A Secondary Metabolic Enzyme Functioned as an Evolutionary Seed of a Primary Metabolic Enzyme

Jun Kawaguchi, Hikaru Mori, Noritaka Iwai, and Masaaki Wachi (ID)*

Department of Life Science and Technology, Tokyo Institute of Technology, Yokohama 226-8501, Japan

*Corresponding author: E-mail: mwachi@bio.titech.ac.jp.
Associate editor: Miriam Barlow

Abstract

The antibiotic alaremycin has a structure that resembles that of 5-aminolevulinic acid (ALA), a universal precursor of porphyrins, and inhibits porphyrin biosynthesis. Genome sequencing of the alaremycin-producing bacterial strain and enzymatic analysis revealed that the first step of alaremcyin biosynthesis is catalysed by the enzyme, AlmA, which exhibits a high degree of similarity to 5-aminolevulinate synthase (ALAS) expressed by animals, protozoa, fungi, and α-proteobacteria. Site-directed mutagenesis of AlmA revealed that the substitution of two amino acids residues around the substrate binding pocket transformed its substrate specificity from that of alaremycin precursor synthesis to ALA synthesis. To estimate the evolutionary trajectory of AlmA and ALAS, we performed an ancestral sequence reconstitution analysis based on a phylogenetic tree of AlmA and ALAS. The reconstructed common ancestral enzyme of AlmA and ALAS exhibited alaremycin precursor synthetic activity, rather than ALA synthetic activity. These results suggest that ALAS evolved from an AlmA-like enzyme. We propose a new evolutionary hypothesis in which a non-essential secondary metabolic enzyme acts as an 'evolutionary seed' to generate an essential primary metabolic enzyme.

Key words: 5-aminolevulinate synthase, 5-aminolevulinic acid, ancestral sequence reconstruction, antibiotics, primary metabolic enzyme, secondary metabolic enzyme.

Introduction

Antibiotics are substances produced by micro-organisms antagonistic to the growth of other micro-organisms. The origin of antibiotics is ancient and some antibiotic production occurred over hundreds of million years ago (Wright 2007). Moreover, antibiotic resistance genes occurred over 2 billion years ago, indicating that antibiotic production also existed at that time (Hall and Barlow 2004; D'Costa et al. 2011). It is an interesting question to know whether antibiotics influenced the evolution of ancient life.

Antibiotics' production is often considered 'secondary metabolism' because it is unessential for the growth of the organism. In the natural environment, antibiotics kill or prevent the growth of competitive organisms. Thus, the role of antibiotics is to inhibit the growth of competitors; however, sub-lethal levels of antibiotics provide various effects on organisms, such as transcriptional modulation or enhancement of genetic variability (Goh et al. 2002; Couce and Blázquez 2009; Kohanski et al. 2010; Gutierrez et al. 2013; Bleich et al. 2015). Here we report a novel evolutionary role for antibiotic metabolism as a secondary metabolic enzyme that contributes to the origination of a primary metabolic enzyme.

Previously, we discovered a novel antibiotic, "alaremycin", from the soil bacterium *Streptomyces* sp. A012304 (Awa et al. 2005) (fig. 1). The molecular structure of alaremycin is related to 5-aminolevulinic acid (ALA); therefore, it was designated alaremycin, from "ALA related antibiotic" produced by *Streptomyces* sp. ALA is a universal precursor of porphyrin in almost all living organisms and is converted to several important cofactors, such as heme and chlorophyll (Stojanovski et al. 2019; Bryant et al. 2020) (fig. 1). Alaremycin competitively binds to the substrate binding pocket of porphobilinogen synthase, which condenses two molecules of ALA to produce porphobilinogen for porphyrin biosynthesis (Heinemann et al. 2010). In our attempt to identify the biosynthetic genes of alaremycin, we unexpectedly found that one of the alaremycin synthetic enzymes is a close relative of ALA synthase (ALAS).

ALAS synthesizes ALA by condensing glycine and succinyl-co-enzyme A (CoA). This single-step pathway for ALA biosynthesis, known as the Shemin pathway, occurs in animals, protozoa, fungi and α-proteobacteria (Shemin and Rittenberg 1946; Stojanovski et al. 2019). In humans, ALAS catalyses the rate-limiting step for heme biosynthesis and its dysfunction is responsible for serious diseases, such as X-linked sideroblastic anemia (XLSA) and X-linked protoporphyria (XLPP) (Balwani 2019; Peoc'h et al. 2019; Stojanovski et al. 2019). Most bacteria, archaea and plants synthesize ALA through an alternative pathway known as the C_5 pathway (Jahn et al. 1992;
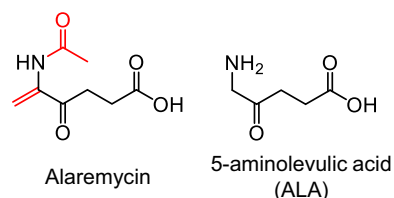
Open Access

Article

Fɪɢ. 1. Structures of alaremycin and ALA. The unique structures in alaremycin are highlighted in red.

Brzezowski et al. 2015; Layer 2021). This pathway includes three-step reactions beginning with glutamate (Nogaj and Beale 2005). Interestingly, the diversity of ALA biosynthetic pathways is in contrast to the porphyrin biosynthetic pathway wherein most steps are highly conserved among nearly all living organisms, although some procaryotes have been known to possess alternative pathways (Dailey et al. 2017). Since the C_5 pathway is widely distributed in prokaryotes, it is considered a primitive pathway for ALA synthesis, whereas the Shemin pathway is modern. However, it is still unknown how and why the Shemin pathway developed and replaced the established C_5 pathway in some ancient organisms.

In the process of elucidating the alaremycin biosynthetic pathway, we discovered a significant similarity and phylogenetic relationship between ALAS and an alaremycin synthetic enzyme named AlmA. This led us to the idea that ALAS may have originated from an AlmA-like enzyme. We surmised that a feature of an antibiotic enzyme, which is non-essential but advantageous for survival, may enable an antibiotic enzyme to serve as an excellent evolutionary seed for a novel enzyme. This hypothesis contends that antibiotic synthetic enzymes play a role in enhancing evolution in a unique way that has not been previously considered.

Here, we describe the results of genome sequencing of the alaremycin producer strain and subsequently characterize the alaremycin synthetic enzymes. Then, we demonstrate an evolutionary relationship between the alaremycin synthetic enzyme, AlmA and ALAS by two distinct approaches: a functional transformation of AlmA to ALAS by site-directed mutagenesis and the characterization of a predicted AlmA/ALAS common ancestor by ancestral sequence reconstruction (ASR). We discussed the evolutionary trajectory of ALAS and the evolutionary role of secondary metabolic enzymes, including AlmA-like proteins.

Results and Discussion

Identification of the Biosynthetic Gene Cluster of Alaremycin

The draft genome sequencing of the alaremycin producer strain, *Streptomyces* sp. A012304, was performed to identify gene(s) involved in the biosynthesis of alaremycin. Therefore, we obtained 3.93 Gbp and 38.6 million reads, and 36.3 million reads were applied for an assembly. Consequently, we established a draft genome of 9.67 Mbp. The number of contigs was 281, and the average length

was 34.4 kbp. The minimum and maximum length of the contigs was 131 bp and 432 kb, respectively. A total of 28,768 open reading frames (ORFs) with a length of 30 a.a. or more were found, of which 10,407 ORFs were positive by basic local alignment search tool (BLAST) analysis. Of these, we selected two ALA synthetic gene homologues as candidates for the biosynthetic gene of alaremycin. We assumed that either one is responsible for ALA synthesis, whereas the other is involved in alaremycin biosynthesis. One showed homology with the prokaryotic *hemA* genes encoding glutamyl-tRNA reductase within the C_5 pathway of ALA biosynthesis. The other exhibited homology with the *ALAS* genes encoding ALAS within the Shemin pathway. This gene was later named *almA*. Although most bacteria do not contain ALAS, some *Streptomyces* species have a homologue enzyme, known as cyclic ALAS (cALAS), which catalyses the cyclization of ALA to produce the C_5N unit moiety found in several antibiotics (Petrícek et al. 2006; Zhang et al. 2010; Petříčková et al. 2015; Hanh et al. 2018). Thus, we initially suspected that AlmA is a member of the cALAS family. However, the phylogenetic tree comprised of ALAS, cALAS and AlmA showed that the branch of AlmA is out of the cALAS clade and is closer to the ALAS clade (fig. 2A). This phylogenetic topology suggests that AlmA has a distinct function compared with that of cALAS. Next, we attempted to determine which one was responsible for ALA biosynthesis to exclude the ALA synthetic gene and to identify the alaremycin synthetic gene.

