Title

Identification, cloning and heterologous expression of biosynthetic gene cluster for desertomycin

Authors

Takuya Hashimoto[1], Ikuko Kozone[2], Junko Hashimoto[2], Hikaru Suenaga[1], Manabu Fujie[3], Noriyuki Satoh[3], Haruo Ikeda[4], Kazuo Shin-ya[1,5,6]*

[1]National Institute of Advanced Industrial Science and Technology, 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan

[2]Japan Biological Informatics Consortium, 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan

[3]Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna-son, Kunigami-gun, Okinawa 904-0495, Japan.

[4]Kitasato Institute for Life Sciences, Kitasato University, 1-15-1 Kitasato, Minami-ku, Sagamihara, Kanagawa 252-0373, Japan

[5]The Biotechnology Research Center, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

[6]Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

*E-mail: k-shinya@aist.go.jp

Abstract

From our in-house microbial genome database of secondary metabolite producers, we identified a candidate biosynthetic gene cluster for desertomycin from *Streptomyces nobilis* JCM4274. We report herein the cloning of the 127-kb entire gene cluster for desertomycin biosynthesis using bacterial artificial chromosome (BAC) vector. The entire biosynthetic gene cluster for desertomycin was introduced in the heterologous host, *Streptomyces lividans* TK23, with an average yield of more than 130 mg L^{-1}.

Screening for bioactive microbial secondary metabolites has been carried out for more than 70 years. Although the findings of bioactive natural products have resulted in great benefits to humankind, the rate of discovery of skeletally-novel compounds from microorganisms has significantly decreased over time (ref. 1, 2). Recent advances in sequencing technology gave us massive genome sequences, including biosynthetic gene clusters for microbial secondary metabolites, in public databases. However, most of those biosynthetic gene clusters remain unidentified to date. For the efficient discovery of novel compounds from such unused biosynthetic gene clusters, accumulation of datasets for the identified biosynthetic gene clusters, which can be used as reference databases such as MiBIG (ref. 3) and DoBuiscuit (ref. 4), is important. To accomplish this purpose, heterologous expression of the entire biosynthetic gene clusters is the most reliable way to characterize the exact biosynthetic gene clusters. We have determined bacterial genome sequences of over 100 microorganisms including actinomycetes, *Pseudomonas*, *Burkholderia*, *Bacillus*, etc., for genome mining of biosynthetic gene clusters. In this genome database, we focused on the YM-216391 producing-*Streptomyces nobilis* JCM4274, because its genome contains not only the biosynthetic gene cluster for YM-216391, but also several others for secondary metabolites including the candidate gene of desertomycin.

Desertomycins have been reported to show antifungal activities (ref. 5) and are members of marginolactone compounds, which contain amino or guanidino groups, and possess a ring size of 31 carbon atoms or more, such as mathemycin, primycin and kanchanamycin (ref. 6). The previous biosynthetic studies on desertomycin have only identified the last step of the biosynthesis catalyzed by DstH (ref. 7). The amidinohydrolase homologue encoded by *dstH* catalyzes the conversion of de-amidination of desertomycin B to desertomycin A during post-PKS tailoring process. Although the region encoding desertomycin biosynthesis was proposed in a previous study (ref. 7), the experimental validation of the entire biosynthetic gene cluster has not been examined due to the difficulties in cloning and transformation of such huge size DNA. We have established a technique for the cloning and heterologous expression of large biosynthetic gene clusters for type-I PKS compounds using bacterial artificial chromosome (BAC) vectors and have succeeded in the heterologous production of mediomycin, neomediomycin (ref. 8), JBIR-156 (ref. 9) and quinolidomicin (ref. 10) whose biosynthetic gene clusters are 161, 183, 137 and 213 kbp in length, respectively. Here we report the identification of the entire biosynthetic gene cluster for desertomycin spanning 127 kb region.

