Cloning and sequencing of the kedarcidin biosynthetic gene cluster from *Streptoalloteichus* sp. ATCC 53650 revealing new insights into biosynthesis of the enediyne family of antitumor antibiotics

Jeremy R. Lohman,[a] Sheng-Xiong Huang,[a] Geoffrey P. Horsman,[b] Paul E. Dilfer,[a] Tingting Huang,[a] Yihua Chen,[b] Evelyn Wendt-Pienkowski,[b] and Ben Shen[*,a,b,c,d]

[a]Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, USA
[b]Division of Pharmaceutical Sciences, School of Pharmacy, University of Wisconsin-Madison, Madison, Wisconsin, 53705, USA
[c]Department of Molecular Therapeutics, The Scripps Research Institute, Jupiter, Florida 33458, USA
[d]Natural Products Library Initiative at The Scripps Research Institute, The Scripps Research Institute, Jupiter, Florida 33458, USA
** The Scripps Research Institute, 130 Scripps Way, #3A1, Jupiter, Florida 33458, USA; Tel: (561) 228-2456; Fax: (561) 228-2472; Email:shenb@scripps.edu*

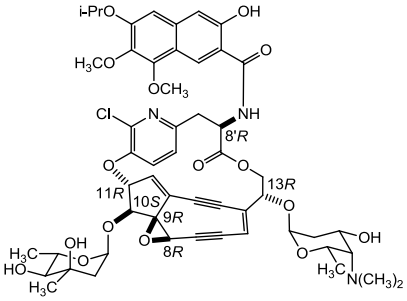
Supplemental Information

Fig S1. The amino acid sequence of the KED apoprotein deduced from *kedA* and comparison with the KED apoprotein determined by Edman degradation[1] as well as other known apoproteins.[2] KedA, deduced from *kedA* (the *kedA* gene has also been recently deposited to the GeneBank under accession no. ADG01894[3]); KedA*, determined by Edman degradation[1]; CagA (accession no. D10827), C-1027; AxnA (D11457), the enediyne structure is not known (AxnA is identical to CagA except for the D122A variation, which could be resulted from sequence error); NcsA (D10996), neocarzinostatin; McmA (D90006), auromomycin (whose structure has not been determined). EFE72430 (from *Streptomyces ghanaensis* ATCC 14672), ACZ98704 (from *Streptosporangium roseum*), EDX26137 (from *Streptomyces* sp. Mg1), and ADJ47157 (from *Amycolatopsis mediterranei*, strain U-32) are predicted apoproteins from putative enediyne biosynthetic gene clusters unveiled during genome sequencing efforts. Neither their enediyne chromoprotein production nor the enediyne chromophore structures have been established. The leader peptides are shown in italic and the first amino acid of the mature apoproteins are shown in red. ACZ98704 was predicted to have two cut sites. In addition to the predicted N-terminus of Ala-Ser-Ala-Ala-Val, two variants of the KedA apoprotein with a Ser-Ala-Ala-Val or an Ala-Ala-Val terminus were isolated.[1]