We performed a complementation test of the *Escherichia coli* ΔhemA mutant strain MGHA01, which is deficient in the C_5 pathway because of a lack of *hemA* and shows ALA auxotrophy. As shown in fig. 2B, MGHA01 carrying a plasmid vector did not grow on an L agar plate in the absence of ALA. Under these conditions, a plasmid carrying the *hemA* homologue of A012304 restored the growth of MGHA01 in the absence of ALA, whereas that carrying *almA* did not. This result clearly indicates that the *hemA* homologue is responsible for ALA synthesis in A012304 as with most bacteria. This is also supported by the observation that the A012304 genome contains a *hemL* gene, which is an essential component of the C_5 pathway. Additionally, several genes involved in heme biosynthesis are also found downstream of the *hemA* homologue, that is, *hemB*, *hemC*, and *hemD*, which encode porphobilinogen synthase, porphobilinogen deaminase and uroporphyrinogen III synthase, respectively (fig. 2C). It was suggested that the *ALAS* homologue is involved in alaremycin biosynthesis. Notably, three serial genes were found downstream of the *ALAS* homologue, annotated as *N*-acetyltransferase, oxidoreductase and major facilitator superfamily (MFS) transporter, respectively (fig. 2D), which may be responsible for alaremycin biosynthesis together with the *ALAS* homologue. To test this, the gene cluster was introduced into *E. coli* JM109 and the ability of the cells to produce alaremycin was examined. As shown in fig. 2E, high-performance liquid chromatography (HPLC) followed by MS analysis revealed that the *E. coli*
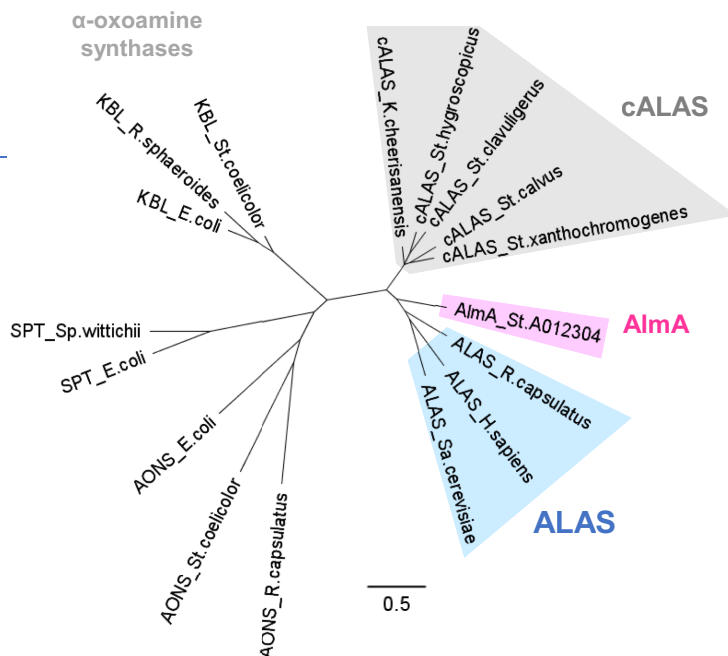
A

Abbreviations

α-Oxoamine synthases
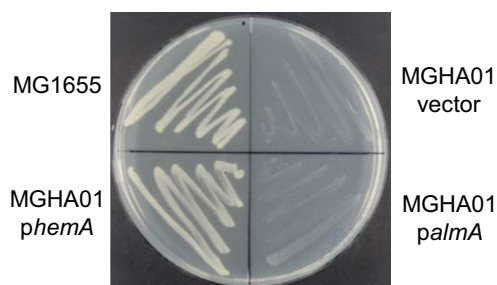
AONS   8-Amino-7-oxononanoate synthase
KBL   Glycine C-acetyltransferase
SPT   Serine C-palmitoyltransferase

Genus

A   Amycolatopsis
E   Escherichia
H   Homo
K   Kitasatospora
R   Rhodobacter
Sa   Saccharomyces
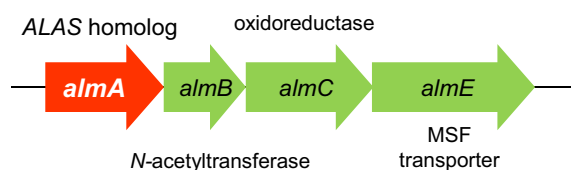Sp   Sphingomonas
Ss   Streptosporangium
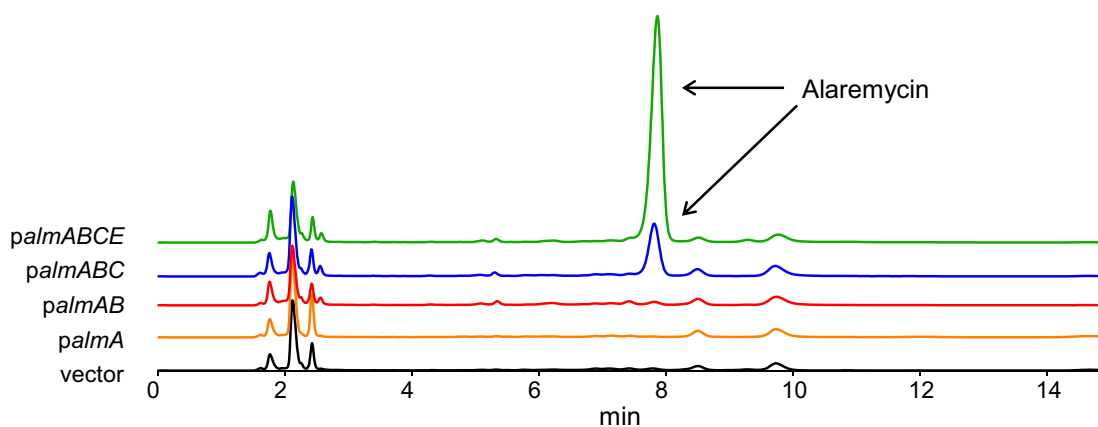St   Streptomyces

B

C

D

E

FIG. 2. Identification of the biosynthetic gene cluster of alaremycin. (A) Amino acid sequence-based phylogenetic tree, which comprises ALAS, cyclic ALAS (cALAS), AlmA and α-oxoamine synthases. Abbreviations of α-oxoamine synthases and the organism's genus are shown at the left. The clades of ALAS and cALAS are highlighted in blue and gray, respectively, and AlmA is highlighted in magenta. The tree was constructed by the method based on Bayesian inference. (B) Complementation of the E. coli ΔhemA mutant by hemA and ALAS homologues (almA) of A012304. The plasmid carrying hemA or almA gene was introduced into the ALA auxotrophic E. coli MGHA01. E. coli wild-type strain MG1655 was used as a positive control. (C) Gene alignment of the hemA homologue and its flanking region. The hemB, hemC and hemD genes encode porphobilinogen synthase, porphobilinogen deaminase and uroporphyrinogen III synthase, respectively. (D) Gene alignment of the ALAS homologue and its flanking region. These four genes are named almA (ALAS homologue), almB (N-acetyltransferase), almC (oxidoreductase) and almE (MFS transporter). (E) Production of alaremycin by E. coli JM109 transformed with the plasmid palmABC and palmABCE. Ethyl acetate extracts of the culture supernatants were analysed by HPLC as described in the Materials and Methods section.
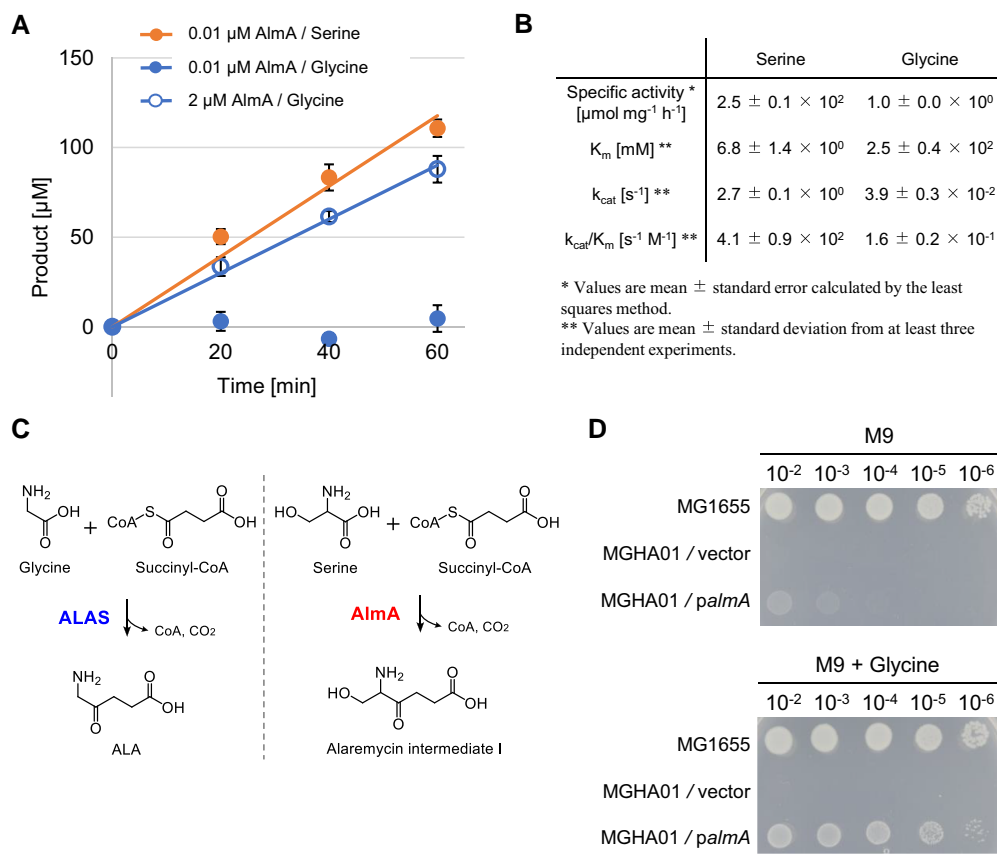
A

