Supplementary Material

Title: Isolation and Identification of an Anthracimycin Analogue from *Nocardiopsis kunsanensis*, a Halophile from a Saltern, by Genomic Mining Strategy

Authors: Fernanda L. Sirota[1,#,*], Falicia Goh[1,#], Kia-Ngee Low[1,#], Lay-Kien Yang[1], Sharon C. Crasta[1], Birgit Eisenhaber[1], Frank Eisenhaber[1,2], Yoganathan Kanagasundaram[1,*], Siew Bee Ng[1,*]

Table S1

Suppl. Table S1 - List of top organisms with the highest number of domain hits

Organisms	no. domain hits	
	psi-blast	blast
Streptomyces sp. CNH365	41	41
Streptomyces sp. NRRL F-5065	41	41
Streptomyces sp. T676	41	41
Streptomyces sp. TP-A0875	41	41
Nocardiopsis kunsanensis	41	40
Sorangium cellulosum	40	39
Bacillus amyloliquefaciens	39	36
Bacillus subtilis	39	36
Streptococcus pneumoniae	39	33

Table S2

Suppl. Table 2 – Blast 2 sequences results with E-value 0.0 for all of them

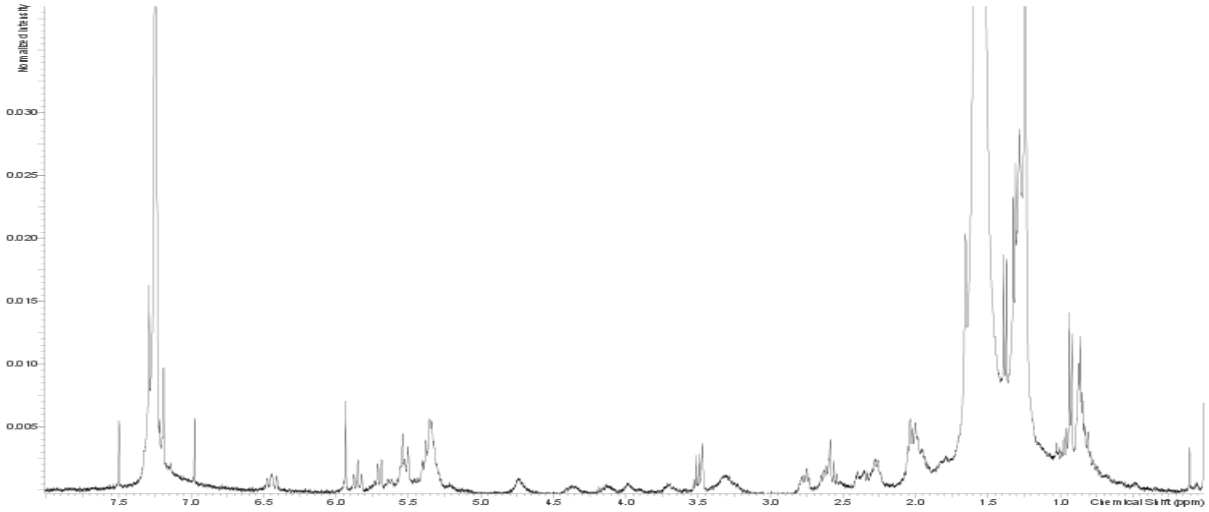
	accession no. (length)		query cover	identity
	query *Streptomyces sp.* T676	subject *N. kunsanensis*		
atcC[1]	CTQ34880 (1108)	WP_017577639[#] (1084)	98%	60%
atcD[2]	CTQ34881 (6561)	WP_017577638* (6332)	99%	50%
atcE[3]	CTQ34882 (4468)	WP_017577637* (4336)	99%	54%
atcF[4]	CTQ34883 (3869)	WP_020480252* (3763)	99%	55%

[1] atcC; *trans*-AT; acylhydrolase, acyltransferase (malonyl-CoA), FMN-dependent enoylreductase
[2] atcD; polyketide synthase, modules 1-4
[3] atcE; polyketide synthase, modules 5-7
[4] atcF; polyketide synthase, modules 8-10
[#] [acyl-carrier-protein] S-malonyltransferase
* hypothetical protein

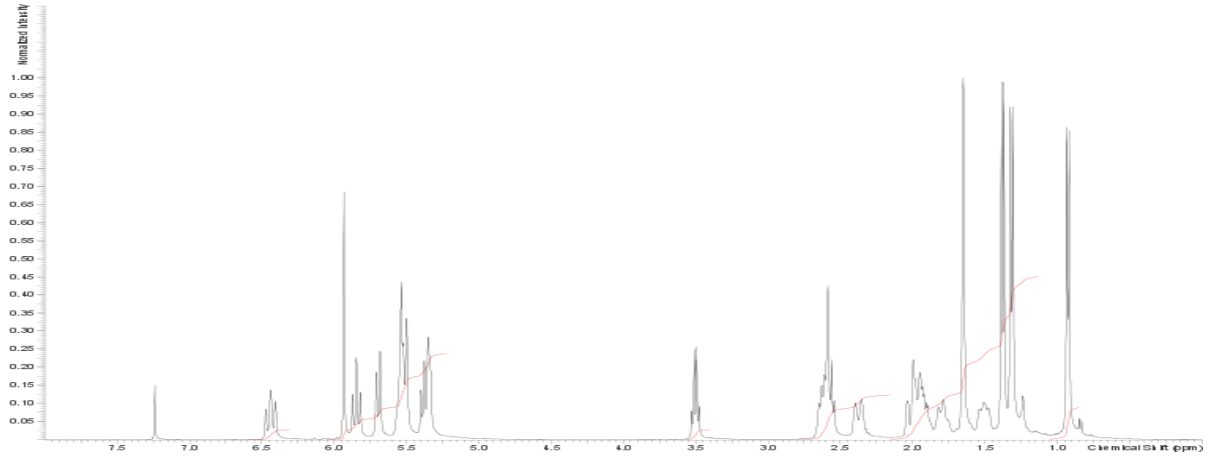
Figure S1

Structure elucidation of (a) anthracimycin (1) in *N. kunsanensis*, (b-c) anthracimycin standard in *Streptomyces* sp. T676 and (d) their comparison with anthracimycin BII-2619.

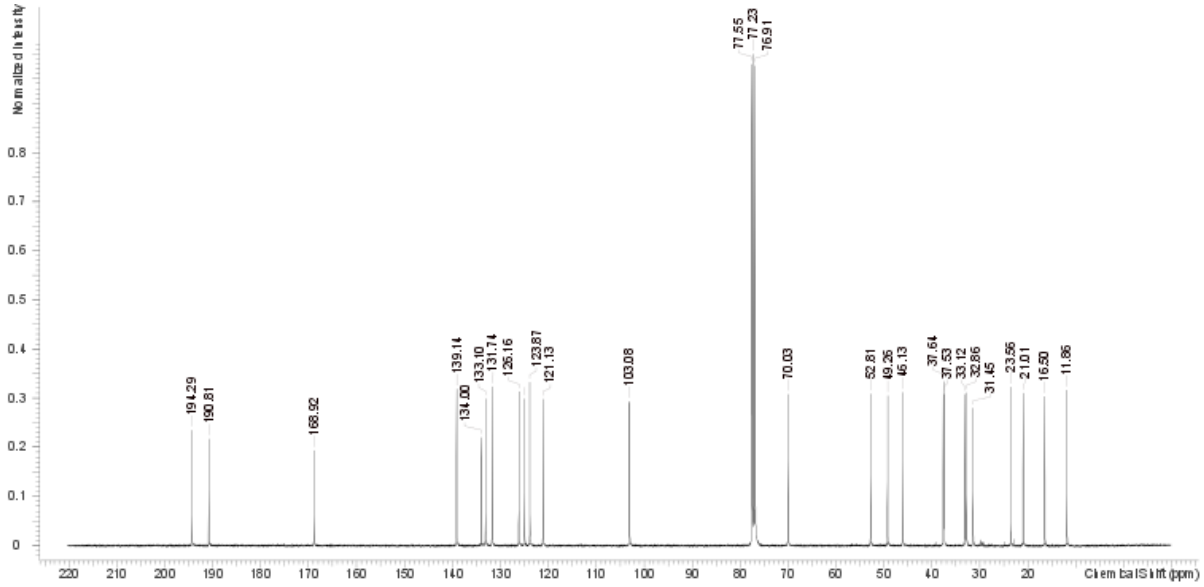
(a) ^{1}H NMR spectrum of anthracimycin (**1**) from *N. kunsanensis* in $CDCl_3$ at 400 MHz.



