

Characterization of the Maduropeptin Biosynthetic Gene Cluster from *Actinomadura madurae* ATCC 39144 Supporting a Unifying Paradigm for Eneidyne Biosynthesis

Steven G. Van Lanen,[†] Tae-jin Oh,[†] Wen Liu,[†] Evelyn Wendt-Pienkowski,[†] and Ben Shen^{*,†,‡,§}

Contribution from the Division of Pharmaceutical Sciences, University of Wisconsin National Cooperative Drug Discovery Group, and Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53705

Received May 21, 2007; E-mail: bshen@pharmacy.wisc.edu

Abstract: The biosynthetic gene cluster for the enediyne antitumor antibiotic maduropeptin (MDP) from *Actinomadura madurae* ATCC 39144 was cloned and sequenced. Cloning of the *mdp* gene cluster was confirmed by heterologous complementation of enediyne polyketide synthase (PKS) mutants from the C-1027 producer *Streptomyces globisporus* and the neocarzinostatin producer *Streptomyces carzinostaticus* using the MDP enediyne PKS and associated genes. Furthermore, MDP was produced, and its apoprotein was isolated and N-terminal sequenced; the encoding gene, *mdpA*, was found to reside within the cluster. The biosynthesis of MDP is highlighted by two iterative type I PKSs — the enediyne PKS and a 6-methylsalicylic acid PKS; generation of (S)-3-(2-chloro-3-hydroxy-4-methoxyphenyl)-3-hydroxypropionic acid derived from L- α -tyrosine; a unique type of enediyne apoprotein; and a convergent biosynthetic approach to the final MDP chromophore. The results demonstrate a platform for engineering new enediynes by combinatorial biosynthesis and establish a unified paradigm for the biosynthesis of enediyne polyketides.

Introduction

Maduropeptin (MDP) from *Actinomadura madurae* ATCC 39144 belongs to the enediyne family of antitumor antibiotics.^{1–4} Similar to all chromoprotein enediynes, which include C-1027⁵ and neocarzinostatin (NCS),⁶ MDP is isolated as a noncovalent complex consisting of an apoprotein and a reactive nine-membered enediyne chromophore (Figure 1). Upon release from the apoprotein, the enediyne chromophore can undergo a rearrangement to form a benzenoid diradical species that can abstract hydrogen atoms from DNA to initiate a cascade leading to DNA breaks, ultimately leading to cell death.^{7,8} While this mode of action has attracted interest in using the enediynes for cancer treatment, high cytotoxicity has delayed the application of enediynes as therapeutic agents.⁹ However, the development

of NCS conjugates with poly(styrene-*co*-maleic acid) or its various alkyl esters¹⁰ and a monoclonal-antibody derivative of the 10-membered enediyne calicheamicin (CAL)¹¹ under the trade name Mylotarg have alleviated toxicity allowing the drugs to reach the markets in Japan and the USA, respectively, providing the inspiration that the engineering and discovery of new enediynes will facilitate anticancer drug development.

The central feature of the enediyne chromophores is their exotic 9- or 10-membered ring structure consisting of a characteristic diyne conjugated to a double bond, otherwise referred to as the enediyne core. The majority of the nine-membered enediyne chromophores, like NCS,¹² C-1027,¹³ and MDP (Figure 1), contain additional structural features that enhance the chemistry, which in turn modulates the biological activity. The MDP chromophore, the structure of which was solved in 1994,¹⁴ also contains a 3,6-dimethylsalicylic acid moiety, an aminosugar previously named madurose, and an (S)-3-(2-chloro-3-hydroxy-4-methoxyphenyl)-3-hydroxypropionic acid moiety. Similar to NCS and C-1027, MDP is isolated as a chromoprotein complex, although a new method had to be

[†] Division of Pharmaceutical Sciences.

[‡] University of Wisconsin National Cooperative Drug Discovery Group.

[§] Department of Chemistry.

- (1) Xi, Z.; Goldberg, I. H. In *Comprehensive Natural Products Chemistry*; Barton, D., Nakanishi, K., Meth-Cohn, O., Eds.; Elsevier: New York, 1999; pp 533–592.
- (2) Jones, G. B.; Fouad, F. S. *Curr. Pharm. Des.* **2002**, *8*, 2415–2440.
- (3) Thorson, J. S.; Shen, B.; Whitwam, R. E.; Liu, W.; Li, Y.; Ahlert, J. *Bioorg. Chem.* **1999**, *27*, 172–188.
- (4) Galm, U.; Hager, M. H.; Van Lanen, S. G.; Ju, J.; Thorson, J. S.; Shen, B. *Chem. Rev.* **2005**, *105*, 739–758.
- (5) Otani, T.; Yasuhara, T.; Minami, Y.; Shimazu, T.; Zhang, R.; Xie, M. *Agric. Biol. Chem.* **1991**, *55*, 407–415.
- (6) Gibson, B. W.; Herlihy, W. C.; Samy, T. S. A.; Hahm, K.; Maeda, H.; Meienhofer, J.; Biemann, K. *J. Biol. Chem.* **1984**, *259*, 10801–10806.
- (7) Dedon, P. C.; Goldberg, I. H. *J. Biol. Chem.* **1990**, *265*, 14713–14716.
- (8) Koide, Y.; Ito, A.; Edo, K.; Ishida, N. *Chem. Pharm. Bull.* **1986**, *34*, 4425–4427.
- (9) Zhen, Y.; Ming, X.; Yu, B.; Otani, T.; Saito, H.; Yamada, Y. *J. Antibiot.* **1989**, *42*, 1294–1298.

- (10) *Enediyne Antibiotics as Antitumor Agents*; Doyle, T. W., Borders, D. B., Eds.; Marcel-Dekker: New York, 1995; pp 1–459.
- (11) Sievers, E. L.; Appelbaum, F. R.; Spielberger, R. T.; Forman, S. J.; Flowers, D.; Smith, F. O.; Shannon-Dorcy, K.; Berger, M. S.; Bernstein, I. D. *Blood* **1999**, *93*, 3678–3684.
- (12) Edo, K.; Mizugaki, M.; Koide, Y.; Seto, H.; Furihata, K.; Otake, N.; Ishida, N. *Tetrahedron Lett.* **1985**, *26*, 331–340.
- (13) Yoshida, K.; Minaomi, Y.; Azuma, R.; Saeki, M.; Otani, T. *Tetrahedron Lett.* **1993**, *34*, 2637–2640.
- (14) Schroeder, D. R.; Colson, K. L.; Klotz, S. E.; Zein, N.; Langley, D. R.; Lee, M. S.; Matson, J. A.; Doyle, T. W. *J. Am. Chem. Soc.* **1994**, *116*, 9351–9352.

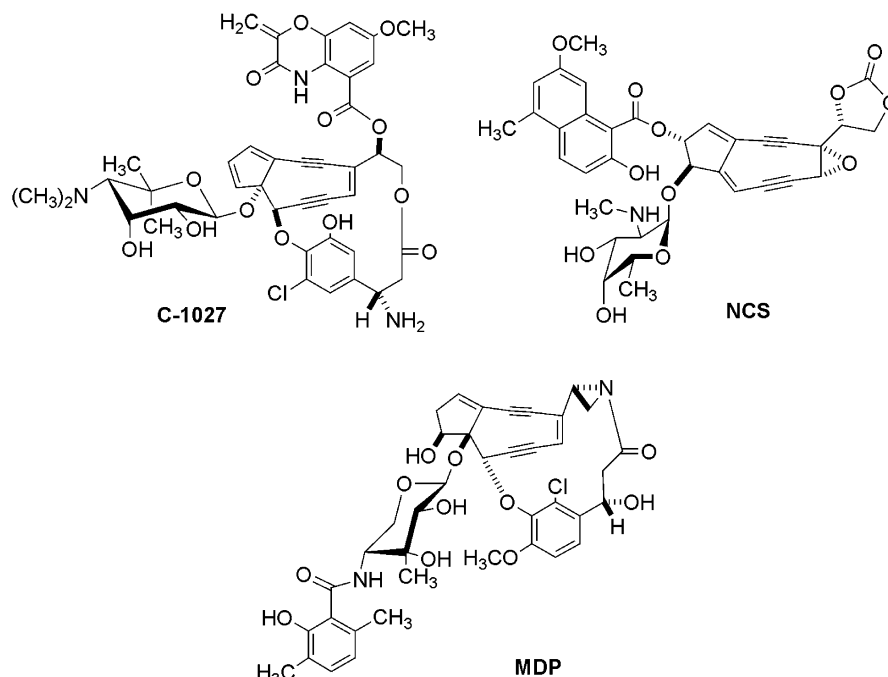


Figure 1. Structures of the nine-membered enediynes C-1027, NCS, and MDP.

developed to release the chromophore from the apoprotein for structural characterization.¹⁵ Furthermore, the MDP apoprotein was deemed a new type of enediyne apoprotein with no homology to CagA⁶ or NcsA,⁵ the C-1027 and NCS apoproteins, respectively, although only the approximate amino acid composition, and not the primary sequence, was reported.¹⁶ The so-called histone-specific proteolytic activity of the apoproteins for the chromoprotein enediynes, which included the MDP apoprotein,¹⁷ has now been generally accepted as an artifact that resulted from adventitious proteases contaminating the apoprotein or chromoprotein preparation.^{17–20}

The 9-membered NCS or 10-membered CAL enediyne chromophores require an external trigger, such as a base or reducing agent, respectively, to initiate an electronic rearrangement to form the benzenoid diradical species (via a Bergman or Myers rearrangement), which in turn abstracts hydrogen atoms from the deoxyribose of DNA, leading to the eventual DNA damage.¹⁰ In contrast, the MDP enediyne chromophore, like that of C-1027, is relatively unstable, and attempts to isolate the nascent MDP enediyne chromophore were unsuccessful.¹⁴ Instead, a C5-methoxy adduct was structurally elucidated, along with the aromatized product (Figure 2B). Comparisons of these structures and computational modeling led to a proposal where the MDP–OCH₃ adduct, hypothesized to be an artifact of the purification process, is first reconverted to MDP by intramolecular ring contraction to form the aziridine, while simultaneously eliminating methanol via double bond migration and

anti-elimination. The conjugated diyne form of MDP is now primed to undergo Bergman rearrangement, leading to formation of the benzenoid diradical intermediate that, upon coupling with hydrogen atoms, results in the aromatized product. This proposal was in part substantiated upon chemical synthesis of a MDP–OCH₃ adduct mimic, which verified that the exocyclic C4–C13 double bond is labile to rearrangement, resulting in the characteristic aziridine moiety of MDP.²¹

Concurrent with our studies regarding the chromoproteins C-1027²² and NCS,²³ we chose to clone and sequence the *mdp* biosynthetic gene cluster to facilitate biosynthetic studies of the enediynes, to establish the apoprotein sequence for MDP, and to solidify the unifying paradigm for enediyne biogenesis.^{24,25} In this paper, we report the identification of the MDP biosynthetic gene cluster from *A. madurae* ATCC 39144 and propose a convergent biosynthetic pathway leading to the MDP chromophore by sequence analysis of the open reading frames (ORFs) and comparisons to the other cloned enediyne gene clusters, C-1027,^{22,26} NCS,²³ and CAL.^{27,28} To expedite the confirmation of the cluster identity, a C-1027 enediyne polyketide synthase (PKS) mutant strain *Streptomyces globisporus* SB1005²² and an NCS enediyne PKS mutant strain *Streptomyces carzinostaticus* SB5002²³ were utilized for heterologous complementation with the MDP enediyne PKS. Furthermore, the MDP

- (15) Schroeder, D. R.; Lam, K. S.; Veitch, J. M. U.S. Patent 5,281,417, 1994.
- (16) Hanada, M.; Ohkuma, H.; Yonemoto, T.; Tomita, K.; Ohbayashi, M.; Kamei, H.; Miyaki, T.; Konishi, M.; Kawaguchi, H. *J. Antibiot.* **1991**, *44*, 403–414.
- (17) Zein, N.; Solomon, W.; Colson, K. L.; Schroeder, D. R. *Biochemistry* **1995**, *34*, 11591–11594.
- (18) Zein, N.; Casazza, A. M.; Doyle, T. W.; Leet, J. E.; Schroeder, D. R.; Solomon, W.; Nadler, S. G. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8009–8012.
- (19) Zein, N.; Reiss, P.; Bernatowicz, M.; Bolgar, M. *Chem. Biol.* **1995**, *2*, 451–455.
- (20) Heyd, B.; Lerat, G.; Adjadj, E.; Minard, P.; Desmadril, M. *J. Bacteriol.* **2000**, *182*, 1812–1818.