- 0.01 µM AlmA / Serine
- 0.01 µM AlmA / Glycine
- 2 µM AlmA / Glycine

B

	Serine	Glycine
Specific activity * [µmol mg^{-1} h^{-1}]	$2.5 \pm 0.1 \times 10^2$	$1.0 \pm 0.0 \times 10^0$
K_m [mM] **	$6.8 \pm 1.4 \times 10^0$	$2.5 \pm 0.4 \times 10^2$
k_{cat} [s^{-1}] **	$2.7 \pm 0.1 \times 10^0$	$3.9 \pm 0.3 \times 10^{-2}$
k_{cat}/K_m [s^{-1} M^{-1}] **	$4.1 \pm 0.9 \times 10^2$	$1.6 \pm 0.2 \times 10^{-1}$

* Values are mean $\pm$ standard error calculated by the least squares method.
** Values are mean $\pm$ standard deviation from at least three independent experiments.

C

D

Fig. 3. Enzymatic activity of AlmA. (*A*) Time course analysis and kinetic parameters of His-tagged AlmA. The activity against serine was measured at 0.01 µM His-AlmA and against glycine at 0.01 µM and 2 µM His-AlmA. Solid lines show the slope of the time course reaction for glycine calculated using the least squares method. (*B*) Kinetic parameters of AlmA as determined by the Michaelis–Menten equation. (*C*) Catalytic reactions of ALAS and that of the ALAS homologue enzyme AlmA (see also supplementary fig. S2). (*D*) Effect of glycine on complementation of MGHA01 by the p*almA* plasmid. Cells grown in L medium were washed with saline and spotted after dilution on an M9 agar plate (top) or M9 agar plate supplemented with 1-mM glycine (bottom).

JM109 strain expressing these four genes produced alaremycin extracellularly. When the MFS transporter homologue was deleted, extracellular levels of alaremycin decreased, suggesting that it functions as an alaremycin exporter and may contribute to self-resistance. Conversely, *E. coli* strains expressing *almA* encoding the ALAS homologue and its downstream gene encoding *N*-acetyltransferase or *almA* did not produce alaremyicn. Based on these results, we concluded that *almA* and the two genes encoding *N*-acetyltransferase and oxidoreductase are required for alaremycin biosynthesis. The last gene encoding the MFS transporter is required for alaremycin extrusion. Accordingly, these four genes were designated *almA*, *almB*, *almC*, and *almE* (fig. 2D).

Enzymatic Reaction Catalysed by AlmA

In the biosynthetic pathway of alaremycin, it is expected that the AlmA catalyses the condensation of succinyl-CoA and amino acid. Considering the structural similarity between ALA and alaremycin, amino acids having one-carbon side chain may serve as a substrate, such as alanine or serine. To confirm the enzyme reaction catalysed by AlmA, we conducted an *in vitro* enzyme assay using purified AlmA protein. AlmA was fused with a His-tag on its N-terminus, and the His-tagged AlmA protein was purified. Next, its enzymatic activity was measured using glycine, alanine, serine and threonine as a substrate (fig. 3A and B). Therefore, the

AlmA efficiently catalysed the condensation of serine and succinyl-CoA. Conversely, AlmA showed no detectable activity for alanine or threonine (supplementary fig. S1, Supplementary Material online). The K_m value for serine is 6.8 ± 1.4 mM, which is comparable to that of *Rhodopseudomonas palustris* ALAS for glycine (2.31 ± 0.12 mM) (Tan et al. 2019) and *Mus musculus* ALAS for glycine (8 ± 0.7 mM) (Stojanovski et al. 2016). This indicates that AlmA adopts serine as a substrate. Moreover, the k_{cat} value of AlmA for serine and succinyl-CoA condensation is 2.7 ± 0.1 s^{-1}, which is almost the same as that of the *Rp. palustris* ALAS for glycine and succinyl-CoA condensation (2.74 ± 0.28 s^{-1}) (Tan et al. 2019). Meanwhile, the k_{cat} value of AlmA for serine and succinyl-CoA condensation is 10-fold higher than that of *M. musculus* ALAS for glycine (0.25 ± 0.004 s^{-1}) (Stojanovski et al. 2016). These kinetic parameters clearly indicate that AlmA catalyses the condensation of serine and succinyl-CoA to synthesize 5-amino-6-hydroxy-4-oxohexanoic acid, which is the intermediate I of alaremycin (fig. 3C). Considering the catalytic reaction of AlmA and the annotated function of AlmB and AlmC, we have proposed a biosynthetic pathway and efflux mechanism for alaremycin (shown in supplementary fig. S2, Supplementary Material online).

We also examined the activity of AlmA for glycine, which is a substrate of ALAS. Unexpectedly, AlmA also exhibited a slight but significant activity for glycine at higher enzyme concentrations (2 µM). The K_m and k_{cat} values for

glycine were $2.5 \pm 0.4 \times 10^2$ mM and $3.9 \pm 0.3 \times 10^{-2}$ s^{-1}, respectively, which is 37-fold higher and 69-fold lower than that for serine. These results suggest that AlmA can synthesize ALA, albeit with lower efficiency. This is supported by the complementation test for the ALA auxotrophic MGHA01. A plasmid carrying the *almA* gene slightly restored the growth of MGHA01 on the M9 minimal agar plate in the absence of ALA and it was greatly improved by supplementation with 1 mM glycine (fig. 3D), indicating that AlmA indeed synthesizes ALA *in vivo* using glycine as a substrate.

Site-directed Mutagenesis on AlmA towards ALAS

The fact that AlmA exhibits ALA synthesis activity led us to the idea that ALAS may have evolved from an AlmA-like protein. As described in the Introduction, it is likely that the Shemin pathway emerged in an ancestral species of α-proteobacteria after the C$_5$ pathway had already been distributed throughout the bacterial domain. We inferred that the AlmA-like antibiotic-synthesizing enzyme, which produces ALA as a by-product, developed into a robust ALA synthetic enzyme and was adopted as a novel ALA synthetic pathway within the ancestral species of α-proteobacteria.

To estimate the evolutionary relationship of AlmA and ALAS, we first compared the amino acid sequences of AlmA and several ALAS from α-proteobacteria, yeast and human, focussing on the amino acid residues composing the substrate binding pocket (fig. 4A; supplementary fig. S3, Supplementary Material online). We found that the 82nd residue of AlmA is Ser, but the corresponding residue is conserved as Thr among all ALAS. This residue corresponds to the 83rd Thr of *R. sphaeroides* ALAS, which is responsible for the amino acid substrate selectivity (Shoolingin-Jordan et al. 2003). Furthermore, the S83T replacement in cALAS of *S. nodosus* subsp. *asukaensis* reportedly resulted in the loss of its ability to produce a C$_5$N unit containing antibiotic asukamycin, suggesting that the 83rd residue is responsible for its enzymatic activity (Rui et al. 2010). The 3D structure of *R. capsulatus* ALAS (RcALAS: Protein Data Bank [PDB] number 2BWP) and that of AlmA constructed by 3D homology modeling show that these amino acid residues are close to the substrate amino acid, glycine. Distances between the substrate and the amino acid residues (4.9 Å in AlmA and 3.6 Å in RcALAS) suggest that the space of the substrate binding pocket of AlmA is larger than that of ALAS (fig. 4B). This may explain why AlmA prefers the bulkier substrate serine over glycine. Additionally, we focused on the 86th Gly of AlmA wherein the corresponding residues among ALAS are conserved as bulkier Ser or Ala. This residue is located in a glycine-rich loop, which is important for succinyl-CoA binding (Astner et al. 2005). Based on these findings, we suspected that the 82nd Ser and 86th Gly of AlmA are responsible for substrate specificity and are related to the divergence of AlmA and ALAS. To confirm the effect of these two residues on substrate specificity, we constructed five AlmA variants by introducing site-specific substitutions: S82T,

G86A, G86S, S82T/G86A and S82T/G86S. We examined their ability to synthesize ALA *in vivo*.

We performed the complementation test for the *E. coli* Δ*hemA* mutant MGHA01 to simply evaluate the ALA synthetic activity of these AlmA variants. As shown in fig. 4C, the *almA*-S82T, *almA*-G86S, *almA*-S82T/G86A and *almA*-S82T/G86S genes restored the growth of MGHA01 more effectively than the original *almA* gene on an M9 minimal agar plate in the absence of ALA. Especially, growth recovery was markedly improved for the *almA*-S82T/G86S gene. These results suggest that S82T/G86S simultaneous substitution is the most responsive to the change in substrate specificity of AlmA.

Next, we performed an *in vitro* enzyme assay with AlmA-S82T and AlmA-S82T/G86S. As shown in fig. 4D, AlmA-S82T exhibited a markedly decreased specific activity for serine compared with the original AlmA, which was only 0.3% of the original AlmA activity. Unexpectedly, the specific activity for glycine did not increase but slightly decreased to 31% of that of the original AlmA. In the case of the double substituent AlmA-S82T/G86S, the activity for serine was similar to the single substituent, AlmA-S82T, but the activity of glycine increased compared with AlmA-S82T to a level comparable to the original AlmA.

The results of the *in vitro* enzyme assay indicate that the single substitution of S82T caused a drastic decrease in activity for serine but only a marginal decrease for glycine. Consequently, AlmA-S82T lost most of its synthetic activity for the intermediate I but still retained weak ALA synthetic activity. The additional G86S substitution increased the affinity for glycine, while maintaining low affinity for serine. Consequently, AlmA-S82T/G86S produced ALA more efficiently than AlmA-S82T. The behaviour in complementation tests may be explained by these *in vitro* results. The antibacterial activity of 5-amino-6-hydroxy-4-oxohexanoic acid (intermediate I) was reported by Perlman et al. (1981). The improvement of growth recovery by *almA*-S82T resulted from the reduction of toxic intermediate I production though ALA production was also marginally reduced. Additional G86S substitution improved growth by producing more ALA. This is the reason why AlmA-S82T exhibited improved growth recovery compared with AlmA and why AlmA-S82T/G86S showed better growth recovery than AlmA-S82T.

In conclusion, we revealed that the 82nd and 86th residues of AlmA are determinants of whether it produces intermediate I or ALA. The fact that only two mutations transform the function of AlmA towards ALAS suggests that AlmA could have evolved into ALA synthase.

ASR of AlmA and ALAS

Using a BLAST search, we found that AlmA homologues are distributed among several species or strains of Streptomyces, Bacillus and γ-Proteobacteria. They usually utilize the C$_5$ pathway to synthesize ALA. We found that 82nd and 86th residues were highly conserved as Ser and Gly, respectively, among AlmA homologues (supplementary

A


```
                              82   86
AlmA    St. sp. A012304    AMHRAVDEYG AGSGGSRNIG GTHAHYARLE SSLAEWHGKE AALVFPTGYG 116
     ⎧  Rb. capsulatus      AMHEALEAVG AGSGGTRNIS GTTAYHRRLE AEIADLHQKE AALVFSSAYN 117
     │  Rb. sphaeroides     AMHEALDSTG AGSGGTRNIS GTTLYHKRLE AELADLHGKE AALVFSSAYI 117
ALAS ⎨  Rp. palustris       AMVETTTRIG TGAGGTRNIA GTHHPLVQLE QEIASLHGKE AALLFTSGYV 116
     │  Sa. cerevisiae      AMHKTIDKYG CGAGGTRNIA GHNIPTLNLE AELATLHKKE GALVFSSCYV 184
     ⎩  H. sapiens          ATQETLQRHG AGAGGTRNIS GTSKFHVELE QELAELHQKD SALLFSSCFV 260
                            *        *  * ** *** *           **    *   *  *     **
```
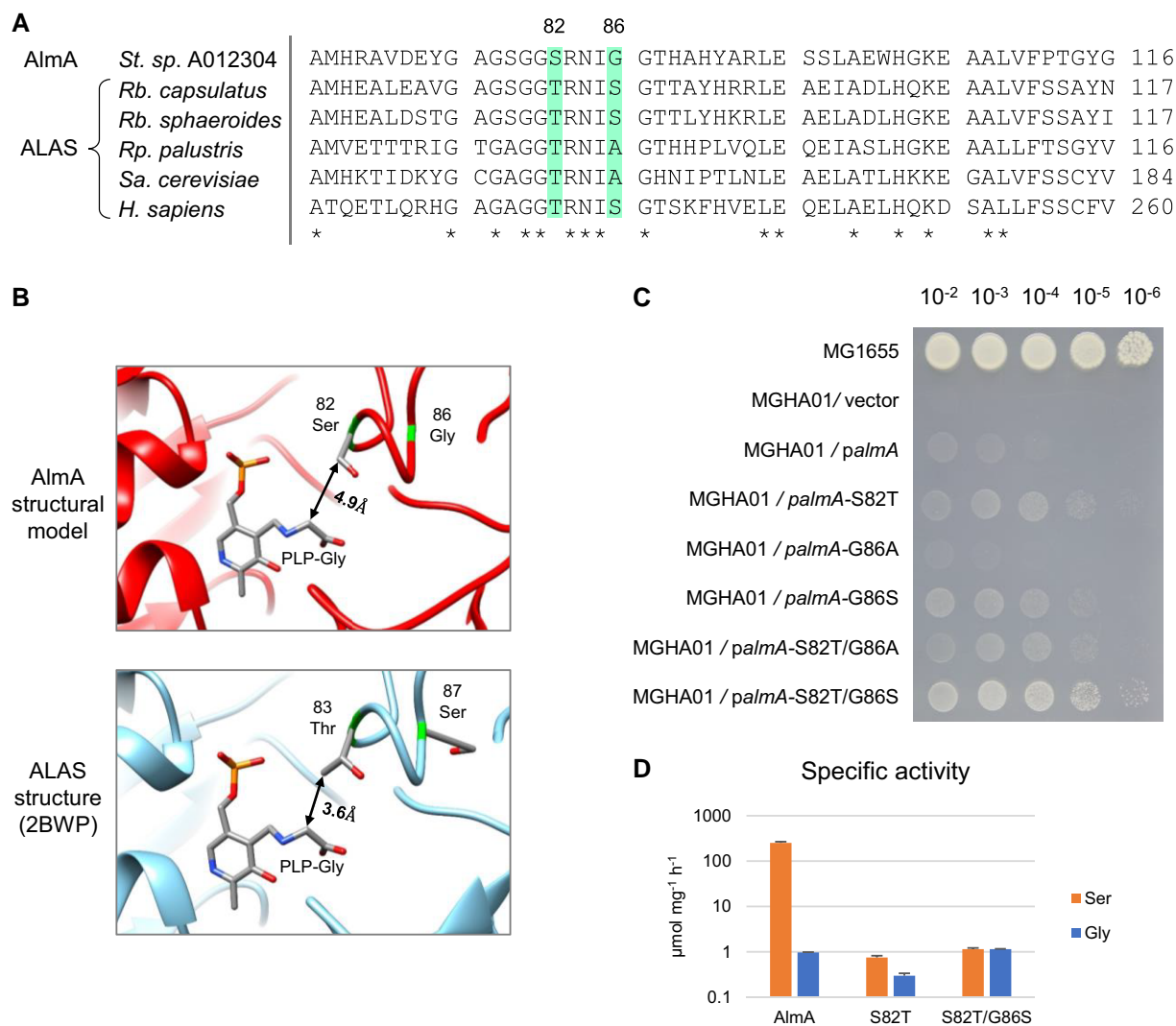
B

C

D

FIG. 4. Effects of site-specific mutations at the 82nd and 86th amino acid residues in AlmA on the complementation ability of MGHA01. (A) Sequence alignment of AlmA and ALAS from various species. The genus abbreviations are *St: Streptomyces, Rb: Rhodobacter, Rp: Rhodopseudomonas, Sa: Saccharomyces* and *H: Homo*. The residues corresponding to 82nd and 86th of AlmA are highlighted in green. Asterisks represent conserved residues. (B) Comparison of active sites between ALAS from *Rb. capsulatus* (PDB ID: 2BWP) and AlmA. 3D structure of AlmA was predicted by SWISS-MODEL using ALAS from *Rb. capsulatus* as a template. The shortest distance between the 82nd residue and the substrate glycine is shown by arrows. (C) Complementation of MGHA01 by AlmA variants. Cells grown in L medium were washed with saline and spotted after dilution on an M9 agar plate. (D) Specific activities of His-AlmA and its variants. Specific activity of His-AlmA against serine was measured using 0.01 μM enzyme and the others with 2 μM.

fig. S4, Supplementary Material online). Thus, these AlmA homologues may produce the alaremycin intermediate I or other ALA analogues in a manner similar to AlmA. These AlmA homologues enabled us to perform an ASR analysis. ASR is a phylogenetic analysis to estimate the extinct ancestral sequences (Jermann et al. 1995; Thornton 2004). ASR has been successfully applied to reveal the evolutionary trajectory of a protein of interest (Voordeckers et al. 2012; Risso et al. 2013; Blanquart et al. 2021).

To estimate the evolutionary trajectory of AlmA and ALAS, we performed an ASR analysis. The phylogenetic tree was constructed from 65 amino acid sequences: 31 of AlmA homologues and 31 of ALAS from α-Proteobacteria (supplementary Table S1, Supplementary Material online). We used only α-proteobacterial ALAS for ASR because

α-Proteobacteria are the most primitive taxa among the organisms utilizing the Shemin pathway and their ALAS are probably closer to the common ancestor of AlmA/ALAS. Three cALAS sequences were used as an out-group (supplementary Table S1, Supplementary Material online). Each sequence was collected from broad orders of organisms: AlmA homologues from nine orders and ALAS from eight orders of α-proteobacteria. The constructed tree revealed three large clades: AlmA homologues, ALAS and an out-group. We reconstructed three ancestral sequences: 1) ancestral AlmA (ancAlmA) as an ancestral node of the AlmA homologues clade; 2) ancestral ALAS (ancALAS) as an ancestral node of ALAS clade; and 3) a common ancestor of AlmA and ALAS (CA) as a common ancestral node of ancAlmA and ancALAS (fig. 5A). The degree of sequence
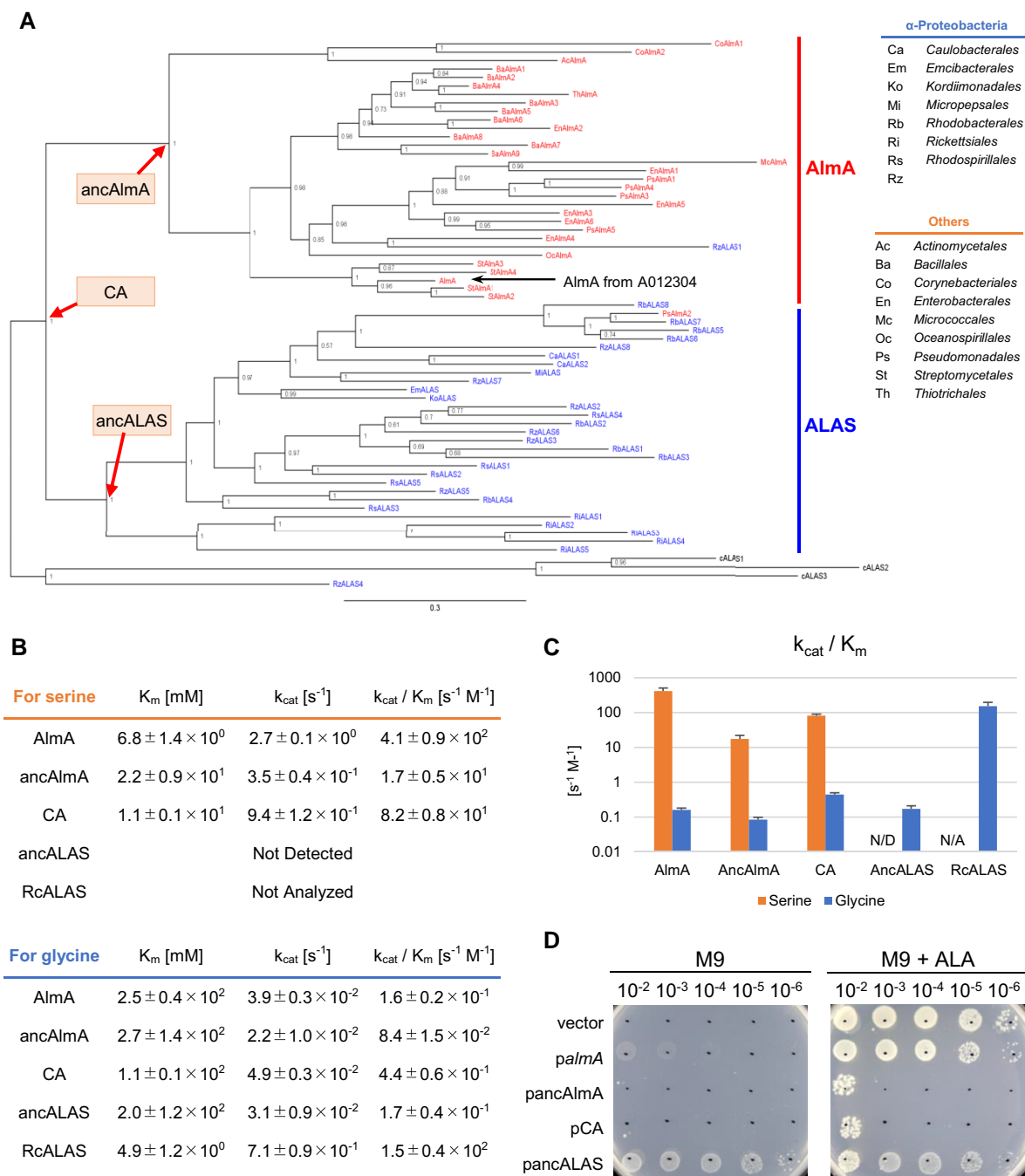
FIG. 5. Kinetic parameters of reconstructed ancestral enzymes and modern AlmA and ALAS. (*A*) Phylogenetic tree based on the amino acid sequences of the AlmA homologues and ALAS. Accession numbers and class of the host of each sequence are summarized in supplementary Table S1, Supplementary Material online. The organism orders are described by two letters, and the abbreviations are shown on the right side. ALAS from α-proteobacteria are shown in blue, and AlmA are shown in orange. Cyclic ALAS (cALAS) sequences are used as the out-group. Reconstructed ancestral nodes are indicated by red arrows, and AlmA from A012304 is indicated by a black arrow. (*B*) Kinetic parameters of AlmA, RcALAS and reconstructed enzymes as determined by the Michaelis–Menten equation. All enzymes were his-tagged on the N-terminus, and detailed conditions of the enzymatic assay are described in the Materials and Methods section. (*C*) Graphical representation of k_{cat}/K_m values of each enzyme. Error bars show the standard deviation calculated from at least three experiments. (*D*) Complementation or growth suppression of MGHA01 by reconstructed genes. Cells grown in L medium were washed with saline and spotted after dilution on an M9 agar plate (left) and an M9 agar plate supplemented with 5 μg/mL ALA (right).

B

For serine	K_m [mM]	k_{cat} [s^{-1}]	k_{cat} / K_m [s^{-1} M^{-1}]
AlmA	$6.8 \pm 1.4 \times 10^0$	$2.7 \pm 0.1 \times 10^0$	$4.1 \pm 0.9 \times 10^2$
ancAlmA	$2.2 \pm 0.9 \times 10^1$	$3.5 \pm 0.4 \times 10^{-1}$	$1.7 \pm 0.5 \times 10^1$
CA	$1.1 \pm 0.1 \times 10^1$	$9.4 \pm 1.2 \times 10^{-1}$	$8.2 \pm 0.8 \times 10^1$
ancALAS		Not Detected	
RcALAS		Not Analyzed	

For glycine	K_m [mM]	k_{cat} [s^{-1}]	k_{cat} / K_m [s^{-1} M^{-1}]
AlmA	$2.5 \pm 0.4 \times 10^2$	$3.9 \pm 0.3 \times 10^{-2}$	$1.6 \pm 0.2 \times 10^{-1}$
ancAlmA	$2.7 \pm 1.4 \times 10^2$	$2.2 \pm 1.0 \times 10^{-2}$	$8.4 \pm 1.5 \times 10^{-2}$
CA	$1.1 \pm 0.1 \times 10^2$	$4.9 \pm 0.3 \times 10^{-2}$	$4.4 \pm 0.6 \times 10^{-1}$
ancALAS	$2.0 \pm 1.2 \times 10^2$	$3.1 \pm 0.9 \times 10^{-2}$	$1.7 \pm 0.4 \times 10^{-1}$
RcALAS	$4.9 \pm 1.2 \times 10^0$	$7.1 \pm 0.9 \times 10^{-1}$	$1.5 \pm 0.4 \times 10^2$

similarity between AlmA and the reconstituted ancestral enzymes was 62%–66%, whereas that between RcALAS and the reconstituted ancestral enzymes was 59%–61%. Meanwhile, the ancestral enzymes exhibited a greater extent of sequence similarity at 87%–93% between each other. Notably, the 82nd and 86th residues of ancAlmA and CA are the AlmA type, that is Ser and Gly, respectively, whereas those of ancALAS are the ALAS type, that is Thr and Ser, respectively (supplementary fig. S5, Supplementary Material online).

We purified these three ancestral enzymes and determined their kinetic parameters *in vitro* (fig. 5B and C). RcALAS was used as a representative for the existing ALAS. Regarding ancAlmA and CA, their K_m values were smaller for serine than for glycine, whereas their k_{cat} values were larger for serine than for glycine. Thus, the k_{cat}/K_m values for ancAlmA and CA indicate a greater catalytic efficiency for serine. The values for serine were 17 ± 5 and 82 ± 8 s^{-1}M^{-1}, respectively, and those for glycine were 0.084 ± 0.015 and 0.44 ± 0.06 s^{-1}M^{-1}, respectively. Therefore, ancAlmA and CA prefer serine to glycine as a substrate and efficiently produce intermediate I. Conversely, ancALAS exhibited catalytic activity only with glycine but not with serine. The k_{cat}/K_m value for glycine of ancALAS (0.17 ± 0.04 s^{-1}M^{-1}) was only 0.11% of that of RcALAS (150 ± 40 s^{-1}M^{-1}), which resulted from a larger K_m value and a smaller k_{cat} value. High concentrations of RcALAS (2 μM) showed a slight but significant activity using serine as a substrate, suggesting that RcALAS can synthesize intermediate I as a by-product (supplementary fig. S6, Supplementary Material online). The apparent specific activity for serine calculated from the initial reaction rate was 1.2 μmol mg^{-1} h^{-1}, which is 0.48% of that of AlmA for serine (Figs. 3B and 4D). However, the intermediate I production by RcALAS showed non-linearity probably because of product inhibition. We were unable to adopt the Michaelis–Menten equation to determine the kinetic parameters of RcALAS with serine.

We also performed a complementation test for MGHA01 to examine the *in vivo* activity of these reconstructed enzymes *in vivo*. The ancALAS restored the growth of MGHA01 on the M9 agar plate without ALA, whereas ancAlmA and CA did not (fig. 5D). Notably, ancAlmA and CA severely suppressed the growth of MGHA01 on the M9 ager plate even in the presence of ALA (fig. 5D). The growth recovery by ancALAS may be simply explained by its ALA synthetic activity. Meanwhile, severe growth inhibition by ancAlmA and CA probably resulted from the toxicity of intermediate I. However, the reason why ancAlmA and CA suppressed growth more severely than the original AlmA will require further analysis.

According to the results of the kinetic assay and complementation test, we concluded that ancAlmA and CA may be categorized as AlmA-like enzymes, which primarily synthesize intermediate I with slight ALA synthesis activity, whereas the ancALAS may be categorized as ALAS, although its activity was much lower than the existing ALAS, such as RcALAS. Overall, our ASR analysis strongly suggests that ALAS originated from an AlmA-like antibiotic enzyme.

Probable Evolutionary Trajectory of AlmA and ALAS

In this section, we discuss how ALAS has evolved from an AlmA-like enzyme by combining our site-directed mutagenesis results and ASR analysis. Site-directed mutagenesis on AlmA revealed that the 82nd and 86th residues affect substrate specificity. The kinetic assay showed that the S82T/G86S substitution on AlmA resulted in a loss of function for the intermediate I synthetic activity, whereas it still sustained ALA synthetic activity (fig. 4D). The highest growth recovery by AlmA-S82T/G86S indicated that its low ALA synthetic activity was sufficient to restore the growth of the ALA auxotrophic *E. coli* when intermediate I synthetic activity was low enough (fig. 4C). A similar tendency was observed in case of CA (82nd Ser/86th Gly) and ancALAS (82nd Thr/86th Ser). The ancALAS lost the intermediate I synthetic activity substantially, whereas it exhibited ALA synthetic activity comparable with CA (figs. 5B and C). Consequently, ancALAS showed significant growth recovery in the complementation test.

In summary, we suggest the following evolutionary model to show how ALAS evolved from an AlmA-like enzyme (fig. 6). First, an ancestral bacterium having the C$_5$ pathway acquired an AlmA-like enzyme as an antibiotic-synthesizing enzyme. Then, the ancestor of ALAS appeared from an AlmA-like enzyme through mutations that resulted in the loss of intermediate I synthetic activity while sustaining ALA synthetic activity, and thus an ancestral α-proteobacterial organism acquired ancestral ALAS. Afterwards, the ancestral α-proteobacteria lost the C$_5$ pathway, whereas ancestral ALAS with low ALA synthetic
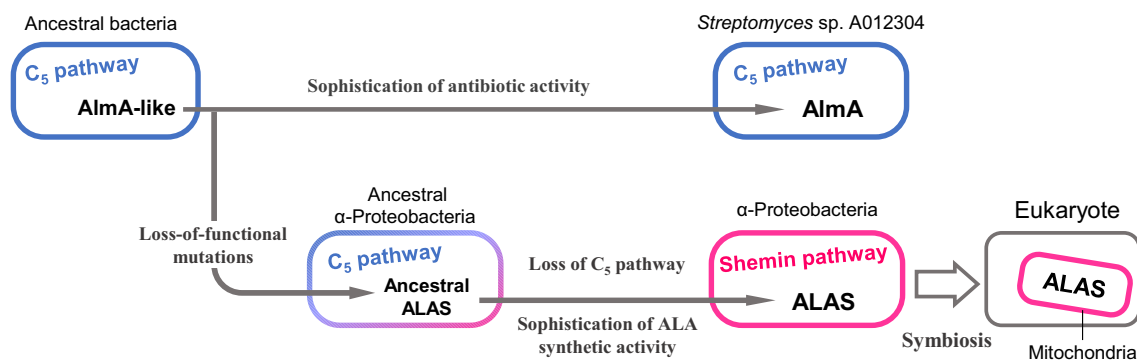

FIG. 6. Hypothetical model for the evolutionary trajectory of AlmA and ALAS. Boxes indicate bacterial organisms and the colors of their frames and words show whether they utilize the C$_5$ pathway (blue) or Shemin pathway (magenta). Evolutionary events are represented by gray arrows and gray words. A thick open arrow with a gray frame indicates the symbiotic event.

activity was able to develop an efficient ALA synthetic enzyme through mutations and natural selection. Thereafter, ALAS may have been inherited to an ancestral eukaryotic organism via symbiosis, when ancestral eukaryotic organisms acquired mitochondria. This is based on reports that suggest that α-proteobacterial symbiosis is the origin of mitochondria (Zimorski et al. 2014; Roger et al. 2017) and human ALAS localizes within the mitochondria (Munakata et al. 2004; Nomura et al. 2021). Overall, our evolutionary model provides new insight into how the divergence of the ALA biosynthetic pathway occurred and how the AlmA-like antibiotic synthase contributed to the origin of ALAS.

This hypothesis raises a question why ALAS holders are limited to α-proteobacteria in the bacterial domain. Unfortunately, our results were not sufficient to answer this question. The α-proteobateria involve obligate intracellular parasites, such as *Rickettsia* and *Chlamydia*, and photosynthetic bacteria such as *Rhodobacter*. The Shimin pathway is possibly advantageous for these organisms to survive in their habitats; alternatively, the C_5 pathway is disadvantageous. However, this point needs further investigation.

Hypothesis that Antibiotic Synthase Contributes to Primary Metabolic Enzyme Origination

We examined the phylogenetic relationship of AlmA and ALAS and demonstrated their evolutionary trajectory, which indicates that a nonessential AlmA-like enzyme is the probable origin of the essential ALAS. In this case, the AlmA-like enzyme had acted as an "evolutionary seed," which created a novel enzyme. We assume that the antibiotic-synthesizing enzyme, like AlmA, may be an excellent evolutionary seed for two reasons. First, since the antibiotic gene is unessential for the host, it is capable of accepting radical mutations that cause a loss of function. Secondly, antibiotics are often substrate analogues, which inhibit a target enzyme by binding to its substrate binding pocket competitively. Thus, such analogue synthesizing enzymes may be easily converted to a novel enzyme producing the substrate itself, which may be how the AlmA-like enzyme evolved to ALAS. If an evolved enzyme is dominant to a primitive one, it may be incorporated as a new part of a biosynthetic pathway.

Here, we propose a hypothesis that non-essential secondary metabolic enzymes act as evolutionary seeds to generate essential primary metabolic enzymes. This would represent a novel role for antibiotics that have contributed to the evolution of life. Further studies may uncover the origin of other primary metabolic genes based on this novel evolutionary principle.

Materials and Methods

Culture Media

E. coli cells were grown in L broth (1% polypeptone, 0.5% yeast extract, 0.5% NaCl and 0.1% glucose, pH 7.2) or M9 minimal medium (6.8 g/L Na_2HPO_4, 3.0 g/L KH_2PO_4, 0.5 g/L NaCl, 1.0 g/L NH_4Cl, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$ and 4 g/L glucose). For agar plates, 1.5% agar was added to the medium. For culturing the ALA auxotrophic MGHA01, 50 μg/mL ALA was included. For culturing plasmid-containing strains, 50 μg/mL ampicillin or 20 μg/mL kanamycin was added.

Draft Genome Sequencing

The genomic DNA of the alaremycin-producing strain, *Streptomyces* sp. A012304, was extracted using the phenol–chloroform extraction method. Draft genome sequencing was performed by Hokkaido System Science Co., Ltd. using Illumina Hiseq with the Paired-End method. *De novo* assembly was performed using the assembly programme, Velvet (Zerbino and Birney 2008). ORF prediction was performed using the programme GetORF constituted in Jemboss.

Plasmid Construction

Plasmids for complementation tests and HPLC analysis were constructed by ligating DNA fragments containing each gene with pTrc99A in which the trc promoter induces overexpression of cloned genes when isopropyl β-D-1-thiogalactopyranoside (IPTG) is added to the culture media. DNA fragments from *Streptomyces* sp. A012304 were polymerase chain reaction (PCR) amplified using genomic DNA as a template. DNA fragments containing reconstructed ancestral genes were excised from the plasmids prepared by Eurofins.

Plasmids for His-tagged enzyme overproduction were constructed using the pET28b(+) vector. Ligated DNA fragments containing the AlmA and RcALAS genes were PCR amplified using genomic DNA of A012304 or *Rhodobacter capsulatus* as a template, respectively, and all other fragments were PCR amplified using pTrc99A-based plasmids as a template. All purified proteins were designed to be fused with a His-tag at the N-terminus.

Site-directed mutagenesis was performed by overlapping PCR (Ho et al. 1989) using point-substituted primer sets. Upstream and downstream sequences from the mutated site were PCR amplified with a 20–30 nt overlap, respectively, then the amplified DNA fragments were mixed and subjected to second round PCR. pT-*almA* was used for the PCR-template to construct single substituted *almA*, and pT-*almA*-S82T was used for the PCR-template to construct the double substituted, *almA*.

The primer sets for PCR and restriction sites for cloning are shown in supplementary Table S2, Supplementary Material online. All plasmid insertions were confirmed by sequencing. All constructed plasmids were introduced into *E. coli* JM109 by electroporation. Plasmids were purified using a FastGene Plasmid Mini Kit (NIPPON Genetics).

Complementation Test

The *E. coli* Δ*hemA* mutant strain, MGHA01, was constructed by P1 phage-mediated transduction using a wild-

type K-12 strain MG1655 as a recipient and an *hemA*::Km[R] mutant FN102 (Fukai et al. 2003) as a donor. Plasmids for the complementation assay were introduced into MGHA01 using electroporation. A complementation test was performed by streaking overnight cultures of transformants, which were grown on L broth. For complementation tests on M9 media, overnight cultures grown on L broth were washed with saline. Serial dilutions of the cell suspension were spotted onto an M9 agar plate. Then, 1 mM IPTG was added for gene induction. Each plate was incubated at 37°C overnight for the L plate or 48 h for the M9 plate, and the resulting growth of the transformants was monitored.

Extraction of Alaremycin, HPLC and MS Analysis

The *E. coli* strain JM109 carrying plasmids expressing the *almA*, *almAB*, *almABC* or *almABCE* genes were cultivated in M9 medium at 30°C for 4 days. Gene expression was induced by adding 1 mM IPTG at 0.3–0.5 OD_{660}. Cells were removed via centrifugation and the supernatants of the culture medium were collected. The supernatant was extracted by repeated solvent dissolution with ethyl acetate/water (1:1, v/v) at pH 2, 9 and 2. The resultant organic phase was evaporated to remove solvent and the extract was resolved in acetonitrile. Samples were then subjected to HPLC analysis using a Prominence system (Shimadzu) equipped with a COSMOSIL HILIC packed column (nacalai tesque). The separation was done with a mobile phase of acetonitrile/30 mM ammonium acetate (4:1, v/v) at a flow rate of 1 ml/min. Alaremycin was detected by absorption at 253 nm. Mass spectrometry analysis was performed using electrospray time-of-flight mass spectrometry (ESI-TOF-MS) using a micrOTOF II (Bruker). For sample preparation, the HPLC fraction around 7.9 min was collected and extracted with ethyl acetate, pH 2, and the resulting organic phase was evaporated. ESI-TOF-MS was performed in positive mode.

Protein Purification

The *E. coli* strain BL21(DE3), carrying a plasmid expressing the protein of interest, was grown in 100 or 150 ml of L broth at 30°C for 2 h. Then, 0.5 mM IPTG was added and the culture was continued for another 22 h at 20°C. The cells were collected by centrifugation and then lysed by sonication in 300 mM KCl, 50 mM KH_2PO_4 and 5 mM imidazole. The lysates were centrifuged at 15,300 g for 20 min to remove insoluble materials. His-tagged proteins were purified using the Profinia system (Bio-Rad). His-tagged proteins were eluted with 300 mM KCl, 50 mM KH_2PO_4 and 250 mM imidazole. Eluted fractions were desalted with 20 mM HEPES-NaOH, pH 7.5. The protein concentrations were determined by a Bradford protein assay kit and the protein solutions were adjusted to 10 μM with 20 mM HEPES-NaOH pH 7.5, 20 μM PLP and 10% glycerol. The purified proteins were stored at −30°C. The purity of the proteins was confirmed by sodium dodecyl sulphate–polyacrylamide gel electrophoresis.

Enzymatic Assay

The ALA synthesis activity assay was performed using the spectrophotometric method reported by Hunter and Ferreira (1995) with minor modifications. The reaction solution was 20 mM HEPES-NaOH, pH 7.5, containing 20 μM PLP, 3 mM $MgCl_2$, 250 μM thiamine pyrophosphate, 1 mM α-ketoglutarate, 1 mM nicotinamide adenine dinucleotide (NAD^+), 0.25 U/ml α-ketoglutarate dehydrogenase complex (Sigma), amino acid substrate, succinyl-CoA and purified protein. In this assay, free CoA is simultaneously released with ALA or intermediate I of alaremycin production. Free CoA and NAD^+ are then converted to succinyl-CoA and NADH by α-ketoglutarate dehydrogenase, and the NADH concentration is determined by its specific absorption at 340 nm. The amino acids tested included glycine, serine, alanine and threonine. The reactions were performed at 37°C and the absorbance was measured by a spectrophotometer (NanoDrop 2000, Thermo). The molar extinction coefficient of NADH is $6560 \ M^{-1}cm^{-1}$. Reaction mixtures without amino acids were used as controls.

For kinetic parameter determination, the reaction time was fixed at 60 min. Enzyme concentrations were adjusted according to their activity levels to maintain a surplus supply of succinyl-CoA by α-ketoglutarate dehydrogenase: 0.01 μM AlmA for serine, 0.1 μM ancAlmA for serine, 0.05 μM CA for serine, 0.05 μM. RcALAS for glycine and all others were 2 μM. Kinetic parameters were determined by nonlinear regression fitting using GraphPad Prizm9.

3D Homology Modeling and Site-Directed Mutagenesis

The 3D homology modeling of AlmA was done by SWISS-MODEL (Waterhouse et al. 2018) using the RcALAS 3D structure (PDB id: 2BWP) as a template. Amino acid sequences were aligned by ClustalW. The 3D structures were treated using UCSF Chimera to create images or to calculate atomic distance.

Sequence Data Collection

Amino acid sequences used for phylogenetic analysis were collected from the NCBI RefSeq database (https://www.ncbi.nlm.nih.gov/refseq/), which comprised non-redundant sequences, to avoid strain level redundancies. cALAS sequences used in fig. 2A were selected from bacterial species, known as producers of antibiotics containing a C_5N unit and previously discussed in phylogenetic analysis by Petříčková et al. (2015). ALAS sequences used in fig. 5A were obtained using its EC number "2.3.1.37" as a query and filtering only α-proteobacterial proteins. To avoid complex calculations, similar sequences indicated by a neighbour joining tree were removed, and finally, 31 ALAS sequences were selected. AlmA homologue sequences were obtained by BLAST search using the AlmA sequence as a query. Very similar sequences were also removed, and 30 AlmA homologous sequences were selected. Three cALAS sequences were selected as an

out-group. Data collection was performed from July to August 2020.

Phylogenetic Tree Construction

Selected sequences for phylogenetic tree construction were aligned by ClustalW. The amino acid based phylogenetic trees were constructed by Bayesian inference using MrBayes 3.2.7 (Ronquist et al. 2012). The adopted evolutionary model was gamma distribution and the amino acid substitutional model was WAG (Whelan and Goldman 2001). The Markov chain Monte Carlo runs were performed for 0.4 million generations for the tree shown in fig. 2A and 2 million generations for the tree in fig. 5A, with sampling at every 500 generations. After deletion of one-fourth burn-in, consensual topologies were computed from sampled trees. Trees were depicted using FigTree software.

ASR

ASR was performed using PAML software (phylogenetic analysis by maximum likelihood) 4.9j (Yang 2007). The adopted evolutionary model was gamma distribution, and the amino acid substitutional model was WAG. Reconstructed sequences were aligned, and extra sequences, which were not conserved in most of the extant sequences, were removed. The codons of the reconstructed amino acid sequences were optimized for *E. coli*. The whole nucleotide sequences were artificially synthesized by Eurofins and the sequences subcloned into the pEX-K4J2 vector. These plasmids were introduced into *E. coli* JM109 using electroporation.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Data Availability

Draft genome sequence of *Streptomyces* sp. A012304 was deposited to DDBJ (accession numbers BQUN 01000001-BQUN01000228).

References

Astner I, Schulze JO, van den Heuvel J, Jahn D, Schubert WD, Heinz DW. 2005. Crystal structure of 5-aminolevulinate synthase, the first enzyme of heme biosynthesis, and its link to XLSA in humans. *EMBO J*. **24**:3166–3177.

Awa Y, Iwai N, Ueda T, Suzuki K, Asano S, Yamagishi J, Nagai K, Wachi M. 2005. Isolation of a new antibiotic, alaremycin, structurally related to 5-aminolevulinic acid from *Streptomyces* sp. A012304. *Biosci Biotechnol Biochem*. **69**:1721–1725.

Balwani M. 2019. Erythropoietic protoporphyria and X-linked protoporphyria: pathophysiology, genetics, clinical manifestations, and management. *Mol Genet Metab*. **128**:298–303.

Blanquart S, Groussin M, Le Roy A, Szöllosi GJ, Girard E, Franzetti B, Gouy M, Madern D. 2021. Resurrection of ancestral malate dehydrogenases reveals the evolutionary history of Halobacterial proteins: deciphering gene trajectories and changes in biochemical properties. *Mol Biol Evol*. **38**:3754–3774.

Bleich R, Watrous JD, Dorrestein PC, Bowers AA, Shank EA. 2015. Thiopeptide antibiotics stimulate biofilm formation in Bacillus subtilis. *Proc Natl Acad Sci U S A*. **112**:3086–3091.

Bryant DA, Hunter CN, Warren MJ. 2020. Biosynthesis of the modified tetrapyrroles-the pigments of life. *J Biol Chem*. **295**: 6888–6925.

Brzezowski P, Richter AS, Grimm B. 2015. Regulation and function of tetrapyrrole biosynthesis in plants and algae. *Biochim Biophys Acta*. **1847**:968–985.

Couce A, Blázquez J. 2009. Side effects of antibiotics on genetic variability. *FEMS Microbiol Rev*. **33**:531–538.

Dailey HA, Dailey TA, Gerdes S, Jahn D, Jahn M, O'Brian MR, Warren MJ. 2017. Prokaryotic heme biosynthesis: multiple pathways to a common essential product. *Microbiol Mol Biol Rev*. **81**: e00048-00016.

D'Costa VM, King CE, Kalan L, Morar M, Sung WW, Schwarz C, Froese D, Zazula G, Calmels F, Debruyne R, et al. 2011. Antibiotic resistance is ancient. *Nature*. **477**:457–461.

Fukai Y, Nihei C, Kawai K, Yabu Y, Suzuki T, Ohta N, Minagawa N, Nagai K, Kita K. 2003. Overproduction of highly active trypanosome alternative oxidase in *Escherichia coli* heme-deficient mutant. *Parasitol Int*. **52**:237–241.

Goh EB, Yim G, Tsui W, McClure J, Surette MG, Davies J. 2002. Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. *Proc Natl Acad Sci U S A*. **99**:17025–17030.

Gutierrez A, Laureti L, Crussard S, Abida H, Rodríguez-Rojas A, Blázquez J, Baharoglu Z, Mazel D, Darfeuille F, Vogel J, et al. 2013. β-lactam antibiotics promote bacterial mutagenesis via an RpoS-mediated reduction in replication fidelity. *Nat Commun*. **4**:1610.

Hall BG, Barlow M. 2004. Evolution of the serine beta-lactamases: past, present and future. *Drug Resist Updat*. **7**:111–123.

Hanh NPK, Hwang JY, Oh HR, Kim GJ, Choi H, Nam DH. 2018. Biosynthesis of 2-amino-3-hydroxycyclopent-2-enone moiety of bafilomycin in *Kitasatospora cheerisanensis* KCTC2395. *J Microbiol*. **56**:571–578.

Heinemann IU, Schulz C, Schubert WD, Heinz DW, Wang YG, Kobayashi Y, Awa Y, Wachi M, Jahn D, Jahn M. 2010. Structure of the heme biosynthetic *Pseudomonas aeruginosa* porphobilinogen synthase in complex with the antibiotic alaremycin. *Antimicrob Agents Chemother*. **54**:267–272.

Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*. **77**:51–59.

Hunter GA, Ferreira GC. 1995. A continuous spectrophotometric assay for 5-aminolevulinate synthase that utilizes substrate cycling. *Anal Biochem*. **226**:221–224.

Jahn D, Verkamp E, Söll D. 1992. Glutamyl-transfer RNA: a precursor of heme and chlorophyll biosynthesis. *Trends Biochem Sci*. **17**: 215–218.

Jermann TM, Opitz JG, Stackhouse J, Benner SA. 1995. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. *Nature*. **374**:57–59.

Kohanski MA, DePristo MA, Collins JJ. 2010. Sublethal antibiotic treatment leads to multidrug resistance via radical-induced mutagenesis. *Mol Cell*. **37**:311–320

Layer G. 2021. Heme biosynthesis in prokaryotes. *Biochim Biophys Acta Mol Cell Res*. **1868**:118861.

Munakata H, Sun JY, Yoshida K, Nakatani T, Honda E, Hayakawa S, Furuyama K, Hayashi N. 2004. Role of the heme regulatory motif in the heme-mediated inhibition of mitochondrial import of 5-aminolevulinate synthase. *J Biochem*. **136**:233–238.

Nogaj LA, Beale SI. 2005. Physical and kinetic interactions between glutamyl-tRNA reductase and glutamate-1-semialdehyde aminotransferase of *Chlamydomonas reinhardtii*. *J Biol Chem*. **280**:24301–24307.

Nomura K, Kitagawa Y, Aihara M, Ohki Y, Furuyama K, Hirokawa T. 2021. Heme-dependent recognition of 5-aminolevulinate synthase by the human mitochondrial molecular chaperone ClpX. *FEBS Lett*. **595**:3019–3029.

Peoc'h K, Nicolas G, Schmitt C, Mirmiran A, Daher R, Lefebvre T, Gouya L, Karim Z, Puy H. 2019. Regulation and tissue-specific expression of δ-aminolevulinic acid synthases in non-syndromic sideroblastic anemias and porphyrias. *Mol Genet Metab*. **128**:190–197.

Perlman KL, Schömer U, Williams TH, Perlman D. 1981. Microbial production of vitamin B12 antimetabolites. IV. Isolation and identification of 4-keto-5-amino-6-hydroxyhexanoic acid. *J Antibiot (Tokyo)*. **34**:483–488.

Petrícek M, Petrícková K, Havlícek L, Felsberg J. 2006. Occurrence of two 5-aminolevulinate biosynthetic pathways in *Streptomyces nodosus* subsp. *asukaensis* is linked with the production of asukamycin. *J Bacteriol*. **188**:5113–5123.

Petříčková K, Chroňáková A, Zelenka T, Chrudimský T, Pospíšil S, Petříček M, Krištůfek V. 2015. Evolution of cyclizing 5-aminolevulinate synthases in the biosynthesis of actinomycete secondary metabolites: outcomes for genetic screening techniques. *Front Microbiol*. **6**:814.

Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM. 2013. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases. *J Am Chem Soc*. **135**:2899–2902.

Roger AJ, Muñoz-Gómez SA, Kamikawa R. 2017. The origin and diversification of mitochondria. *Curr Biol*. **27**:R1177–R1192.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. **61**:539–542.

Rui Z, Petríčková K, Skanta F, Pospíšil S, Yang Y, Chen CY, Tsai SF, Floss HG, Petríček M, Yu TW. 2010. Biochemical and genetic insights into asukamycin biosynthesis. *J Biol Chem*. **285**:24915–24924.

Shemin D, Rittenberg D. 1946. The biological utilization of glycine for the synthesis of the protoporphyrin of hemoglobin. *J Biol Chem*. **166**:621–625.

Shoolingin-Jordan PM, Al-Daihan S, Alexeev D, Baxter RL, Bottomley SS, Kahari ID, Roy I, Sarwar M, Sawyer L, Wang SF. 2003. 5-Aminolevulinic acid synthase: mechanism, mutations and medicine. *Biochim Biophys Acta*. **1647**:361–366.

Stojanovski BM, Breydo L, Uversky VN, Ferreira GC. 2016. Murine erythroid 5-aminolevulinate synthase: truncation of a disordered N-terminal extension is not detrimental for catalysis. *Biochim Biophys Acta*. **1864**:441–452.

Stojanovski BM, Hunter GA, Na I, Uversky VN, Jiang RHY, Ferreira GC. 2019. 5-Aminolevulinate synthase catalysis: the catcher in heme biosynthesis. *Mol Genet Metab*. **128**:178–189.

Tan Z, Zhao J, Chen J, Rao D, Zhou W, Chen N, Zheng P, Sun J, Ma Y. 2019. Enhancing thermostability and removing hemin inhibition of *Rhodopseudomonas palustris* 5-aminolevulinic acid synthase by computer-aided rational design. *Biotechnol Lett*. **41**:181–191.

Thornton JW. 2004. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat Rev Genet*. **5**:366–375.

Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, Verstrepen KJ. 2012. Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biol*. **10**:e1001446.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, *et al*. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. **46**:W296–W303.

Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. **18**:691–699.

Wright GD. 2007. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat Rev Microbiol*. **5**:175–186.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. **24**:1586–1591.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. **18**:821–829.

Zhang W, Bolla ML, Kahne D, Walsh CT. 2010. A three enzyme pathway for 2-amino-3-hydroxycyclopent-2-enone formation and incorporation in natural product biosynthesis. *J Am Chem Soc*. **132**:6402–6411.

Zimorski V, Ku C, Martin WF, Gould SB. 2014. Endosymbiotic theory for organelle origins. *Curr Opin Microbiol*. **22**:38–48.