



Published in final edited form as:

Org Lett. 2017 November 17; 19(22): 6192–6195. doi:10.1021/acs.orglett.7b03120.

Genome Mining of *Micromonospora yangpuensis* DSM 45577 as a Producer of an Anthraquinone-Fused Eneidyne

Xiaohui Yan^{#†}, Jian-Jun Chen^{#†}, Ajeeth Adhikari[†], Dong Yang^{†,§}, Ivana Crnovcic[†], Nan Wang[†], Chin-Yuan Chang[†], Christoph Rader^{||}, and Ben Shen^{*,†,‡,§}

[†]Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, United States

[‡]Department of Molecular Medicine, The Scripps Research Institute, Jupiter, Florida 33458, United States

[§]Natural Products Library Initiative at the Scripps Research Institute, The Scripps Research Institute, Jupiter, Florida 33458, United States

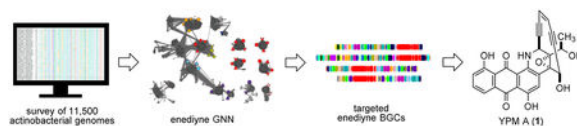
^{||}Department of Immunology and Microbiology, The Scripps Research Institute, Jupiter, Florida 33458, United States

[#] These authors contributed equally to this work.

Abstract

A new anthraquinone-fused eneidyne, yangpumicin A (YPM A, 1), along with four Bergman cyclization congeners (YPM B–E, 2–5), was isolated from *Micromonospora yangpuensis* DSM 45577 after mining eneidyne biosynthetic gene clusters from public actinobacterial genome databases and prioritizing the hits by an eneidyne genome neighborhood network analysis for discovery. YPM A is potent against a broad spectrum of human cancer cell lines. The discovery of 1 provides new opportunities for the functionalization of eneidyne to develop new conjugation chemistries for antibody—drug conjugates.

Graphical Abstract



Since the determination of the chromophore structure of neocarzinostatin (NCS) in 1985, 12 eneidyne natural products have been isolated and structurally characterized, with four additional compounds isolated in their cycloaromatized form.¹ The eneidyne feature an unsaturated core containing two acetylenic groups conjugated to a double bond or incipient

*Corresponding Author: shenb@scripps.edu.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.orglett.7b03120. Experimental procedures, Tables S1–S4, and Figures S1–S36 (PDF)

The authors declare no competing financial interest.

double bond.² Based on the size of the enediyne core, the enediynes are classified into two subcategories: 9-membered and 10-membered enediynes. The 9-membered enediynes include C-1027, NCS, kedarcidin, maduropeptin, N1999A2, sporolide, cyanosporasides, and the fijiolides. The 10-membered enediynes can be further divided into two families: the calicheamicin (CAL)-like enediynes, i.e., CAL, esperamicin, shishijimicin, and namenamicin, and the anthraquinone-fused enediynes, i.e. dynemicin (DYN), uncialamycin (UCM), and tiancimycin (TNM).³ DYN and UCM were isolated from *Micromonospora chersina* in 1989⁴ and from *Streptomyces uncialis* in 2005,⁵ respectively, using conventional activity-guided discovery methods, while TNM was discovered recently from *Streptomyces* sp. CB03234 using an efficient strain prioritization and genome mining approach from a strain collection containing 3400 actinomycetes.¹