- (21) Suffert, J.; Toussaint, D. *Tetrahedron Lett.* **1997**, *38*, 5507–5510.
- (22) Liu, W.; Christenson, S. D.; Standage, S.; Shen, B. *Science* **2002**, *297*, 1170–1173.
- (23) Liu, W.; Nonaka, K.; Nie, L.; Zhang, J.; Christenson, S. D.; Bae, J.; Van, Lanen, S. G.; Zazopoulos, E.; Farnet, C. M.; Yang, C. F.; Shen, B. *Chem. Biol.* **2005**, *12*, 1–10.
- (24) Liu, W.; Ahlert, J.; Gao, Q.; Wendt-Pienkowski, E.; Shen, B.; Thorson, J. S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11959–11963.
- (25) Zazopoulos, E.; Huang, K.; Staffa, A.; Liu, W.; Bachmann, B. O.; Nonaka, K.; Ahlert, J.; Thorson, J. S.; Shen, B.; Farnet, C. M. *Nat. Biotechnol.* **2003**, *21*, 187–190.
- (26) Liu, W.; Shen, B. *Antimicrob. Agents Chemother.* **2000**, *44*, 382–392.
- (27) Ahlert, J.; Shepard, E.; Lomovskaya, N.; Zazopoulos, E.; Staffa, A.; Bachmann, B. O.; Huang, K.; Fonstein, L.; Czisny, A.; Whitwam, R. E.; Farnet, C. M.; Thorson, J. S. *Science* **2002**, *297*, 1173–1176.
- (28) Sambrook, J.; Russell, D. W. *Molecular Cloning: A Laboratory Manual*, 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2002.

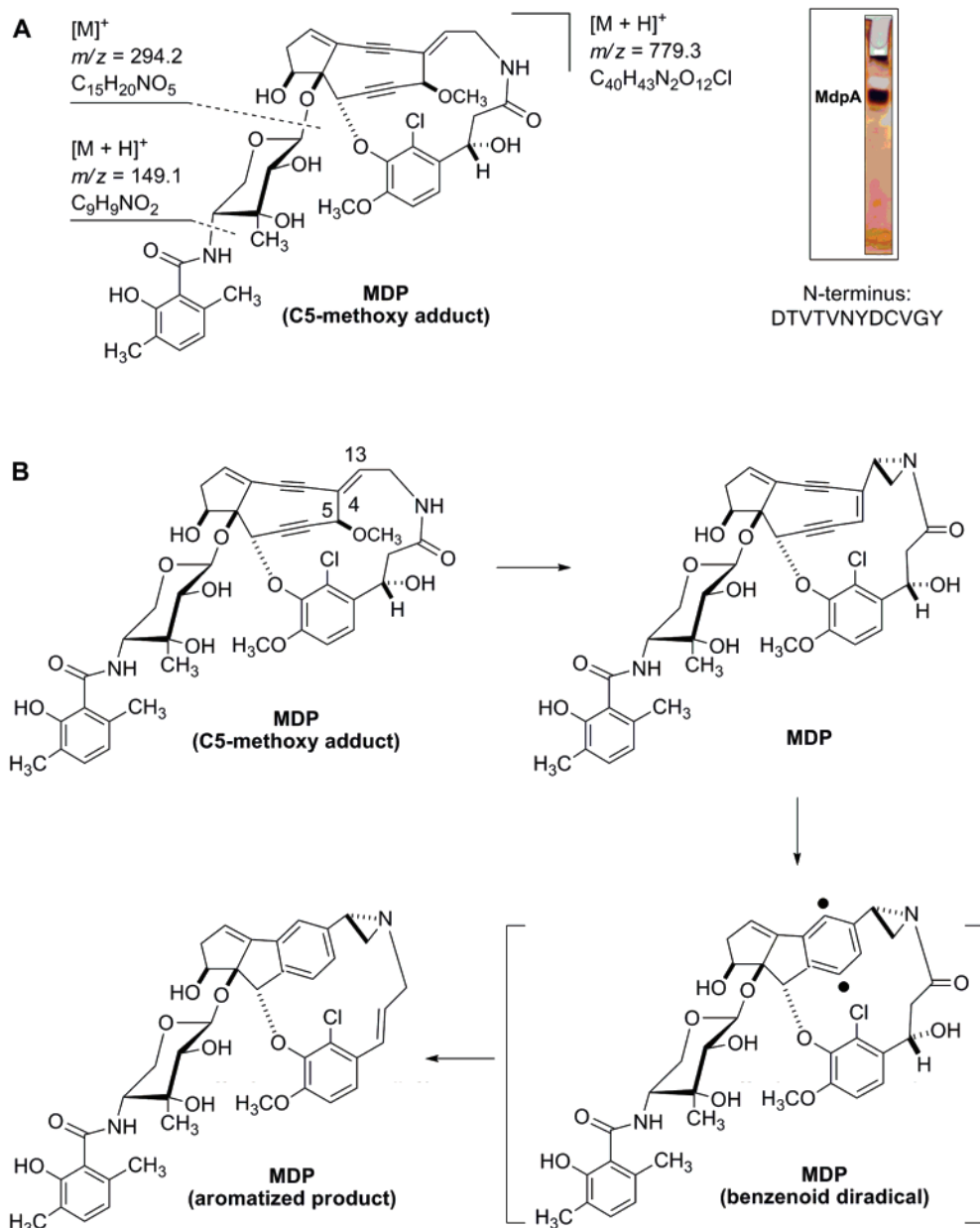


Figure 2. Isolation of MDP chromophore and binding protein, MdpA. (A) The chromophore was purified and confirmed by MS giving the characteristic MDP–OCH₃ adduct and associated mass fragments as reported previously;¹⁴ the MDP binding protein was purified and confirmed by N-terminal sequencing. (B) Isolation of MDP as a MDP–OCH₃ adduct, its rearrangement to the aziridine form MDP, subsequent Bergman rearrangement to afford the MDP–benzenoid diradical, and final formation of the MDP–aromatized product after addition of two hydrogen atoms as described previously.¹⁴

chromoprotein complex was isolated from *A. madurae* ATCC 39144. The MDP chromophore was confirmed as the MDP–OCH₃ adduct by mass spectrometry (MS) analysis, and the MDP apoprotein was N-terminal sequenced, confirming that the *mdpA* gene, which encodes the apoprotein, was indeed localized within the cloned *mdp* cluster. The results provide the first experimental evidence for a unified enediyne biosynthetic paradigm and establish a new enediyne binding protein, MdpA, with no sequence homologues in the protein databanks. Together the results lay the groundwork for future enzymatic characterizations of the MDP biosynthetic pathways and provide the basis for combinatorial biosynthesis to generate new enediyne natural products.

Experimental Procedures

Bacterial Strains, Plasmids, and Sequence Analysis. *A. madurae* ATCC 39144 and *S. carzinostaticus* ATCC 15944 wild-type strains

producing MDP and NCS, respectively, were obtained from American Type Culture Collection (Rockville, MD). The *S. globisporus* wild-type²² and *S. globisporus* SB1005²² and *S. carzinostaticus* SB5002²³ mutant strains, in which the *sgcE* and *ncsE* enediyne PKS genes have been inactivated, respectively, were described previously. Gigapack III XL and *E. coli* XL1-Blue MR cells were obtained from Stratagene (La Jolla, CA). pGEM-T Easy and pGEM-3zf(+) were from Promega (Madison, WI). SuperCos1 was from Stratagene. pANT841,²³ pSET152,²⁹ pUWL201PW,²⁹ and pWHM79³⁰ were described previously. The C-1027 and NCS enediyne PKS expression constructs pBS1019²² (that overexpresses *sgcE* under the *ErmE** promoter²⁹) and pBS5020²³ (that overexpress *ncsE5*, *ncsE*, and *ncsE10* under the *ErmE** promoter²⁹) were reported previously. A DIG-labeling kit and calf intestinal

(29) Kieser, T.; Bibb, M.; Buttner, M.; Chater, K. F.; Hopwood, D. A. *Practical Streptomyces Genetics*; The John Innes Foundation: Norwich, U.K., 2000.

(30) Shen, B.; Hutchinson, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6600–6604.

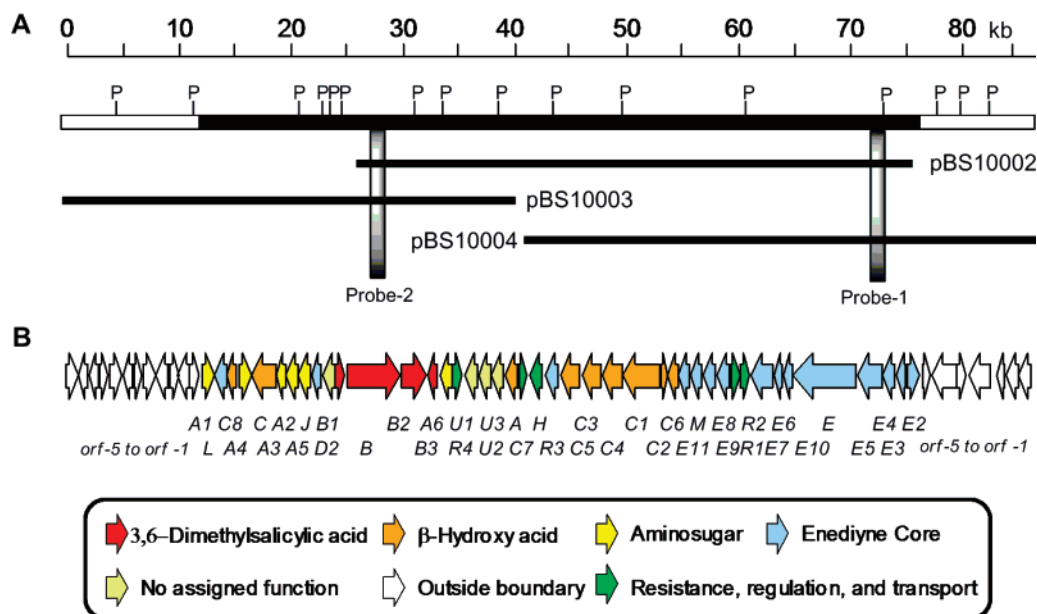


Figure 3. MDP biosynthetic gene cluster from *A. madurae* ATCC 39144. (A) Restriction map of the 85-kb DNA region encompassed by three overlapping cosmids represented by pBS10002, pBS10003, and pBS10004. Solid black indicates the region whose gene products are predicted to be involved in MDP biosynthesis (62.8 kb). P, *Pst*I. Probe-1 and -2 were used to localize the overlapping cosmids covering the MDP cluster. (B) Genetic organization of the MDP biosynthetic gene cluster. Proposed functions for individual *orfs* are color-coded and summarized in Table 1.

phosphatase were from Roche (Indianapolis, IN). Ligations were performed with T4 DNA ligase from Promega. Restriction enzymes were from New England Biolabs (Ipswich, MA) or Invitrogen (Carlsbad, CA). The shotgun library of pBS10002 for DNA sequencing of the *mdp* cluster was prepared by Lucigen (Madison, WI). DNA sequencing was carried out at the University of Wisconsin Biotechnology Center (Madison, WI). Sequence analysis was carried out using BLASTN available from NCBI, and contiguous DNA was compiled using Lasergene (DNASTAR Inc., Madison, WI). ORFs were predicted using ORFinder from NCBI, and protein sequences were analyzed using PSI-BLAST (NCBI), SignalP, and Vector NTI (Invitrogen). Protein sequence identity and similarity were calculated using the Vector NTI program AlignX.