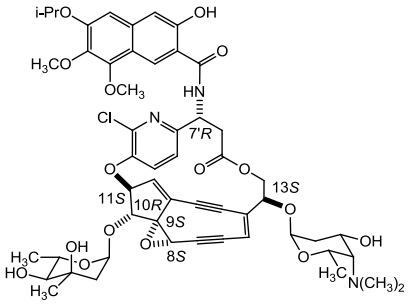
```
KedA        --MLQWKTALKTSAALVAAAGFALTLGTPA-SAAASA-----------AVSVSPATGLADGATVTVSASGFATSTSATALQCAILADGRG 76
KedA*       ----------------------------------ASA-----------AVSVSPATGLADGATVTVSASGFATSTSATALQCAILADGRG 45
CagA        -MSLRHMSRRASRFGVVAVASIGLAAAAQS-VAFAAP-----------AFSVSPASGLSDGQSVSVSVSGAAAGETYYIAQCAPVG-GQD 76
AxnA        -MSLRHMSRRASRFGVVAVASIGLAAAAQS-VAFAAP-----------AFSVSPASGLSDGQSVSVSVSGAAAGETYYIAQCAPVG-GQD 76
NcsA        MVPISIIRNRVAKVAVGSAAVLGLAVGFQTPAVAAAP-----------TATVTPSSGLSDGTVVKVAGAGLQAGTAYDVGQCAWVDTGVL 79
McmA        --MLQNTSRFLARAGATVGVAAGLAFSLPADRDGAPG------------VTVTPATGLSNGQTVTVSATGLTPGTVYHVGQCAVVEPGVI 76
EFE72430    -MTVKNKLGLIARVSATAVIATGVAVAFQPAAMAASP-----------VVTVSPATGLKDGDTVTVTGTGLTPGVVYHVSQCESVTSGTY 78
ACZ98704    -MQIANKGKLLTKLGATAALALGLGLAFQPAAGATTTPASSGASISAAAISAAPSTGLSDNTTITISATGLQPGSVYHLGQCAAVRPDAF 89
EDX26137    -MTVKNTLAAITRLGAVAAIATGLAVAVQPAAMAAPAAA---------AVTVTPATGLSDNQQVTVTGTGLTPGTVYHVGQCAVVAPNTF 80
ADJ47157    ---MSARTPIGTKAVVAAGFAVALACTGAATASAAPAAP---------ALAASPSSDLADGQVVDVSGTGYTAGSTIVLLECDAAQPAGR 78
Consensus                                      :.             :.:*::.* :.  : :: :*  ..      :*

KedA        ACNVAEFHDFSLSG-GEGTTSVVVRRSFTGYVMPDGPEVGAVDCDTAPGGCQIVVGGNTGEYG-NAAISFG 145
KedA*       ACNVAEFHDFSLSG-GEGTTSVVVRRSFTGYVMPDGPEVGAVDCDTAPGGCEIVVGGNTGEYG-NAAISFG 114
CagA        ACNPATATSFTTDASGAASFSFVVRKSYTG-STPEGTPVGSVDCATAA--CNLGAGNSGLDLG-HVALTFG 143
AxnA        ACNPATATSFTTDASGAASFSFVVRKSYAG-STPEGTPVGSVDCATDA--CNLGAGNSGLDLG-HVALTFG 143
NcsA        ACNPADFSSVTADANGSASTSLTVRRSFEG-FLFDGTRWGTVDCTTAA--CQVGLSDAAGNGPEGVAISFN 147
McmA        GCDATTSTDVTADAAGKITAQLKVHSSFQAVVGANGTPWGTVNCKVVSCSAGLGSDS---GEGAAQAITFA 144
EFE72430    GCDPTTVIDIAADAQGKVSTQFVVRKTFQAVKGAEGIPSGTVDCTVSACAVGMGDDQ---GVGGGQRITFG 146
ACZ98704    ACNAATNVDVTASATGTITKTLTVRSSFTGSA-ADGSTW-AINTATTPTVIAVFNNA---FDGGTTPLSF- 154
EDX26137    GCDKTTSLDVVADAQGKVTAQLRVHVSFSAVVGASSTPWGTVDAKATPTQVGLGSDA---GEGGGQLISFK 148
ADJ47157    ACDKAALVATVAGADGTLAAKLTVHQAFQAVDLSTGAAGTTVDCATAHCVIASADTSNTGTEGAGVSITFG 149
Consensus   .*: :       .. *  :  . *: :: .     .    :::  .                    ::*
```

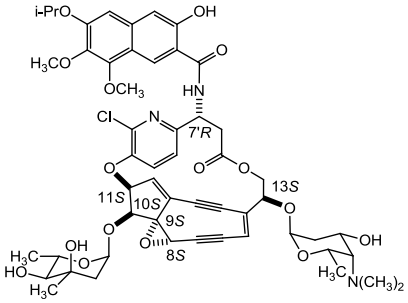
Fig. S2. The original (**1**, proposed in 1992[4,5]) and revised structures (**2**, revised in 1997[6] and **3**, revised in 2007,[7] is considered to be the final correct structure) of kedarcidin (KED) chromophore.



