J|A|C|S ARTICLES

Published on Web 04/15/2010

A Three Enzyme Pathway for 2-Amino-3-hydroxycyclopent-2-enone Formation and Incorporation in Natural Product Biosynthesis

Wenjun Zhang, Megan L. Bolla, Daniel Kahne, and Christopher T. Walsh*

Department of Biological Chemistry & Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, and Department of Chemistry & Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

Received January 12, 2010; E-mail: Christopher_Walsh@hms.harvard.edu

Abstract: A number of natural products contain a 2-amino-3-hydroxycyclopent-2-enone five membered ring, termed C_5N, which is condensed via an amide linkage to a variety of polyketide-derived polyenoic acid scaffolds. Bacterial genome mining indicates three tandem ORFs that may be involved in C_5N formation and subsequent installation in amide linkages. We show that the protein products of three tandem ORFs (ORF33−35) from the ECO-02301 biosynthetic gene cluster in *Streptomyces aizunenesis* NRRL-B-11277, when purified from *Escherichia coli*, demonstrate the requisite enzyme activities for C_5N formation and amide ligation. First, succinyl-CoA and glycine are condensed to generate 5-aminolevulinate (ALA) by a dedicated PLP-dependent ALA synthase (ORF34). Then ALA is converted to ALA-CoA through an ALA-AMP intermediate by an acyl-CoA ligase (ORF35). ALA-CoA is unstable and has a half-life of ~10 min under incubation conditions for off-pathway cyclization to 2,5-piperidinedione. The ALA synthase can compete with the nonenzymatic decomposition route and act in a novel second transformation, cyclizing ALA-CoA to C_5N. C_5N is then a substrate for the third enzyme, an ATP-dependent amide synthetase (ORF33). Using octatrienoic acid as a mimic of the C_{56} polyenoic acid scaffold of ECO-02301, formation of the octatrienyl-C_5N product was observed. This three enzyme pathway is likely the general route to the C_5N ring system in other natural products, including the antibiotic moenomycin.

Introduction

A variety of natural products with a vast range of biological activities have polyketide and nonribosomal peptide fragments joined together. Frequently, these scaffolds arise from hybrid nonribosomal peptide synthetase (NRPS)-polyketide synthase (PKS) assembly lines. In some cases, the polyketide backbone predominates as in rapamycin and FK506,[1,2] where a single NPRS-derived pipecolate is embedded in a polyketide framework. The reverse can also occur as in bleomycin and its congeners,[3] where a single polyketide fragment interrupts the nonribosomal peptide backbone. During the biosynthesis of some NRPS/PKS derived natural products, the non-NRPS/PKS machinery is enlisted to carry out the condensation between scaffold fragments. In one example, a polyketide acid, coronofacic acid, is enzymatically ligated to a nonproteinogenic amino acid, coronamic acid, by a *trans*-acting amide synthetase to yield the phytohormone antagonist coronatine.[4,5]

Of special note are cyclic five-membered nitrogen-containing ring structures that generate conformational constraints and offer hydrogen bonding possibilities for interaction with target proteins. Examples of these ring systems include pyrrolidine-2,5-diones as in the methylsuccinamide terminus of andrimid,[6] and the pyrrolidine-2,4-dione (tetramic acid) moieties in a wide range of natural products including equisetin[7] and cyclopiazonate.[8] In these instances, the pyrrolidine-diones are generated by the chain termination domains of hybrid PKS−NRPS assembly lines during product release. Typically, the ketone at C4 is enolized with the 4-hydroxy-3-ene tautomer predominating.

A distinct type of nitrogen-containing cyclic dione serves as a hydrogen bond donor/acceptor pharmacophore in more than 30 members of the manumycin family.[9] Termed the C_5N unit, this is formally a 2-aminocyclopentanedione unit, but it too predominates as the enol tautomer, 2-amino-3-hydroxycyclopent-2-enone. In most cases, including the manumycins, limocrocin,[10] Sch725424[11] and ECO-02301,[12] the amino group of the C_5N unit is acylated through an amide bond to a polyenoic acid component of polyketide origin (Figure 1). This suggests an

(1) Graziani, E. I. *Nat. Prod. Rep.* **2009**, *26*, 602–9.
(2) Schwecke, T.; Aparicio, J. F.; Molnar, I.; Konig, A.; Khaw, L. E.; Haydock, S. F.; Oliynyk, M.; Caffrey, P.; Cortes, J.; Lester, J. B.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 7839–43.
(3) Shen, B.; Du, L.; Sanchez, C.; Edwards, D. J.; Chen, M.; Murrell, J. M. *J. Ind. Microbiol. Biotechnol.* **2001**, *27*, 378–85.
(4) Bender, C.; Palmer, D.; PenalozaVazquez, A.; Rangaswamy, V.; Ullrich, M. *Arch. Microbiol.* **1996**, *166*, 71–75.
(5) Parry, R. J.; Lin, M. T.; Walker, A. E.; Mhaskar, S. *J. Am. Chem. Soc.* **1991**, *113*, 1849–1850.
(6) Jin, M.; Fischbach, M. A.; Clardy, J. *J. Am. Chem. Soc.* **2006**, *128*, 10660–10661.
(7) Sims, J. W.; Fillmore, J. P.; Warner, D. D.; Schmidt, E. W. *Chem. Commun.* **2005**, 186–188.
(8) Liu, X.; Walsh, C. T. *Biochemistry* **2009**, *48*, 8746–8757.
(9) Sattler, I.; Thiericke, R.; Zeeck, A. *Nat. Prod. Rep.* **1998**, *15*, 221–240.
(10) Hanajima, S.; Ishimaru, K.; Sakano, K.; Roy, S. K.; Inouye, Y.; Nakamura, S. *J. Antibiot.* **1985**, *38*, 803–805.
(11) Yang, S. W.; Chan, T. M.; Terracciano, J.; Patel, R.; Loebenberg, D.; Chen, G. D.; Patel, M.; Gullo, V.; Pramanik, B.; Chu, M. *J. Antibiot.* **2005**, *58*, 192–195.

10.1021/ja1002845

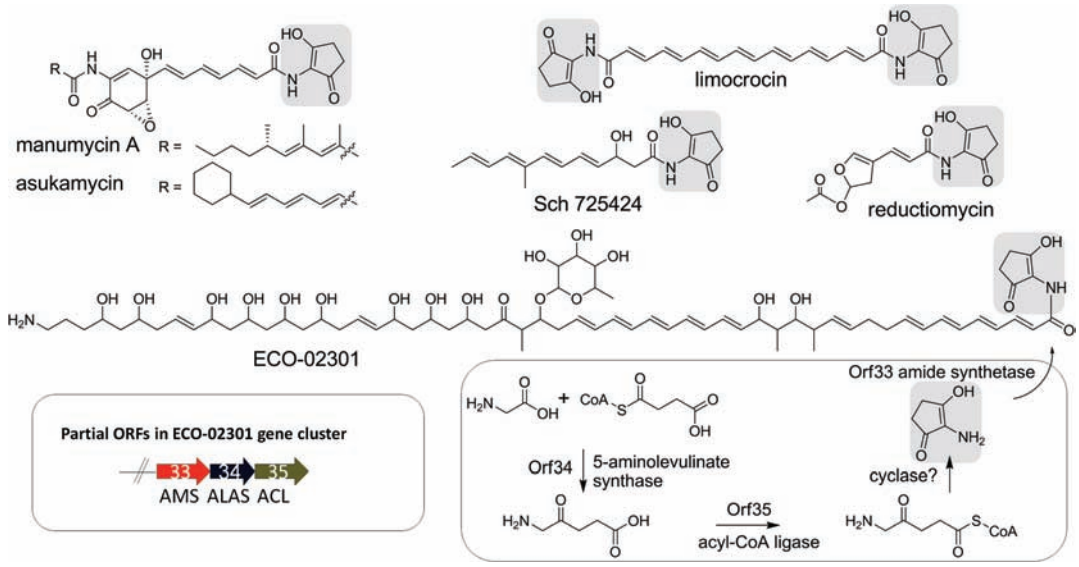

Figure 1. Examples of natural products containing C_5N (shaded). Candidate ORFs and a putative biosynthetic pathway for C_5N formation during ECO-02301 biosynthesis are boxed. ALAS, 5-aminolevulinate synthase; ACL, acyl-CoA ligase; AMS, amide synthetase.

analogy to the biosynthesis of coronatine and the aminocoumarin antibiotics, where the acid and amine components are connected late in the biosynthetic pathway by an amide-forming ligase.[13,14] An unusual variant is found in moenomycin biosynthesis where the C_5N unit is in amide linkage to a hexuronic acid ring.[15]

Initial inspection of the C_5N scaffold suggested it could arise from 5-aminolevulinic acid (ALA) by an unusual cyclization process (Figure 1).[16−18] ALA is a primary metabolite in many organisms as the immediate precursor to the monopyrrolic building block, porphobilinogen, for heme and corrin biogenesis.[19] While ALA can arise from reduction of glutamate on glutamyl-tRNA$_{glu}$, ALA can also be generated via the Shemin pathway in which PLP-dependent condensation of glycine and succinyl-CoA results in net loss of CO_2 and CoASH to form the five carbon 5-amino-4-keto-levulinic acid.[20] In 2006, genetic studies of the asukamycin producer *Streptomyces nodosus* subsp. *asukaensis* revealed genes for both pathways to ALA.[21] The knockout of gene *hemA-asuA* abolished asukamycin production. Furthermore, expression and purification of the encoding PLP-binding enzyme in *Escherichia coli* gave low but detectable activity of ALA synthesis. These results validated that the glycine and succinyl-CoA condensing enzyme is responsible for ALA formation and likely involved in the biosynthesis of C_5N.