The enediynes are some of the most potent natural products in existence to date, with subnanomolar inhibitory concentrations (IC₅₀s) against a broad panel of cancer cell lines.^{3,6} Although too toxic for direct use as chemotherapeutic agents, the enediynes have proven highly efficacious for anticancer therapy when delivered using polymer-based or antibody-directed systems. NCS as a polymer drug conjugate (SMANCS) has been marketed since 1994 for the treatment of certain types of leukemia and hepatoma.⁷ A CAL derivative was used as the payload of the first FDA-approved antibody—drug conjugate (ADC), Mylotarg (gemtuzumab ozogamicin), in 2000. (Pfizer voluntarily withdrew Mylotarg in the U.S. in 2010 due to the lack of clinical benefit in comparison with chemotherapy.) In September 2017, Mylotarg was reapproved by the FDA for the treatment of newly diagnosed and relapsed or refractory CD33-positive acute myeloid leukemia (AML). In August 2017, the FDA also approved Besponsa (inotuzumab ozogamicin), an ADC composed of a CD22-targeting monoclonal antibody and a CAL derivative, for the treatment of adult patients with relapsed or refractory B-cell AML. The approval of Mylotarg and Besponsa doubles the number of ADCs on the market and greatly encourages the search for new enediynes for anticancer therapy. Other ADCs using C-1027 (also known as lidamycin)⁸ and UCM as payloads are currently in various stages of preclinical and clinical studies.⁶ It is remarkable that four of the 12 known enediynes have been translated into clinical drugs or are in preclinical development, representing an extraordinary success rate of 33%.

Recent efforts on genome mining of enediyne natural products have demonstrated that the potential of bacteria, especially actinobacteria, to produce enediyne natural products was greatly underappreciated. By a high-throughput genome survey of 3400 strains from the actinomycetes strain collection at The Scripps Research Institute (TSRI), we identified 81 potential enediyne producers and discovered TNM from *Streptomyces* sp. CB03234.¹ In 2015, we conducted a virtual survey of bacterial genomes from the Genbank and the Joint Genome Institute (JGI) genome databases and characterized 87 enediyne biosynthetic gene clusters from 78 different bacterial strains.⁹ It is remarkable that 68 of the 78 strains belong to the actinobacteria phylum, supporting actinobacteria as prolific enediyne producers. An enediyne genome neighborhood network (GNN) consisting of all the proteins from the potential enediyne gene clusters enabled us to rapidly annotate the gene clusters, and to predict the possible structures of the enediynes.⁹

Herein we report the survey of 11500 actinobacterial genomes in the NCBI and JGI genome databases. Using proteins of the enediyne PKS cassette as probes, 137 distinct enediyne gene clusters were identified. GNN analysis of all the enediyne gene clusters facilitated the characterization of *Micromonospora yangpuensis* DSM 45577¹⁰ as a producer of a new anthraqui-nine-fused enediyne (Figure 1).

As actinobacteria are the most prolific enediyne producers, we aimed at mining the potential enediyne producers in the actinobacteria phylum. BlastP search of the 11500 actinobacterial genomes (as of January 5, 2017) in the JGI and NCBI genome databases, using proteins of the C-1027 enediyne PKS cassette (SgcE, SgcE3, SgcE4, SgcE5, and SgcE10) as queries,^{8,11} resulted in 322 enediyne biosynthetic gene clusters (BGCs). As many of the gene clusters are highly similar or duplicates, we dereplicated identical gene clusters using a 90% amino acid identity cutoff to afford 137 distinct enediyne gene clusters from 129 actinobacteria strains. An enediyne GNN consisting of 9775 proteins from the 137 enediyne BGCs was generated and used to annotate the enediyne BGCs. From the GNN analysis of the *dyn*,¹² *ucm*, and *tnm* gene clusters,¹ we noticed that many proteins from the *M. yangpuensis* DSM 45577 enediyne BGC (the *ypm* gene cluster) clustered with proteins from the three known ones (Figures 2B and S1), suggesting that it may produce an anthraquinone-fused enediyne.

Bioinformatic analysis of the *ypm* gene cluster revealed 34 genes spanning 40 kb. The encoded proteins showed identities ranging from 32% to 92% with proteins encoded by the *dyn*, *ucm*, and *tnm* gene clusters (Figure S1 and Tables S1 and S2). Like the *ucm* and *tnm* clusters, the *ypm* gene cluster harbors a gene encoding a Rieske (2Fe-2S) iron—sulfur protein (YpmM) and a gene encoding an oxidoreductase (YpmI), which are absent in the *dyn* gene cluster.¹ On the other hand, the *ypm* gene cluster contains four genes (*ypmT5*, *ypmT8*, *ypmU20*, and *ypmU21*) that encode proteins homologous to proteins from the *dyn* gene cluster¹¹ but are absent in the *tnm* and *ucm* clusters (Figure 2A,B and Tables S1 and S2). YpmT5 and YpmT8 were predicted to be subunits for ABC-type transporters, while the functions of YpmU20 and YpmU21 remain unknown. Furthermore, the *ypm* cluster contains a gene encoding a cytochrome P450 monooxygenase (YpmL), with amino acid identity of 42%, 44%, and 56%, to TnmL, Dynorf19, and DynE10, respectively.