Biochemicals, Chemicals, and Media. Biochemicals, chemicals, media, and other molecular biology reagents were from standard commercial sources and used without further purification.

Cosmid DNA Library Construction. Recombinant DNA manipulations in *E. coli*²⁸ and *Streptomyces*²⁹ were carried out following standard protocols. The general procedure for cloning the enediynes PKS gene and associated ORFs was previously described.²⁵ A 3.5 kb-DNA fragment covering most of *mdpE5/E4/E3* was PCR amplified from total genomic DNA using Platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA). Reactions were performed following the manufacturer's protocol with 5% DMSO added, an annealing temperature of 62 °C for 30 s, and the following primers: (forward) 5'-GGMTTCCACCAG-GCGTACTTC-3'/(reverse) 5'-SCCGARRTTCGASGGGTTCAT-3', where M = A or C; S = C or G; and R = A or G. The DNA fragment was inserted into pGEM-T Easy following the provided protocol to yield pBS10001. After confirmation of the insert by DNA sequencing, pBS10001 was digested with *Pst*I to yield a 1.2-kb fragment (one *Pst*I site was within the MCS of pGEM-T Easy, and the other, within *mdpE4*), which was used to prepare a DIG-labeled probe (Probe-1) following the provided instructions (Roche) (Figure 3A).

A cosmid library was constructed by partially digesting *A. madurae* ATCC 39144 chromosomal DNA with *Sau*3A1 followed by dephosphorylation with calf intestinal phosphatase. A modified version of SuperCos1 was prepared by digestion with *Eco*47III and religation to eliminate the kanamycin resistance gene to yield SuperKos as described

previously.³¹ After an overnight ligation at 16 °C, the mixture was packaged using Gigapack III XL and used to transfect *E. coli* XL1 Blue MR cells following the instructions provided (Stratagene). Colony hybridization was performed using Probe-1 with Hybond-N from Amersham Biosciences (Piscataway, NJ). Ten clones that hybridized with Probe-1 were chosen, and the cosmids were isolated and their inserts were end-sequenced using the primers (forward) 5'-GG-GAATAAGGGCGACACGG-3' and (reverse) 5'-GCTTATCGAT-GATAAGCGGTC-3' to determine the best candidate for complete sequencing.

A single cosmid (pBS10002) was identified to harbor DNA regions encoding the enediynes PKS²⁴ and a second, distinct PKS (Figure 3A) and was used to generate a shotgun library for complete DNA sequence determination. The resultant DNA sequences were compiled and assembled into contigs, and gaps were filled in by primer walking or by subcloning fragments covering the gaps into pGEM-3zf(+) and subsequently sequencing the cloned fragments using the M13 universal primers.

After complete sequencing of pBS10002, a second probe (Probe-2) was prepared using PCR amplification with pBS1002 as a template and the following primers: (forward) 5'-GGTGCTGCTGACCACCT-TCC-3'/(reverse) 5'-CGACGTCCACCGTGACGACGACC-3' to give a 1.5-kb fragment near the upstream end of pBS1002 (Figure 3A) and used to rescreen the *A. madurae* cosmid library. Cosmid pBS10003 was chosen as the best candidate to contain the most upstream boundary on the basis of end-sequencing of the DNA insert. Alternatively, pBS1004, which hybridized with Probe-1, was chosen to likely contain the downstream boundary based on end-sequencing of the DNA insert. The final DNA sequence of the entire *mdp* cluster was obtained by chromosomal walking from pBS10002 directly into pBS10003 and pBS10004, combined with subcloning of selected fragments from pBS10003 and pBS10004 into pGEM-3zf(+) and sequencing using M13 universal primers.

Culturing of *A. madurae* ATCC 39144. The culture conditions of *A. madurae* ATCC 39144 for genomic DNA isolation were previously

(31) Wendt-Pienkowski, E.; Huang, Y.; Zhang, J.; Li, B.; Jiang, H.; Kwon, H.-J.; Hutchinson, C. R.; Shen, B. *J. Am. Chem. Soc.* **2005**, *127*, 16442–16452.

described.²⁴ *A. madurae* spores were isolated according to standard procedures upon culturing on ISP4 plates at 28 °C for 5–7 days.²⁹ Production and isolation of the MDP chromoprotein complex were essentially the same as those described previously.^{14,15} Purification of the MDP chromophore from the chromoprotein was carried out according to the literature procedure,¹⁴ and the resultant MDP chromophore was subjected to atmospheric pressure chemical ionization-mass spectrometry (APCI-MS) analysis using an Agilent 1100 LC-MSD mass spectrometer (Agilent Technologies, Palo Alto, CA) with a Phenomenex Luna C18 column (5 μ , 4.6 mm \times 100 mm) (Torrance, CA) and an isocratic mobile phase of 80% methanol in water at 1 mL/min.

Heterologous Complementation. To construct the *mdp* enediyne PKS gene cassette expression vector pBS10005, a 9.5-kb *Pst*I-*Bam*HI fragment, containing the *mdpE5*, *mdpE*, *mdpE10*, and *mdpE6* genes, was recovered from pBS10004 and cloned into the same sites of pANT841 to yield pBS10006. This cassette was then isolated as a 9.5-kb *Hind*III-*Bam*HI fragment from pBS10006 and coligated with a 0.45-kb *Eco*RI-*Hind*III fragment of the *Er*mE* promoter from pWHM79 into the *Bam*HI and *Eco*RI sites of pSET152 to yield pBS10005.

To cross-complement the *sgcE:ermE* mutation in *S. globisporus* with *mdpE*, pBS10005 was introduced into *S. globisporus* SB1005²² by intergenetic conjugation according to the previously described procedure with pBS1019 similarly introduced as a positive control.²² To cross-complement the *ncsE:ermE* mutation in *S. carzinostaticus* with *mdpE*, pBS10005 was introduced into *S. carzinostaticus* SB5002²³ by PEG-mediated protoplast transformation as described previously with pBS5020 similarly introduced as a positive control.²³ The resultant recombinant strains *S. globisporus* SB1011 [i.e., SB1005 (pBS10005)] and SB1010 [i.e., SB1005 (pBS1019)] were cultured and analyzed for C-1027 production by HPLC analysis as described previously²² with the *S. globisporus* wild-type strain as a control (Figure 7A). C-1027 production was confirmed by matrix-assisted laser desorption/ionization-mass spectrometry (MALDI-MS), yielding a characteristic $[M + H]^+$ at $m/z = 844.259$ and 846.273 , consistent with the molecular formula of $C_{43}H_{42}N_3O_{13}Cl$ (enediynes form calculated 843.241) and $C_{43}H_{44}N_3O_{13}Cl$ (aromatized form calculated 845.256), respectively. Similarly, the *S. carzinostaticus* recombinant strains SB5007 [i.e., SB5002 (pBS10005)] and SB5003 [i.e., SB5002 (pBS5020)] were cultured and analyzed for NCS production by HPLC analysis following the published procedure²³ with the *S. carzinostaticus* ATCC 15944 wild-type strain as a control (Figure 7B). The identity of NCS was confirmed by liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS), yielding a characteristic $[M + H]^+$ at $m/z = 660.3$, consistent with the molecular formula of $C_{35}H_{33}NO_{12}$ (calculated 659.2).

Isolation of the MDP Apoprotein from *A. madurae* ATCC 39144 and Confirmation of Its Identity as MdpA. The MDP apoprotein MdpA was purified from *A. madurae* ATCC 31944 following the described procedure except for the inclusion of 5 mM EDTA in all solutions.¹⁵ After binding to DEAE cellulose anion exchange resin, the MDP chromoprotein complex was eluted with 50 mM Tris-HCl, pH 8.0, containing 0.5 M NaCl, and the resultant chromoprotein complex was precipitated by addition of $(NH_4)_2SO_4$ to 70% saturation. After centrifugation at $30\,000 \times g$ at 4 °C for 30 min, the pellet was collected, resuspended in 50 mM HEPES, pH 7.5, with 5 mM EDTA, and dialyzed overnight in the same buffer. The MDP chromoprotein complex was further purified using anion exchange chromatography as described except a 5 mL HiTrapQ column (GE Healthcare; Piscataway, NJ) was used to substitute for the TSK Gel DEAE 3SW column.¹⁶ The fractions were analyzed by SDS-PAGE and native PAGE with 15% acrylamide and visualized using nondestructive silver staining.³² The gel was soaked with gentle agitation for 24 h increments in H_2O , acetic acid/MeOH/ H_2O (10:30:60, v/v), acetic acid/ H_2O (10:

90, v/v), and H_2O . The protein band was electroeluted onto Hybond-P (GE Healthcare) and sequenced at Tufts University Core Facility (Boston, MA). The first 12 amino acid residues of the purified apoprotein were determined as DTVTVNYDXVG Y using automated Edman Degradation. The residue "X" is predicted to be a cysteine based on the encoding DNA sequence and was not detected by N-terminal sequencing.

Cloning of *mdpC4* and *mdpC1* for *E. coli* Expression. The genes for *mdpC4* and *mdpC1* were amplified by PCR using Expand Long Template PCR System from Roche (Indianapolis, IN) with 1X supplied buffer, 250 μ M dNTPs, 5% DMSO, 20 ng pBS10002, 5 U DNA polymerase, and 15 μ M each of the following primer pairs: *mdpC4* (forward) 5'-GGTATTGAGGGTCGCATGAAGGTGACTCAGACC-GAG-3'/(reverse) 5'-AGAGGAGAGTTAGAGCCTCAGCGCAGCTC-GACGCC-3'; and *mdpC1* (forward) 5'-GGTATTGAGGGTCGCAT-GACTGCGATCGGAACGGC-3'/(reverse) 5'-AGAGGAGAGTTA-GAGCCTCACCTTCACTCCGATCGCTC-3'. The PCR program included an initial hold at 96 °C for 2 min, followed by 30 cycles of 96 °C for 10 s, 56 °C for 30 s, and 68 °C for 40 s per kb of desired product. The gel-purified PCR product was inserted into pET-30 Xa/LIC using ligation-independent cloning as described by Novagen (Madison, WI), affording pBS10007 (for *mdpC4*) and pBS10008 (for *mdpC1*), and sequenced to confirm PCR fidelity.