Based on the chemical structure of desertomycin A, we predicted its biosynthetic gene cluster. The deduced biosynthetic gene cluster for desertomycin including nineteen genes which encode eight type-I polyketide synthases (PKSs), *dstA1-A8*, seven accessory enzymes, three transporters and a LuxR-family transcriptional regulator was found

in *S. novilis* JCM4274 (Table S1). The eight PKS genes are transcribed in one direction. Deduced amino acid sequences of the eight PKSs (DstA1-A8) were compared with those of known type-I PKSs to predict their domain organizations (Figure S1). The deduced gene products of the eight PKSs include twenty-one modules for decarboxylative chain-elongation reactions in addition to an acyl-carrier protein (ACP) located at N-terminus of DstA1. Genes, *dstB* and *dstD*, encoding arginine 2-monooxygenase and acyl-CoA ligase, respectively, which are responsible for the formation of a starter unit of polyketide synthase DstA1, 4-guanidinobutyryl-CoA, also lie in the biosynthetic gene cluster. (ref. 11). A gene encoding an amide hydrolase, which acts at the second step in the biosynthesis of the starter unit, is responsible for conversion from 4-guanidinobutyramide to 4-guanidinobutyric acid, was not found in the biosynthetic gene cluster but the corresponding gene lies at a 6.5 Mb distance from the biosynthetic gene cluster. A discrete acyltransferase (AT) gene, *dstE* was also located in the biosynthetic gene cluster, which is responsible for the loading of 4-guanidinobutyryl-CoA onto the N-terminus ACP of DstA1. Substrate specificity of AT domains in each module was also predicted by comparing their amino acid sequences. The AT domains in module 2, 3, 4, 5, 8, 9, 11, 14, 16, 17 and 20, and in module 1, 6, 7, 12, 13, 15, 18, 19 and 21 are specific for malonyl transferring domains, and methylmalonyl transferring domains, respectively (Figure S2) (ref. 12). Surprisingly, the catalytic serine residue (-GH**S**-) was not conserved in the AT domain in module 10. In addition, this AT domain is shorter than the other AT domains: the second conserved motif was deleted (Figure S2). AT domain in module 10 might be categorized as dysfunctional domain (ref. 13). Similar AT domains were reported in the biosynthetic gene cluster for gephyronic acid (module 2 and 4) (ref. 14), conglobatin (module 3) (ref. 15), and bengamide (module 3) (ref. 16). These AT domains also do not contain the conserved serine residue in their active sites and are shorter than typical AT domain. These AT domains look like dysfunctional, but the chain-extension reaction is not terminated at the corresponding module thereafter the chain extension reaction might be continued to final module and the acyl chain is generated and released because desired macrolactone compounds were produced from these producers. The active site and the reaction mechanism in these shorter AT domains are still unclear. Ketoreductase (KR) domains of each module were analyzed for prediction of stereochemical outcome of hydroxy groups. The KR domains in module 5, 7, 10, 13, 15, 18 and 19, and in module 1, 2, 3, 4, 6, 8, 9, 11, 12, 14, 16, 17, 20 and 21 are A types responsible for the L-configuration of hydroxy groups and B type for the D-configuration, respectively (Figure 2) (ref. 17). The thioesterase (TE) domain resides in the C-terminus of DstA8, which releases the acyl intermediate from DstA8 then a carboxyl residue at C-1 position of the acyl chain released would be regioselectively cyclized with a hydroxyl residue at C-41 position of the

acyl chain.

For the cloning and heterologous expression of the biosynthetic gene cluster, the BAC library of *S. nobilis* JCM4274 was constructed as previously described (ref. 18). Partially *Bam*HI-digested genome DNA fragments were cloned into *Bam*HI site of an integrating BAC vector pKU518 (ref. 19). BAC clones containing the entire biosynthetic gene cluster for desertomycin were selected by PCR screening with primers targeting the flanking region of the biosynthetic gene cluster. After PCR screening, four desired clones were obtained from 1,536 clones in the BAC library and among these, pKU518JCM4274_P1_P1-M4, was introduced into *S. lividans* TK23. The resulting transformant was cultivated in seed medium at 320 rpm, at 27 ˚C for 2 days. Aliquots (450 μL) of pre-cultures were inoculated into 15 mL of 0.3 x BPS medium (ref. 20) and were cultured on a rotary shaker (180 rpm) at 27 ˚C for 5 days.