To investigate whether *M. yangpuensis* DSM 45577 produces an anthraquinone-fused enediyne, we first cultivated the strain on a small scale (2 × 50 mL). The fermentation broth was extracted with EtOAc, dried in vacuo, dissolved in MeOH, and then analyzed by HPLC, antibacterial assay, and biochemical induction assay (BIA).¹³ The crude extract showed strong antibacterial activity against *Kocuria rhizophila* ATCC 9341 and DNA-cleavage activity by BIA using *E. coli* BR513 as an indicator. Followed by bioactivity-guided isolation, an antibacterial metabolite, yangpumicin A (YPM A, 1), with an [M + H]⁺ ion at *m/z* 456.1070 was detected by HRESIMS, with a molecular formula of C₂₆H₁₇NO₇ (calcd for the [M + H]⁺ ion at *m/z* 456.1078) (Figure S2). The molecular formula of 1 differs from that of TNM A (C₂₇H₁₉NO₈) by the loss of a CH₂O group and from that of UCM (C₂₆H₁₇NO₆) by the presence of an additional oxygen atom. The UV spectrum of 1 was also similar to that of UCM and TNM A (Figure S3). These results suggested that 1 could be an analogue of TNM A and UCM. To establish the structure of 1, we performed large-scale

fermentation (18 L) of *M. yangpuensis* DSM 45577 and isolated 1, together with four additional congeners (2–5), from the fermentation broth (Figures 1 and 2C).

YPM A (1) was isolated as a purple powder. The high similarity between the ^1H and ^{13}C NMR spectra of 1 (Tables S3 and S4 and Figures S4 and S5) and TNM A confirmed 1 as an anthraquinone-fused enediyne. Comparison of the NMR data of 1 with those of TNM A revealed an oxygenated aromatic carbon C-6 (δ 162.2) in 1. This assignment was confirmed by the absence of H-6, the coupling pattern for H-7 ($J = 7.7$ Hz, 1.4 Hz), and the HMBC correlations from H-7 to C-5 and C-9, and from H-8 to C-6 and C-10. The complete structure of 1 was assigned on the basis of ^1H – ^1H COSY, HSQC, and HMBC data (Figures 1, 3A, and S6–S9).

The relative configuration of the stereogenic centers of 1, except for C-26, was the same as TNM A, according to the ROESY experiment (Figures 3B and S9) and molecular model. The relative stereochemistry at C-26 was established as R* by comparing the chemical shift of H-26 (δ 4.32) and H-27 (δ 1.32) in DMSO- d_6 with those of (26R*)-UCM H-26 (δ 4.33), and H-27 (δ 1.31), and (26S*)-UCM H-26 (δ 4.20), and H-27 (δ 1.34) (Figure S10).^{14,15} The ECD curves of 1 (Figure 3C) showed good consistency with TNM A and an authentic standard of UCM, supporting the absolute configuration of 1.