Subcloning of *mdpC4* and *mdpC1* for *Streptomyces lividans* Expression. Plasmids pBS10007 and pBS10008 were digested with *Nde*I-*Eco*RI, and the 1.7-kb (*mdpC4*) and 3.5-kb (*mdpC1*) DNA fragments were purified and ligated to the identical sites of pUW201PW to give pBS10009 and pBS10010, respectively. Plasmids were transformed into *S. lividans* TK-64 by PEG mediated protoplast transformation following standard procedure. After 16 h, the transformation plates were overlaid with 1 mL of water supplemented with 50 μ g/mL thiostrepton. Single colonies were picked 3 days after overlay and transferred to a fresh R2YE-agar plate containing 50 μ g/mL thiostrepton for a second round of antibiotic selection. For protein production, a loopful of mycelium was inoculated into 5 mL of liquid R2YE medium supplemented with 25 μ g/mL thiostrepton and grown for 2 days at 28 °C. After homogenizing the mycelia, 1 mL was used to inoculate 50 mL of liquid R2YE medium in a 250-mL baffled flask containing 4 g of glass beads and 25 μ g/mL thiostrepton, and the resulting culture was incubated at 28 °C for 3 days at 250 rpm prior to harvesting.

Purification of His₆-MdpC4 from *S. lividans* (pBS1007) and His₆-MdpC1 from *S. lividans* (pBS1008). Cells were resuspended in 100 mM Tris-HCl, pH 8.0, 300 mM KCl, and 5 mM β -mercaptoethanol and lysed using a French Press with 2 passes at 15 000 psi. After centrifugation, the supernatant was loaded onto a column containing Ni-NTA agarose, and the proteins were purified as described by Qiagen (Valencia, CA) using a 20 mM imidazole wash and elution buffer containing 250 mM imidazole. Protein was desalted into 20 mM Tris-HCl, pH 8, and 0.5 mM DTT using PD-10 columns (GE Healthcare) and stored as 50% glycerol stocks at -20 °C. A dominant band corresponding to the expected molecular weight of MdpC4 (63 kD) and MdpC1 (122 kD) was detected using 10% SDS-PAGE.

Activity of His₆-MdpC4 and His₆-MdpC1. Activity assays for MdpC1 were performed and analyzed as described for SgcC1.^{54,55} Activity assays for MdpC4 were performed as described for SgcC4,^{51,52} and the reactions were analyzed using C18 reversed-phase HPLC with an analytical Apollo C18 column (250 mm \times 4.6 mm, 5 μ ; Grace Davison, Deerfield, IL). A series of linear gradients was developed from 0.1% trichloroacetic acid in 10% acetonitrile (A) to 0.1% trichloroacetic acid in 90% acetonitrile (B) in the following manner (beginning time and ending time with linear increase to % B): 0–2 min, 5% B; 2–19 min, 50% B; 19–24 min, 100% B; 24–32 min, 100% B; and 32–35 min, 5% B. The flow rate was kept constant at 1.0 mL/min, and elution was monitored at 280 nm. Peaks corresponding to L- α -tyrosine and (S)- β -tyrosine were identified by comparison to authentic standards (Sigma-Aldrich, St. Louis, MO and Peptech

(32) Gharahdaghi, F.; Weinberg, C. R.; Meagher, D. A.; Imai, B. S.; Mische, S. M. *Electrophoresis* **1999**, *20*, 601–605.

Corporation, Burlington, MA, respectively) and with comparisons to SgcC4 activity assays.

Results

Cloning, Sequencing, and Organization of the *mdp* Gene Cluster. Degenerate primers designed according to conserved regions found in the known NCS, C-1027, and CAL enediyne PKS and the associated upstream ORF were used to amplify the MDP enediyne PKS gene fragment from *A. madurae* ATCC 39144 genomic DNA.²⁴ The amplified fragment, exemplified by Probe-1 containing portions of *mdpE* and *mdpE5* (Figure 3A), was subsequently used to screen a cosmid library of *A. madurae*, with several cosmids resulting in positive hits from Southern hybridization. Of the positive clones, a single cosmid pBS10002 was chosen because end-sequencing of the DNA insert revealed sequences whose deduced gene products, MdpB at one end and MdpE2 at the other end, showed similarities to the chlorothricin 6-methylsalicylic acid synthase (an iterative type I PKS)⁴³ and an unknown protein associated with all known enediyne PKSs, respectively (Figure 3A). pBS10002 was fully sequenced by a shotgun approach, and the primary sequence of the DNA insert was reconstructed. Analysis of the potential ORFs revealed numerous gene products with a similarity to proteins contained within the C-1027²² and NCS²³ gene clusters. A second cosmid, pBS10003, was subsequently identified using Probe-2 from pBS10002 to rescreen the *A. madurae* cosmid library to obtain the upstream boundary, and a third cosmid (pBS10004) hybridizing to Probe-1 was chosen to complete the downstream boundary. Together the three overlapping cosmids cover 85 kb of contiguous DNA encompassing the entire *mdp* gene cluster (Figure 3A).

Functional Assignments of Genes within the *mdp* Cluster. The *mdp* gene cluster consists of minimally 42 *orfs* spanning a 52-kb contiguous DNA region responsible for biosynthesis, resistance, and transport of the MDP chromoprotein complex (Figure 3B and Table 1). Sequence analysis by BLAST comparison of gene products within the *mdp* cluster provided functional assignments that are consistent with five ORFs (MdpA1 through MdpA5) responsible for aminosugar production, three ORFs (MdpB through MdpB2) for 3,6-dimethylsalicylyl-CoA biosynthesis, eight ORFs (MdpC through MdpC8, excluding MdpC5) for (*S*)-3-(2-chloro-3-hydroxy-4-methoxyphenyl)-3-hydroxypropionic acid production, twelve ORFs (MdpE through MdpE11) for production of the enediyne core, three ORFs (MdpA6, MdpB3, and MdpC5) for the final convergence of the four building blocks into the MDP chromophore, and five ORFs (MdpR1 through MdpR4) responsible for transport, regulation, and resistance, including one putative ORF (MdpA) containing the expected amino acid composition for the MDP apoprotein. The DNA sequence of the entire cluster, including predicted genes and functional assignments for the gene products, has been deposited in GenBank under accession number AY271660.

The *mdp* Cluster Boundaries and Unassigned Gene Products. The *mdp* gene cluster is proposed to start at *mdpA1* and conclude at *mdpE2* (Table 1). The *orfs* upstream of *mdpA1* have gene products predicted to be involved in cell membrane biogenesis or possibly secondary metabolism of a distinct natural product, and the *orfs* downstream of *mdpE2* have gene products with similarities to enzymes involved in ATP biosynthesis. Furthermore, the gene products of *mdpA1* and *mdpE2* also

represent the furthest upstream and downstream sequence homologues with respect to the C-1027,²² NCS,²³ or CAL²⁷ gene clusters, although *orfs* outside the proposed boundary cannot be discounted precluding the development of a genetic system for *A. madurae*. Within the gene cluster, two P-450 oxidoreductases (MdpU1 and MdpU3) and an unknown protein MdpU2 that form a putative operon have not been assigned specific functions with respect to MDP biosynthesis. However, these ORFs are likely involved in aziridine formation to further process the enediyne core.

Resistance, Regulation, and Transport. Four genes (*mdpR1* through *mdpR4*) are predicted to be involved in resistance, regulation, and transport. Three of the four gene products, MdpR1, MdpR2, and MdpR3, have homologues within the C-1027²² and NCS²³ gene clusters. MdpR1 has sequence homology to the StrR family of regulatory proteins,³³ and MdpR2, to the large family of AraC-type transcriptional regulatory proteins.³⁴ MdpR3 is homologous to Pur8, which is a transmembrane protein that, when expressed in *S. lividans*, was implicated in late stage efflux of the antibiotic puromycin.^{35,36} MdpR4 has sequence similarity to conserved hypothetical proteins involved in DNA alkylation repair, although all of these protein homologues have putative assignments that are not based on definitive functions.

The MDP Apoprotein MdpA. Bioinformatics analysis of the *mdp* gene cluster revealed a single gene product, MdpA, had characteristics of the MDP apoprotein,¹⁶ including no homology to CagA⁵ or NcsA,⁶ a predicted leader peptide of 34 amino acids; no Y, H, R, and K residues in the mature protein; and a low isoelectric point (predicted pI of 3.7 with and 3.0 without the leader peptide). The *mdpA* gene is translated as a 166-amino acid protein, and SignalP analysis predicted a leader peptide that is cleaved between A32 and D33, resulting in a 133 amino acid protein with a predicted molecular weight of 12 928 Da and having no significant similarities to any protein deposited at NCBI.

The genuine identity of the MDP apoprotein was elucidated by fermentation of the *A. madurae* ATCC 39144 strain, isolation of the MDP chromoprotein complex as described previously using a bioassay against *Micrococcus luteus* as guidance,¹⁵ and direct characterization of the MDP chromoprotein complex. Thus, a portion of the purified MDP chromoprotein complex was extracted with methanol–ethyl acetate (1:1 ratio), and the resultant MDP chromophore was subjected to APCI-MS analysis yielding a pair of $[M + H]^+$ ions at $m/z = 779.3$ and $m/z = 781.3$ with a 3:1 ratio, consistent with the molecular formula of $C_{40}H_{43}N_2O_{12}Cl$ expected for the monochlorinated MDP–OCH₃ adduct (calculated 779.3 and 781.3).¹⁴ Also, two prominent $[M]^+$ fragment ions were observed at $m/z = 149.1$ and 294.2, consistent with the benzamide (molecular formula $C_9H_9NO_2$, calculated 149.0) and the entire ribopyranose moiety (molecular formula $C_{15}H_{20}NO_5$, calculated 294.0), respectively (Figure 2A).¹⁴ The remaining complex was analyzed by native PAGE, and the only detected band was acid-washed and eluted

(33) Distler, J.; Ebert, A.; Mansouri, K.; Pissowotzki, K.; Stockmann, M.; Piepersberg, W. *Nucleic Acids Res.* **1987**, *15*, 8041–8056.

(34) Rhee, S.; Martin, R. G.; Rosner, J. L.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 10413–10418.

(35) Tercero, J. A.; Espinosa, J. C.; Lacalle, R. A.; Jimenez, A. *J. Biol. Chem.* **1996**, *271*, 1579–1590.

(36) Tercero, J. A.; Lacalle, R. A.; Jimenez, A. *Eur. J. Biochem.* **1993**, *218*, 963–971.