KED chromophore (**1**)
(1992, ref 4 and 5)



KED chromophore (**2**)
(1997, ref 6)



KED chromophore (**3**)
(2007, ref 7)

Fig. S3. Enediyne natural products whose structures[8] have been determined: (A) 9-membered enediynes C-1027, neocarzinostatin (NCS), maduropetin (MDP), kedarcidin (KED), N1999A2, sporolides A and B, cyanosporasides A and B, and fijiolides A and B. C-1027, NCS, MDP, and KED were isolated as chromproteins. N1999A2 was isolated as the enedyne chromophore alone. The sporolides, cyanosporasides, and fijiolides were isolated as aromatized products, whose enediyne forms were hypothetical; no apoprotein has been associated with these compounds. (B) 10-membered enediynes, all of which were isolated as discrete small molecules.

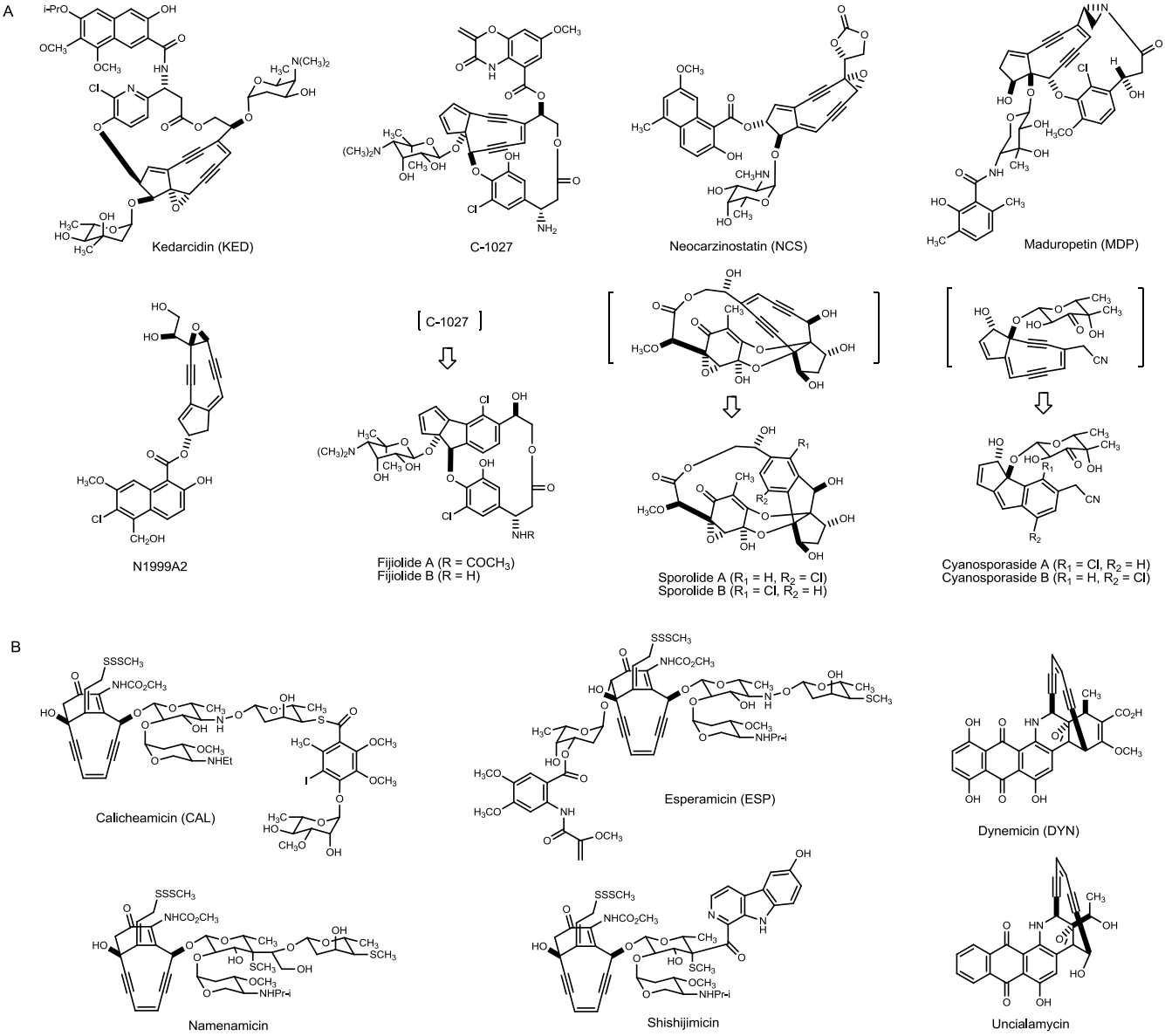


Fig S4. HPLC and HRESIMS analysis of the KED chromophore confirming the enediyne form and supporting the aromatized forms.[9,10] (I) HPLC chromatograms showing freshly prepared EtOAc extract of the KED chromophore from a KED chromoprotein sample (top panel) and the same EtOAc extract kept at room temperature overnight (bottom panel). Enediyne form (●); aromatized form (♦). (II) Structures of the enediyne and aromatized forms of the KED chromophore supported by HRESIMS analysis. HRESIMS spectra of (III) the KED chromophore enediyne form with $[M + H]^+$ ion at *m/z* 1030.37338 and (IV) the KED chromophore aromatized form with $[M + H]^+$ ion at *m/z* 1032.39109.