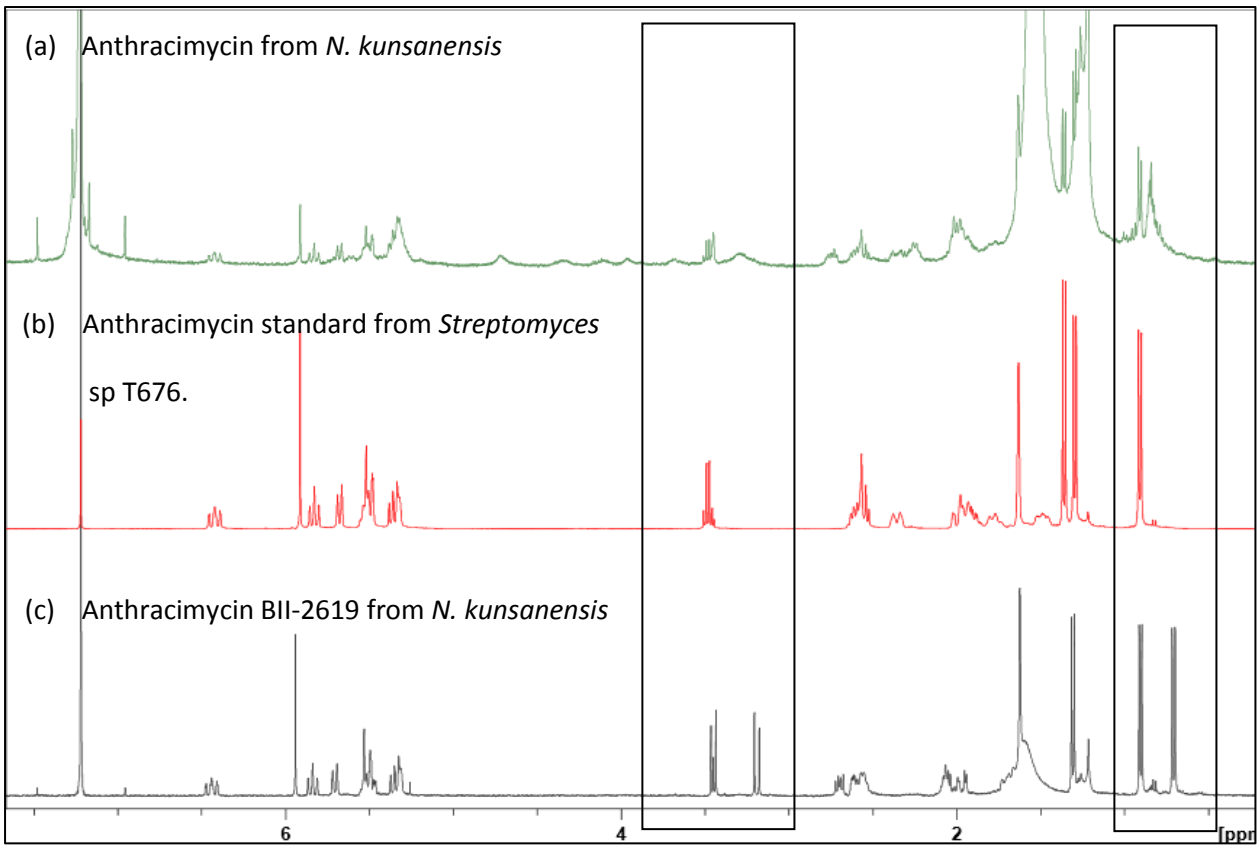
(b) ^{1}H NMR spectrum of anthracimycin standard from *Streptomyces* sp. T676 in $CDCl_3$ at 400 MHz.



(c) ^{13}C NMR spectrum of anthracimycin standard from *Streptomyces* sp. T676 in $CDCl_3$ at 100 MHz.



(d) ^{1}H NMR data derived from the anthracimycin standard as well as from anthracimycin (**1**) and anthracimycin BII-2619 (**2**) from our purification from *N. kunsanensis* shown in comparison along the same *x*-axis.

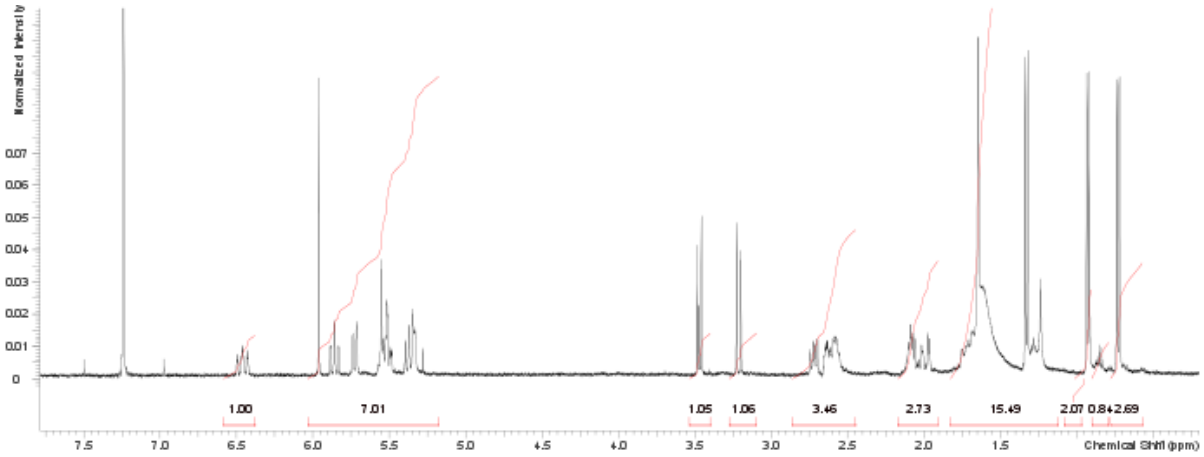


(a) Anthracimycin (**1**) isolated from *N. kunsanensis*, (b) anthracimycin standard isolated from *Streptomyces* sp. T676 and (c) anthracimycin BII-2619 (**2**) isolated from *N. kunsanensis*.

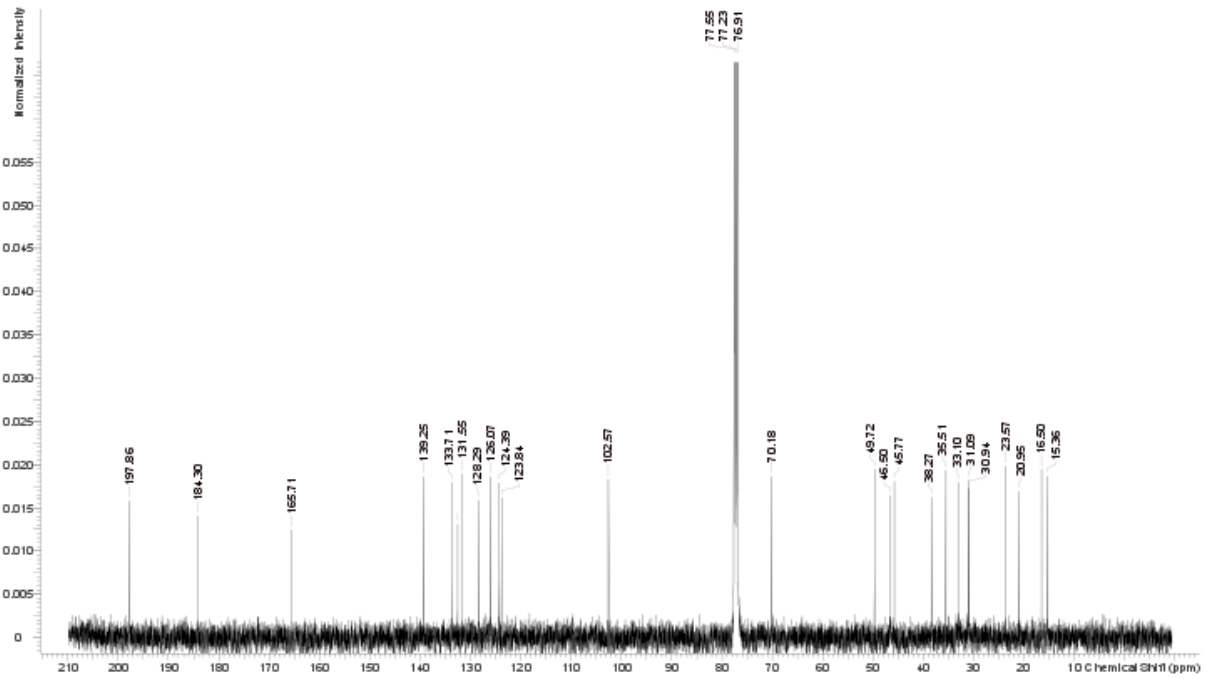
Figure S2

Structure elucidation of anthracimycin BII-2619 (2).