Table 1. Deduced Functions of ORFs in the Maduropeptin Biosynthetic Gene Cluster

gene	size ^a	putative function	homologue ^b	identity%/similarity%
<i>orf(-5)</i>	444	dioxygenase	SimC5 (AAK06796)	41/52
<i>orf(-4)</i>	403	dioxygenase	SimC5 (AAK06796)	31/42
<i>orf(-3)</i>	226	transcriptional repressor	CmaC (AAK44709)	37/50
<i>orf(-2)</i>	557	methylmalonyl-CoA decarboxylase	Sim11 (NP_823479)	64/76
<i>orf(-1)</i>	162	glyoxylase/dioxygenase superfamily	SCO2237 (NP_626487)	47/62
Predicted Upstream Boundary of the <i>mdp</i> Cluster				
<i>mdpA1</i>	344	dTDP-1-glucose synthase	SgcA1 (AAL06657)	53/74
<i>mdpL</i>	460	epoxide hydrolase	UbiH (AAM78003)	51/64
<i>mdpC8</i>	340	Zn-dependent alcohol dehydrogenase	AdhP (NP_396172)	54/68
<i>mdpA4</i>	406	C-methyltransferase	McmC (CAK50790)	58/74
<i>mdpC</i>	539	monooxygenase	SgcC (AAL06674)	71/81
<i>mdpA3</i>	328	glucuronic acid decarboxylase	UXS1 (Q96V00)	53/62
<i>mdpA2</i>	447	NDP-glucose dehydrogenase	AlgD (AAB01487)	52/64
<i>mdpA5</i>	403	aminotransferase	ArmB (Q8ZNF3)	54/68
<i>mdpJ</i>	148	unknown	SgcJ (AAL06676)	63/80
<i>mdpD2</i>	394	oxidoreductase	SgcD2 (AAL06685)	67/79
<i>mdpB1</i>	348	C-methyltransferase	RemG (CAE51172)	28/43
<i>mdpB</i>	1747	6-methylsalicylic acid synthase	ChlB1 (AAZ77673)	51/63
<i>MdpB2</i>	548	adenylate ligase	NcsB2 (AAM77987)	44/53
<i>MdpB3</i>	285	putative hydrolase	PlaP6 (ABB69740)	33/42
<i>mdpA6</i>	466	glycosyltransferase	SgcA6 (AAL06670)	43/57
<i>mdpR4</i>	245	DNA alkylation repair	RBE_0692 (YP_537862)	36/56
<i>mdpU1</i>	409	P-450 oxidoreductase	CypX (AAC25766)	42/57
<i>mdpU2</i>	381	hypothetical protein	nfa43920 (BAD59241)	33/47
<i>mdpU3</i>	411	P-450 oxidoreductase	Cyp28 (NP_828645)	42/56
<i>mdpC7</i>	461	aminotransferase	BioA (Q47862)	38/55
<i>mdpA</i>	167 ^c	apoprotein	No	---
<i>mdpR3</i>	525	transmembrane efflux protein	Pur8 (P42670)	48/xx
<i>mdpH</i>	442	hydroxylase	SgcH (AAL06669)	48/63
<i>mdpC5</i>	460	NRPS condensation enzyme	SgcC5 (AAL06671)	45/58
<i>mdpC3</i>	518	halogenase	SgcC3 (AAL06656)	54/65
<i>mdpC4</i>	537	aminomutase	SgcC4 (AAL06680)	72/82
<i>mdpC1</i>	1115	NRPS adenylation enzyme	SgcC1 (AAL06681)	50/59
<i>mdpC2</i>	95	NRPS peptidyl carrier protein	SgcC2 (AAL06679)	38/52
<i>mdpC6</i>	198	O-methyltransferase	FkbG (AAF86386)	42/54
<i>mdpE11</i>	267	unknown	SgcE11 (AAL06691)	60/73
<i>mdpM</i>	365	unknown	SgcM (AAL06686)	29/35
<i>mdpE9</i>	552	oxidoreductase	SgcE9 (AAL06693)	74/81
<i>mdpE8</i>	189	unknown	SgcE8 (AY048670)	53/66
<i>mdpR1</i>	337	regulatory protein	StrR (CAA55579)	48/60
<i>mdpR2</i>	251	transcription regulatory protein	SgcR2 (AAL06696)	49/64
<i>mdpE7</i>	464	P-450 hydroxylase	SgcE7 (AAL06697)	60/69
<i>mdpE6</i>	183	flavin reductase	SgcE6 (AAL06698)	50/59
<i>mdpE10</i>	152	type II thioesterase	SgcE10 (AAL06692)	72/84
<i>mdpE</i>	1944	enediye PKS	SgcE (AAL06699)	56/63
<i>mdpE5</i>	330	unknown	SgcE5 (AAL06700)	70/78
<i>mdpE4</i>	678	unknown	SgcE4 (AAL06701)	56/67
<i>mdpE3</i>	300	hypothetical	SgcE3 (AAL06702)	47/56
<i>mdpE2</i>	331	unknown	SgcE2 (AAL06703)	55/61
Predicted Downstream Boundary of the <i>mdp</i> Cluster				
<i>orf1</i>	500	Aldehyde dehydrogenase	PutA (CAB71240)	51/65
<i>orf2</i>	162	acetyltransferase	blr0463 (BAC45728)	21/30
<i>orf3</i>	525	alcohol dehydrogenase	BetA (CAB51051)	43/57
<i>orf4</i>	208	unknown	COG2096 (INO_G_A)	27/43
<i>orf5</i>	523	predicted peptidase	Iap (NP_251629)	36/48

^a Numbers are in amino acids. ^b Given in parentheses are NCBI accession numbers; references, if not in the text, can be found therein. ^c Contains a 34 amino acid leader peptide.

onto a PVDF membrane for N-terminal sequencing using standard Edman degradation (Figure 2A). The N-terminal sequence of the first 12 residues matched that of the predicted mature protein, confirming MdpA as the elusive MDP apoprotein and simultaneously providing the first evidence that the *mdp* gene cluster was localized and cloned.

Aminosugar Biosynthesis. The biosynthesis of the aminosugar of MDP consists of five enzymatic steps beginning from D-glucose-1-phosphate prior to attachment to the enediye core (Table 1 and Figure 4A). MdpA1 is highly homologous to SgcA1, a biochemically characterized α -D-glucopyranosyl-1-

phosphate thymidyltransferase, and catalyzes the formation of TDP-glucose.³⁷ MdpA2, similar to the family of UDP-glucose dehydrogenases that includes *Pseudomonas aeruginosa* AlgD, catalyzes the net four-electron oxidation of TDP-glucose to form TDP-glucuronic acid.³⁸ MdpA3 is similar to the family of UDP-glucuronic acid decarboxylases and catalyzes oxidative decarboxylation of TDP-glucuronic acid to form thymidyl 5'- β -L-threo-pentapyranosyl-4''-ulose diphosphate.^{39,40} MdpA4 is

(37) Murrell, J. M.; Liu, W.; Shen, B. *J. Nat. Prod.* **2004**, *67*, 206–213.

(38) Naught, L. E.; Gilbert, S.; Imhoff, R.; Snook, C.; Beamer, L.; Tipton, P. *Biochemistry* **2002**, *41*, 9637–9645.

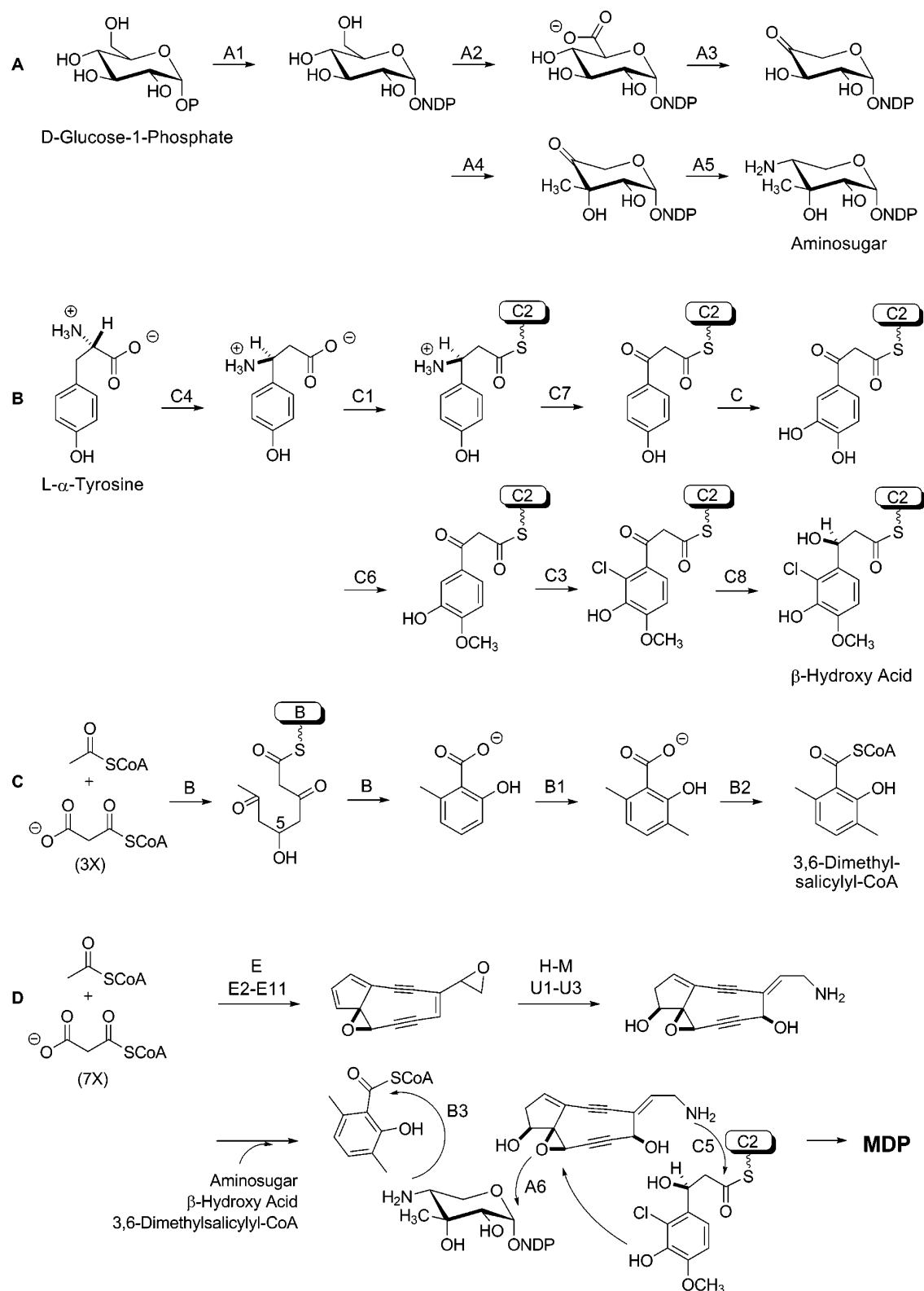


Figure 4. Proposed biosynthetic pathway for MDP chromophore: (A) aminosugar; (B) β -hydroxy acid; (C) 3,6-dimethylsalicylyl-CoA; and (D) enediyne core and convergent biosynthesis of four components to yield MDP chromophore.

homologous to a C-methyltransferase family that includes MmcC and TylC3, which are necessary for mycarose biosynthesis in mithramycin from *Streptomyces argillaceus* and tylosin from *Streptomyces fradiae*, respectively.⁴¹ Therefore, MdpA4

catalyzes the incorporation of a methyl group from S-adenosylmethionine (SAM) with concomitant hydroxyl epimerization to form thymidyl 5'-3-methyl- α -D-erythro-pentapyranosyl-4''-ulose diphosphate. Finally, MdpA5, similar to *Salmonella*

(39) Breazeale, S. D.; Ribeiro, A. A.; Raetz, C. R. H. *J. Biol. Chem.* **2002**, *277*, 2886–2896.

(40) Bar-Peled, M.; Griffith, C. L.; Doering, T. L. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12003–12008.