YPM B (2) was isolated as a dark blue powder with a molecular formula of $\text{C}_{26}\text{H}_{21}\text{O}_7$ by the HRESIMS with an $[\text{M} + \text{H}]^+$ ion at m/z 460.1402 (calcd at 460.1391) (Figure S2). The ^1H and ^{13}C NMR spectra of 2 showed signals that could be assigned to the anthraquinone motif (C-2 to C-15) seen previously in TNM A and 1 (Tables S3 and S4 and Figures S11 and S12). However, the signals attributed to the enediyne core and epoxide units in 1 were replaced by a 1,2,3,4-tetrasubstituted benzene ring, a tertiary alcohol, and a methine group in the NMR spectrum of 2. It was suggested that the enediyne core moiety (C-18–C-23) was changed into a 1,2,3,4-tetrasubstituted benzene ring, and the epoxide unit (C-16 and C-25) was replaced by a tertiary alcohol (C-25) and a methine group (C-16), due to the Bergman cycloaromatization (Figure S13). This conclusion was supported by the HMBC correlations in 2 (Figure 3A). All of the stereogenic centers of 2 remain the same during the Bergman cycloaromatization from 1 to 2, with the exception of C-16 and C-25 (Figure S13).¹⁶ Because of the *cis*-fused benzene- containing bridge, indicated by molecular modeling, H-16 and H-24 must have a *cis* relationship. This is supported by the ROESY correlations between H-16 (δ 3.44) and H-14 (δ 7.24), H-17 (δ 5.17), and CH_3 -27 (δ 1.39) (Figures 3B and S14), as well as the W-type coupling pattern between H-16 and H-24 ($J = 1.4$ Hz) (Table S3 and Figure S11). The latter also indicates that H-16 and H-24 have the same orientation, thus confirming the 16R configuration.¹⁶ As for C-25, during the process of the Bergman cycloaromatization, the 25-OH and H-16 should retain the same orientation, thus assigning the 25R configuration in 2. Combined with other 2D NMR data (Figures S15–S17), the structure of 2 was established (Figure 1).

YPM C (3) was isolated as a dark blue powder that gave a $[\text{M} + \text{H}]^+$ ion at m/z 476.1330 in the HRESIMS, with a molecular formula of $\text{C}_{26}\text{H}_{21}\text{NO}_8$ (calcd for the $[\text{M} + \text{H}]^+$ ion at m/z 476.1340) (Figure S2). Comparison of the NMR data of 3 (Tables S3 and S4 and Figures S18–S23) with those of 2 showed that they were structural congeners, and 3 differed from 2

by the addition of an oxygen atom. The difference between an oxygenated tertiary carbon C-16 (δ 77.6) in 3 and the C-16 methine (δ 48.7) in 2 indicated the presence of 16-OH in 3. This assignment was supported by the HMBC correlations in 3 (Figure 3A) as well as the downfield shift of H-14 and H-27 from δ 7.24 and δ 1.39 in 2 to δ 7.78 and δ 1.56 in 3, respectively, in the ^1H NMR spectrum.¹⁷ The 16-OH was further assigned to be in the α -orientation according to the Bergman cycloaromatization and the ROESY correlations (Figure S23). Considering the absolute configuration of 1 and the proposed biogenesis of 3 by Bergman cycloaromatization (Figure S13), the absolute configuration of 3 was assigned as 16*S*,17*R*,24*S*,25*R*,26*R*.

YPM D (4) and YPM E (5) were isolated as dark blue powders with molecular formula of $\text{C}_{26}\text{H}_{21}\text{NO}_8$, based on the $[\text{M} + \text{H}]^+$ ion at m/z 476.1343 (calcd at 476.1340), and $\text{C}_{26}\text{H}_{21}\text{NO}_9$, based on the $[\text{M} + \text{H}]^+$ ion at m/z 492.1282 (calcd at 492.1289), respectively, by the HRESIMS analysis (Figure S2) and ^{13}C NMR spectroscopic data (Table S4). Comparison of the 1D and 2D NMR data suggested high structural similarity between 4 and 2 (or 5 and 3) (Tables S3 and S4 and Figures S24–S35). The main difference between 4 and 2 (or 5 and 3) was that the methyl group at C-26 in 2 and 3 was replaced by a hydroxymethyl group in 4 and 5, respectively. This conclusion was supported by the HMBC correlations from H-27 to C-26 and C-25 either in 4 or in 5 (Figure 3A). Thus, the structures of 4 and 5 were assigned as shown in Figure 1.