The culture extract was subjected to LC/MS and a novel peak showing UV absorption peak at 220 nm was detected in transformants (Figure 1). The *m/z* of the new peak was found to be 1192.7577 ($C_{61}H_{110}NO_{21}^+$, $[M+H]^+$: + 0.7 mmu) by HRMS analysis. The product was identified as desertomycin A by comparison with an authentic standard. Desertomycin A was also detected from culture extracts of *S. nobilis* JCM4274. The productivity of desertomycin A in the *S. lividans* TK23 transformant was 139 ± 29 mg L^{-1}, while that of the original strain *S. nobilis* JCM4274 was only 28 ± 6 mg L^{-1}. The productivity of desertomycin A from transformants was higher than that of the original producing strain under the same cultivation conditions (Figure S3).

Although the absolute stereochemistry of desertomycin A had been determined by Kishi *et al* (ref. 21), the stereochemistry of a methyl residue at C-32 was inconsistent with prediction by classification of KR domain (in module 6), which is concerned with the stereochemistry at C-32 methyl residue. In module 6, which corresponds to the biogenesis of the C-32 methyl group, we recognized an additional DH domain that may be dysfunctional. Recently, a biochemical study on dysfunctional DH domains, NanDH1, NanDH5 and NigDH1, revealed that some dysfunctional DH domains have epimerization activity for α-methyl groups (ref. 22). Therefore, we considered that the stereochemical output at C-32 might come from the epimerizing activity of the dysfunctional DH domain in module 6. To confirm stereochemistry at C-32, we analyzed the relative configuration between C-32 and C-33 by the *J*-based configuration analysis (JBCA) (ref. 23) using vicinal ^1H–^1H (from DQF-COSY spectra) and long-range ^1H–^{13}C coupling constants (from *J*-resolved HMBC spectra) (ref. 24-27). A large coupling constant value between H-31 and H-32 ($^3J_{H31\text{-}H32}$ = 10 Hz) revealed that they are in the *anti* orientation, while that between H-32 and H-33 was small

($^3J_{H32\text{-}H33}$ ~ 1.2 Hz) showing that they are in the *gauche* orientation (Figure S4). The small coupling constant value between H-32 and C-33 ($^2J_{H32\text{-}C33}$ = ~ 0 Hz) was consistent with the *anti* orientation between H-32 and O-33. In addition, the large coupling constant value between H (H-33) and CH_3 (C-32; $^3J_{33H\text{-}32CH3}$ = 6 Hz) indicates *anti* orientation between H (33-H) and CH_3 (C-32). Taking into consideration these results, when assuming the stereochemistry of C-33 as *R** configuration, the configuration of C-32 is elucidated as *S**. These results support the reported stereochemistry, which proved the credibility of the inactive DH domain function to elucidate the absolute configuration of methyl residues on type-I PKS compounds. Thus, stereochemistry of acyl intermediate bound to ACP in Module 21 was considered to be (6*S*, 7*R*, 8*S*, 9*S*, 14*S*, 15*S*, 18*S*, 19*R*, 23*S*, 25*R*, 27*S*, 29*S*, 30*S*, 31*R*, 32*S*, 33*R*, 35*S*, 37*S*, 41*S*, 42*R*, 43*R*) and is illustrated as shown in Figure 2.

In this study, we have identified the entire biosynthetic gene cluster for desertomycin by cloning and heterologous expression experiments. The size of the inserted DNA in the BAC clone, pKU518JCM4274_P1_P1-M4, was 181 kb which is sufficiently large to accommodate the entire biosynthetic gene cluster for 42-membered large macrolides, desertomycin (127 kb). The production of desertomycin was confirmed in *S. lividans* TK23 despite of the absence of a gene encoding amide hydrolase in the BAC clone pKU518JCM4274_P1_P1-M4. Therefore, it is likely that an endogenous gene product of amide hydrolase in the *S. lividans* TK23 (accession # EOY46468; hydrolase, 262 aa) could complement the function for the hydrolysis of amide residue. Accumulation of such experimental data of known compounds would be a clue for the production of compounds derived from unknown biosynthetic gene clusters and skeletally novel compounds by genome mining.