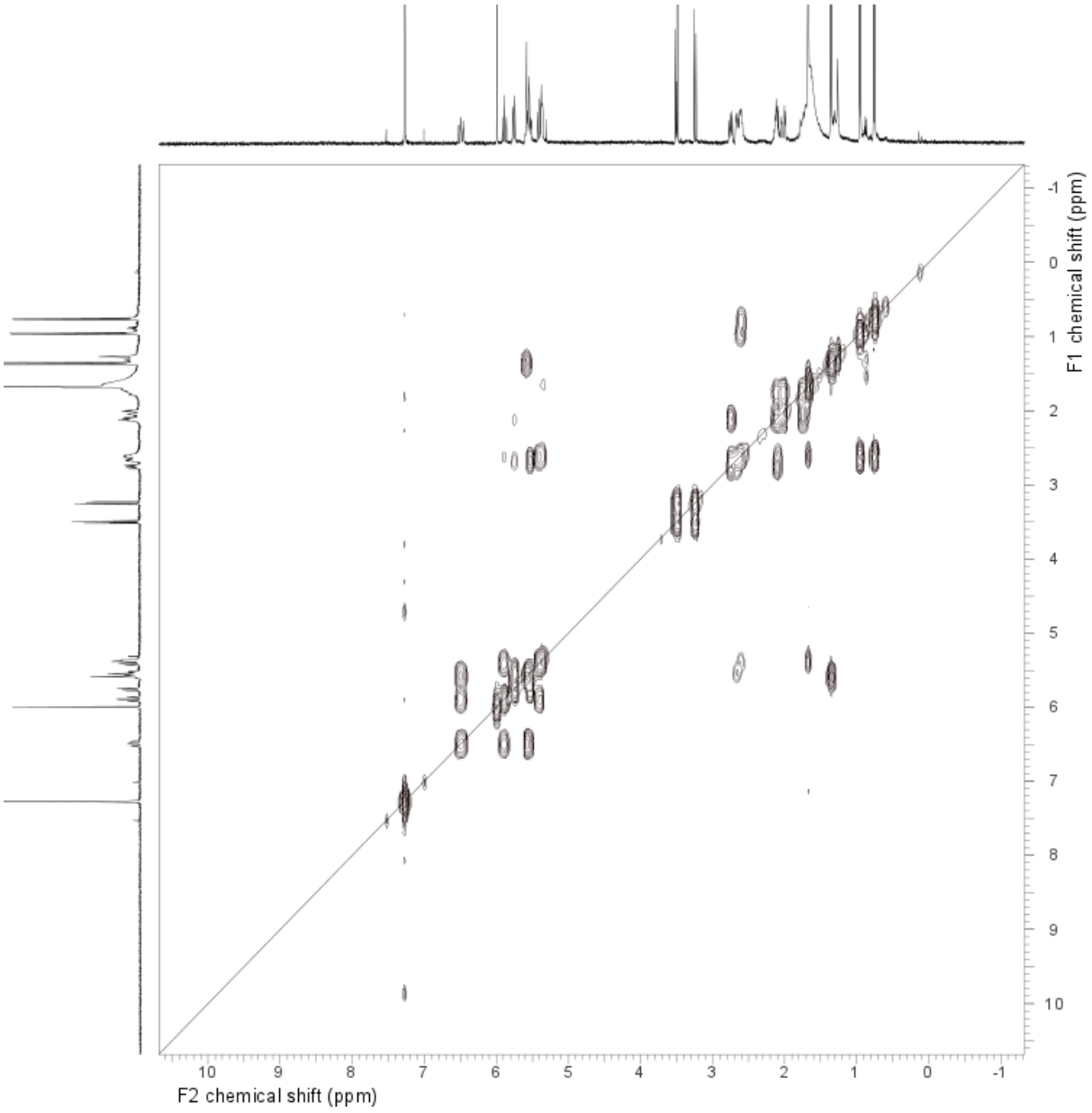
(a) ^{1}H NMR spectrum of anthracimycin BII-2619 (**2**) in $CDCl_3$ at 400 MHz.



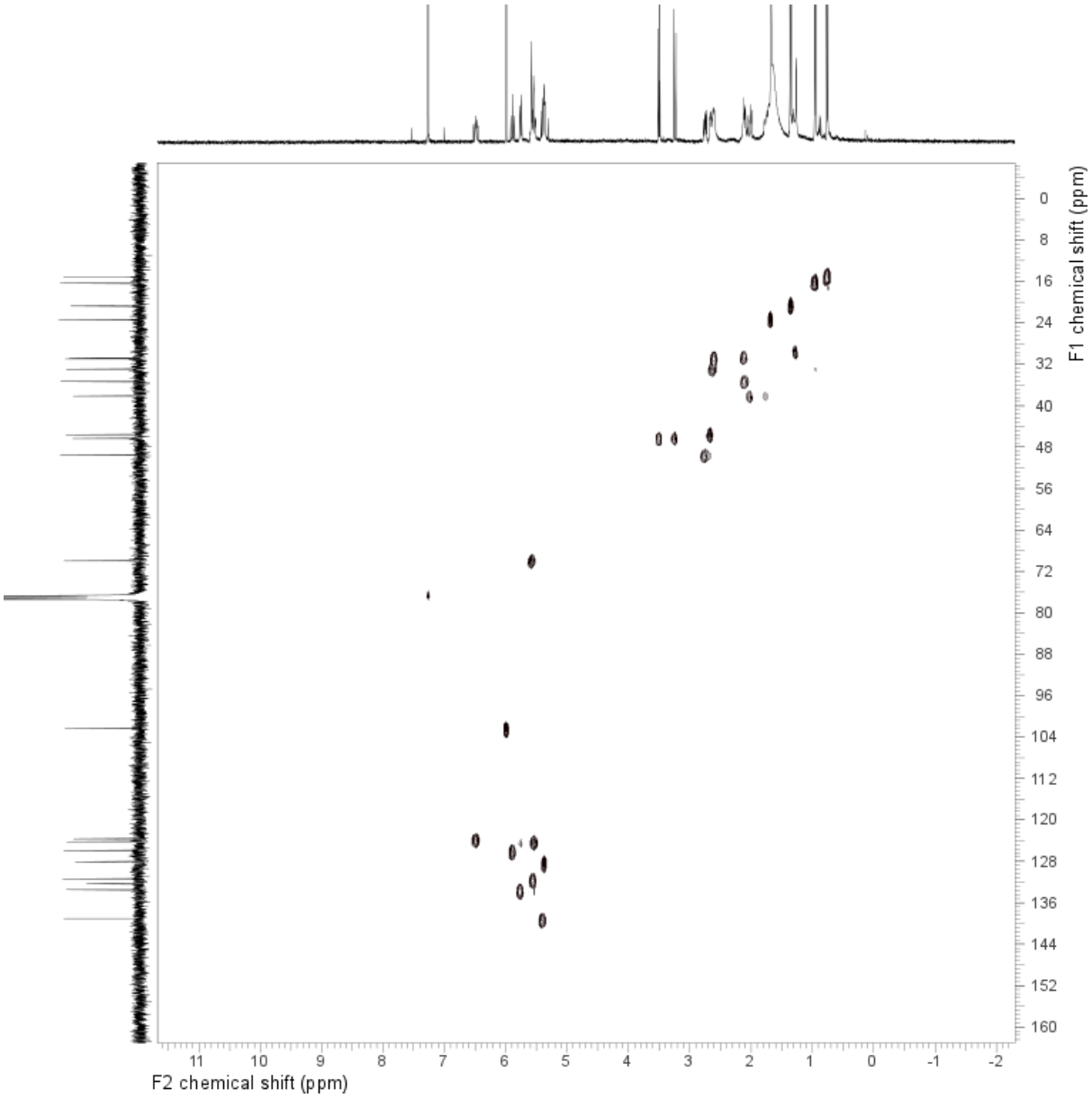
(b) ^{13}C NMR spectrum of anthracimycin BII-2619 (**2**) in $CDCl_3$ at 100 MHz.