YPM A (1) shares an identical enediyne core with UCM and TNM A, with a slight difference in the A-ring of the anthraquinone moiety. UCM has no functional groups on the A-ring, 1 has one hydroxy group at C-6 position, and TNM A has one hydroxy group at C-6 and one methoxy group at C-7. These three compounds were tested for their cytotoxic activity against five different human cancer cell lines, including melanoma (SKMEL-5), breast (MDA-MB-231 and SKBR-3), central nervous system (SF-295), and nonsmall cell lung (NCI-H226). YPM A exhibited high potency against all the tested cancer cell lines, with IC_{50} s values ranging from 0.26 to 2.9 nM (Table 1 and Figure S36). In each of the human cancer cell lines examined, the potency of 1 was between that of TNM A and UCM.

Microbes have been a prominent source of therapeutic agents, and their extraordinary capacity to produce new natural products is increasingly recognized. With the explosive growth of microbial genome data, it is of great importance to develop new approaches to mine the microbial genomes in a rapid and efficient way. GNN analysis is a powerful bioinformatics tool that allows the annotation and comparison of multiple gene clusters in a single step, which greatly facilitates strain prioritization for targeted natural products from large strain collections or public genome databases. Combination of a high-throughput real-time PCR screening method¹⁸ with an enediyne GNN resulted in the prioritization of *Streptomyces* sp. CB03234 from the TSRI strain collection and the discovery of TNMs.¹ A genome survey of 11500 actinobacterial genomes from the public databases and analysis of the constructed enediyne GNN in this study enabled the mining of *M. yangpuensis* DSM 45577 and the discovery of YPMs. This technology could be readily applied to the targeted discovery of natural products with other featured scaffolds.

The discovery of YPM A expands the family of the anthraquinone-fused enediynes. The multiple anthraquinone-fused enediyne producers offer more opportunities to produce the enediyne compounds by microbial fermentation. As microbial fermentation is the major route of supply for the enediynes, strain improvement and fermentation optimization of the enediyne producers will increase the titer of these compounds, and enable their sustainable supply for preclinical studies and eventual clinical applications. The discovery of YPM A also provides precious opportunity to study the SAR within the anthraquinone-fused enediynes. Various substitutions on the anthraquinone moiety allow the functionalization of the enediyne scaffold to develop new conjugation chemistry for the attachment to antibodies and other entities. Finally, comparison of the biosynthetic machineries for the anthraquinone-fused enediynes sets the stage to apply combinatorial biosynthetic strategies for enediyne structural diversity or the production of designer analogues.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank K. C. Nicolaou (Rice University) for providing an authentic sample of UCM. This work is supported in part by the State Key Laboratory of Applied Organic Chemistry, College of Chemistry and Chemical Engineering, Lanzhou University, and the Chinese Scholarship Council (201606185009 to J.C.) and by the Institute of Applied Ecology, Chinese Academy of Sciences, and a scholarship from Chinese Scholarship Council (201504910034 to N.W.), a German Research Foundation (DFG) postdoctoral fellowship (to I.C.), and the National Institutes of Health GM115575 (to B.S.) and CA204484 (to B.S. and C.R.). This is manuscript no. 29597 from The Scripps Research Institute.

