokyo, Japan). We thank Charles Melançon III for valuable help and discussion. This work was supported in part by a grant from the National Institutes of Health (GM 040541).

Supporting Information Available: Experimental details. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Tsuchida, T.; Iinuma, H.; Kinoshita, N.; Ikeda, T.; Sawa, R.; Takahashi, Y.; Naganawa, H.; Sawa, T.; Hamada, M.; Takeuchi, T. *J. Antibiot.* **1993**, *46*, 1772–1774.
(2) Tsuchida, T.; Sawa, R.; Takahashi, Y.; Iinuma, H.; Sawa, T.; Naganawa, H.; Takeuchi, T. *J. Antibiot.* **1995**, *48*, 1148–1152.
(3) Tsuchida, T.; Inuma, H.; Kinoshita, N.; Ikeda, T.; Sawa, T.; Hamada, M.; Takeuchi, T. *J. Antibiot.* **1995**, *48*, 217–221.
(4) Kuo, M. S.; Yurek, D. A.; Mizsak, S. A. *J. Antibiot.* **1989**, *42*, 357–360.
(5) Nagaoka, K.; Matsumoto, M.; Oono, J.; Yokoi, K.; Ishizeki, S.; Nakashima, T. *J. Antibiot.* **1986**, *39*, 1527–1532.
(6) Nakao, Y.; Fujita, M.; Warabi, K.; Matsunaga, S.; Fusetani, N. *J. Am. Chem. Soc.* **2000**, *122*, 10462–10463.
(7) Wolkenberg, S. E.; Boger, D. L. *Chem. Rev.* **2002**, *102*, 2477–2495.
(8) Alcaro, S.; Coleman, R. S. *J. Med. Chem.* **2000**, *43*, 2783–2788.
(9) Mao, Y.; Varoglu, M.; Sherman, D. H. *Chem. Biol.* **1999**, *6*, 251–263.
(10) Zhao, Q.; He, Q.; Ding, W.; Tang, M.; Kang, Q.; Yu, Y.; Deng, W.; Zhang, Q.; Fang, J.; Tang, G.; Liu, W. *Chem. Biol.* **2008**, *15*, 693–705.
(11) Hornemann, Y.; Kehrer, J. P.; Nunez, C. S.; Ranieri, R. L. *J. Am. Chem. Soc.* **1974**, *96*, 320–322.
(12) Liu, C.; Kelly, G. T.; Watanabe, C. M. *Org. Lett.* **2006**, *8*, 1065–1068.
(13) Corre, C.; Lowden, P. A. *Chem. Commun. (Camb)* **2004**, 990–991.
(14) Rawlings, B. J. *Nat. Prod. Rep.* **1997**, *14*, 523–556.
(15) Wawrik, B.; Kerkhof, L.; Zylstra, G. J.; Kukor, J. J. *Appl. Environ. Microbiol.* **2005**, *71*, 2232–2238.
(16) Metsa-Ketela, M.; Halo, L.; Munukka, E.; Hakala, J.; Mantsala, P.; Ylihonko, K. *Appl. Environ. Microbiol.* **2002**, *68*, 4472–4479.
(17) Cheng, Y. Q.; Tang, G. L.; Shen, B. *J. Bacteriol.* **2002**, *184*, 7013–7024.
(18) Ogasawara, Y.; Katayama, K.; Minami, A.; Otsuka, M.; Eguchi, T.; Kakinuma, K. *Chem. Biol.* **2004**, *11*, 79–86.
(19) Ota, Y.; Tamegai, H.; Kudo, F.; Kuriki, H.; Koike-Takeshita, A.; Eguchi, T.; Kakinuma, K. *J. Antibiot.* **2000**, *53*, 1158–1167.
(20) Llewellyn, N. M.; Li, Y.; Spencer, J. B. *Chem. Biol.* **2007**, *14*, 379–386.
(21) Gullon, S.; Olano, C.; Abdelfattah, M. S.; Brana, A. F.; Rohr, J.; Mendez, C.; Salas, J. A. *Appl. Environ. Microbiol.* **2006**, *72*, 4172–4183.
(22) Kim, H. J.; Pongdee, R.; Wu, Q.; Hong, L.; Liu, H.-W. *J. Am. Chem. Soc.* **2007**, *129*, 14582–14584.
(23) Gunsior, M.; Breazeale, S. D.; Lind, A. J.; Ravel, J.; Janc, J. W.; Townsend, C. A. *Chem. Biol.* **2004**, *11*, 927–938.
(24) Glawischnig, E.; Gierl, A.; Tomas, A.; Bacher, A.; Eisenreich, W. *Plant Physiol.* **2001**, *125*, 1178–1186.
(25) Chang, Z.; Vining, L. C. *Microbiology* **2002**, *148*, 2135–2147.
(26) Mootz, H. D.; Marahiel, M. A. *Curr. Opin. Chem. Biol.* **1997**, *1*, 543–551.
(27) Rausch, C.; Weber, T.; Kohlbacher, O.; Wohlleben, W.; Huson, D. H. *Nucleic Acids Res.* **2005**, *33*, 5799–5808.
(28) Grindey, G. B.; Nichol, C. A. *Anal. Biochem.* **1970**, *33*, 114–119.
(29) Tillett, D.; Dittmann, E.; Erhard, M.; von Dohren, H.; Borner, T.; Neilan, B. A. *Chem. Biol.* **2000**, *7*, 753–764.

JA907307H