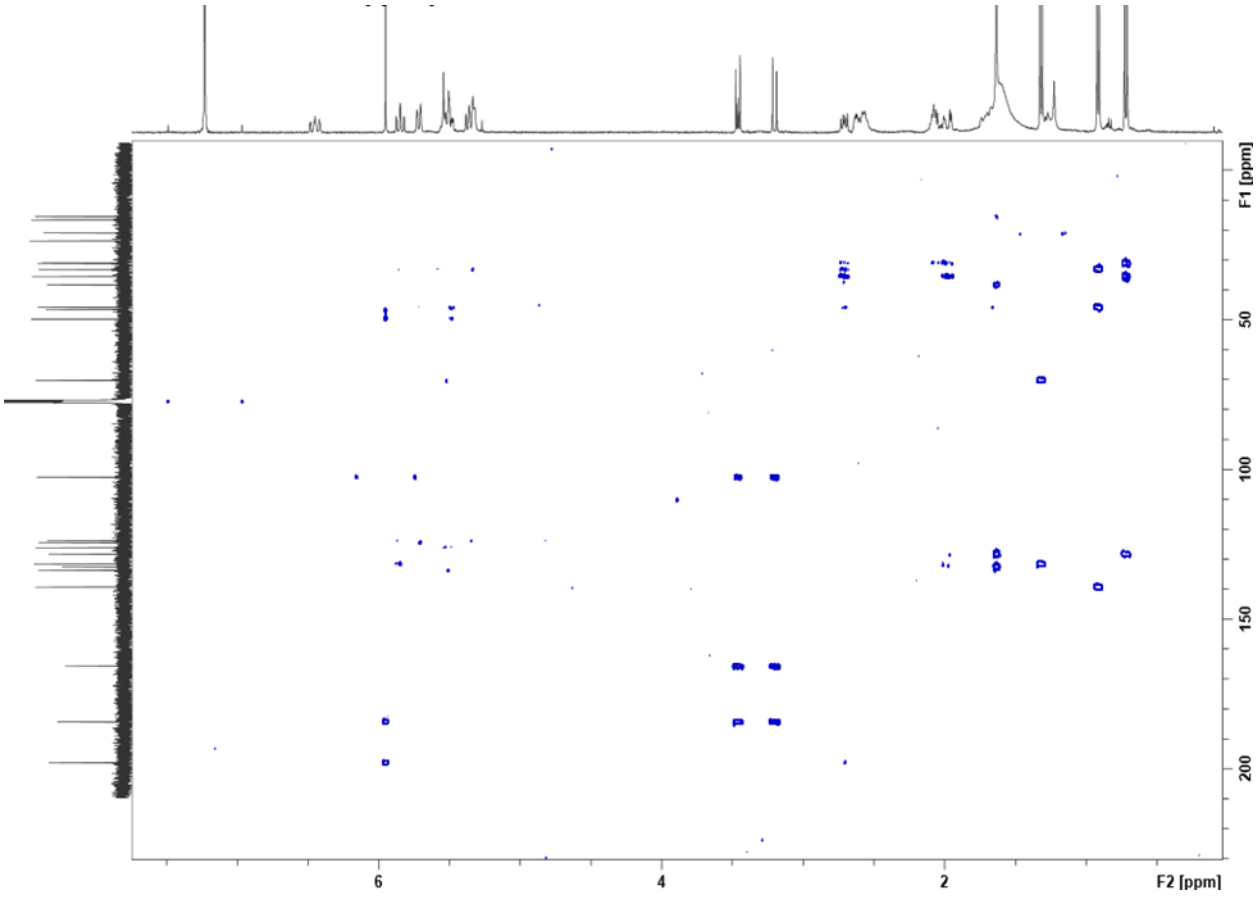
(c) COSY spectrum of anthracimycin BII-2619 (**2**) in $CDCl_3$ at 400 MHz.



(d) HSQCED spectrum of anthracimycin BII-2619 (**2**) in $CDCl_3$ at 400 MHz.



(e) HMBC spectrum of anthracimycin BII-2619 (**2**) in $CDCl_3$ at 400 MHz.



(f) NOESY spectrum of anthracimycin BII-2619 (**2**) in $CDCl_3$ at 400 MHz.

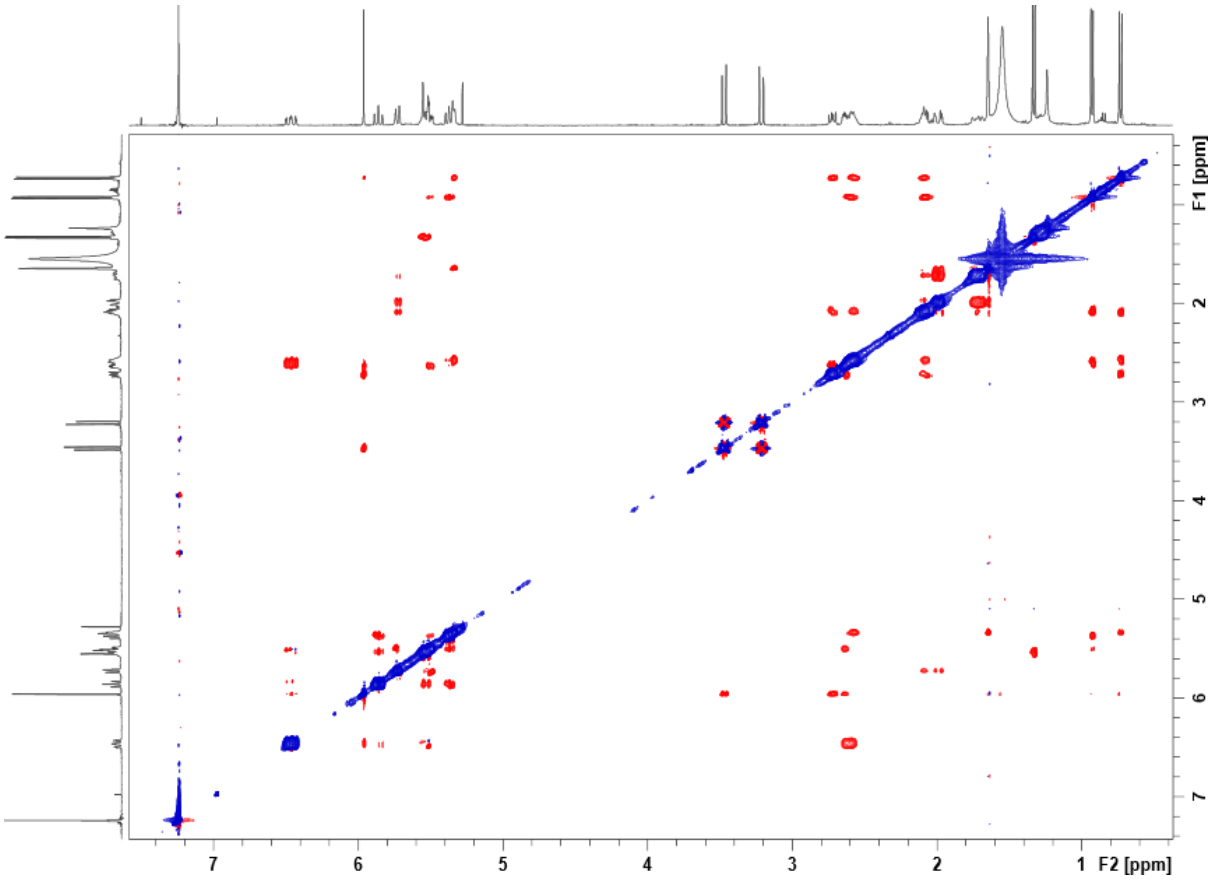
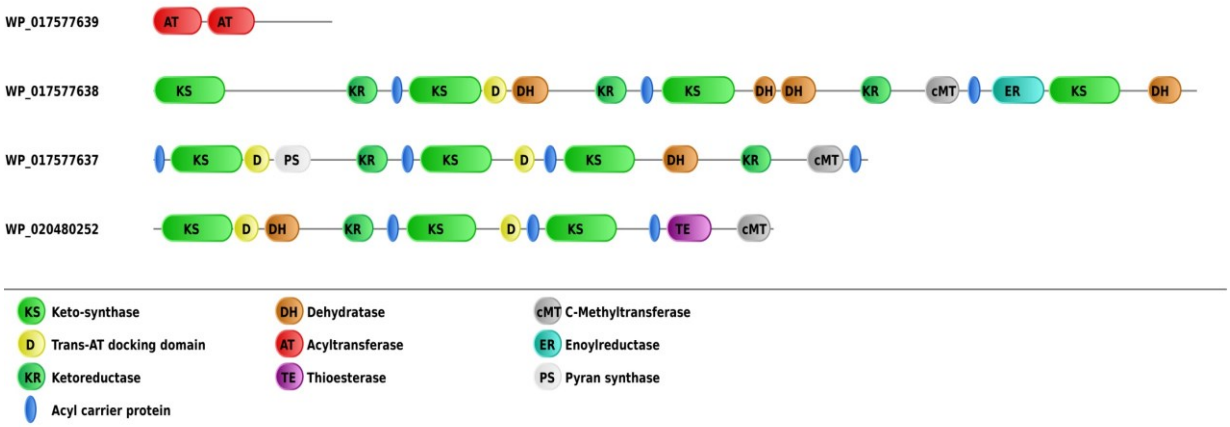