REFERENCES

- (1). Yan X; Ge H; Huang T; Hindra; Yang D; Teng Q; Crnovcic I; Li X; Rudolf JD; Lohman JR; Gansemans Y; Zhu X; Huang Y; Zhao L-X; Jiang Y; Nieuwerburgh FV; Rader C; Duan Y; Shen B *mBio* 2016, 7, e02104–16. [PubMed: 27999165]
- (2). Van Lanen SG; Shen B *Curr. Top. Med. Chem* 2008, 8, 448–459. [PubMed: 18397168]
- (3). Liang Z-X *Nat. Prod. Rep* 2010, 27, 499–528. [PubMed: 20336235]
- (4). Konishi M; Ohkuma H; Matsumoto K; Kamei H; Miyaki T; Oki T; Kawaguchi H; VanDuyne GD; Clardy JJ *Antibiot* 1989, 42, 1449–1452.
- (5). Davies J; Wang H; Taylor T; Warabi K; Huang X-H; Andersen RJ *Org. Lett* 2005, 7, 5233–5236. [PubMed: 16268546]
- (6). Shen B; Hindra; Yan X; Huang T; Ge H; Yang D; Teng Q; Rudolf JD; Lohman JR *Bioorg. Med. Chem. Lett* 2015, 25, 9–15. [PubMed: 25434000]
- (7). Abe S; Otsuki M *Curr. Med. Chem.: Anti-Cancer Agents* 2002, 2, 715–726. [PubMed: 12678722]
- (8). Chen Y; Yu D; Zhang C; Shang B; He H; Chen J; Zhang H; Zhao W; Wang Z; Xu X; Zhen Y; Shao RG *Mol. Carcinog* 2015, 54, 1–8. [PubMed: 23857500]
- (9). Rudolf JD; Yan X; Shen BJ *Ind. Microbiol. Biotechnol* 2016, 43, 261–276.
- (10). Zhang L; Xi L; Ruan J; Huang Y *Int. J. Syst. Evol. Microbiol* 2012, 62, 272–278. [PubMed: 21398502]
- (11). Liu W; Ahlert J; Gao Q; Wendt-Pienkowski W; Shen B; Thorson JS *Proc. Natl. Acad. Sci. U. S. A* 2003, 100, 11959–11963. [PubMed: 14528002]
- (12). Gao Q; Thorson JS *FEMS Microbiol. Lett* 2008, 282, 105–114. [PubMed: 18328078]
- (13). Elespuru RK; White RJ *Cancer Res* 1983, 43, 2819–2930. [PubMed: 6221794]

- (14). Nicolaou KC; Zhang H; Chen JS; Crawford JJ; Pasunoori L *Angew. Chem., Int. Ed* 2007, 46, 4704–4707.
- (15). Nicolaou KC; Chen JS; Zhang H; Montero A *Angew. Chem., Int. Ed* 2008, 47, 185–189.
- (16). Williams DE; Bottriell H; Davies J; Tietjen I; Brockman MA; Andersen RJ *Org. Lett* 2015, 17, 5304–5307. [PubMed: 26465962]
- (17). Richter C; Helaly SE; Thongbai B; Hyde KD; Stadler MJ *Nat. Prod* 2016, 79, 1684–1688.
- (18). Hindra; Huang T; Yang D; Rudolf JD; Xie P; Xie G; Teng Q; Lohman JR; Huang Y; Zhao L-X; Jiang Y; Duan Y; Shen B J. *Nat. Prod* 2014, 77, 2296–2303. [PubMed: 25238028]

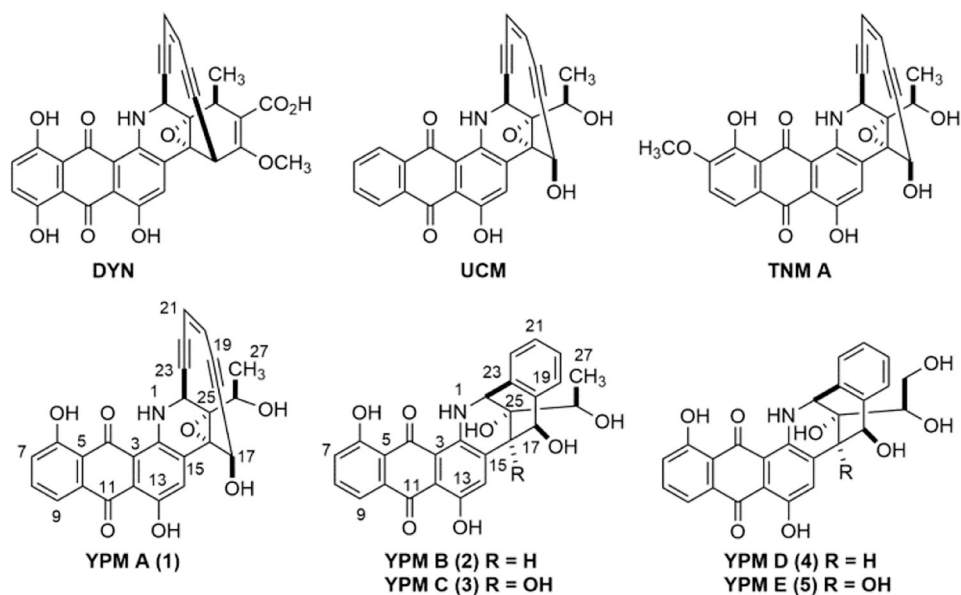


Figure 1.
Structures of anthraquinone-fused enediynes DYN, UCM, and TNM A in comparison to YPM A (1) and its congeners (2–5) discovered in this study.