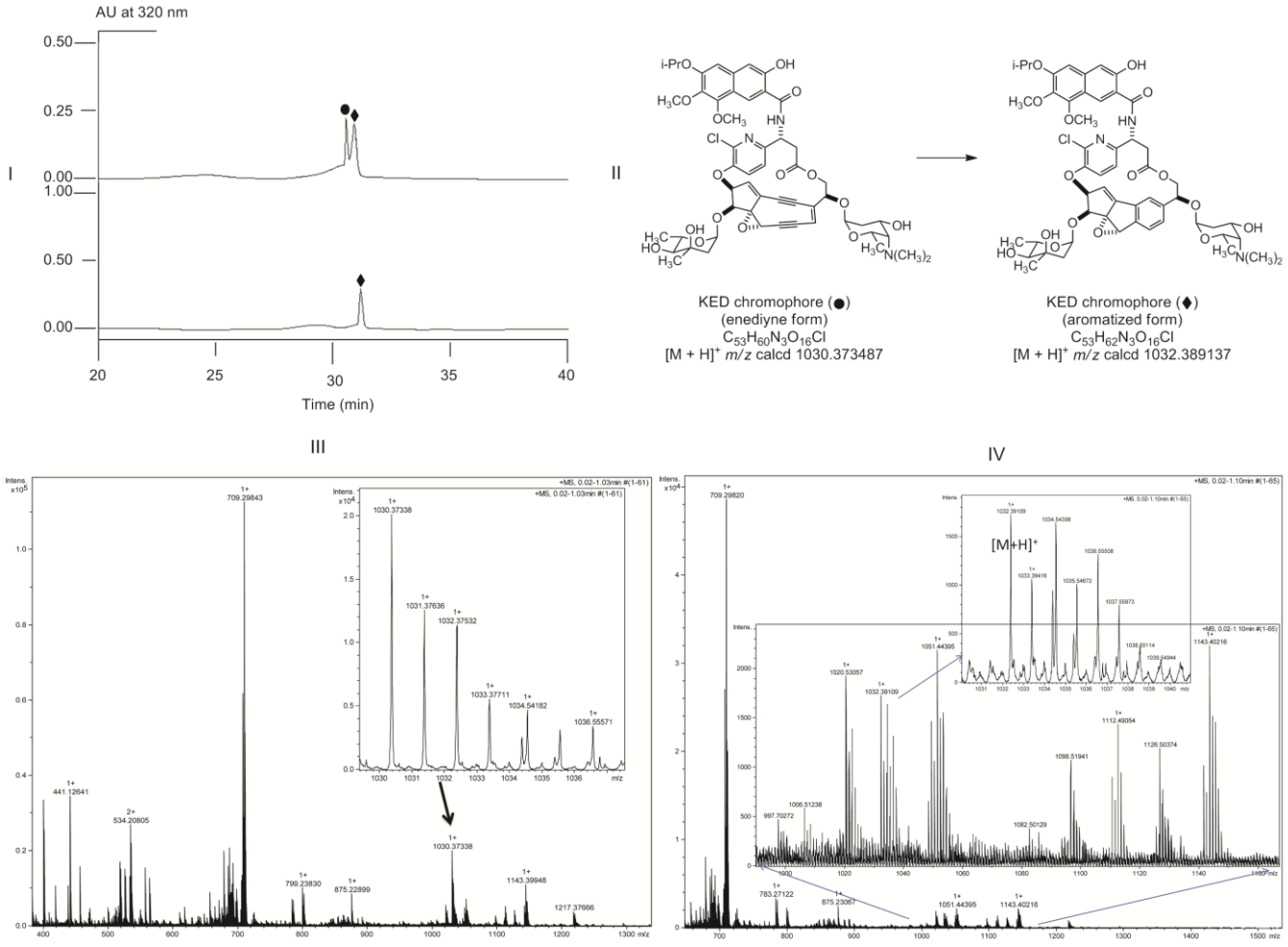


Fig S5. SDS-PAGE (5-15%) analysis of purified KedF used in characterizing KedF as an epoxide hydrolase: lane 1, MW standard and lane 2, purified KedF. The predicted molecular weight of the recombinant KedF is 44.8 kDa.

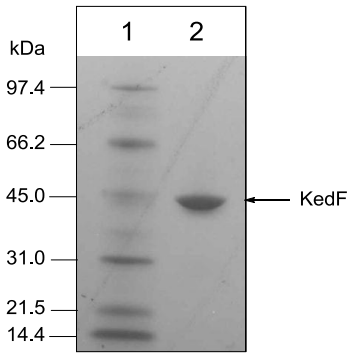


Fig S6. Comparative analysis of the KED, C-1027, and MDP biosynthetic gene clusters supporting the proposed pathway for (*R*)-2-aza-3-chloro-β-tyrosine in KED biosynthesis: (A) genes encoding enzymes with high sequence homology among the three clusters for the L-tyrosine-derived moieties in C-1027 and MDP biosynthesis and for (*R*)-2-aza-3-chloro-β-tyrosine in KED biosynthesis and (B) their proposed pathways.

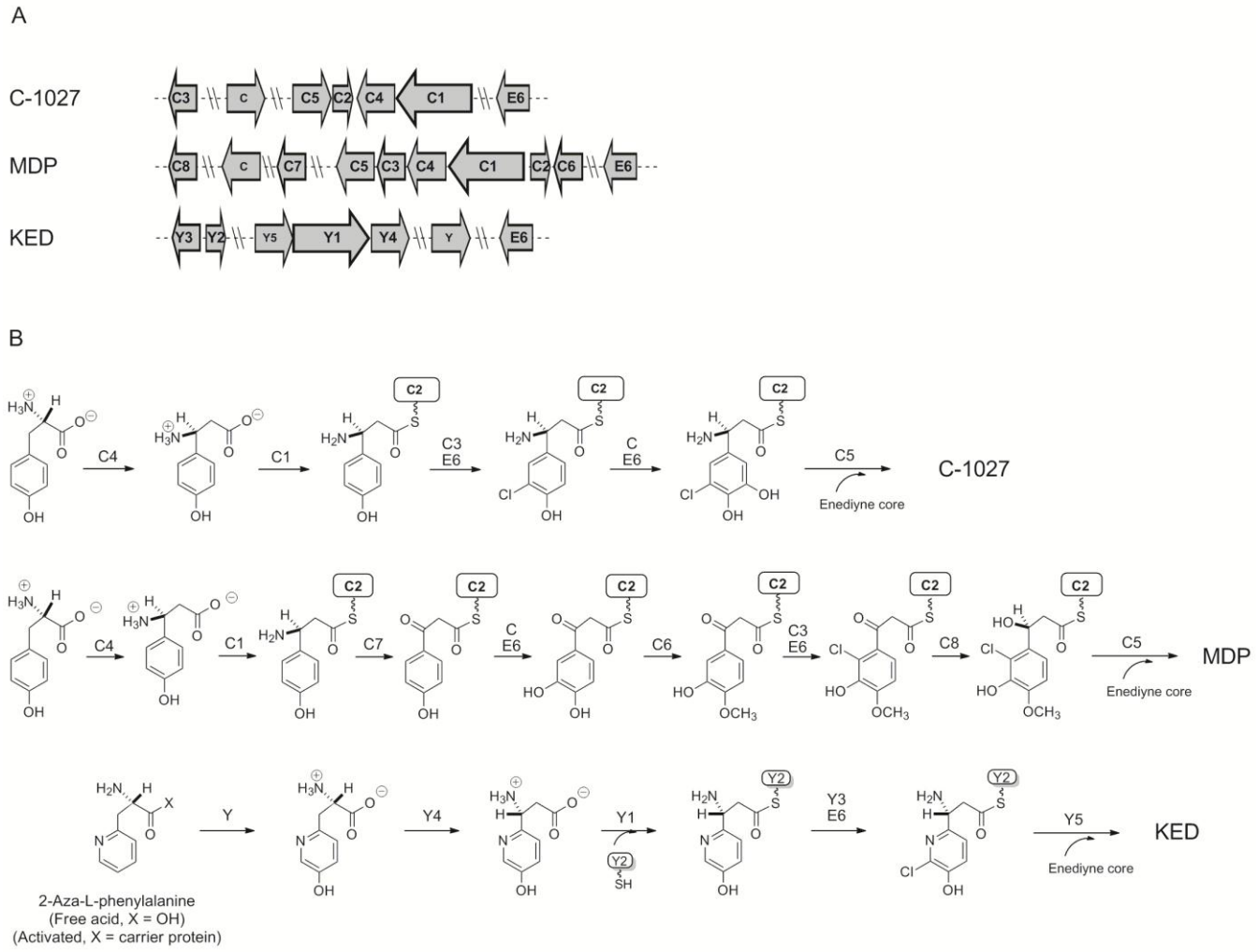
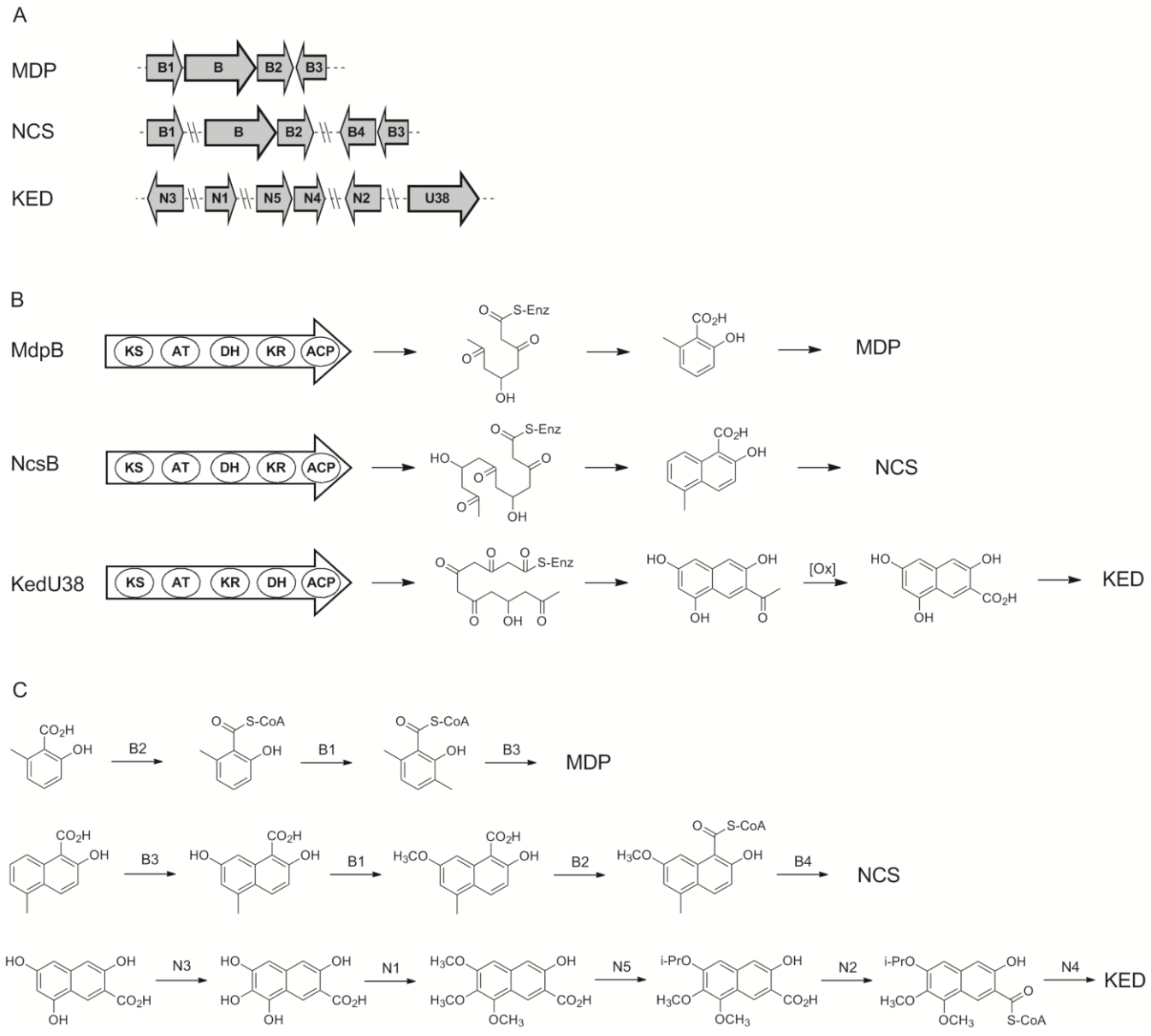


Fig. S7. Comparative analysis of the KED, NCS, and MDP biosynthetic gene clusters supporting the proposed pathway for 3-hydroxy-7,8-dimethoxy-6-isopropoxy-2-naphthoic acid in KED biosynthesis: (A) genes encoding enzymes with high sequence homology among the three clusters for aromatic acid moieties in KED, NCS and MDP biosynthesis, (B) domain organization of the three iterative type I PKSs MdpB, NcsB, and KedU38 (KS, ketosynthase; AT, acyl transferase; DH, dehydratase; KR, ketoreductase; ACP, acyl carrier protein) and their proposed pathways for the biosynthesis of the nascent aromatic acids, and (C) tailoring steps for converting the nascent aromatic acids to fully modified aromatic acid moieties and their activation and incorporation in MDP, NCS, and KED biosynthesis.



References

1. S. J. Hofstead, J. A. Matson, A. R. Malacko and H. Marquardt, *J. Antibiot.* 1992, **45**, 1250-1254.
2. J. S. Thorson, B. Shen, R. E. Whitwam, W. Liu and Y. Li, *Bioorg. Chem*. 1999, **27**, 172-188.
3. V. T. T. Hang, T. S. Kim, T.-J. Oh and J. K. Sohng, *Biotechnol. Bioproc. Eng*. 2011, **16**, 462-469.
4. J. E. Leet, D. R. Schroeder, S. J. Hofstead, J. Golik, K. L. Colson, S. Huang, S. E. Klohr, T. W. Doyle and J. A. Matson, *J. Am. Chem. Soc.* 1992, **114**, 7946-7948.
5. J. E. Leet, D. R. Schroeder, D. R. Langley, K. L. Colson, S. Huang, S. E. Klohr, M. S. Lee, J. Golik, S. J. Hofstead, T. W. Doyle and J. A. Matson, *J. Am. Chem. Soc.* 1993, **115**, 8432-8443.
6. S. Kawata, S. Ashizawa and M. Hirama, *J. Am. Chem. Soc.* 1997, **119**, 12012-12013.
7. F. Ren, P. C. Hogan, A. J. Anderson and A. G. Myers, *J. Am. Chem. Soc.* 2007, **129**, 5381-5383.
8. S. G. Van Lanen and B. Shen, *Curr. Topics Med. Chem.* 2008, **8**, 448-459.
9. A. G. Myers, A. R. Hurd and P. C. Hogan, *J. Am. Chem. Soc.* 2002, **124**, 4583-4585.
10. K. Ogawa, Y. Koyama, I. Ohashi, I. Sato and M Hirama, *Angew. Chem. Int. Ed.* 2009, **48**, 1110-1113.