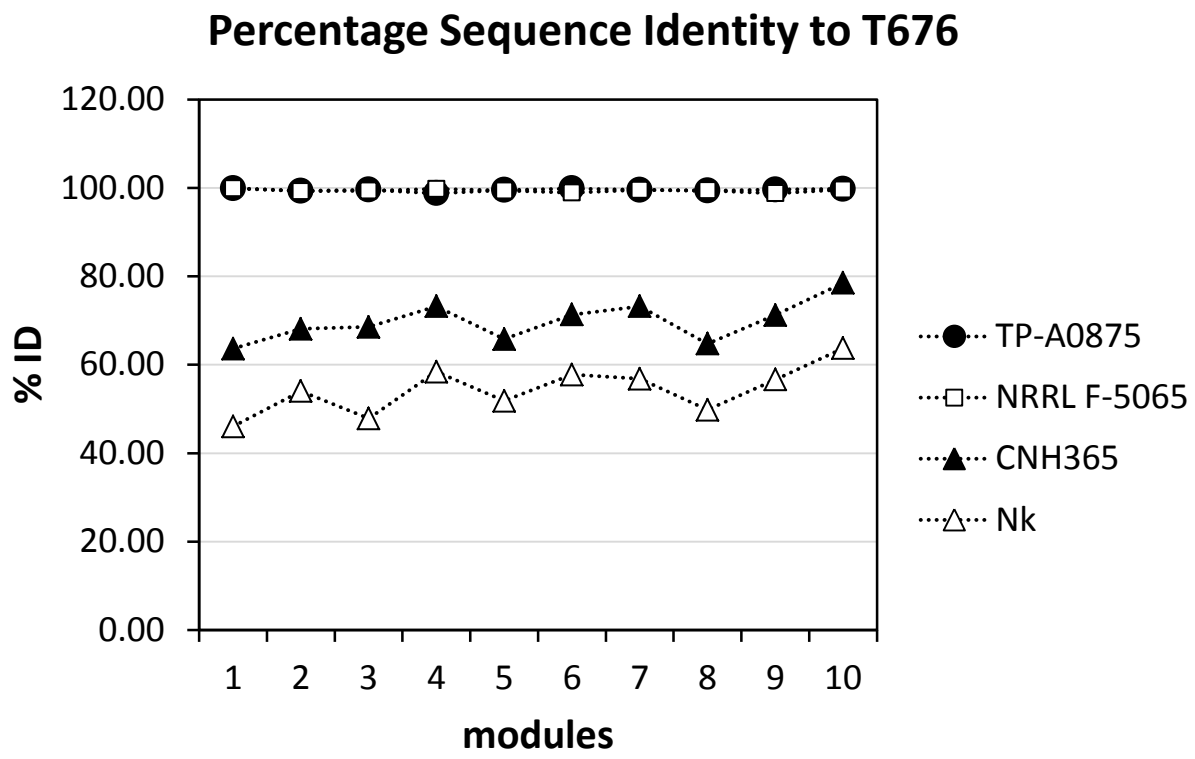
Figure S3

AntiSMASH prediction for *N. kunsanensis* contig 32 (NZ_ANAY01000032) (ActC-ActF).



Schematic representation of the four main functional proteins (AtcC, WP_017577639; AtcD, WP_017577638; AtcE, WP_017577637; and AtcF, WP_020480252) from the anthracimycin biosynthetic gene cluster of *N. kunsanensis* based on antiSMASH. The different functional domains are labelled accordingly.

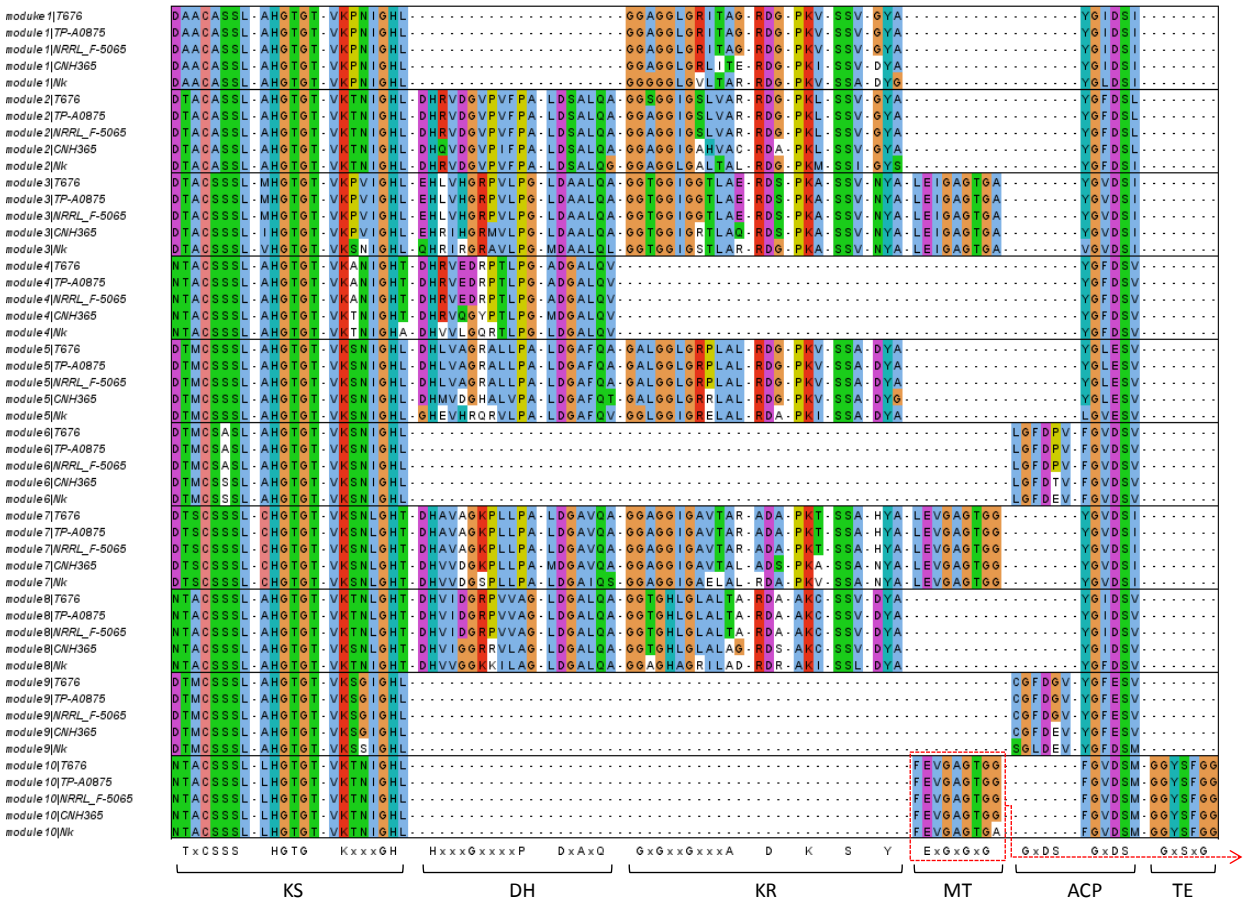
Figure S4



Protein sequences from functional modules belonging to *Streptomyces* sp. T676, *Streptomyces* sp. TP-A0875, *Streptomyces* sp. NRRL F-5065, *Streptomyces* sp. CNH365 and *Nocardiopsis kunsanensis* (Nk) were aligned with MAFFT E-INS-i algorithm as described in methods. Module boundaries were in agreement with antiSMASH prediction and all regions in between domains were maintained. The percentage sequence identities of T676 against all other four strains were calculated per module. This result is in agreement with the phylogenetic tree obtained from the complete protein sequence alignment (Figure 3 in the main text), where T676 is closest to TP-A0875 and NRRL F-5065, followed by CNH365 and the furthest organism *N. kunsanensis*.

Figure S5

(a) Sequence alignments of the core regions of functional domains.



Sequence alignment of regions containing conserved characteristic residues of the different PKS domains: keto-synthase (KS), dehydratase (DH), ketoreductase (KR), methyltransferase (MT), acyl carrier protein (ACP) and thioesterase (TE). To facilitate visualization, sequence motifs were grouped into their PKS functional modules ranging from 1 to 10 and dash symbols (-) were added to represent gaps of unspecified length. The different sequences belong to 4 anthracimycin producing *Streptomyces* strains (T676, TP-A0875, NRRL-F5065 and CNH365) and *N. kunsanensis* (Nk). ClustalX color scheme was used. The red box highlighting the MT motif in module 10 indicates that in its primary sequence, this domain occurs at the C-terminal end. Here, it was shifted between KR and ACP to facilitate alignment visualization to other motifs in MT domains from different modules.

(b) Phylogeny of KS domains

KS domains from the chlorotonil biosynthetic gene cluster from *Sorangium cellulosum* So ce1525 (red), the anthracimycin biosynthetic gene cluster from *N. kunsanensis* (blue) and all other predicted KS domains from *N. kunsanensis* by antiSMASH (black). Clade classification was done according to [9] based on the work of Nguyen *et al*. [3]. For the additional predicted KS domains by antiSMASH in black, we added the name of the predicted cluster type and its contig in the draft genome. Sequence alignment was done with MAFFT and the tree with MEGA6 using the Neighbor-Joining method. The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). The analysis involved 27 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 388 positions in the final dataset.

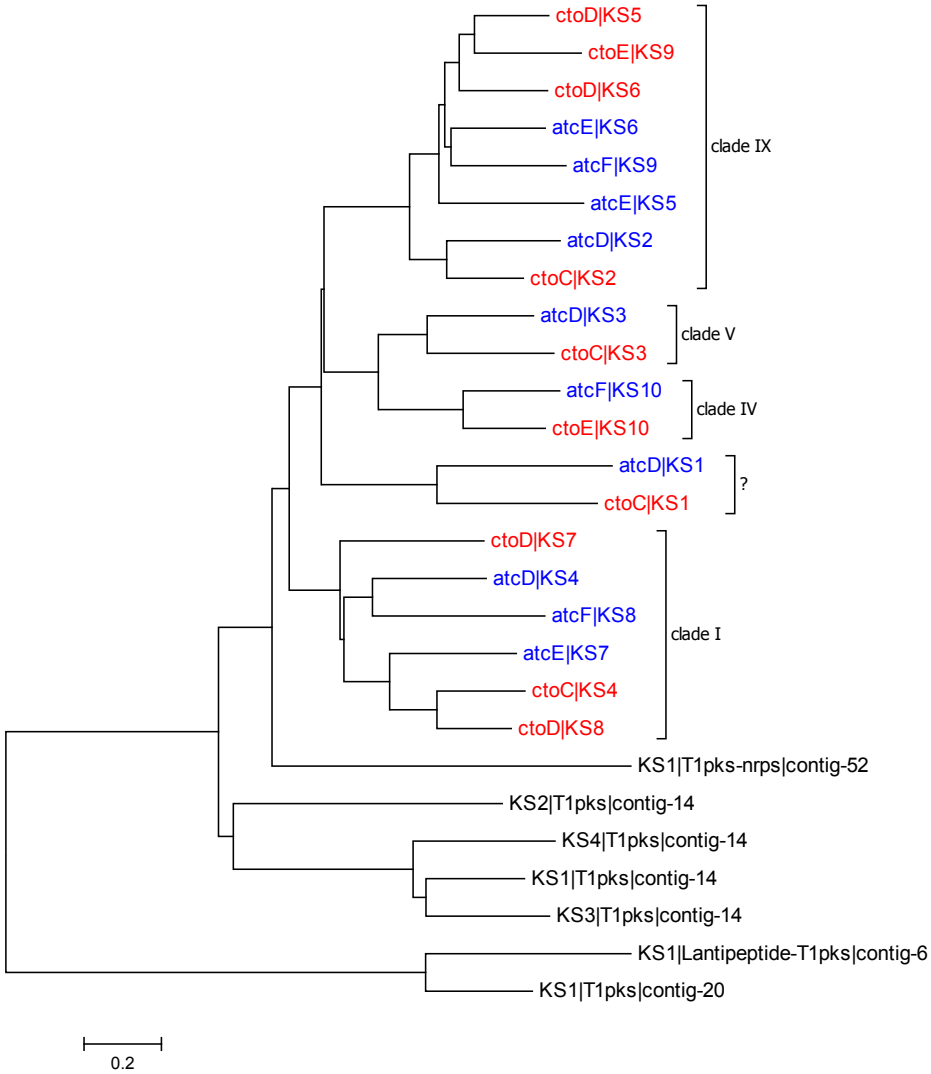
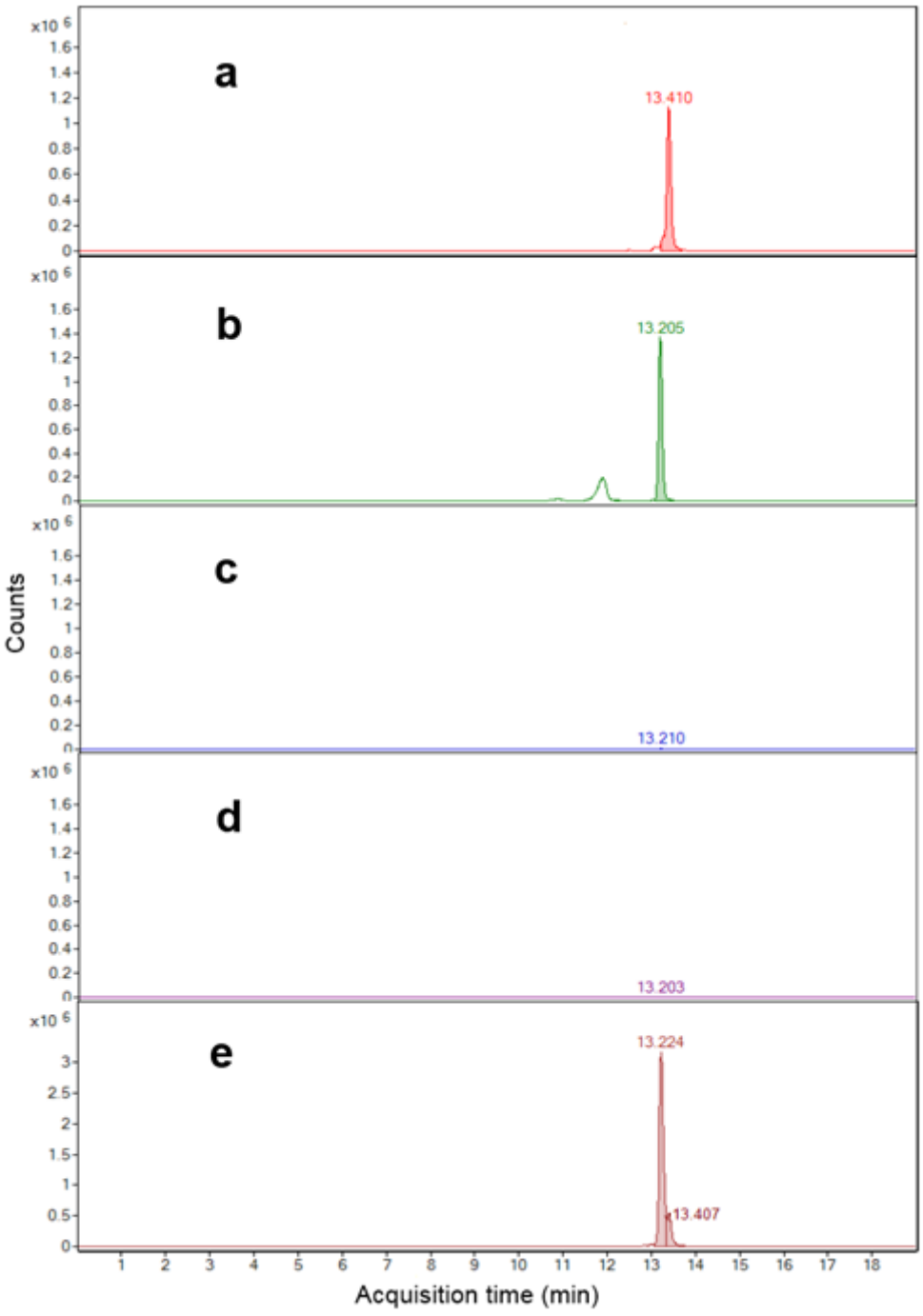


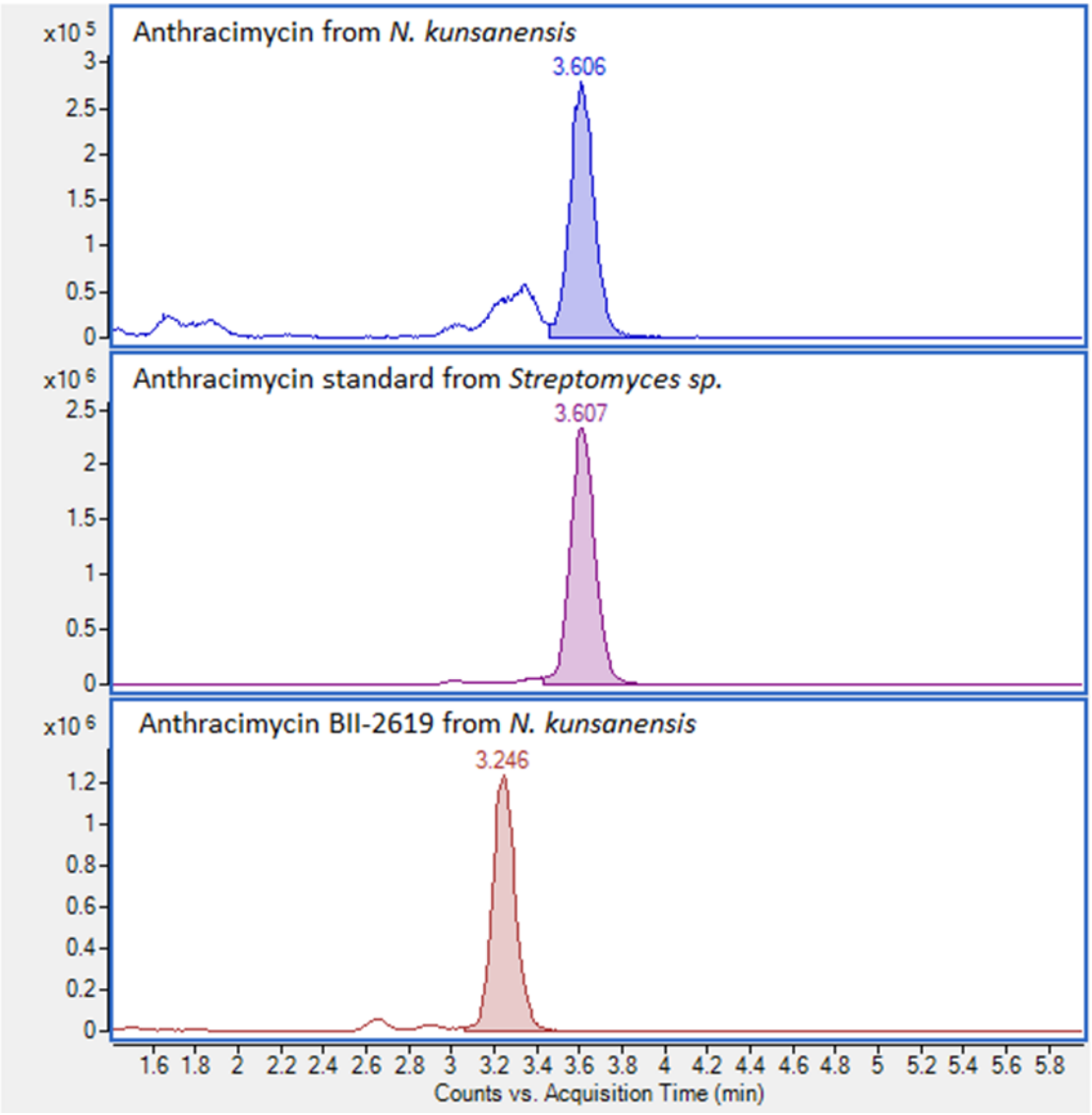
Figure S6 Chromatographic studies of extracts

(a) EIC (extracted ion chromatogram) of *m/z* 397.2379.



Anthracimycin standard (13.4 min) isolated from *Streptomyces* sp. strain T676 (a). Methanol extracts of *N. kunsanensis* cultivated in 50 mL of CM (b), CA02LB (c), and CA07LB (d). (e) Extract of *N. kunsanensis* biomass derived from the 4 liter fermentation in CM.

(b) Extracted ion chromatogram at m/z 397.2379 re-analyzed using an isocratic gradient.



In order to ascertain that the 0.2 min difference in retention time in Figure S6a is not a measurement error between anthracimycin (**1**) and anthracimycin BII-2619 (**2**), the three NMR samples (anthracimycin (**1**), anthracimycin standard and anthracimycin BII-2619 (**2**)) were re-analyzed using an isocratic gradient. Extracted ion chromatogram of m/z 397.2379 showed anthracimycin (**1**) and anthracimycin standard at retention time 3.6 min, whereas anthracimycin BII-2619 (**2**) is at 3.2 min. To note, the expected error rate for our HPLC device is clearly below 0.05 min (see also J.W. Dolan LCGC North America v.32 N8 2014 pp.546-551, reference [22] in paper).

Figure S7

CID MS/MS fragmentation spectra of anthracimycin (1)

(a) at energy 10 eV (b) at energy 20 eV (c) at energy 40 eV.

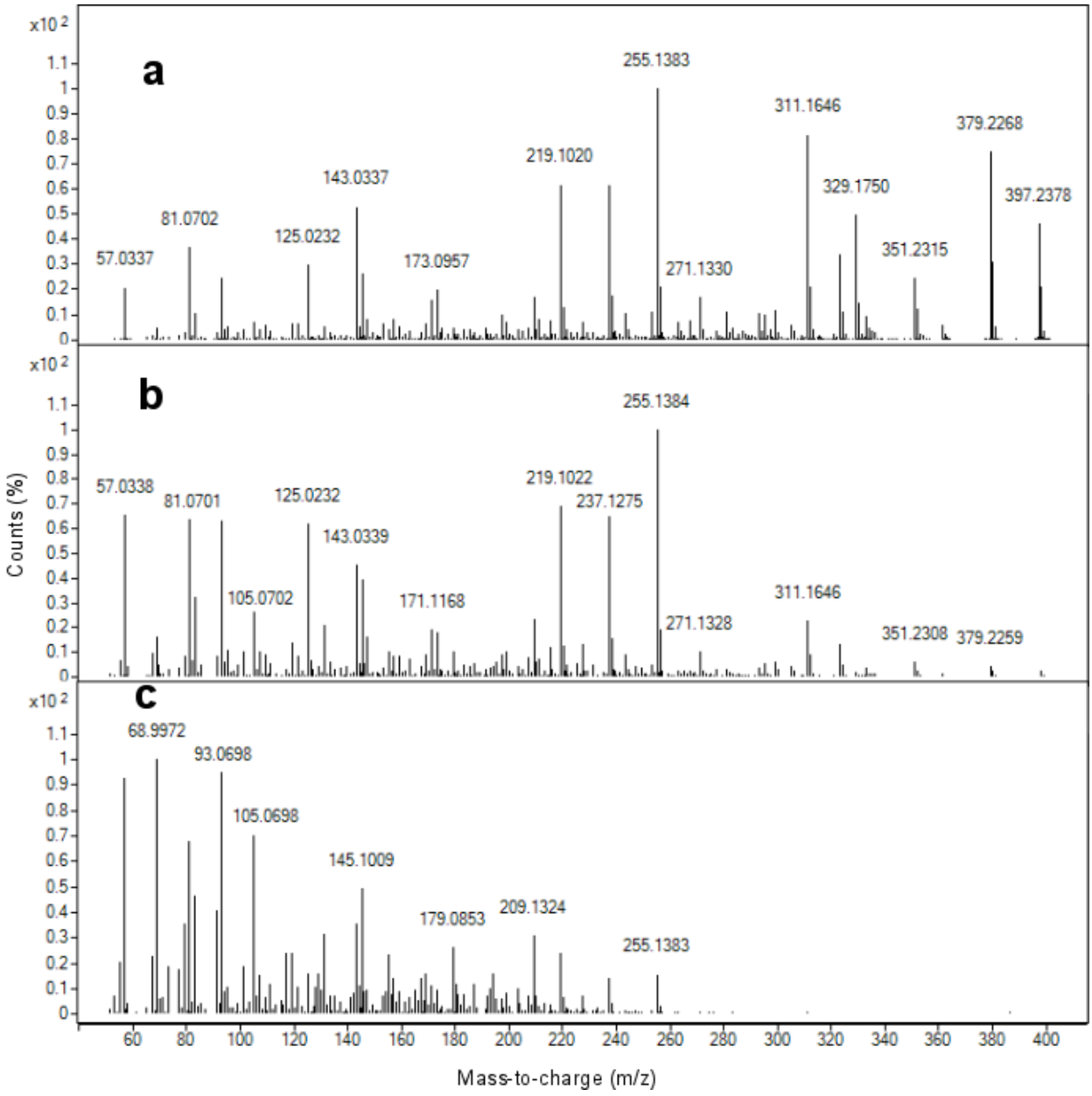


Figure S8

CID MS/MS fragmentation spectra of anthracimycin BII-2619 (2)

(a) at energy 10 eV (b) at energy 20 eV (c) at energy 40 eV.

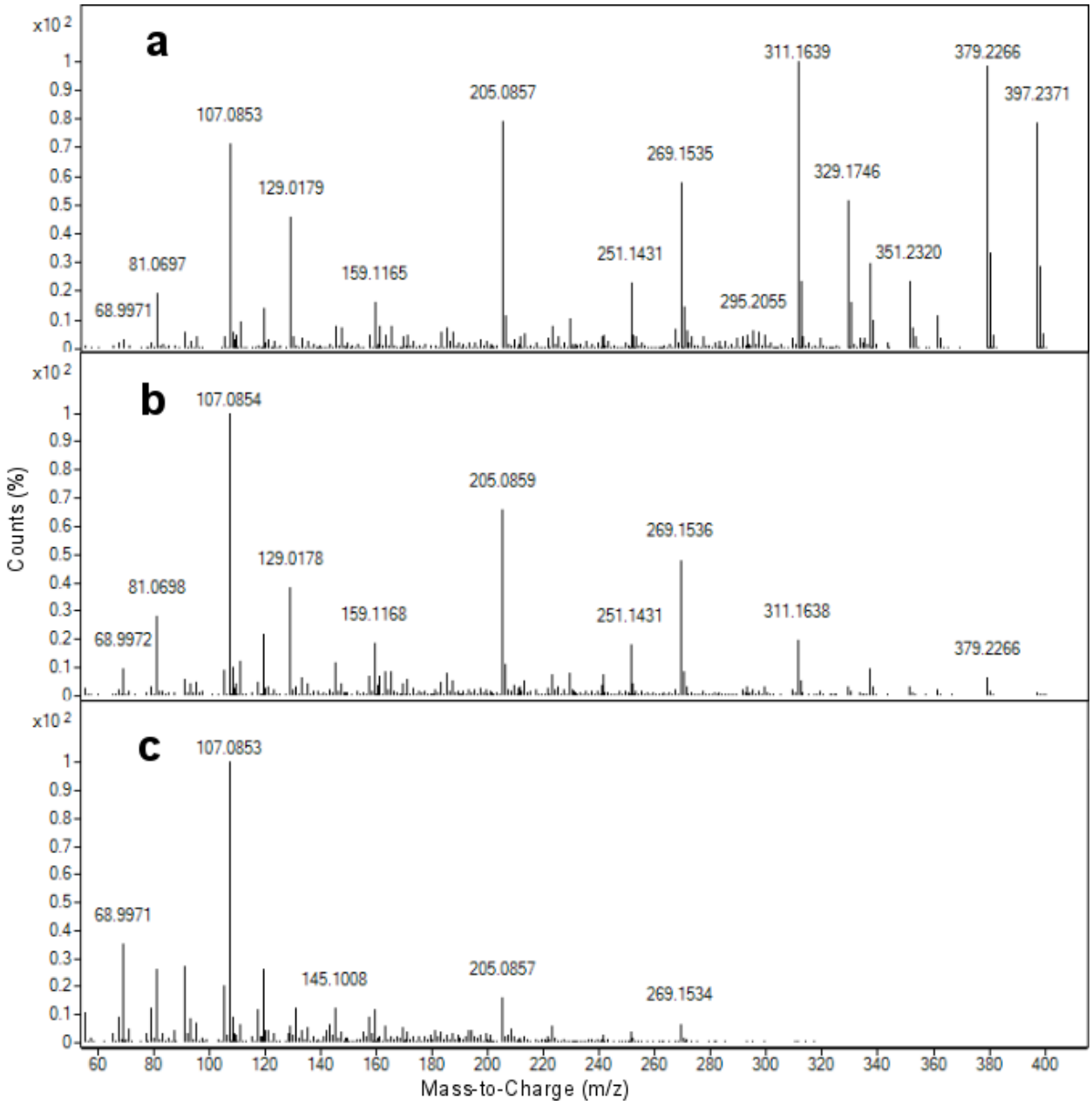
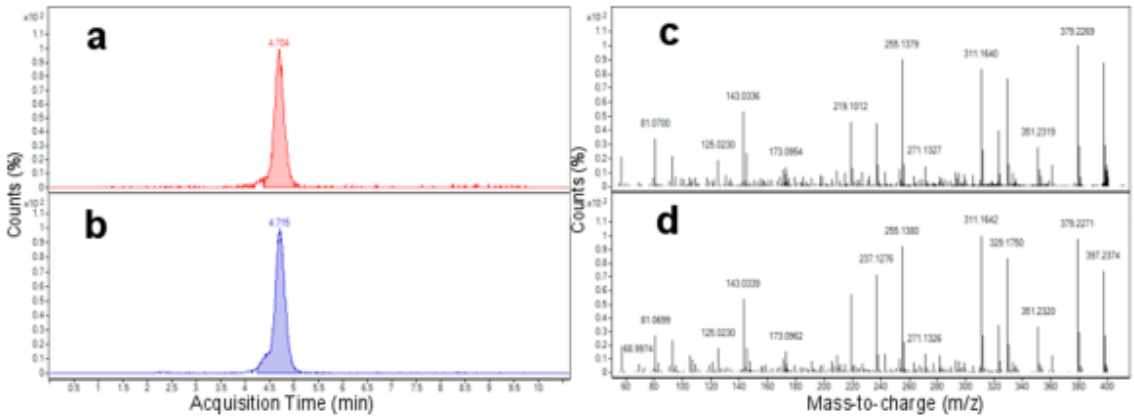


Figure S9

EIC of m/z 397.2379 on a chiral column.

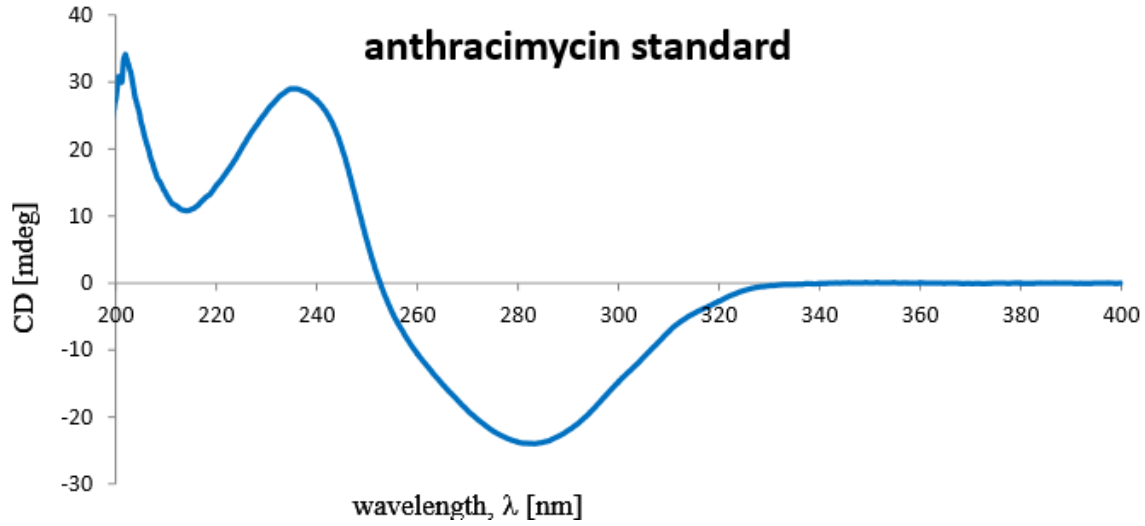


(a) anthracimycin (**1**) and (b) anthracimycin standard; MS/MS fragmentation spectra of m/z=397.2379 from (c) anthracimycin (**1**) and (d) anthracimycin standard.

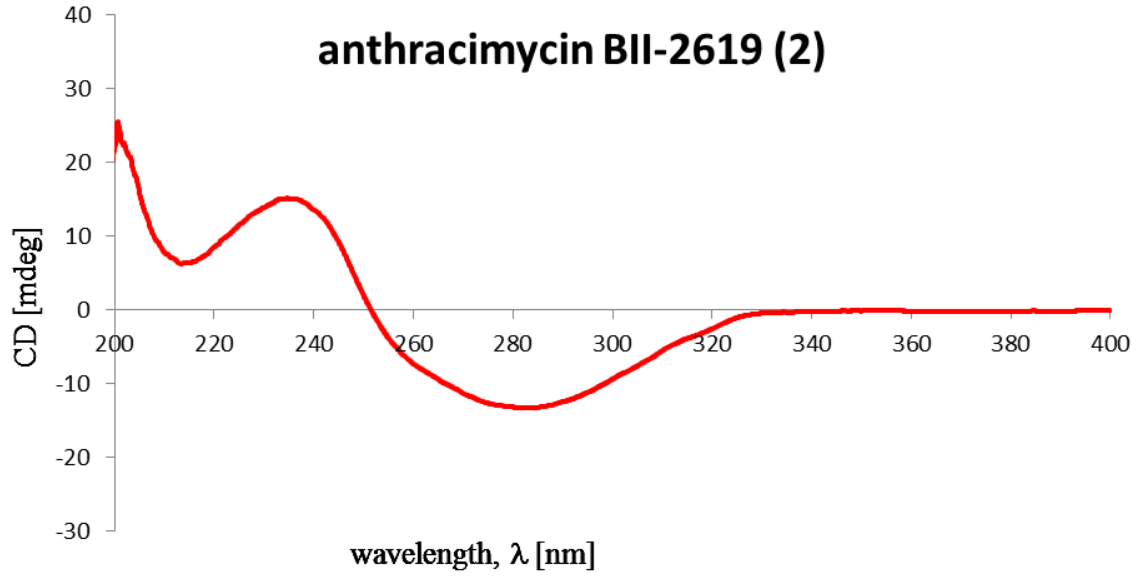
Figure S10

Experimental electronic circular dichroism spectrum of anthracimycin standard and anthracimycin BII-2619 (2) in methanol.

(a)



(b)



(a) The ECD results for our anthracimycin standard are in good agreement with those of the recently reported anthracimycin (Supplementary page S20 in ref. Jang *et al.* Anthracimycin, a potent anthrax antibiotic from a marine-derived actinomycete. Angew. Chem. Int. Ed Engl. 2013;52:7822–4). (b) The ECD results for anthracimycin BII-2619 (**2**) are provided for future reference. The form of the graph is the same as in (a).