Two attributes of assembly of C_5N-containing amide natural products intrigued us most: (1) the unusual cyclization of ALA to C_5N and (2) the ability of the amide synthetase to activate the presumably weak nitrogen nucleophile of C_5N as well as its specificity for the polyenoic or hexuronic acid partner substrates. To this end, we have examined the biosynthetic gene cluster of antifungal agent ECO-02301 for possible enzyme candidates responsible for the formation and ligation of the C_5N unit to the C_{56} polyenoic acid moiety.[12] The production of ECO-02301 had been predicted by genomic analysis of *Streptomyces aizunensis* NRRL B-11277, and then ECO-02301 was successfully identified as a 1297 Da linear polyketide product from the fermentation broth. Notable to our interests is the terminal C_5N moiety and the prediction by bioinformatic analysis that ORF33−35 in the ECO-02301 gene cluster could be involved in C_5N formation and ligation (Figure 1). ORF33 could be a putative amide synthetase (AMS, ligation of C_5N to polyenoate), while ORF34 has high sequence similarity to a PLP-dependent ALA synthase (ALAS) and ORF35 has signature sequences predicted for an acyl-CoA ligase (ACL). Through *in vitro* biochemical characterization of all three enzymes, we present experimental validation of the C_5N biosynthetic pathway, including identification of the novel C_5N cyclase activity.

Materials and Methods

Bacterial Strains, Plasmids, Materials, and Instrumentation. *S. aizunensis* NRRL B-11277 was obtained from USDA ARS Culture Collection. Oligonucleotide primers were synthesized by Integrated DNA Technologies, and PCR was performed with Phusion High-Fidelity PCR Master Mix (NEB). Cloning was performed using the Xa/LIC Cloning Kit (Novagen). Recombinant plasmid DNA was purified with a QIAprep kit (Qiagen). DNA sequencing was performed at the Molecular Biology Core Facilities of the Dana Farber Cancer Institute. *E. coli* BL21 Gold (DE3) Cells (Stratagene) transformed with pET-derived vectors were used for overexpression of proteins in LB medium supplemented with kanamycin. Nickel-nitrilotriacetic acid agarose (Ni-NTA) superflow resin and SDS-PAGE gels were purchased from Qiagen and Biorad, respectively. Protein samples were concentrated using 30 kDa MMCO Amicon Ultra filters (Millipore). DNA and protein concentrations were determined by Nanodrop 1000 spectrophotometer (Thermo Scientific). All chemicals were purchased from Sigma-Aldrich unless stated

(12) McAlpine, J. B.; Bachmann, B. O.; Piraee, M.; Tremblay, S.; Alarco, A. M.; Zazopoulos, E.; Farnet, C. M. *J. Nat. Prod.* **2005**, *68*, 493–496.
(13) Liyanage, H.; Penfold, C.; Turner, J.; Bender, C. L. *Gene* **1995**, *153*, 17–23.
(14) Galm, U.; Dessoy, M. A.; Schmidt, J.; Wessjohann, L. A.; Heide, L. *Chem. Biol.* **2004**, *11*, 173–183.
(15) Ostash, B.; Saghatelian, A.; Walker, S. *Chem. Biol.* **2007**, *14*, 257–267.
(16) Nakagawa, A.; Wu, T. S.; Keller, P. J.; Lee, J. P.; Omura, S.; Floss, H. G. *J. Chem. Soc., Chem. Commun.* **1985**, 519–521.
(17) Beale, J. M.; Lee, J. P.; Nakagawa, A.; Omura, S.; Floss, H. G. *J. Am. Chem. Soc.* **1986**, *108*, 331–332.
(18) Endler, K.; Schuricht, U.; Hennig, L.; Welzel, P. *Tetrahedron Lett.* **1998**, *39*, 13–16.
(19) Jaffe, E. K. *Bioorg. Chem.* **2004**, *32*, 316–325.
(20) Avissar, Y. J.; Ormerod, J. G.; Beale, S. I. *Arch. Microbiol.* **1989**, *151*, 513–519.
(21) Petricek, M.; Petrickova, K.; Havlicek, L.; Felsberg, J. *J. Bacteriol.* **2006**, *188*, 5113–5123.

otherwise. [4-^{13}C]5-Aminolevulinic acid and NMR solvents were purchased from Cambridge Isotope Laboratories, Inc. 2,5-Piperidinedione was purchased from Small Molecules. Boc-5-aminolevulinic acid was purchased from AnaSpec, Inc.

NMR spectra were recorded on a Varian Inova 500 (500 MHz ^{1}H, 125 MHz ^{13}C) instrument in $CDCl_3$. LC-MS analysis was performed on an Agilent Technologies 6520 Accurate-Mass Q-TOF LC-MS instrument or an Agilent Technologies 6210 Accurate-Mass TOF LC-MS instrument. HPLC analysis was performed on a Beckman Coulter System Gold with an Inertsil ODS-4 C18 column (4.6 × 250 mm) from GL Sciences, Inc. A Phenomenex Luna C18 column (21.2 × 250 mm) was used for preparative HPLC. A SPECTRAmax plus 384 96-well plate reader from Molecular Devices was used for continuous spectrophotometric assays.

Cloning, Overexpression, and Purification of ORF33−35. ORF33−35 were PCR amplified from genomic DNA extracted from *S. aizunensis*. For ORF33, the forward primer was 5′-GGTATTGAGGGTCGCATGACCCCGCAGGACCATTGGTG-3′ and the reverse primer was 5′-AGAGGAGAGTTAGAGCCTTAGGAGTCGAGCAGCTGCAGCC-3′. For ORF34, the forward primer was 5′-GGTATTGAGGGTCGCATGAACCTGCACCTGGAATCGTA-3′ and the reverse primer was 5′-AGAGGAGAGTTAGAGCCTTACGAAAGCCAGTTCCTGTCGG-3′. For ORF35, the forward primer was 5′-GGTATTGAGGGTCGCATGACCCGGTCGGTGGCGGCCGT-3′ and the reverse primer was 5′- AGAGGAGAGTTAGAGCCTTACGCGTAGCGGTGTGCCAGCT-3′. Purified PCR products were ligated to pET-30 Xa/LIC following the standard protocol and confirmed by DNA sequencing. The resulting expression constructs were transformed into *E. coli* BL21 cells for protein expression. Expression and purification for all three proteins followed the same general procedure and is detailed as follows. In 1 L of liquid culture, the cells were grown at 37 °C in LB medium with 50 μg/mL kanamycin to an OD_{600} of 0.4. The cells were cooled on ice for 10 min and then induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 16 h at 16 °C. The cells were harvested by centrifugation (6000 rpm, 6 min, 4 °C), resuspended in 30 mL of lysis buffer (20 mM HEPES, pH 8.0, 0.5 M NaCl, and 5 mM imidazole), and lysed by sonication on ice. Cellular debris was removed by ultracentrifugation (35 000 rpm, 30 min, 4 °C). Ni-NTA agarose resin was added to the supernatant (1 mL/L of culture) and the solution was nutated at 4 °C for 1 h. The protein resin mixture was loaded into a gravity flow column, and proteins were eluted with increasing concentrations of imidazole in Buffer A (50 mM HEPES, pH 8.0, and 2 mM EDTA). Purified proteins were concentrated and buffer exchanged into Buffer A and 10% glycerol using Amicon Ultra filters. Gel filtration chromatography was performed on a Superdex 200 10/300 column connected to an Amersham automated FPLC system at 4 °C (50 mM phosphate buffer, and 150 mM NaCl, pH 7.8). The final proteins were flash-frozen in liquid nitrogen and stored at −80 °C.