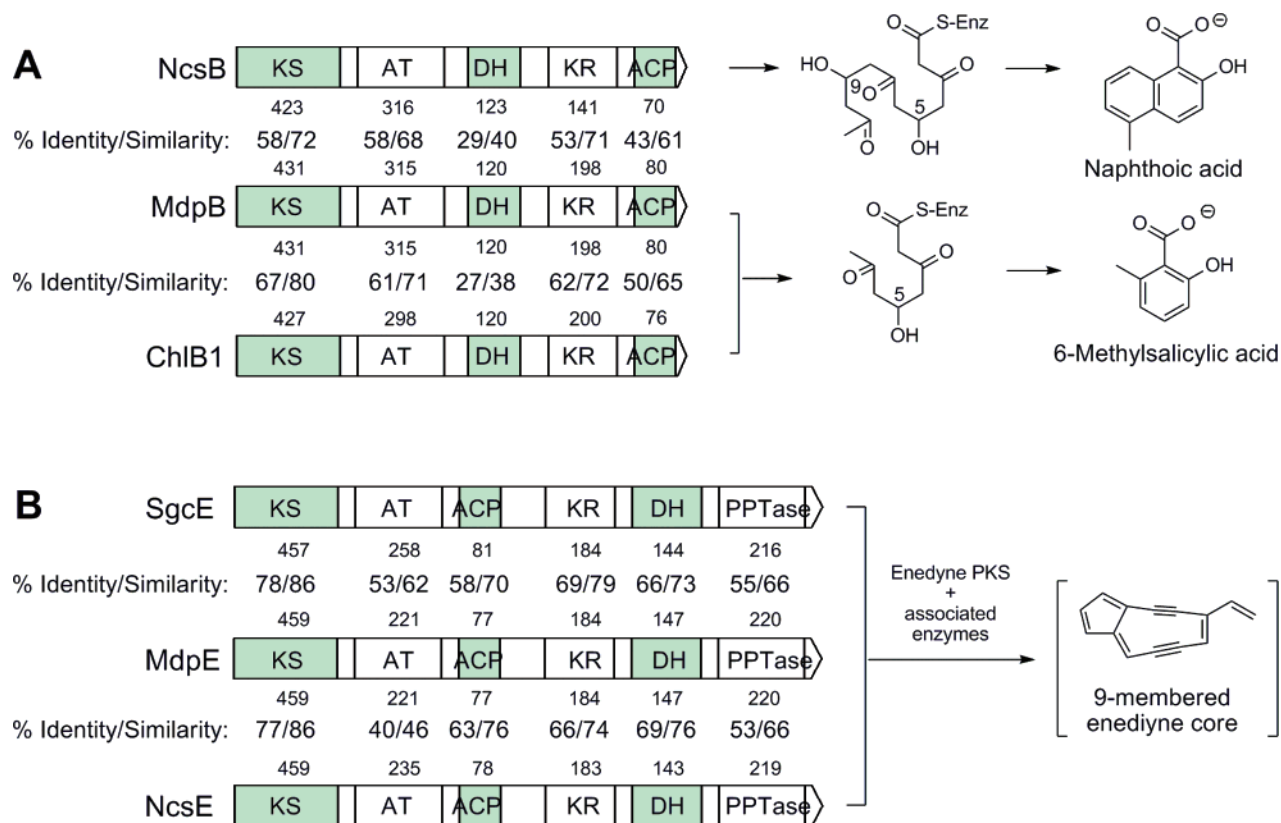


Figure 5. Comparisons of MdpB with selected bacterial iterative type I PKSs and MdpE with selected enediyne PKSs. (A) Domain organization of MdpB, a 6-methylsalicylic acid synthase (6-MSAS), and comparisons to NcsB involved in naphthoic acid biosynthesis and ChlB1 involved in 6-methylsalicylic acid biosynthesis. (B) Domain organization of MdpE and comparison to the enediyne PKS SgcE and NcsE. Abbreviations are KS, ketosynthase; AT, acyltransferase; ACP, acyl carrier protein; KR, ketoreductase; DH, dehydratase; and PPTase, phosphopantetheinyl transferase. The number of amino acids used for comparisons are given above and below the numbers of % identity/% similarity.

typhimurium ArnB implicated in lipid A biosynthesis,⁴² catalyzes a transamination leading to TDP-4-amino-4-deoxy-3-methyl- β -D-ribose.

3,6-Dimethylsalicylyl-CoA Biosynthesis. Three genes, *mdpB* through *mdpB2*, were identified within the *mdp* gene cluster that encode proteins with functions that would support their involvement in the biosynthesis of the 3,6-dimethylsalicylic acid moiety of MDP (Table 1 and Figure 4C). The gene for *mdpB* encodes a PKS with five known domains: a ketosynthase (KS), an acyltransferase (AT), a dehydratase (DH), a ketoreductase (KR), and an acyl carrier protein (ACP). MdpB is most similar to ChlB1, an iterative type I PKS involved in 6-methylsalicylic acid production during chlorothricin biosynthesis,^{43,44} but is also similar to NcsB, responsible for the production of 2-hydroxy-5-methyl-1-naphthoic acid moiety of neocarzinostatin (Figure 5A).^{23,45} This newly discovered family of bacterial iterative type I PKSs⁴⁶ has identical domain organization to that of *Aspergillus terreus* and other fungal 6-methylsalicylic acid synthases (6-

MSASs) that catalyze 3 rounds of decarboxylative condensation to afford a tetraketide.⁴⁷ As a consequence of these comparisons, MdpB is proposed to catalyze the synthesis of 6-methylsalicylic acid through 3 iterations with the regiospecific ketoreduction at position 5 that is seen with 6-MSAS. MdpB1 contains an *S*-adenosylmethionine binding motif found in several O-, N-, and C-methyltransferases⁴⁸ and is proposed to catalyze C-methyltransferase activity to generate 3,6-dimethylsalicylic acid. The closest homologues of MdpB1, RemG and RemH, are implicated in the unprecedented geminal bis-methylation of the aromatic polyketide resistomycin from *Streptomyces resistomycificus*,⁴⁹ supporting the proposed role for MdpB1. Finally, MdpB2 utilizes the free acid to generate the 3,6-dimethylsalicylyl-CoA thioester. The closest homologue of MdpB2, NcsB2, has recently been characterized as a naphthoic acid-CoA ligase supporting the proposed assignment for this enzyme.⁵⁰

(S)-3-(2-Chloro-3-hydroxy-4-methoxyphenyl)-3-hydroxypropionic Acid Biosynthesis. Eight genes (*mdpC* to *mdpC8* excluding *mdpC5*) were identified to process L- α -tyrosine to form the (S)-3-(2-chloro-3-hydroxy-4-methoxyphenyl)-3-hydroxypropionic acid moiety of MDP (Table 1 and Figure 4B). MdpC through MdpC5 have closest sequence homology with enzymes involved in the biosynthesis of (S)-3-chloro-5-hydroxy- β -tyrosine of C-1027, with sequence identities ranging from 38%

(41) González, A.; Remsing, L. L.; Lombó, F.; Fernández, M. J.; Prado, L.; Braña, A. F.; Künzel, E.; Rohr, J.; Méndez, C.; Salas, J. A. *Mol. Gen. Genet.* **2004**, *264*, 827–835.

(42) Noland, B. W.; Newman, J. M.; Hendle, H.; Badger, J.; Christopher, J. A.; Tresser, J.; Buchanan, M. D.; Wright, T. A.; Rutter, M. E.; Sanderson, W. E.; Müller-Dieckmann, H.-J.; Gajiwala, K. S.; Buchanan, S. G. *Structure* **2002**, *10*, 1569–1580.

(43) Shao, L.; Qu, X.; Jia, X.; Zhao, Q.; Tian, Z.; Wang, M.; Tang, G.; Liu, W. *Biochem. Biophys. Res. Commun.* **2006**, *345*, 133–139.

(44) Jia, X.-Y.; Tian, Z.-H.; Shao, L.; Qu, X.-D.; Zhao, Q.-F.; Tang, J.; Tang, G.-L.; Liu, W. *Chem. Biol.* **2006**, *13*, 575–585.

(45) Sthapit, B.; Oh, T. J.; Lamichhane, R.; Liou, K.; Lee, H. C.; Kim, C. G.; Sohng, J. K. *FEBS Lett.* **2004**, *566*, 201–206.

(46) Shen, B. *Curr. Opin. Chem. Biol.* **2003**, *7*, 285–295.

(47) Fujii, I.; Ono, Y.; Tada, H.; Gomi, K.; Ebizuka, Y.; Sankawa, U. *Mol. Gen. Genet.* **1996**, *253*, 1–10.

(48) Kozbial, P. Z.; Mushegian, A. R. *BMC Struct. Biol.* **2005**, *5*, 1–26.

(49) Jakobi, K.; Hertweck, C. *J. Am. Chem. Soc.* **2004**, *126*, 2298–2299.

(50) Cooke, H. A.; Zhang, J.; Griffin, M. A.; Van, Lanen, S. G.; Shen, B.; Bruner, S. D. *J. Am. Chem. Soc.* **2007**, *129*, 7728–7729.

to 72%.²² The initial step is catalyzed by MdpC4, a 4-methylideneimidazole-5-one (MIO) containing aminomutase similar to SgcC4, to generate (*S*)- β -tyrosine from L- α -tyrosine.^{51–53} Loading of (*S*)- β -tyrosine to a free-standing peptidyl carrier protein (PCP) (MdpC2) is then achieved by a discrete adenylation enzyme, MdpC1, which has closest homology to the C-1027 biosynthetic homologue SgcC1 that has been recently characterized and shown to prefer (*S*)- β -tyrosine as a substrate.^{54,55} Following activation and tethering of the (*S*)- β -Tyr as the (*S*)- β -tyrosyl-*S*-MdpC2 intermediate, a transamination occurs to eliminate the β -amino group catalyzed by the PLP-dependent transaminase MdpC7, homologous to the aminotransferase BioA family involved in cofactor metabolism.⁵⁶ The 4'-hydroxyphenyl-3-ketopropionic acid, thioester-tethered to MdpC2, is next modified by a hydroxylase MdpC (similar to SgcC),²² an *O*-methyltransferase MdpC6,⁵⁷ and a halogenase MdpC3 (similar to SgcC3).⁵⁸ The (*S*)- β -hydroxy functionality is subsequently formed by stereospecific reduction of the β -carbonyl group, the most likely candidate for which is MdpC8, a homologue of a family of hypothetical zinc-dependent alcohol dehydrogenases with unknown function.

MdpC4 and MdpC1 Enzyme Activity. The activity of MdpC4 and MdpC1 was tested to confirm the origin of the (*S*)-3-(2-chloro-3-hydroxy-4-methoxyphenyl)-3-hydroxypropionic acid moiety of MDP and to demonstrate this biosynthetic pathway mimics the assembly of the (*S*)-3-chloro-4,5-dihydroxy- β -phenylalanine moiety found in C-1027. Recombinant MdpC4 was first purified from *S. lividans*, and the purified enzyme was incubated with L- α -tyrosine under the assay conditions as described for SgcC4.^{51,52} Using HPLC analysis with the free acids, a new peak appeared eluting before the substrate, and this new peak had an identical retention time to that of authentic (*S*)- β -tyrosine, which was confirmed by co-injections and comparisons using SgcC4-catalyzed reactions (Figure 6A). Similar to SgcC4, the MdpC4-catalyzed reaction is reversible with a 7:3 ratio of (*S*)- β -tyrosine/L- α -tyrosine observed at equilibrium (Figure 6A). All hypothetical α -amino acid substrates, such as L-4-methoxyphenylalanine, L-3-chlorotyrosine, and L-3-hydroxytyrosine, were not recognized by MdpC4.

Recombinant MdpC1 was partially purified from *S. lividans*, and the enzyme was subjected to the same preliminary analysis as that described for SgcC1.^{54,55} Using the amino acid dependent ATP-PP_i exchange assay, MdpC1 was unable to recognize L- α -tyrosine, but activity was observed when MdpC1 was incubated with (*S*)- β -tyrosine (Figure 6B). Similarly to SgcC1, MdpC1 was active with other β -amino acid substrates including (*R*)- β -tyrosine (75%), 3-hydroxy- β -tyrosine (39%), and 3-chloro- β -tyrosine (24%).

Enediynes Core Biosynthesis and Convergent Biosynthesis for the MDP Chromophore. The MDP enediynes core is

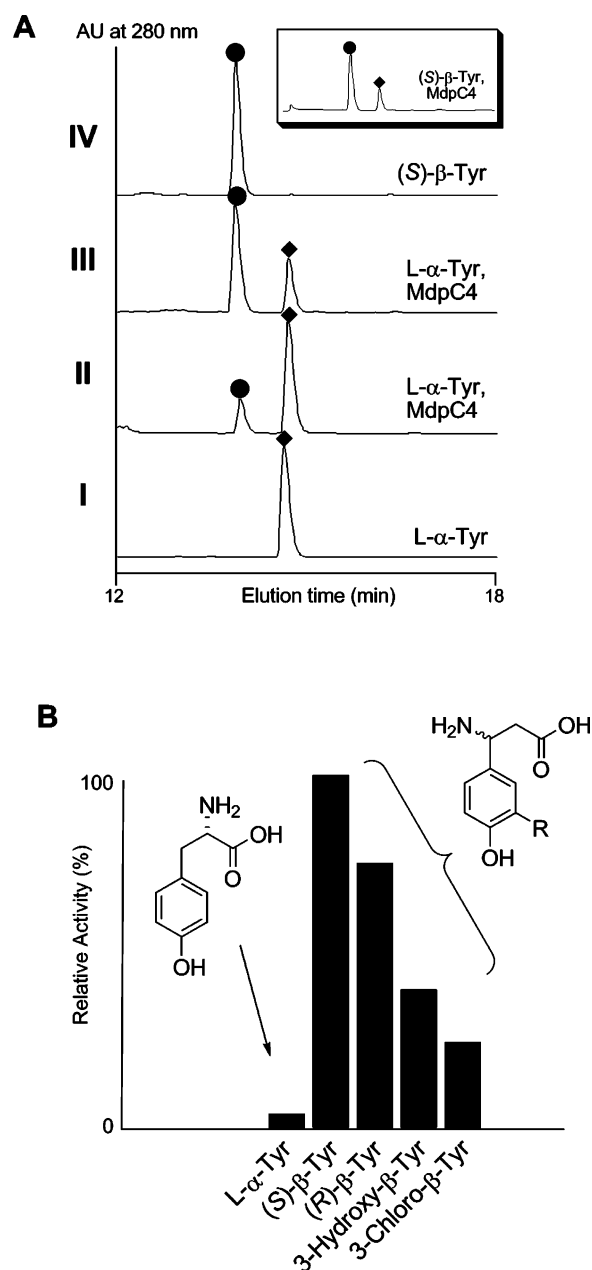


Figure 6. Activity of MdpC4 and MdpC1. (A) HPLC profile of authentic L- α -tyrosine (I), L- α -tyrosine incubated 30 min with MdpC4 (II), L- α -tyrosine incubated overnight with MdpC4 (III), and authentic (*S*)- β -tyrosine (IV). Shown in the inset is the overnight incubation of (*S*)- β -tyrosine with MdpC4, which gives the identical profile and peak ratio compared to the forward reaction (III). Symbols are (●) for L- α -tyrosine and (◆) for (*S*)- β -tyrosine. (B) Amino acid dependent ATP-PP_i exchange assay for MdpC1 comparing L- α -tyrosine with β -tyrosine analogues where R is the respective substituent.

assembled by an iterative type I PKS, MdpE, that has high sequence homology and identical domain organization to other known enediynes PKSs (Figure 5B).^{22–25} Four domains are evident by sequence analysis, including a KS, AT, dehydratase (DH), and KR. Two additional domains have been proposed for the enediynes PKS family, a central ACP and a C-terminal phosphopantetheinyltransferase (PPTase) domain.

MdpE is envisioned to assemble a nascent linear polyunsaturated polyketide backbone that is further modified by tailoring enzymes, including a minimum of 10 additional proteins (MdpE2 through MdpE11) (Table 1 and Figure 4D).^{24,25,46}

- (51) Christenson, S. D.; Liu, W.; Toney, M. D.; Shen, B. J. *Am. Chem. Soc.* **2003**, *125*, 6062–6063.
- (52) Christenson, S. D.; Wu, W.; Spies, M. A.; Shen, B.; Toney, M. D. *Biochemistry* **2003**, *42*, 12708–12718.
- (53) Christianson, C. V.; Montavon, T. J.; Van Lanen, S. G.; Shen, B.; Bruner, S. D. *Biochemistry* **2007**, *46*, 7205–7214.
- (54) Van, Lanen, S. G.; Dorrestein, P. C.; Christenson, S. D.; Liu, W.; Ju, J.; Kelleher, N. L.; Shen, B. J. *Am. Chem. Soc.* **2005**, *127*, 11594–11595.
- (55) Van, Lanen, S. G.; Dorrestein, P. C.; Lin, S.; Kelleher, N. L.; Shen, B. J. *Biol. Chem.* **2006**, *281*, 29633–29640.
- (56) Wu, C. H.; Chen, H. Y.; Shiu, D. *Gene* **1996**, *174*, 251–258.
- (57) Wu, K.; Chung, L.; Revill, W. P.; Katz, L.; Reeves, C. D. *Gene* **2000**, *251*, 81–90.
- (58) Lin, S.; Van Lanen, S. G.; Shen, B. J. *Am. Chem. Soc.*, published online Sep 22, 2007 <http://dx.doi.org/10.1021/ja072311g>.

Comparison to the C-1027²² and NCS²³ gene clusters suggests MdpH, MdpJ, MdpL, and MdpM — a hydroxylase, an unknown protein, an epoxide hydrolase, and an unknown protein, respectively — are also required for processing the MDP enediynes core. Upon formation of the enediynes core intermediate, the aminosugar madurose is attached via MdpA6, a glycosyltransferase; the 3,6-dimethylsalicylic acid via MdpB3, a putative acyltransferase/esterase; and the (*S*)-3-(2-chloro-3-hydroxy-4-methoxyphenyl)-3-hydroxypropionic acid via MdpC5, a condensation enzyme, although the timing of the coupling steps is still unclear.

Heterologous Complementation Using MdpE. To circumvent the difficulty in developing a genetic system for *A. madurea*, heterologous complementation of the respective enediynes PKS for C-1027 production in *S. globisporus* and NCS production in *S. neocarzinostaticus* was chosen to confirm the cloning and localization of the *mdp* gene cluster. Enediynes PKS inactivated mutant strains for C-1027, *S. globisporus* SB1005²² and NCS, *S. carzinostaticus* SB2002²³ have been previously constructed and chosen for cross-complementation using *mdpE* and its associate genes. The *mdpE* enediynes PKS expression vector pBS10007 was constructed in the integrative vector pSET152²⁹ that can be stably maintained via site specific integration into the chromosome. pBS10007, in which the expression of *mdpE*, together with its immediate upstream gene *mdpE5* and downstream genes *mdpE10* and *mdpE6*, is under the control of the constitutive *Erme** promoter,²⁹ was introduced into SB1005 and SB2002, with pBS1019 (an *sgcE* expression construct) and pBS2020 (an *ncsE5-ncsE-ncsE10* expression construct) as positive controls, respectively.

The resulting *S. globisporus* SB1010 and SB1011 recombinant strains were cultured and examined for C-1027 production by HPLC analysis with the *S. globisporus* wild-type strain as a positive control (Figure 7A). Similarly, the resulting *S. carzinostaticus* SB5003 and SB5007 recombinant strains were cultured and examined for NCS production by HPLC analysis with the *S. carzinostaticus* wild-type strain as a positive control (Figure 7B). Production of C-1027 in SB1011 and NCS in SB5007 was restored to a detectable level, albeit to around 20% of C-1027 in SB1010 and 50% of NCS in SB5003, respectively (Figure 7). The HPLC fraction containing the expected C-1027 chromophore and its aromatized product from the SB1011 recombinant strain was analyzed by high-resolution MALDI-MS, yielding the characteristic $[M + H]^+$ ions at $m/z = 844.259$ and $m/z = 846.273$, consistent with the molecular formula of $C_{43}H_{42}N_3O_{13}Cl$ (enediynes form calculated 843.241) and $C_{43}H_{44}N_3O_{13}Cl$ (aromatized form calculated 845.256), respectively, and identical to that isolated from the *S. globisporus* wild-type strain.²² Likewise, LC-ESI-MS was used to confirm the NCS chromophore from the SB5007 recombinant strain, giving an $[M + H]^+$ ion at $m/z = 660.3$, consistent with the molecular formula of $C_{35}H_{33}NO_{12}$ (calculated 659.2) and identical to that found from the *S. carzinostaticus* wild-type strain.²³ In total, the data show that the cloned *mdpE* enediynes PKS gene can functionally cross-complement the *sgcE* or *ncsE* enediynes PKS gene for C-1027 or NCS biosynthesis in their native producer, respectively.

Discussion

The biosynthetic gene clusters for C-1027 and NCS, two prominent members of the nine-membered enediynes family,

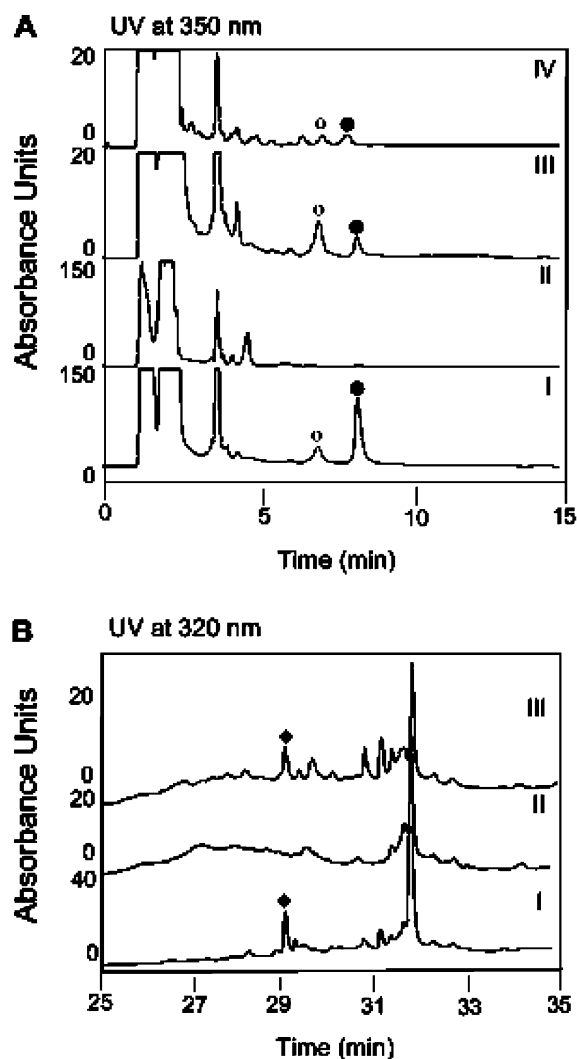


Figure 7. Heterologous complementation of C-1027 and NCS enediynes PKS mutants by the MdpE enediynes PKS cassette. (A) Complementation of C-1027 enediynes polyketide synthase mutant (SB1005) with empty vector (II); *sgcE*, plasmid pBS1019 (III); and *mdpE/E5/E6/E10*, plasmid pBS3010 (IV). For comparisons, production of C-1027 from wild-type producer is shown (I). (B) Complementation of NCS enediynes polyketide synthase mutant (SB5002) with *ncsE*, plasmid pBS5019 (I); empty vector (II); and *mdpE/E5/E6/E10*, plasmid pBS10005 (III). Symbols are (●) for C-1027 in the enediynes form, (○) for C-1027 aromatized product, and (◆) for NCS.

have been cloned and sequenced, and the gene products were functionally assigned based on bioinformatics analysis and biochemical studies.^{22–26,37,50–55,58} A third nine-membered enediynes, MDP, was targeted to provide further evidence for our proposed unifying paradigm for enediynes biosynthesis and to facilitate metabolic engineering of the enediynes family for novel analogues. Despite varying degrees of structural uniqueness, all three members share a virtually identical enediynes core (Figure 1), and therefore the gene clusters were hypothesized to harbor genes encoding a similar enediynes PKS and associated enzymes.^{22–27} Indeed, knowledge of the shared minimal enediynes PKS gene cassettes for the previously cloned enediynes clusters was successfully employed to probe DNA libraries in order to isolate other enediynes producers as well as to locate the MDP enediynes PKS locus.^{24,25} This approach was used here as a starting point for chromosomal walking to ultimately lead to the isolation of the three overlapping cosmids covering the entire *mdp* gene cluster (Figure 3).

Consistent with the characterization of C-1027²² and NCS²³ gene clusters, the biosynthesis of the MDP chromophore occurs by the assembly of individual components before convergence to form the final enediyne chromophore. As expected the biosynthesis of the aminosugar moiety begins by the conversion of glucose-1-phosphate to the nucleotide sugar diphosphate. Based on high sequence identity to SgcA1, which is involved in C-1027 deoxy aminosugar biosynthesis and has been characterized biochemically,³⁷ the nucleotide is proposed to be TTP. The two subsequent steps involve formation of a C6-carboxylic acid followed by oxidative decarboxylation to generate a 4-keto-pentapyranose. These two enzymatic steps are unusual for natural product biosynthesis but have been proposed and characterized for the production of lipid A that is implicated in polymyxin resistance of gram-negative bacteria such as *S. typhimurium* and *E. coli*.^{39,40} After production of the 4-keto-pentapyranose, C-methylation and hydroxyl epimerization at C3 occur through an enolate intermediate, and the overall conversion has been demonstrated for the homologue TylC3 involved in tylosin biosynthesis.⁵⁹ The final step before addition to the aglycon is transamination, an enzymatic step that is commonly seen in the biosynthesis of monosaccharides in secondary metabolism.

The MDP chromophore contains two monocyclic aromatic moieties in addition to the enediyne core (Figure 2), and subsequently the gene cluster was found to contain a second iterative type I PKS that is distinct from the MdpE enediyne PKS. Although MdpB is most similar to ChlB1, a bacterial 6-MSAS in the chlorothricin gene cluster,^{43,44} MdpB, also has sequence similarity to NcsB, which is responsible for the naphthoic acid moiety of NCS^{23,45} (Figure 5A). However, unlike NcsB, MdpB is expected to catalyze only three rounds of decarboxylative condensation with regiospecific ketoreduction at C5 instead of five iterations and ketoreduction at C5 and C9 as proposed for NcsB. Despite this variance in chain length, MdpB has a higher identity to NcsB (48% identity/64% similarity) than the functionally equivalent fungal 6-MSAS counterpart (41% identity/56% similarity). The addition of MdpB to this rapidly growing family of bacterial iterative type I PKS, which now consists of minimally five enzymes,^{43,44,46} should ultimately aid in deciphering the molecular details inherent in the iterative type I PKSs that control the chain length, regiospecific reduction, and hence the structure of final aromatic polyketides in bacteria.

The second monocyclic aromatic moiety of MDP is the (*S*)-3-(2-chloro-3-hydroxy-4-methoxyphenyl)-3-hydroxypropionic acid, the origin of which only became evident using comparative genomic analysis. The moiety originates from L- α -tyrosine, as deciphered from the discovery of homologues with >50% identity (excluding only MdpC2) to the biosynthetic enzymes involved in production of the β -amino acid of C-1027.²² The majority of the pathway to the β -amino acid moiety in C-1027 has now been reconstituted using recombinant enzymes,^{51–55,58} and the two initial steps for MDP biosynthesis have been shown here to be identical: MdpC4 is a L- α -tyrosine aminomutase and MdpC1 is an (*S*)- β -tyrosine activating adenylation enzyme. In both cases, L- α -tyrosine is first converted to the (*S*)- β -tyrosine by a growing family of MIO-containing aminomutases using

no exogenous cofactors, now consisting of four enzymes,^{51–53,60,61} followed by activation and loading of the resultant (*S*)- β -tyrosine onto the free-standing MdpC2 PCP. While further transformations are paralleled, the halogenation and hydroxylation of the (*S*)- β -tyrosyl-S-MdpC2 intermediate differ in regiochemistry; MDP biosynthesis also requires three additional reactions: transamination, carbonyl reduction, and O-methylation. The timing of these events is currently under investigation.

The enediynes of the nine-membered family are isolated as a chomoprotein complex, and the genes encoding the apoprotein for C-1027 and NCS both reside within the boundaries for the respective gene cluster.^{22,23} Likewise, it was hypothesized the MDP apoprotein was also encoded within the gene cluster, although a functional assignment based on bioinformatics alone was inconclusive because (i) the MDP apoprotein is not homologous to either CagA or NcsA, (ii) only the relative abundance of amino acids, not the amino acid sequence of the apoprotein, is known, and (iii) previous reports identified the MDP apoprotein as having molecular weights of 22.5 kD by gel filtration,¹⁶ 32 kD by SDS-PAGE,¹⁴ and 13 kD by mass spectrometry.⁶² Of the 42 ORFs predicted to be involved in MDP biosynthesis, MdpA was the only logical candidate as the apoprotein based on (i) elimination of other ORFs from functional assignments, (ii) the prediction of a 32-amino acid leader peptide, and (iii) the lack of H, K, W, and R residues, the last of which was reported during the initial characterization of the MDP chromoprotein complex.¹⁶

To experimentally identify the MDP apoprotein and to resolve the size ambiguities in the literature, *A. madurae* was fermented and the MDP chromoprotein complex was isolated. The N-terminal sequence of the resultant apoprotein was determined and confirmed to be identical to that of the predicted mature MdpA. Interestingly, while the nine-membered enediyne apoproteins all share ~40% identity, MdpA has no similarities to proteins deposited at NCBI which includes CagA²² and NcsA.²³ Therefore, the discovery of MdpA as the MDP apoprotein not only confirms that we have localized the *mdp* gene cluster but also establishes MdpA as a new type of enediyne apoprotein. Furthermore, based on ORF analysis, MdpA is translated as a 16 212 Da protein, and N-terminal sequencing revealed MdpA is processed to yield a 12 928 Da protein, which is consistent with the data obtained by mass spectrometry.⁶²

To further verify the identification of the *mdp* gene cluster, heterologous complementation was chosen to circumvent the difficulty in developing a genetic system for the MDP producer, *A. madurae*. We have previously shown that (i) inactivation of *sgcE* and *ncsE* afforded *S. globisporus* SB1005 and *S. carzino-staticus* SB5002 mutant strains whose ability to produce C-1027 and NCS was completely abolished and (ii) C-1027 and NCS production could be restored to the SB1005 and SB5002 mutant strains upon expression of a functional copy of *sgcE* and *ncsE5-ncsE-ncsE10* in trans, respectively.^{22,23} Upon introduction of the *mdpE* enediyne cassette expression plasmid pBS10007 to SB1005 or SB5002, production of C-1027 or NCS was indeed successfully restored, respectively, in the resultant

(59) Chen, H.; Zhao, Z.; Hallis, T. M.; Guo, Z.; Liu, H. *Angew. Chem., Int. Ed.* **2001**, 40, 607–610.

(60) Jin, M.; Fischbach, M. A.; Clardy, J. *J. Am. Chem. Soc.* **2006**, 128, 10660–10661.

(61) Walker, K. D.; Klettke, K.; Akiyama, T.; Croteau, R. *J. Biol. Chem.* **2004**, 279, 53947–53954.

(62) Zein, N.; Schroeder, D. R. In *Advances in DNA Sequence-Specific Agents*; Palumbo, M., Ed.; Elsevier: Cambridge, MA, 1998; Vol. 3, pp 201–225.

recombinant strains (Figure 7). Although yields were low, ranging from 20% for C-1027 in SB1011 and 50% for NCS in SB5007 in comparison with that found from complementation with its cognate enediyne PKS *sgcE* in SB1010 and *ncsE5-ncsE-ncsE10* in SB5003, the results definitively demonstrated that the *mdp* enediyne PKS cassette can functionally cross-complement the *sgcE* or *ncsE* mutation in their respective producer, supporting that the cloned gene cluster does indeed encode the proteins for MDP production.

Of utmost significance for our future combinatorial biosynthesis strategies, cross-complementation between *mdpE* and *sgcE* and between *mdpE* and *sgcE* is consistent with the biosynthesis of the enediyne core occurring through a common mechanism for the nine-membered enediyne family. Although this unified mechanism for enediyne core biosynthesis has been speculated previously on the basis of bioinformatics analysis,^{24,25} this is for the first time now supported by experimental evidence. Finally, the ability to cross-complement the enediyne biosynthetic machinery suggests *S. globisporus* and *S. carzinostaticus* have the appropriate genetic and biochemical traits to facilitate rational genetic engineering to control the chemical features and structural diversity of the enediynes. Therefore, these organisms represent model host organisms for combinatorial biosynthesis to generate new enediyne compounds.⁶³

Conclusion

The gene cluster involved in the biosynthesis, resistance, regulation, and transport of the enediyne MDP was localized and cloned from *A. madurea* ATCC 39144 and sequenced and functionally annotated. Predictions from bioinformatics analysis were consistent with the expected enzymology with several unusual findings, including the uncovering of a second iterative type I PKS for aromatic polyketide biosynthesis and a pathway to a β -hydroxy acid that mirrors β -amino acid formation in C-1027, featuring a rare MIO-containing aminomutase and a set of discrete adenylation, PCP, and condensation enzymes for

β -amino or β -hydroxy acid activation and subsequent incorporation into natural products. In a process similar to that for the other enediynes, the MDP chromophore is formed from the convergence of four building blocks: (i) the enediyne core, (ii) a 3,6-dimethylsalicylyl-CoA, (iii) an aminosugar, and (iv) a halogenated β -phenyl- β -hydroxy propionic acid. Cross-complementation of the *sgcE* or *ncsE* enediyne PKS mutants by the *mdpE* enediyne PKS cassette provided convincing evidence that the MDP gene cluster was identified and also supported a unifying paradigm for nine-membered enediyne core biosynthesis. By establishing the biosynthetic machinery for three individual enediyne gene clusters (i.e., C-1027, MDP, and NCS), comparative genetics should undoubtedly facilitate combinatorial biosynthesis for the discovery of new enediynes. Finally, the MDP apoprotein, MdpA, was identified by isolation of the MDP chromoprotein complex and N-terminal sequencing of the purified apoprotein. MdpA, like the other apoproteins for the nine-membered enediyne chromoproteins, is located within the biosynthetic gene cluster. In contrast, however, MdpA represents a new type of apoprotein with no sequence homology to other known proteins, providing new opportunities to explore apoprotein and enediyne chromophore interaction, enediyne chromophore stabilization, and chromoprotein activation.

Acknowledgment. We thank Dr. Y. Li, Institute of Medicinal Biotechnology, CAMS, Beijing, China for the wild-type *S. globisporus* strain and the AIC of the School of Pharmacy, UW-Madison for support in obtaining MS data. This work is supported in part by NIH Grants CA78747 and CA113297. S.V.L. is the recipient of an NIH postdoctoral fellowship (CA1059845). Dedicated to the memory of Prof. Ian Scott for his pioneering work on natural product biosynthesis and enzyme reaction mechanism.

Note Added after ASAP Publication: Due to a production error, Figures 4 and 5 were incorrect in the version published on the Web on October 6, 2007. The corrected version was published on the Web on October 9, 2007.

JA073275O

(63) Kennedy, D. R.; Gawron, L. S.; Ju, J.; Liu, W.; Shen, B.; Beerman, T. A. *Cancer Res.* **2007**, *67*, 773–781.