Recently, a congener of desertomycins, desertomycin G was reported to exhibit anti-tuberculosis activity (ref. 28). Therefore, new desertomycin family compound is expected to show a variety of bioactivities and broaden the antibacterial spectrum. Structural differences between desertomycin A and G are the presence of a double bond at C-4 and a methyl group at C-24 positions of desertomycin A. Nowadays, we develop the innovative technique “module editing”, by which we can produce desired derivatives of type-I PKS compounds. By applying this technique, genetic modification of the type I PKS gene, *dstA1-8*, on the BAC clone employed as the template would lead to novel derivatives of desertomycins. Heterologous expression of desertomycin A with a production yield of more than 100 mg L^{-1} set the stage for the genetic engineering of polyketide synthases of desertomycin A on a BAC clone, pKU518JCM4274_P1_P1-M4, to produce novel desertomycin analogs.

AUTHOR INFORMATION

Corresponding Author

*Phone/fax: +81-3-3599-8305. E-mail: k-shinya@aist.go.jp (K. Shin-ya).

ORCID

Kazuo Shin-ya: 0000-0002-4702-0661

Haruo Ikeda: 0000-0003-3977-7856

Takuya Hashimoto: 0000-0002-6261-453X

Notes

The authors declare no competing financial interest.

Acknowledgements

This work was supported by Japan Agency for Medical Research and Development (AMED) under Grant Number JP17ae0101002 for K.S.

References

1 Pye CR, Bertin MJ, Lokey RS, Gerwick WH, Linington RG. Retrospective analysis of natural products provides insights for future discovery trends. Proc Natl Acad Sci USA. 2017;114:5601-6.

2 Palazzolo AME, Simons CLW, Burke MD. The natural productome. Proc Natl Acad Sci USA. 2017; 114: 5564-6.

3 Kautsar SA, *et al*. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Res. 2020; 48: D454-8.

4 Ichikawa N, *et al*. DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. Nucleic Acids Res. 2013; 41: D408-14.

5 Uri J, Bognar R, Bekesi I, Varga B. Desertomycin, a new crystalline antibiotic with antibacterial and cytostatic action. Nature. 1958; 182: 401.

6 Zerlin M, Thiericke R. Common principles in macrolactone (marginolactone) biosynthesis. Studies on the desertomycin family. J Org Chem. 1994; 59: 6986-93.

7 Hong H, Samborskyy M, Lindner F, Leadlay PF. An amidinohydrolase provides the missing link in the biosynthesis of amino marginolactone antibiotics. Angew Chem Int Ed. 2016; 55: 1118-23.

8 Zhang LH, *et al*. Characterization of giant modular PKSs provides insight into genetic mechanism for structural diversification of aminopolyol polyketides. Angew Chem Int Ed. 2017; 56: 1740-5.

9 Hashimoto T, *et al*. Novel macrolactam compound produced by the heterologous expression of a large cryptic biosynthetic gene cluster of *Streptomyces rochei* IFO12908. J Antibiot. 2020; 73: 171-4

10 Hashimoto T, *et al*. Biosynthesis of quinolidomicin, the largest known macrolide of terrestrial origin: identification and heterologous expression of a biosynthetic gene cluster over 200 kb. Org Lett. 2018; 20: 7996-9.

11 Hong H, Fill T, Leadlay PF. A common origin for guanidinobutanoate starter units in antifungal natural products. Angew Chem Int Ed. 2013; 52: 13096-9.

12 Reeves CD, *et al.* Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. Biochemistry. 2001; 40: 15464-70.

13 Keatinge-Clay AT. The uncommon enzymology of cis-acyltransferase assembly lines. Chem Rev. 2017; 117: 5334-66.

14 Young J, *et al.* Elucidation of gephyronic acid biosynthetic pathway revealed unexpected SAM-dependent methylations. J. Nat. Prod. 2013; 76: 2269-76.

15 Zhou Y, *et al.* Iterative mechanism of macrodiolide formation in the anticancer compound conglobatin. Chem Biol. 2015; 22: 745-54.

16 Wenzel SC, *et al.* Production of the bengamide class of marine natural products in *Myxobacteria*: biosynthesis and structure–activity relationships. Angew Chem Int Ed. 2015; 54: 15560-4.

17 Keatinge-Clay AT. A tylosin ketoreductase reveals how chirality is determined in polyketides. Chem Biol. 2007; 14: 898-908.

18 Komatsu M, *et al*. Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. ACS Synth Biol. 2013; 2: 384-96.

19 Kim JH, Komatsu M, Shin-ya K, Omura S, Ikeda H. Distribution and functional analysis of the phosphopantetheinyl transferase superfamily in *Actinomycetales* microorganisms. Proc Natl Acad Sci USA. 2018; 115: 6828-33.

20 Jørgensen H, *et al*. Biosynthesis of macrolactam BE-14106 involves two distinct PKS systems and amino acid processing enzymes for generation of the aminoacyl starter unit. Chem Biol. 2009; 16: 1109-21.

21 Tan CH, Kobayashi Y, Kishi Y. Stereochemical assignment of the C21–C38 portion of the desertomycin/oasomycin class of natural products by using universal NMR databases: Proof. Angew Chem Int Ed. 2000; 39: 4282-4284.

22 Xie X, Garg A, Khosla C, Cane DE. Elucidation of the cryptic methyl group epimerase activity of dehydratase domains from modular polyketide synthases using a tandem modules epimerase assay.

J Am Chem Soc. 2017; 139: 9507-10.

23 Murata M, Matsuoka S, Matsumori N, Paul GK, Tachibana K. Absolute configuration of amphidinol 3, the first complete structure determination from amphidinol homologues: Application of a new configuration analysis based on carbon−hydrogen spin-coupling constants. J Am Chem Soc. 1999; 121: 870-1.

24 Kobayashi H, Shin-ya K, Furihata K, Hayakawa Y, Seto H. Absolute configuration of a novel glutamate receptor antagonist kaitocephalin. Tetrahedron Lett. 2001; 42: 4021-3.

25 Park HR, Chijiwa S, Furihata K, Hayakawa Y, Shin-ya K. Relative and absolute configuration of versipelostatin, a down-regulator of molecular chaperone GRP78 expression. Org lett. 2007; 9: 1457-60.

26 Umeda Y, *et al.* Absolute structure of prunustatin A, a novel GRP78 molecular chaperone down-regulator. Org Lett. 2007; 9: 4239-42.

27 Kawahara T, Izumikawa M, Takagi M, Shin-ya K. Relative configuration of JBIR-129, a cytotoxic 34-membered glycosidic polyol macrolide from *Streptomyces* sp. RK74. Org Lett. 2012; 14: 4434-7.

28 Brana AF, *et al*. Desertomycin G, a new antibiotic with activity against *Mycobacterium tuberculosis* and human breast tumor cell lines produced by *Streptomyces althioticus* MSM3, isolated from the Cantabrian sea intertidal macroalgae *Ulva* sp. Mar Drugs 2019; 17: 114.

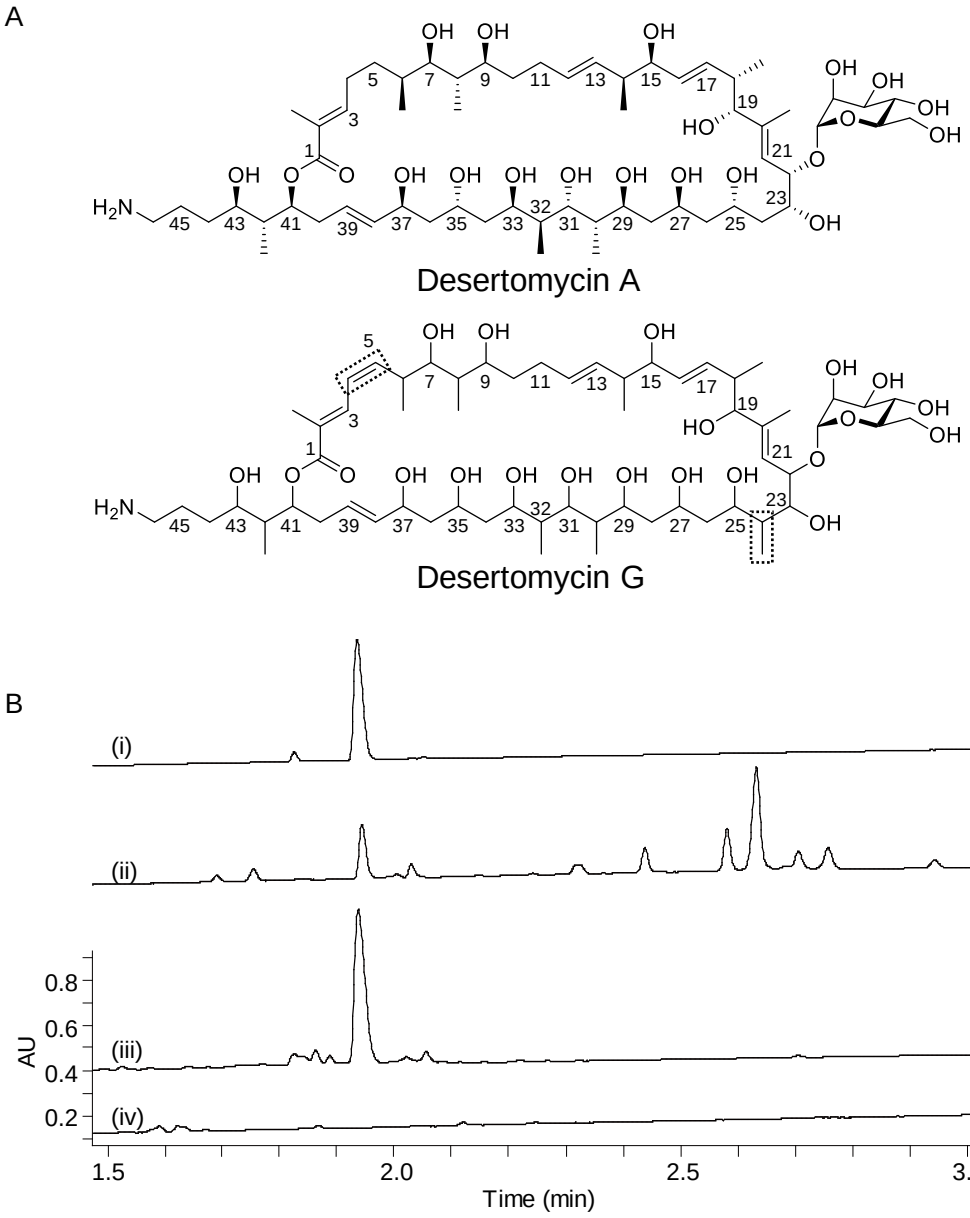

Figure 1. (A) Structure of desertomycins A and G. The differences between the structures are shown by the dotted rectangles. (B) HPLC chromatograms of authentic sample of desertomycin A (i), culture extracts from *S. nobilis* JCM4274 (ii) and *S. lividans* TK23 carrying pKU518JCM4274_P1_P1-M4 (iii) and *S. lividans* TK23 (iv). Chromatograms were traced at 220 nm.

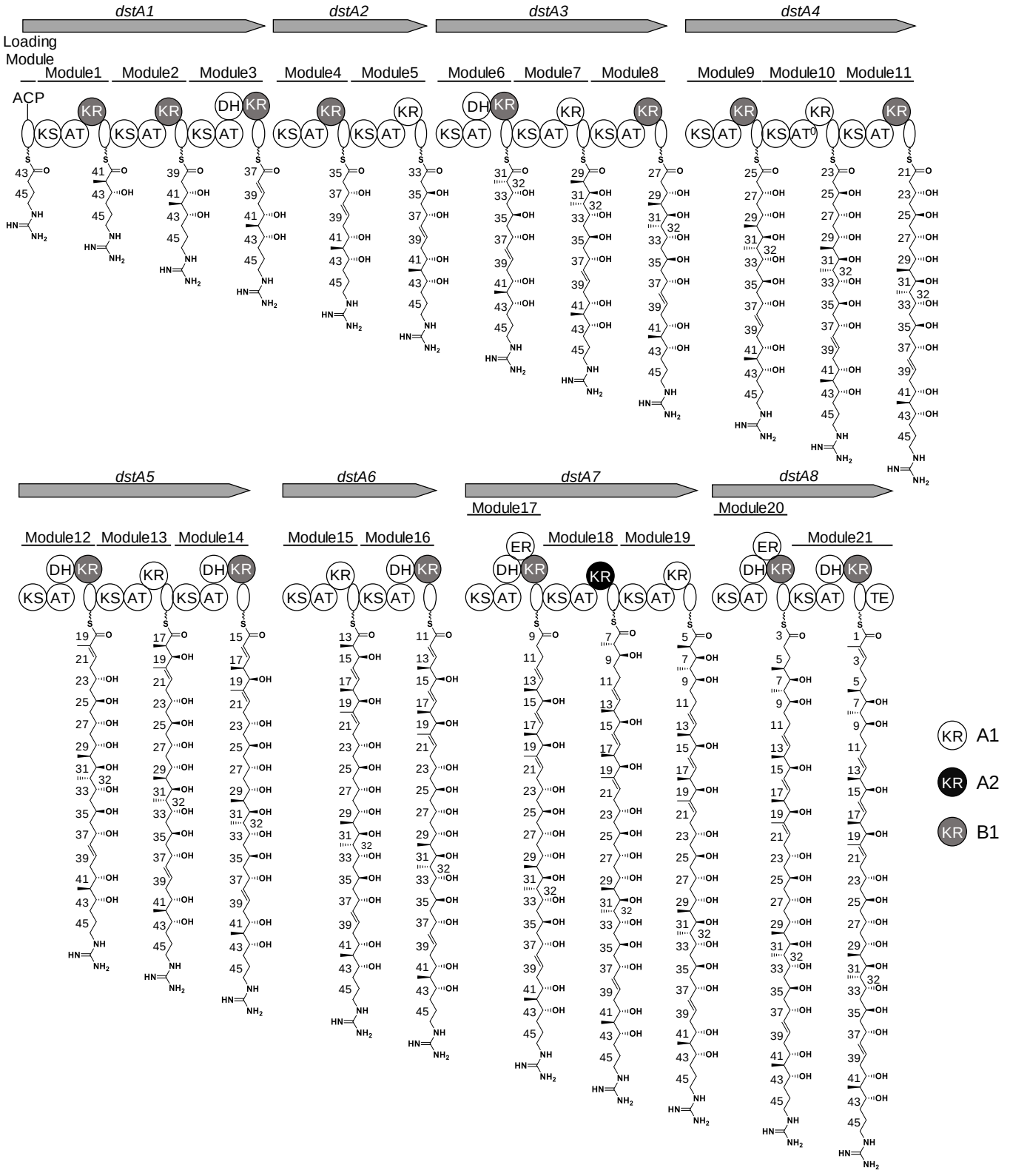

Figure 2. Proposed biosynthetic pathway by polyketide synthases encoded in *dstA1-A8*. KR domains of A1, A2 and B1 type are colored white, black and gray respectively. A methyl group at C-32 may be epimerized by the DH domain in module 6.