HPLC/LC-MS Product Assays. All product assays were performed in 50 μL of 50 mM HEPES (pH 8.0) at 25 °C. LC-MS analysis was normally performed with a linear gradient of 0 to 20% CH_3CN (v/v) over 20 min, 20−95% CH_3CN (v/v) over 5 min, and 95% CH_3CN (v/v) for a further 15 min in H_2O supplemented with 0.1% (v/v) formic acid, at a flow rate of 0.5 mL/min. HPLC analysis was normally performed with a linear gradient of 2−12% CH_3CN (v/v) over 30 min, 12−95% CH_3CN (v/v) over 5 min, and 95% CH_3CN (v/v) for a further 10 min in H_2O supplemented with 0.1% (v/v) trifluoroacetic acid (TFA) at a flow rate of 1 mL/min. For analysis of ALAS activity, 20−50 μM ORF34 was incubated with 5 mM glycine and 1 mM succinyl-CoA for 2 h. The protein was removed by 5 kDa MMCO filter tubes, and the filtered reaction mixture was subjected to LC-MS analysis. OPTA derivatization of amines in the filtered reaction mixture was further performed following the reported protocol.[22] For analysis of ACL activity, 20−50 μM ORF35 was incubated with 2−5 mM acid substrates, 2 mM ATP, 2 mM CoA, and 2 mM $MgCl_2$ for 30 min. The reactions were stopped by addition of tricholoacetic acid (TCA) to a final concentration of 5% (v/v), and the supernatant was subjected to both LC-MS and HPLC analysis. For analysis of cyclase activity, 20−50 μM ORF34 was added to the reaction mixture of ORF35 and incubated for ∼60 min. The reactions were stopped by addition of 100 μL of CH_3CN, and the supernatant was subjected to LC-MS and HPLC analysis. Alternatively, 20−50 μM ORF34 was added to buffer containing synthesized ALA-CoA substrate and reacted for 10 min prior to LC-MS and HPLC analysis. For analysis of AMS activity, 20−50 μM ORF33 was added to buffer containing synthesized 2 mM C_5N, 2 mM 2,4,6-octatrienoic acid, 5 mM ATP, and 2 mM $MgCl_2$. The reactions were stopped after 1 h by addition of CH_3CN and the supernatant was subjected to LC-MS and HPLC analysis. For reconstitution of the entire pathway, reaction mixtures contained 50 μM ORF33−35, 5 mM glycine, 1 mM succinyl-CoA, 5 mM ATP, 2 mM CoA, 2 mM $MgCl_2$, and 2 mM 2,4,6-octatrienoic acid. The reactions were stopped after 2−3 h by addition of 100 μL of CH_3CN and the supernatant was subjected to LC-MS and HPLC analysis.

Preparation of ALA-CoA. Boc-ALA [25 μmol, 1 equiv], CoA sodium salt (25 μmol, 1 equiv), PyBOP (Novobiochem, 50 μmol, 2 equiv) and K_2CO_3 (100 μmol, 4 equiv) were dissolved in tetrahydrofuran/H_2O [1/1 (v/v), 1 mL] and stirred for 2 h at room temperature. The reaction mixture was directly purified by preparative HPLC (two injections, 0.5 mL each) using a linear gradient of 2−60% CH_3CN (v/v) over 30 min in H_2O supplemented with 0.1% (v/v) TFA at a flow rate of 10 mL/min. The HPLC-purified mixture was concentrated on a rotary evaporator and lyophilized to yield Boc-ALA-CoA (30% yield) as a white solid. The Boc protecting group was removed by redissolving Boc-ALA-CoA in 1 mL of 5−10% TFA in H_2O and stirring for 30 min at room temperature. The reaction mixture was again purified by preparative HPLC (same method as above), concentrated, and lyophilized, yielding ALA-CoA (85% yield) as a white solid.

Preparation of 2,4,6-Octatrienyl-C_5N. To a cooled solution (0 °C) of (2*E*,4*E*,6*E*)-octa-2,4,6-trienoic acid[23] (97 mg, 0.700 mmol) in CH_2Cl_2 (7 mL) and DMF (3 drops) was added a solution of oxalyl chloride (100 mg, 0.840 mmol, 1.2 equiv) in $CHCl_2$ (0.5 mL) dropwise over 30 min. The resulting solution was stirred at 0 °C for 2 h, then slowly canulated into a cooled solution (0 °C) of freshly prepared 2-amino-3-hydroxy-2-cyclopenten-1-one hydrochloride[24] (103 mg, 0.700 mmol), *N*,*N*-dimethylaminopyridine (5 mg, 0.041 mmol, 5 mol %) in pyridine (15 mL). The reaction mixture was slowly warmed to room temperature and maintained stirring overnight. The reaction was then concentrated (azeotroping with toluene) and purified by flash column chromatography with 98:2 (CH_2Cl_2/MeOH) to afford a yellow solid (55 mg, 0.236 mmol, 34%). ^{1}H NMR (500 MHz, $CDCl_3$) δ_H 13.6 (s, 1H), 7.45 (broad s, 1H), 7.34 (dd, J = 14.5, 11.5 1H), 6.58 (dd, J = 15.0, 11.0 1H), 6.22 (dd, J = 14.5, 12.0 1H), 6.15−6.20 (m, 1 H), 6.00 (dd, J = 14.5, 9.5 1H), 5.95 (d, J = 15.0 1H), 2.60−2.62 (m, 2H), 2.52−2.55 (m, 2 H), 1.85 (d, J = 7.0 3 H); ^{13}C NMR (125 MHz, $CDCl_3$) δ_C 197.3, 177.6, 173.8, 145.0, 142.5, 136.3, 131.3, 127.2, 119.2, 32.3, 25.7, 18.8.

Kinetic Investigations of ORF34. ORF34 activity was measured at 25 °C in a 100 μL reaction volume containing 100 mM sodium phosphate (pH 7.8), 0.1 mM 5,5′-dithio-bis(2-nitro-benzoic acid) (DTNB), 0.2 mM PLP, and 20 μM ORF34. Kinetic parameters for the succinyl-CoA substrate were determined with the concentration of glycine maintained at 10 mM and the concentration of succinyl-CoA varied from 0 to 425 μM; kinetic parameters for the glycine substrate were determined with the concentration of succinyl-CoA maintained at 0.5 mM and the concentration of glycine varied from 0 to 5 mM. Kinetic parameters for the ALA-CoA substrate in the

(22) Morineau, G.; Azoulay, M.; Frappier, F. *J. Chromatogr.* **1989**, *467*, 209–216.

(23) Lensen, N.; Mouelhi, S.; Bellassoued, M. *Synth. Commun.* **2001**, *31*, 1007–1011.

(24) Ebenezer, W. J. *Synth. Commun.* **1991**, *21*, 351–358.

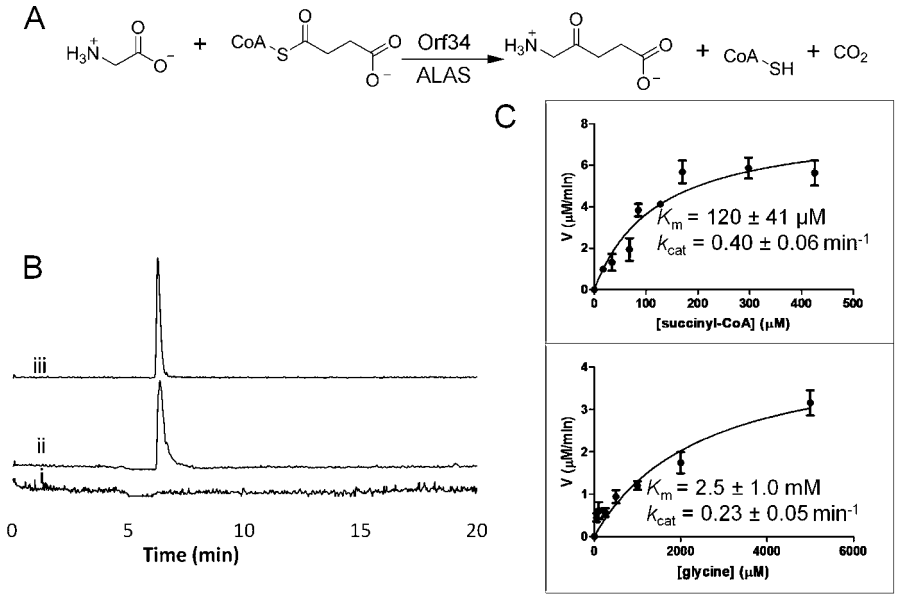

Figure 2. Characterization of ORF34 as a 5-aminolevulinate synthase. (A) Schematic of the condensation of glycine and succinyl-CoA to yield 5-aminolevulic acid catalyzed by ORF34. (B) Extracted ion chromatograms (m/z = 132 $[M + H]^+$) showing ORF34-catalyzed production of ALA (trace ii) as compared to the ALA standard (trace iii). ALA was not detected in controls without ORF34 (shown, trace i) or without any of the substrates (not shown). (C) Determination of ORF34 kinetic parameters by assay with DTNB. Parameters were determined for succinyl-CoA (top panel) by fixing the glycine concentration at 10 mM, while parameters were determined for glycine (bottom panel) by fixing the succinyl-CoA concentration at 500 μM. Error bars represent standard deviations from at least three independently performed experiments.

cyclization reaction were determined by varying the concentration of ALA-CoA from 0 to 400 μM. For each concentration, a control reaction was carried out without enzyme. Assays were initiated by the addition of acyl-CoA. Activity was monitored continuously by following the increase in absorbance at 412 nm resulting from the reaction between the free thiol of CoASH and DTNB. Initial velocities (v_{ini}) were calculated using the extinction coefficient of 14150 M^{-1} cm^{-1} ($v_{ini} = v_{ini}$[with enzyme] $- v_{ini}$[without enzyme]). The v_{ini} data were fitted to the Michaelis−Menten equation in GraphPad Prism to obtain estimates for k_{cat} and K_m.

ATP-[^{32}P]PP$_i$ Exchange Assays for ORF33 and ORF35. A typical assay contained, in a total volume of 800 μL, 5 mM acid substrate, 5 mM ATP, 10 mM $MgCl_2$, 1 mM Na[^{32}P]-pyrophosphate (PP_i) ($\sim 2.5 \times 10^6$ cpm/mL), and 50 mM HEPES, pH 8.0. Reactions were initiated by the addition of enzyme (1 μM of ORF35 or 0.8 μM of ORF33). At regular time intervals, 100 μL aliquots were quenched with 500 μL of a charcoal suspension (100 mM $NaPP_i$, 350 mM $HClO_4$, and 16 g/L charcoal). The mixtures were vortexed and then centrifuged at 13 000 rpm for 6 min. The pellets were washed twice with 500 μL of wash solution (100 mM $NaPP_i$ and 350 mM $HClO_4$). Charcoal-bound radioactivity was measured on a Beckman LS 6500 scintillation counter. Turnover was calculated as (% incorporation of [^{32}P]PP_i)[total PP_i]/[Enz].

ATP-PP$_i$ Release Assays for ORF35. The inorganic pyrophosphate released by enzymatic reaction was measured continuously using the EnzChek Pyrophosphate Assay Kit (Invitrogen). A typical assay contained, in a total volume of 100 μL, 0−5 mM acid substrate, 5 mM ATP, 2 mM $MgCl_2$, and 5 μM ORF35 in 50 mM HEPES, pH 8.0. MESG substrate, purine nucleoside phosphorylase and inorganic pyrophosphatase were added according to the protocol. Reactions were initiated by the addition of acid substrate and monitored at 360 nm. Initial velocities were calculated using the standard curve for inorganic pyrophosphate. The data were fitted to the Michaelis−Menten equation in GraphPad Prism to obtain estimates for k_{cat} and K_m.

Kinetic Investigations of ORF33. The amide synthetase activity of ORF33 was measured at 25 °C in a 100 μL reaction volume containing 50 mM HEPES, pH 8.0, 5 mM ATP, 2 mM $MgCl_2$, 5 μM ORF33 and acid and amine substrates. Activity was monitored continuously by following the increase in absorbance at 355 nm resulting from the production of the ligated product. Kinetic parameters for the 2,4,6-octatrienoic acid substrate were determined with the concentration of C_5N maintained at 1 mM and the concentration of acid varied from 0 to 5 mM. Higher acid concentrations could not be achieved due to poor solubility. Kinetic parameters for the C_5N substrate were determined with the concentration of 2,4,6-octatrienoic acid maintained at 2 mM and the concentration of C_5N varied from 0 to 1.1 mM. Initial velocities (v_{ini}) were calculated using a standard curve of synthetically prepared 2,4,6-octatrienyl-C_5N. The v_{ini} data were fitted to the Michaelis−Menten equation in GraphPad Prism to obtain estimates for k_{cat} and K_m.

Results

Expression and Purification of ORF33−35 in *E. coli*. ORF33−35 from *S. azuinensis* were amplified and cloned into an expression vector encoding an N-terminal His_6-tag. The corresponding recombinant proteins (ORF34, 49 kDa; ORF35, 59 kDa; ORF33, 60 kDa) were overproduced in *E. coli* and purified using Ni-NTA affinity chromatography (Supporting Information Figure S1). The approximate protein yields were 10 mg/L for ORF34, 26 mg/L for ORF35, and 30 mg/L for ORF33.

ORF34 Is a PLP-Dependent Aminolevulinate Synthase (ALAS). Examination of the UV absorption spectra of ORF34 identified a characteristic absorbance maximum at 420 nm, indicating the presence of an enzyme-bound PLP Schiff base (Supporting Information Figure S2).[25] Titration with PLP indicated approximately 60% of the protein was in the holo-form (PLP bound) upon isolation (Supporting Information Figure S3). Incubation of purified ORF34 with glycine and succinyl-CoA led to the formation of a new compound detected by LC-MS, with retention time and mass (m/z = 132 $[M + H]^+$) matching those of the commercial ALA standard (Figure 2A,B). The identity of the product ALA was further confirmed using OPTA

(25) Ferreira, G. C.; Neame, P. J.; Dailey, H. A. *Protein Sci.* **1993**, *2*, 1959–1965.

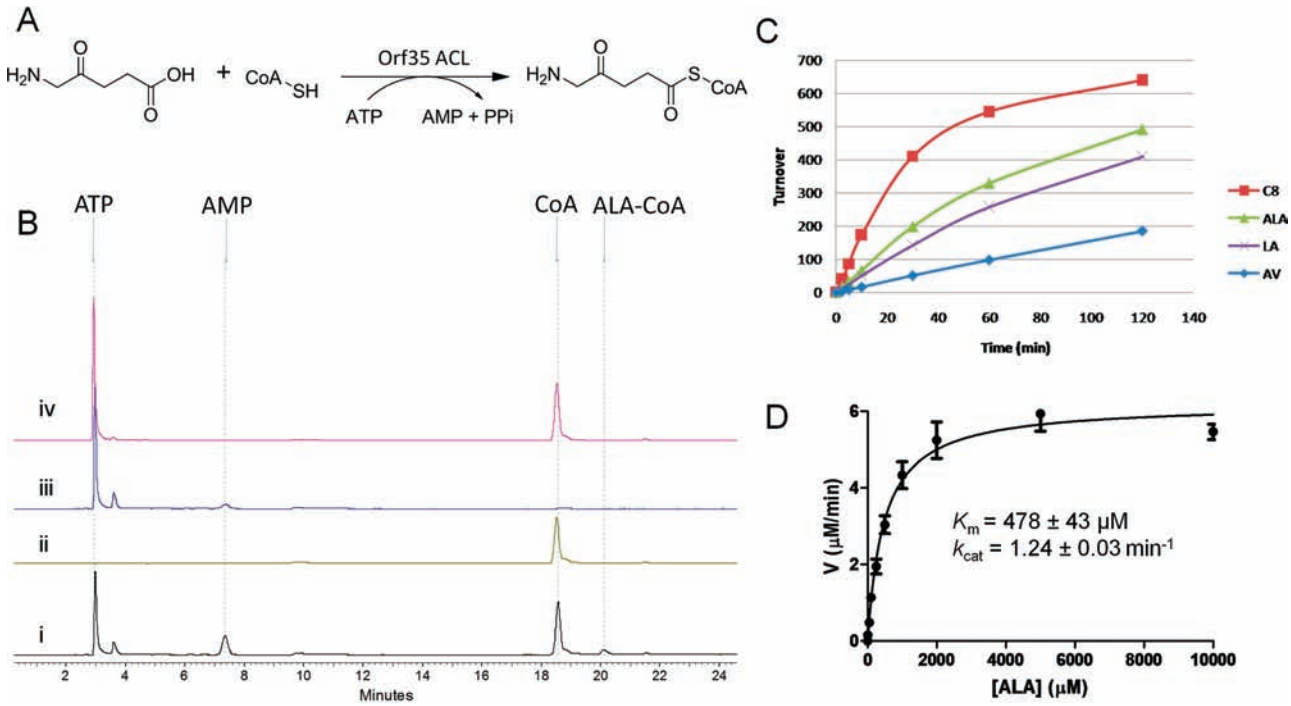

Figure 3. Characterization of ORF35 as an acyl-CoA ligase. (A) Schematic of the ligation of ALA to CoASH catalyzed by ORF35. (B) HPLC traces (260 nm) of ORF35 reaction (trace i) and controls with no ATP (trace ii), no CoA (trace iii), and no enzyme (trace iv). (C) ATP-[^{32}P]PP$_i$ exchange assays monitoring the reversible formation of acyl-AMP as the first half reaction. Calculated rates: k(ALA) = 6.6 min^{-1}, k(octanoic acid [C8]) = 17.5 min^{-1}, k(levulinic acid [LA]) = 5 min^{-1}, k(5-aminovaleric acid [AV]) = 1.7 min^{-1}. (D) Kinetic parameters for the ALA substrate determined by ATP-PP$_i$ release assays. Error bars represent standard deviations from at least three independently performed experiments.

derivatization and LC-MS analysis (Supporting Information Figure S4). The product assays confirmed that ORF34 was an aminolevulinate synthase (ALAS), which is in agreement with the previous report that ALAS is essential for asukamycin biosynthesis in *S. nodosus* subsp. *asukaensis*. Notably, the gene encoding the *S. nodosus* subsp. *asukaensis* ALAS is also adjacent to genes encoding AMS and ACL homologues.[21]

To determine the kinetic parameters of ORF34, the DTNB assay monitoring the free CoA formation was performed (Figure 2C). The K_m values of the enzyme were determined to be 120 μM for succinyl-CoA and 2.5 mM for glycine, which is within range of the reported values for ALASs originating from different organisms (17−257 μM for succinyl-CoA and 1.9−9.7 mM for glycine).[26,27] The turnover number was determined to be ~0.4 min^{-1}, which was ~10-fold less than those of ALASs dedicated to heme biosynthesis, but ~10-fold greater than that of the ALAS involved in the biosynthesis of asukamycin.[21]

ORF35 Is an Acyl-CoA Ligase (ACL). ORF35 shows high sequence similarity to acyl-CoA ligases, and was proposed to ligate ALA synthesized by ORF34 to CoASH. To probe the role of ORF35, *in vitro* assays containing ALA, CoASH, ATP, Mg^{2+} and purified ORF35 were performed in HEPES buffer at pH 8.0 and monitored by LC-MS. The protein readily precipitated at lower pH values (<7.5). ALA-CoA was detected as a new product (m/z = 881.1723 $[M + H]^+$, Δ = 2.1 mmu) (Supporting Information Figure S5), albeit in low yields. Full sets of control experiments confirmed that ORF35 catalyzed the formation of ALA-CoA in an ATP-dependent manner (Figure 3A,B). To determine the substrate specificity of ORF35, the classical ATP-[^{32}P]PP$_i$ exchange assay was used to monitor the reversible formation of acyl-AMP as the first half reaction. The enzyme demonstrated specificity toward ALA over substrate analogs such as levulinic acid and 5-aminovaleric acid (Figure 3C), with a k_{cat} of 6.6 min^{-1} for ALA. Interestingly, ORF35 recognized octanoic acid (k_{cat} = 17.5 min^{-1}) better than ALA. A coupled spectrophotometric assay for PP$_i$ release was further used to kinetically characterize ORF35, and the substrate specificity was in agreement with the ATP-PP$_i$ exchange assays (for ALA, k_{cat} = 1.24 min^{-1}, K_m = 478 μM) (Figure 3D, Table 1 and Supporting Information Figure S6). It is notable that the k_{cat} values obtained from the PP$_i$ release assays were relatively low for all the acids tested, which is likely due to slow, off-pathway release of the tightly bound acyl-AMP intermediates. Regardless of assay used, however, ORF35 showed better specificity toward medium-chain fatty acids, which may indicate that the ancestor of this enzyme is a fatty acid acyl-CoA ligase.

Table 1. ATP-PP$_i$ Release Assays for ORF35 on Selected Acids[a]

acids	V_{max} (mM min^{-1})	K_m (μM)	[ORF35] (μM)	k_{cat} (min^{-1})	k_{cat}/K_m (min^{-1}mM^{-1})
ALA	6.19 ± 0.14	478 ± 43	5	1.24 ± 0.03	2.594
LA	5.43 ± 0.08	1087 ± 46	5	1.08 ± 0.02	0.999
AV	8.07 ± 0.22	7431 ± 370	5	1.61 ± 0.04	0.217
C6	6.56 ± 0.15	187 ± 15	5	1.31 ± 0.03	7.016
C8	6.62 ± 0.23	181 ± 21	5	1.32 ± 0.05	7.316

[a] ALA, 5-aminolevulinic acid; LA, levulinic acid; AV, 5-aminovaleric acid; C6, hexanoic acid; C8, octanoic acid.

To confirm the acyl-CoA formation catalyzed by ORF35, *in vitro* reactions followed by LC-MS analyses were performed with selected acids (Supporting Information Figure S7). As expected, under the same reaction conditions, hexanoyl-CoA was quickly formed in good yield and levulinyl-CoA was formed less efficiently. However, very little ALA-CoA could be detected as noted above, and no 5-aminovaleryl-CoA could be detected by UV or mass ion extraction. Instead, 2,5-

(26) Lin, J.; Fu, W.; Cen, P. *Bioresour. Technol.* **2009**, *100*, 2293–2297.
(27) Bolt, E. L.; Kryszak, L.; Zeilstra-Ryalls, J.; Shoolingin-Jordan, P. M.; Warren, M. J. *Eur. J. Biochem.* **1999**, *265*, 290–299.

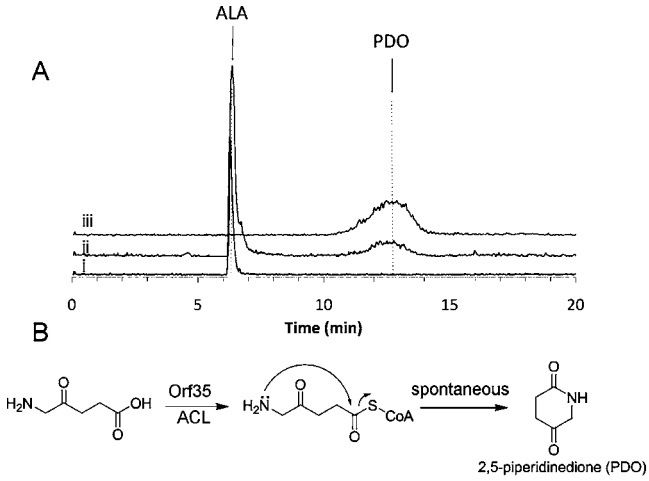

Figure 4. Instability of ALA-CoA. (A) Extracted ion chromatograms (m/z = 114 $[M + H]^+$) showing the formation of PDO in the reaction mixture of ORF35 (trace ii). PDO standard is shown in trace iii and the no enzyme control is shown in trace i. (B) Schematic of the spontaneous cyclization of ALA-CoA to PDO.

piperidinedione and 2-piperidinone were observed as products, respectively, which were reasoned to be formed through nonenzymatic cyclization of the aminoacyl-CoAs. Particularly, at pH 8.0 the -NH_2 group of the product aminoacyl-CoAs is likely reactive enough to serve as an internal nucleophile and attack the activated carbonyl of the CoA thioester terminus, leading to spontaneous formation of the six membered ring (Figure 4). ALA-CoA was found to be stable in acidic conditions, which is consistent with our hypothesis. Considering the instability of ALA-CoA at physiological pH, the mechanism underlying the cyclization of the activated ALA to C_5N instead of the default 2,5-piperidinedione was intriguing. Because no C_5N was detectable in the ORF35 *in vitro* assays, we reasoned that an additional dedicated cyclase was required to direct the synthesis of C_5N from ALA-CoA.

PLP-Dependent ORF34 Is a Bifunctional ALAS and ALA-CoA Cyclase. It has been reported that the keto-form of ALA was in large excess compared to the enolic tautomers at neutral pH. Using ^{13}C and ^1H NMR, no direct observation of the enol forms of ALA had been achieved,[28] arguing against a ready supply of the C5 carbon nucleophile required for spontaneous intramolecular capture of ALA-CoA to C_5N. Therefore, to favor C_5N formation, it seemed likely that activation of C5 adjacent to the amine and sequestration of the reactive terminal amine of ALA-CoA might be required for cyclization. As no putative cyclase was found in the gene cluster, we hypothesized that the PLP-dependent ORF34 ALAS was the best candidate for "cyclase" considering its catalytic mechanism. The ALAS reaction involves binding of glycine to PLP to yield a PLP-glycine complex, and a resonance-stabilized nucleophilic quinonoid intermediate is formed following *pro*-R proton subtraction.[29] Some ALASs can also catalyze labilization of the *pro*-R proton at C5 of the ALA product and generate detectable PLP−ALA quinonoid upon exposure of the enzyme to ALA.[30] These chemical roles of coenzyme PLP in both the forward and reverse reactions meet the prerequisites for C_5N cyclization as mentioned above, particularly the ability to labilize the C5 hydrogen.

To test our hypothesis, purified ORF34 was added to the reaction mixture of ORF35, and the reaction was quenched with acetonitrile and analyzed by LC-MS. C_5N was successfully detected as a new product with UV absorbance maximum at 250 nm (m/z = 114.0557 $[M + H]^+$, Δ = 0.7 mmu), and its mass was shifted by +1 using [4-^{13}C]ALA as an alternative substrate (Supporting Information Figure S8). The product identity was further confirmed by comparison to authentic C_5N synthesized as a standard by reduction of the corresponding nitro-cyclopentanedione as described in the Materials and Methods.[24] Control experiments demonstrated that the formation of C_5N from ALA relied on the presence of both enzymes (ORF34 and ORF35) and all cofactors (CoA, ATP and Mg^{2+}) (Figure 5A). Prior reports indicated synthetic C_5N is unstable upon concentration.[24] We determined by HPLC analysis that, under enzymatic reaction conditions, C_5N displayed a usable half-life of 140 min at pH 8.0. HPLC and LC-MS detection of C_5N from enzymatic reactions validated our hypothesis, demonstrating that the PLP-dependent ORF34 can direct the cyclization of ALA-CoA to yield C_5N (Figure 5B).

To further verify ALA-CoA as the substrate for ORF34 and kinetically characterize the cyclization step, ALA-CoA was synthesized chemically. The synthesis was undertaken with Boc-5-ALA and unblocked CoA, with PyBOP as a condensation reagent to yield Boc-5-ALA-CoA. The Boc group was then removed by TFA and the ALA-CoA product was purified by preparative HPLC. DTNB assays indicated that ALA-CoA had a half-life of ~10 min at pH 8.0, which is consistent with the difficulty in detecting the product ALA-CoA in ORF35 LC-MS assays. C_5N was successfully generated after the incubation of the synthetic ALA-CoA with ORF34, while no C_5N was detectable in the control without enzyme (Figure 5C). These results unequivocally confirmed the role of ORF34 as an ALA-CoA cyclase to afford C_5N. The enzyme demonstrated substantially higher catalytic efficiency for cyclization (k_{cat} = 2.6 min^{-1}, K_m = 36.7 μM, k_{cat}/K_m = 71 min^{-1} mM^{-1}) than for ALA synthesis according to the DTNB assays (Figure 5D).

ORF33 Is a C_5N Amide Synthetase. The condensation presumably between a carboxylic acid and the amine of C_5N constitutes the final step for C_5N incorporation into the polyketide backbone as a chain terminator. This is putatively promoted by ORF33, which shows reasonable sequence similarity (<40%) to amide synthetases (NovL, CloL, CouL, SimL) in the aminocoumarin antibiotic family.[14] To probe the role of ORF33, the purified enzyme was incubated with C_5N, ATP, $MgCl_2$, and 2,4,6-octatrienoic acid (OTEA), which was used to mimic the C-terminal polyolefinic moiety of the polyketide substrate (Figure 6A,B). LC-MS analysis showed production of a new compound with a UV absorbance maximum at 321 nm (λ_{max} = 300 nm for OTEA) (Supporting Information Figures S9 and S10), along with the formation of AMP and attenuation of both amine and acid substrate peaks. HRMS (m/z = 234.1108 $[M + H]^+$, Δ = 1.7 mmu) and mass fragmentation (two major mass fragments: 114.0553, 121.0643) suggested the new product to be 2,4,6-octatrienyl-C_5N (OTEA-C_5N). The identity of the product was further confirmed by comparison to the synthetic standard, produced from synthetic C_5N and OTEA as described in the Materials and Methods (Supporting Information Figure S11). Notably, ORF33 was not able to ligate ALA to OTEA in our assays. These results confirmed the role of ORF33 as a C_5N amide synthetase.

ATP-[^{32}P]PP_i exchange assays demonstrated that ORF33 adenylated carboxylate substrates as predicted on the basis of its homology to AMP-forming enzymes (Figure 6C). The kinetic data further revealed that the enzyme preferred acids of longer

(28) Jaffe, E. K.; Rajagopalan, J. S. *Bioorg. Chem.* **1990**, *18*, 381–394.
(29) Nandi, D. L. *J. Biol. Chem.* **1978**, *253*, 8872–8877.
(30) Hunter, G. A.; Ferreira, G. C. *Biochemistry* **1999**, *38*, 3711–3718.

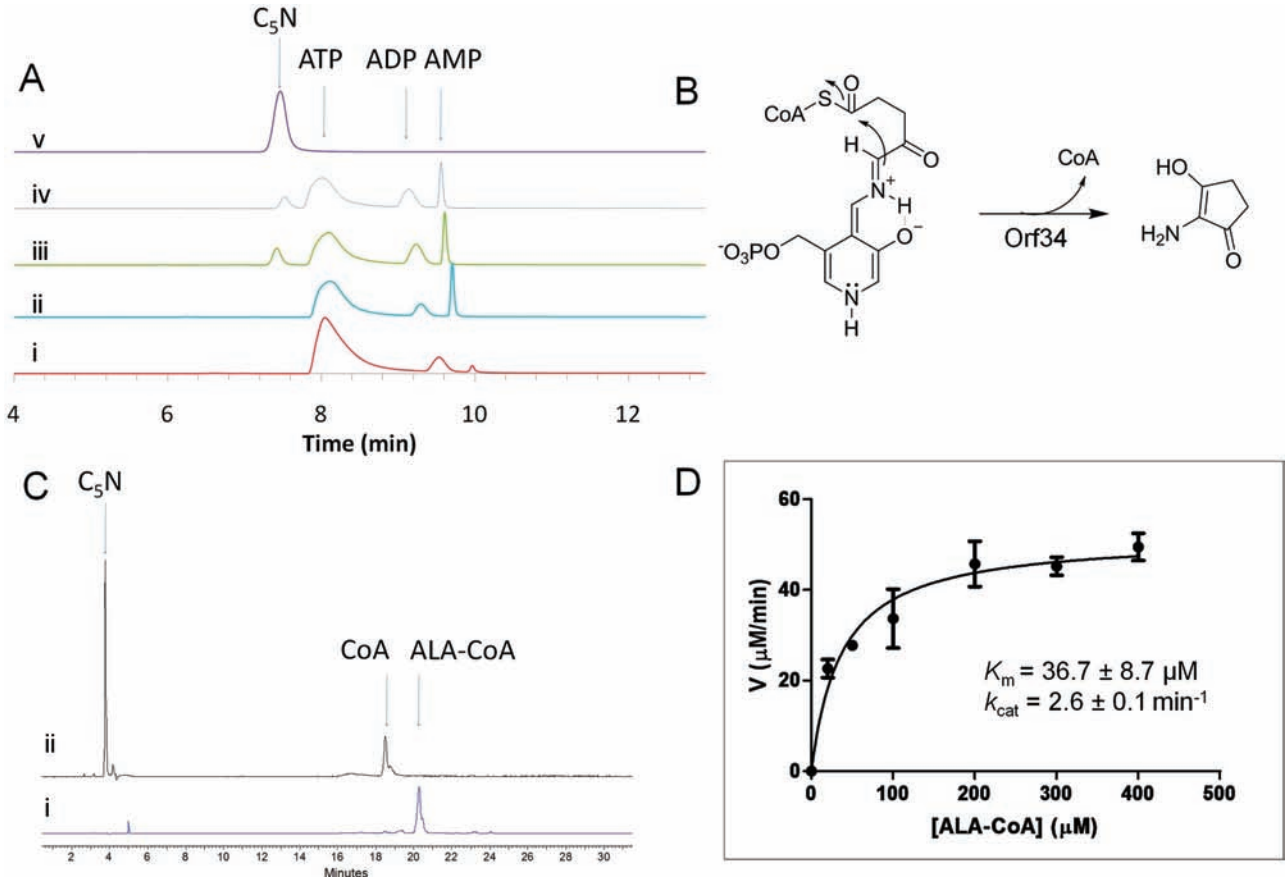

Figure 5. Characterization of PLP-dependent ORF34 as an ALA-CoA cyclase. (A) UV traces (250 nm) during LC-MS analysis of ORF34/ORF35 activity showing reaction with ORF34 only (trace i), ORF35 only (trace ii), ORF34 + ORF35 with [4-^{12}C]ALA (trace iii), ORF34 + ORF35 with [4-^{13}C]ALA (trace iv), and C_5N synthetic standard (trace v). (B) Schematic of cyclization of ALA-CoA to C_5N catalyzed by ORF34 via the PLP−quinonoid intermediate. (C) HPLC trace (250 nm) of ORF34 reaction producing C_5N and CoA using synthetic ALA-CoA as the sole substrate (trace ii). No enzyme control is shown in trace i. (D) Kinetic parameters for the ALA-CoA substrate determined by DTNB assays. Error bars represent standard deviations from at least three independently performed experiments.

chain length and higher degree of unsaturation (OTEA > 2,4-hexadienoic acid ≫2-butenoic acid > octanoic acid) consistent with resemblance to the presumed C_{56} native substrate. It is notable that the turnover rates were slow (k_{cat} = 0.3 min^{-1} for OTEA) in the assays, perhaps limited by slow PP_i off rates. The kinetic parameters of the enzyme toward C_5N and OTEA were then further examined by monitoring the rate of OTEA-C_5N formation at 355 nm (Figure 6D). This wavelength was selected to minimize the inference of the substrate OTEA on UV absorption (Supporting Information Figures S9 and S10). With a C_5N concentration of 1 mM, the enzyme showed a K_m of ∼7.8 mM for OTEA with an apparent k_{cat} of ∼10.2 min^{-1}, indicating the poor affinity of this substrate analog. In contrast, ORF33 had a low K_m of ∼95 μM for C_5N and an apparent k_{cat} of ∼1.8 min^{-1} at an OTEA concentration of 2 mM (poor OTEA solubility precluded reproducible assays at higher concentrations). These results demonstrated the capability of ORF33 in recognizing the presumably weak amine nucleophile of C_5N. The overall turnover rate was an order of magnitude greater in the product assays than in the ATP-PP_i exchange assays performed without the amine donor C_5N, indicating that acid activation and amine binding likely accelerate release of the PP_i coproduct.

Reconstitution of the Entire C_5N Biosynthetic Pathway *in Vitro*. All three enzymes, ORF33, ORF34, and ORF35, were incubated together with the essential cofactors (Mg^{2+}, ATP, CoA) and substrates (glycine, succinyl-CoA and OTEA), and the reaction mixture was monitored by LC-MS. The OTEA-C_5N was successfully produced (Figure 6B, trace ii), and was also detected using ALA as a starting material instead of glycine and succinyl-CoA (data not shown). Omission of any of the three enzymes completely abolished the production of OTEA-C_5N (Figure 6B, trace iv), confirming the necessity of all three enzymes in the biosynthesis and incorporation of the C_5N moiety. In addition, no protein complexes were observed between any of the three enzymes by gel filtration chromatography (data not shown), suggesting that all of the intermediates produced during each transformation are freely diffusible. This is in agreement with the above results that every step can be reconstituted individually *in vitro*.

Discussion

In this work, we have studied the biosynthesis of 2-amino-3-hydroxycyclopent-2-enone (C_5N) and its incorporation via an amide linkage to a partial polyenoate scaffold of the antifungal natural product ECO-02301. The condensation between glycine and succinyl-CoA catalyzed by ORF34 (ALAS) leads to formation of ALA, which is then ligated to CoASH yielding ALA-CoA promoted by the ATP-dependent ORF35 (ACL). ALA-CoA is not stable and can undergo spontaneous intramolecular cyclization with the amine group acting as an internal nucleophile to form the shunt product 2,5-piperidinedione.

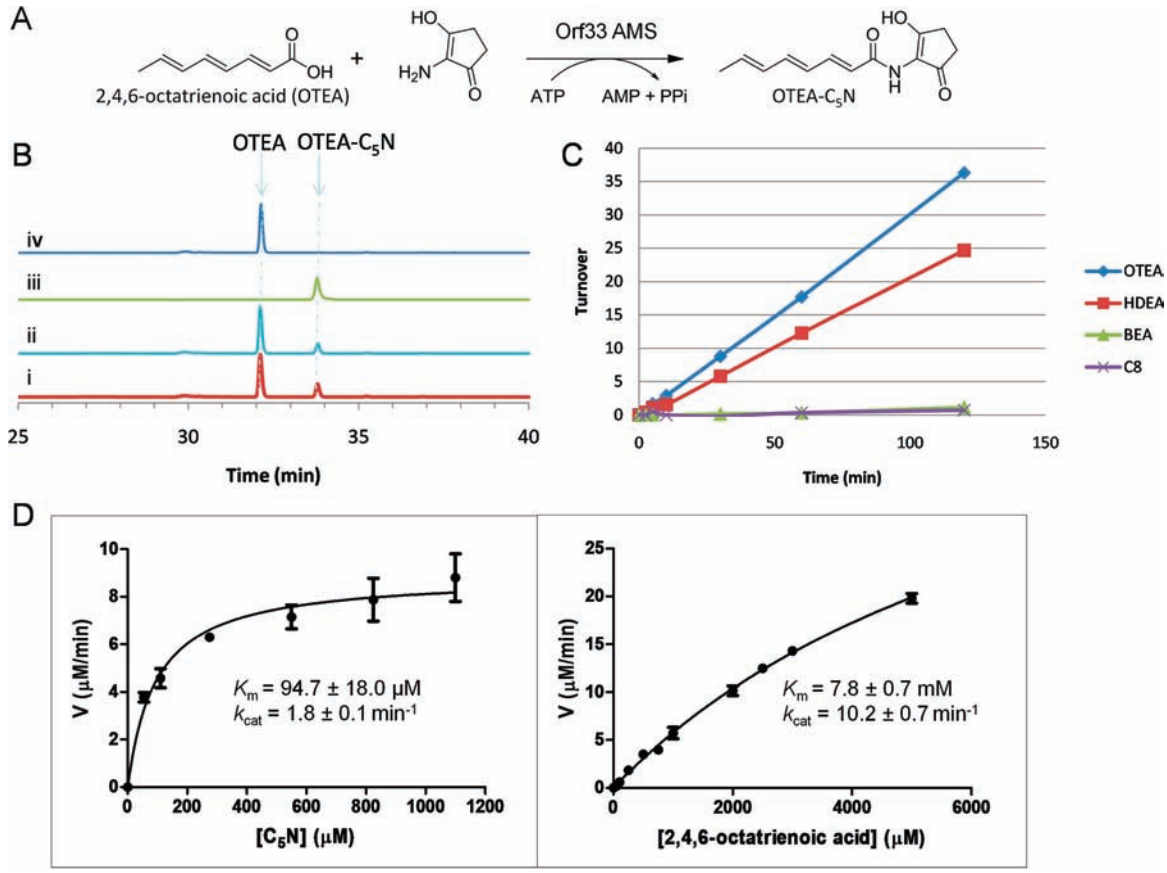

Figure 6. Characterization of ORF33 as an amide synthetase (AMS). (A) Schematic of ligation of C_5N to 2,4,6-octatrienoic acid (OTEA) catalyzed by ORF33. (B) UV traces (321 nm) during LC-MS analysis showing OTEA-C_5N formation using ORF33 and synthetic C_5N (trace i), the entire reconstituted pathway using ORF33−35, glycine, succinyl-CoA and CoA (trace ii). OTEA-C_5N synthetic standard is shown in trace iii and no enzyme control is shown in trace iv. It is notable that all of the no enzyme controls, including the control without ORF33 for trace i and the controls without either of the three enzymes for trace ii, were the same as the one shown in trace iv. (C) ATP-[^{32}P]PP_i exchange assays monitoring the reversible formation of acyl-AMP as the first half reaction of ORF33. Calculated rates: k(OTEA) = 0.30 min^{-1}, k(2,4-hexadienoic acid [HDEA]) = 0.21 min^{-1}, k(2-butenoic acid [BEA]) = 0.009 min^{-1}, k(octanoic acid [C8]) = 0.005 min^{-1}. (D) Kinetic parameters measured for ORF33 by monitoring the formation of OTEA-C_5N at 355 nm. ORF33 concentration was maintained at 5 μM and concentrations of OTEA and C_5N were fixed at 2 mM and 1 mM, respectively. Error bars represent standard deviations from at least three independently performed experiments.

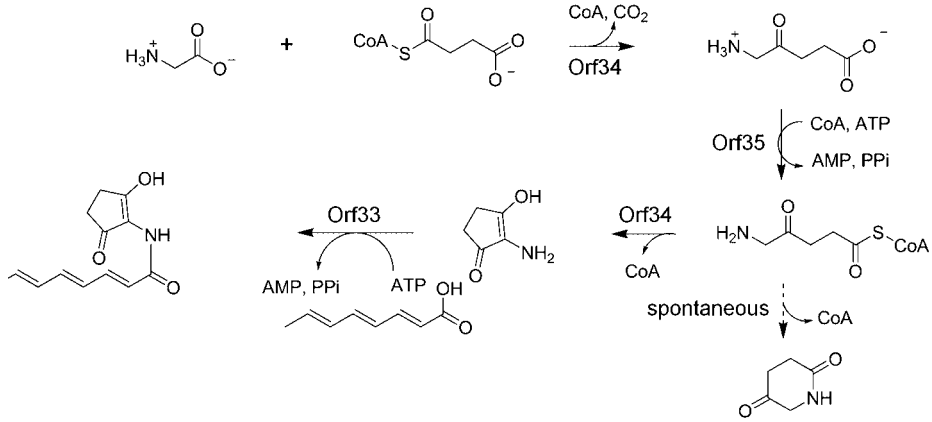

Figure 7. Characterized biosynthetic pathway for C_5N involving ORF33−35 from ECO-02301 gene cluster.

Alternatively, ALA-CoA can be directed by the PLP-dependent ORF34 to form C_5N through promotion of a carbon nucleophile at C5. Finally, ORF33 (AMS) catalyzes ligation of C_5N to a polyenoic acid in an ATP-dependent manner (Figure 7). We suspect that this three enzyme pathway constitutes a general biosynthetic route for C_5N formation and incorporation into other natural products. This supposition is supported by the presence of the same trigene cassette encoding ALAS, ACL, and AMS in reported gene clusters of C_5N-containing natural products, including ECO-0501 and moenomycin.[15,31]

The discovery of ALAS-catalyzed cyclization of ALA-CoA to C_5N solves the mechanistic problem regarding the formation of C_5N from ALA. Interestingly, the ALAS enzyme catalyzes

(31) Banskota, A. H.; McAlpine, J. B.; Sorensen, D.; Ibrahim, A.; Aouidate, M.; Piraee, M.; Alarco, A. M.; Farnet, C. M.; Zazopoulos, E. *J. Antibiot.* **2006**, *59*, 533–542.

two distinct reactions in this pathway: (1) ALA synthesis from glycine and succinyl-CoA, and (2) cyclization of ALA-CoA to afford C_5N. ALAS belongs to the family of á-oxoamine synthases and, because of its long known role in providing building blocks for heme biosynthesis, its PLP-dependent catalytic mechanism has been studied extensively. Catalysis involves two bond cleavage steps (C_α-H and C_α-COO) and one bond forming step (between C_α and the thioester carbonyl of succinyl-CoA).[32] To initiate catalysis, binding of glycine induces transaldimination, leading to the formation of Gly−PLP aldimine complex after deprotonation of the C_α-H_R of glycine by the ε-NH_2 of a lysine residue in the active site. This generates a C_α−PLP carbanion stabilized as the para-quinonoid species. The electrophilic thioester carbonyl is then attacked by the C_α carbanion of the activated Gly−PLP, which yields a proposed α-NH_2-β-ketoadipyl−PLP intermediate upon C−C bond formation from a thio-Claisen type condensation with concomitant release of CoASH. The subsequent decarboxylation of the β-keto acid via C_α-COO^- cleavage and C_α reprotonation lead to the formation of PLP bound ALA that is finally released by transaldimination with the active site lysine (Figure 8A).

We propose a related mechanism for the novel cyclization of ALA-CoA (Figure 8B). The CoA moiety of ALA-CoA is assumed to be a key recognition element, and after its binding in the well-defined CoA binding site,[33] the terminal -NH_2 of ALA-CoA can reach the PLP-Lys internal aldimine and undergo transaldimination (see Supporting Information Figure S12). The C5 proton of the ALA moiety can be abstracted (presumably by the same lysine residue) as proposed for the Gly−PLP intermediate in the first step of ALA formation.[30] Now the C5-ALA−PLP stabilized carbanion can intramolecularly attack the thioester carbonyl, releasing CoASH as a leaving group. The immediate result is the cyclization of the ALA chain to the C_5N group in aldimine linkage to the PLP in the active site. Transaldimination by lysine would release free C_5N and regenerate the resting form of the enzyme. Compared to ALA synthesis by this same PLP enzyme, cyclization of ALA-CoA does not have the decarboxylation step, and the nucleophilic attack is intramolecular instead of intermolecular, which may account for the observed higher catalytic efficiency of the enzyme for cyclization. This is an exquisite example of a PLP-dependent mechanism being employed by Nature to sequester and deactivate the reactive terminal amine of ALA and to simultaneously activate the adjacent carbon to generate a competent carbon nucleophile.

X-ray structures of the *R. capsulatus* ALAS have been recently reported, showing the positions of both Gly−PLP and of succinyl-CoA.[33] The electron density for the 3′,5′-ADP portion of the CoA and the carboxylate of the succinyl moiety was well resolved, allowing the placement of the C_α of the Gly−PLP adduct within 3 Å of the thioester carbonyl when modeled into the active site. With the use of the solved ALAS structure as the template (PDB: 2BWN), the structure of ORF34 was modeled. It showed high similarity to and shared almost identical active site with ALAS from *R. capsulatus* (Supporting Information Figure S12). ALA-CoA was docked into the active site of ORF34 with the CoA adenosine binding site constrained at the same location as that of succinyl-CoA in the *R. capsulatus* ALAS structure. In the modeled ALA-CoA−PLP complex, C5 of the ALA moiety was positioned approximately 3 Å from

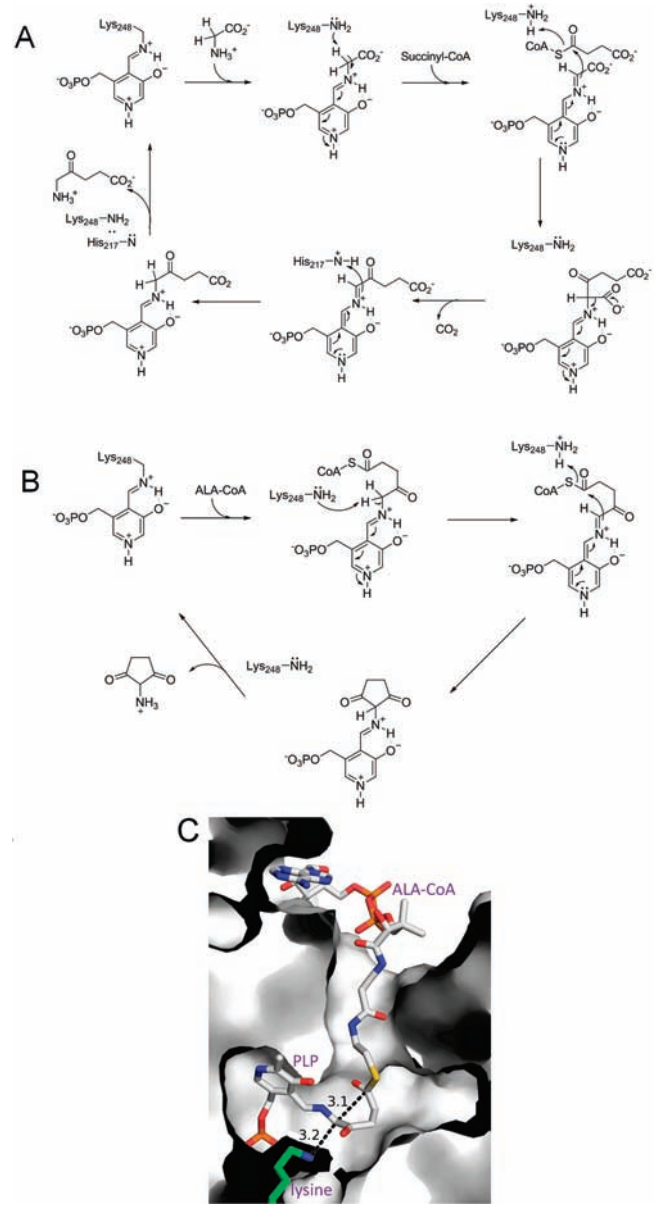

Figure 8. Proposed PLP-dependent catalytic mechanism of ORF34. (A) Mechanism for ALA synthesis. In the substrate-free state, PLP is bound by enzyme (internal aldimine). The incoming glycine induces transaldimination, leading to the PLP−Gly adduct (external aldimine). The electrophilic substrate succinyl-CoA is attacked by the PLP-activated glycine, leading to C−C bond formation with the release of CoASH. Decarboxylation and reprotonation result in formation of the PLP−ALA adduct that is finally released to regenerate the internal aldimine. (B) Mechanism for C_5N synthesis. ALA-CoA induces transaldimination, yielding the ALA-CoA−PLP complex. The thioester carbonyl in ALA-CoA is then attacked by the PLP-activated ALA moiety intramolecularly, leading to C−C bond formation with the release of CoASH. The PLP−C_5N complex is finally released to regenerate internal aldimine. (C) Three-dimensional structural model of the ORF34 active site. The structure was modeled on the *Rhodobacter capsulatus* ALAS using the online program HHpred, and the substrates were docked using the software GOLD.

Lys_{248} and the thioester carbonyl, demonstrating that the ALA moiety can readily fit into the active site pocket for cyclization while retaining the same CoA adenosine binding site at the protein surface (Figure 8C). On the basis of the high similarity of protein structures, we further propose that the "cyclase" activity of the ALAS may be general for all of the ALASs, including those from primary metabolism involved in heme

(32) Zhang, J. S.; Ferreira, C. *J. Biol. Chem.* **2002**, *277*, 44660–44669.
(33) Astner, I.; Schulze, J. O.; van den Heuvel, J.; Jahn, D.; Schubert, W. D.; Heinz, D. W. *EMBO J.* **2005**, *24*, 3166–3177.

biosynthesis. Such an activity would, however, be revealed only in a cell where ALA has been processed to the ALA-CoA thioester.

The two other enzymes in the pathway (ORF35 and ORF33) are both confirmed to be AMP-forming enzymes. ORF35 catalyzes the ligation between acid and CoA substrates to form the C−S thioester bond, while ORF33 generates the connectivity between acid and amine substrates to afford the C−N amide bond found in the natural product. ORF35 showed substrate promiscuity toward various acids and demonstrated higher specificity toward fatty acids than ALA, indicating that the enzyme might have evolved from a fatty acid acyl-CoA ligase. In contrast, ORF33 had low affinity for all the acid substrate analogs tested, suggesting the strong preference of the enzyme toward longer chain polyketide acids. In addition, unlike many AMP-forming enzymes including acyl-CoA ligases and adenylation domains of NRPSs which activate acids to yield acyl-AMP intermediates independently in the first half reaction, ORF33 activates acids much more efficiently upon the binding of the amine, suggesting that C_5N binding likely induces conformation change to render the enzyme in a more favorable state for acid activation.

In addition to the manumycin family compounds, many bioactive metabolites, including reductiomycin (antitumor),[34] senacarcin (antitumor),[35] limocrocin (reverse transcriptase inhibitor),[10] virustomycin (antivirus),[36] enopeptin (antibiotic)[37] and Sch 725424 (antibiotic)[11] also contain the C_5N ring system. Although there has been no direct study of the structure−activity relationship of C_5N to date, the prevalence of this functionality in natural products with a broad range of biological activities suggests a critical role in efficacy. This is especially apparent in the cases such as limocrocin, Sch 725424 and reductiomycin (Figure 1), where C_5N is the major functional group in these structures. It is possible that C_5N serves as a Michael acceptor for covalent inactivation of enzymes, which is a common inhibition mechanism of many other natural and synthetic molecules, including resorcylic acid lactones and acrylamides.[38,39] Our work presented here successfully closes the knowledge gap around the biosynthesis of C_5N, reveals a novel cyclization activity of ALA synthase, and defines a minimal, presumably portable cassette for the formation and amide-mediated incorporation of this unique pharmacophore.

Acknowledgment. This work was supported in part by an NIH Grant GM20011 (C.T.W.) and NIH Grant GM066174 (D.K.). We thank Dr. Albert Bowers for help with HRMS measurements. We also thank Prof. Sheryl Tsai and Dr. Brian Ames for assistance with protein structure modeling and substrates docking.

Supporting Information Available: SDS-PAGE of ORF33-35 purified from *E. coli*; UV profile of purified ORF34; titration of ORF34 with PLP; detection of ALA formation by OPTA derivatization; HRMS of ALA-CoA; determination of ORF35 kinetic parameters for various acid substrates; HPLC traces of ORF35 reactions with selected acid substrates; spectroscopic characterization of C_5N; UV profile of 2,4,6-octatrienoic acid (OTEA); spectroscopic characterization of OTEA-C_5N; ^1H and ^{13}C NMR characterization of synthesized OTEA-C_5N standard; modeled three-dimensional structure of ORF34. This material is available free of charge via the Internet at http://pubs.acs.org.

JA1002845

(34) Cho, H.; Beale, J. M.; Graff, C.; Mocek, U.; Nakagawa, A.; Omura, S.; Floss, H. G. *J. Am. Chem. Soc.* **1993**, *115*, 12296–12304.
(35) Laursen, J. B.; Nielsen, J. *Chem. Rev.* **2004**, *104*, 1663–1685.
(36) Omura, S.; Imamura, N.; Hinotozawa, K.; Otoguro, K.; Lukacs, G.; Faghih, R.; Tolmann, R.; Arison, B. H.; Smith, J. L. *J. Antibiot.* **1983**, *36*, 1783–1786.
(37) Koshino, H.; Osada, H.; Yano, T.; Uzawa, J.; Isono, K. *Tetrahedron Lett.* **1991**, *32*, 7707–7710.
(38) Schirmer, A.; Kennedy, J.; Murli, S.; Reid, R.; Santi, D. V. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4234–9.
(39) Deng, W.; Guo, Z.; Guo, Y.; Feng, Z.; Jiang, Y.; Chu, F. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 469–72.