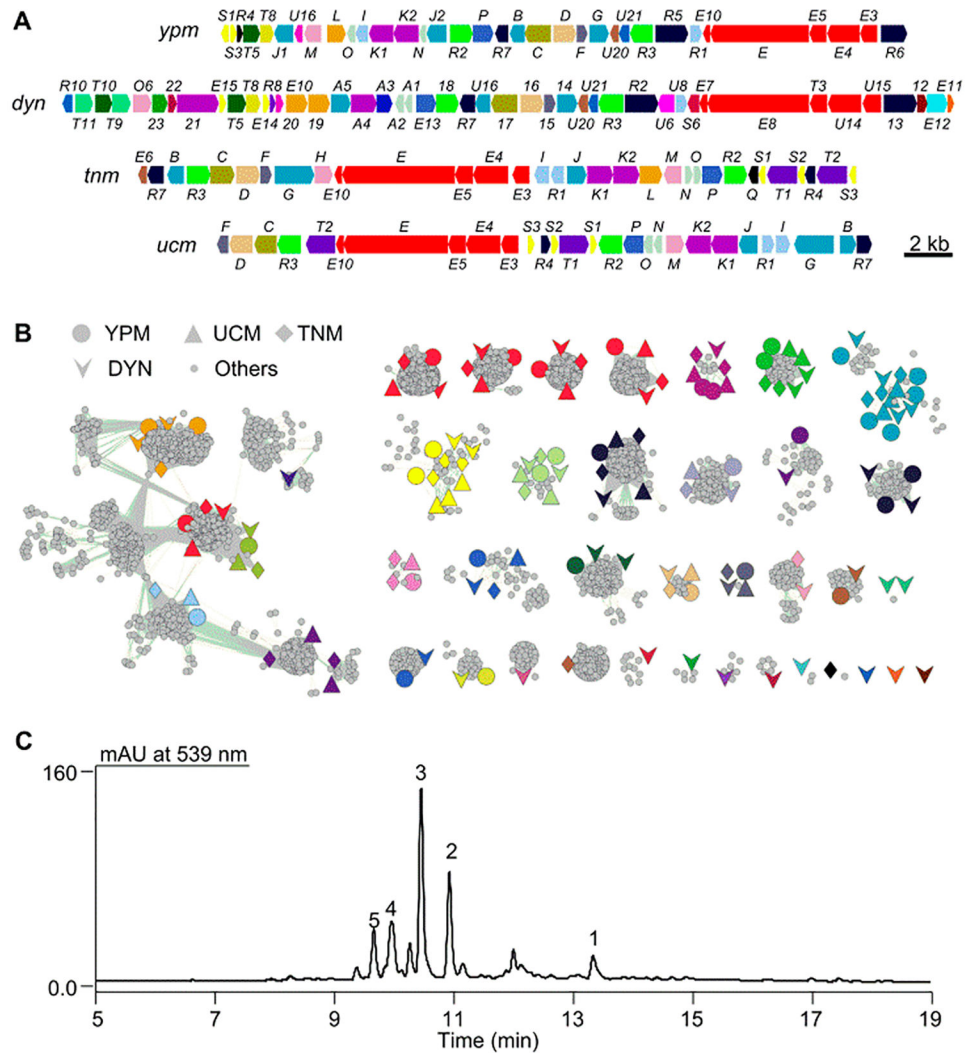


Figure 2. Genome mining of public databases and characterization of *M. yangpuensis* DSM 45577 as a YPM producer. (A) Genetic organization of the *ypm*, *dyn*, *tnm*, and *ucm* gene clusters. Genes are color-coded based on GNN annotation. (B) GNN analysis (E value of 10^{-8}) unveiling functional similarity among the *ypm*, *dyn*, *tnm*, and *ucm* gene clusters. (C) HPLC analysis of fermentation of *M. yangpuensis*, with peaks of YPM A (1) and congeners (2–5) highlighted.

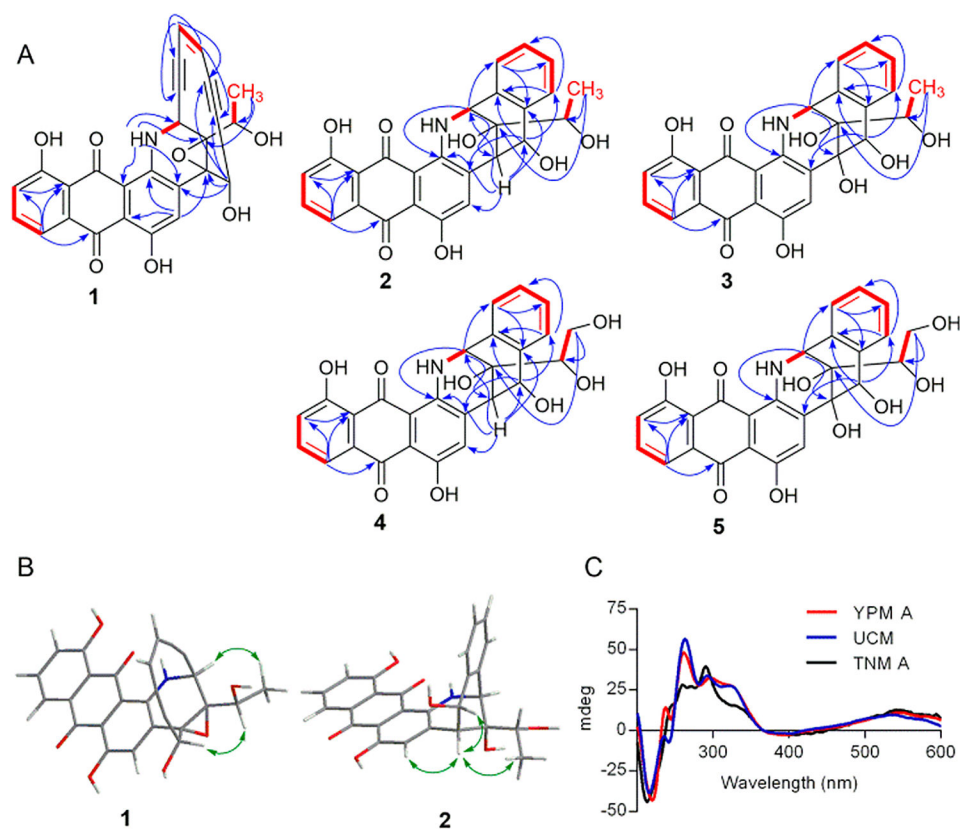


Figure 3. Structural elucidation of YPMs on the basis of 1D and 2D NMR and CD spectroscopic data analysis. (A) Key ^1H - ^1H COSY (red) and HMBC (blue) correlations of 1–5. (B) Key ROESY correlations (green) of 1 and 2. (C) CD spectra of 1 in comparison with authentic stands of UCM and TNM A, supporting the assignment of their absolute stereochemistry.

Table 1.

Cytotoxicity of YPM A against Selected Human Cancer Cell Lines

cell lines	cancer type	IC ₅₀ (nM)		
		YPM A	TNM A	UCM
SK-MEL-5	melanoma	0.40 ± 0.02	0.10 ± 0.02	0.87 ± 0.07
MDA-MB-231	breast	0.57 ± 0.08	0.26 ± 0.04	1.2 ± 0.2
SKBR-3	breast	2.9 ± 0.5	0.50 ± 0.06	3.2 ± 0.6
SF-295	CNS	0.26 ± 0.01	0.19 ± 0.02	0.26 ± 0.01
NCI-H226	NCS lung	2.8 ± 0.4	4.1 ± 0.2	1.9 ± 0.3

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript