Supplementary material

Table of Contents

- **Materials and Methods**
- **Supplementary discussions**

  Supplementary discussion 1. Structure determination of arimetamycin A (**1**)
  Supplementary discussion 2. Structure determination of arimetamycins B (**2**) and C (**3**)
  Supplementary discussion 3. The proposed biosynthesis of arimetamycins A – C (**1** – **3**)
  Supplementary discussion 4. TAR assembly of overlapping clones

- **Supplementary figures**

  Supplementary figure 1. 2D NMR correlations used to determine the structure of arimetamycin A (**1**).
  Supplementary figure 2. 2D NMR correlations used to determine the structures of arimetamycins B (**2**) and C (**3**).
  Supplementary figure 3. The proposed biosynthetic pathway for the dissacharide anthracycline arimetamycin A (**1**) and the monosaccharide anthracyclines arimetamycins B (**2**) and C (**3**).
  Supplementary figure 4. Transformation-associated recombination (TAR) assembly of the overlapping clones AZ493 and AZ129 into a 59 kb insert containing BAC clone.

- **Supplementary tables**

  Supplementary table 1: NMR spectroscopic data of arimetamycin A (**1**) in DMSO-d_6
  Supplementary table 2: NMR spectroscopic data of arimetamycins B (**2**) and C (**3**) in DMSO-d_6
  Supplementary table 3: NMR spectroscopic data of 8-demethoxy-10-deoxysteffimycinone (**4**) in DMSO-d_6
  Supplementary table 4: Gene annotation table of the eDNA clone AZ129

- **References for supplementary material**
- **NMR spectra: supplementary figures 5 – 26**

Materials and methods:

Construction of arrayed libraries: The four archived soil eDNA libraries used in this study were constructed according to previously published protocols.[1] Briefly, 100 – 250 gram aliquots of sifted topsoil were mixed 1:1 (w:v) with lysis buffer [100 mM Tris-HCl, 100 mM EDTA, 1.5 M NaCl, 1% (w/v) CTAB, 2% (w/v) SDS, pH 8.0] and incubated at 70°C for 2 hrs. Soil debris was then removed from the crude lysate by centrifugation (4,000 X *g*, 30 min). 0.7 volumes of isopropanol was added to the supernatant to precipitate crude eDNA, which was subsequently collected by centrifugation (4,000 X *g*, 30 min). eDNA was separated from the remaining soil particulates by gel electrophoresis (1% agarose gel, 20V, 16 hrs). The high molecular weight DNA compression band was electroeluted from the gel and concentrated using a centrifugal concentrator. Purified eDNA was blunt ended (Epicentre, End-It), ligated into the SmaI site of either pWEB or pWEB-TNC (Epicentre), packaged into lambda phage (Epicentre, MaxPlax) and transfected into *E. coli* EC100. Each library was expanded to contain at least 10,000,000 unique cosmid clones. Libraries were arrayed as individual pools of approximately 5,000 unique clones, resulting in approximately 2,000 non-redundant pools for each library. Matching glycerol stock and DNA minipreps corresponding to each sub-pool were created to facilitate clone recovery and PCR screening, respectively.

Library screening and phylogenetic analysis of PCR amplicons: DNA aliquots from the 5,000 membered sub-pools found in each arrayed soil library were screened using degenerate primers designed to amplify min-PKS cassettes containing full-length KS_β genes. Each twenty-five µL PCR reaction contained 50 ng of cosmid DNA, 2.5 µM of each primer (dp:KS_α-TTCGGSGGITTCCAGWSIGCSATG and dp:ACP-TCSAKSAGSGCSAISGASTCGTAICC),[2] 2 mM dNTPs, 1X ThermoPol reaction buffer (New England Biolabs), 0.5 units Taq DNA polymerase and 5% DMSO. PCR was conducted using the following touchdown protocol: denaturation (95 °C, 2 min), 8 touchdown cycles [95 °C, 45 s; 65 °C (−1 °C per cycle), 1 min; 72 °C, 2 min], 35 standard cycles (95 °C, 45 s; 58 °C, 1 min; 72 °C, 2 min), and a final extension step (72 °C, 2 min).[2] The resulting PCR amplicons were gel-purified and sequenced. 400 bp KS_β gene fragments (corresponding to nucleotides 252-652 of *DpsB*) from each amplicon were aligned using ClustalW. The pairwise distances (p-distance) between eDNA-derived KS_β sequences and KS_β sequences from gene clusters encoding for known medicinally relevant aromatic polyketides were calculated using MEGA5.1.[3] eDNA-derived KS_β sequences that fell within a p-distance of 0.3 to a KS_β from a medicinally relevant metabolite were used to generate sub-trees. One thousand bootstrap replicates were used to evaluate the robustness of branches in each neighbor-joining tree. The final tree (Fig. 1 of the manuscript) was constructed with the eDNA-derived KS_β sequences that formed well-supported clades with the KS_β sequences of medicinally relevant metabolites.

Clone recovery and bioinformatics analysis of recovered gene clusters: eDNA clones (AZ129, AZ515 and TX19) containing KS_β sequences that formed well-supported clades with KS_β sequences from the biosynthesis of known anthracyclines were recovered from eDNA libraries by dilution PCR screening of the respective sub-pools from which they were identified. To accomplish this, overnight cultures of each sub-pool were plated into 96-well microtiter plates at a dilution of 10^{-5} or 10^{-6}. After 18 hrs at 37 °C, the diluted cultures were screened by whole-cell PCR using the KS_β screening primers and touch-down PCR protocol described above. PCR positive wells were plated onto solid media to yield distinct colonies that were screened in a second round of whole-cell PCR. The recovered cosmids were sequenced by 454-pyrosequencing and named after the specific library sub-pools from which they were isolated (e.g. AZ129, AZ515 and TX19). Open reading frame predictions were carried out using MetaGeneMark[4], and predicted open reading frames were annotated based on Blast search results. eDNA-derived gene clusters were compared manually to previously reported gene clusters for the biosynthesis of doxorubicin/daunorubicin (GeneBank Accession No.: L35560), aclacinomycin (AF257324), cosmomycin (ABC00728), nogalamycin (AJ224512) and steffimycin (AM156932).

Heterologous expression and cloning of the rhamnose biosynthetic genes: To permit the shuttling of cosmid AZ129 into *Streptomyces* spp. for heterologous expression studies, cosmid AZ129 was digested with *PsiI* and then ligated with the 6.81 kb DraI fragment of pOJ436[5] that contains an origin of transfer (oriT), an apramycin resistance gene [aac (3)IV], and the ϕC31-based integration system. Using *E. coli* S17.1 as a donor strain, the retrofitted cosmid was transferred into *Streptomyces albus* by intergenic conjugation to yield *S. albus*/AZ129. The rhamnose biosynthetic operon (oleL: dNDP-4-keto-6-deoxyglucose 3,5-epimerase, oleS: dNDP-D-glucose synthase, oleE: dNDP-glucose 4,6-dehydratase and oleU: dNDP-4-ketohexulose reductase) was amplified from pRHAM[6] using the forward primer 5'-ATCGCATATGGAGTTACTCGACGTCGACGGGG-3' and the reverse primer 5'-ATCGTTAATTAATCATGCTGCTCCTCGCCGGGTCGGT-3' [restriction sites added for cloning purposes (*NdeI* and *PacI*, respectively) are underlined]. The resulting amplicon was digested with *NdeI*/*PacI* and cloned in front of the *ermE** promoter in the shuttle expression vector pIJ10257 to give pIJRham. This expression vector utilizes an orthogonal phage integration system (ΦBT1) and resistance gene (hygromycin) to those present on the retrofitted AZ129 cosmid clone; thereby, allowing for the co-integration of both the biosynthetic gene cluster and the rhamnose biosynthetic operon into distinct chromosomal sites in *S. albus*, yielding *S. albus*/AZ129/pIJRham. pIJRham was transformed into *E. coli* S17.1 and then shuttled into *S. albus/AZ129* by intergenic conjugation to generate *S. albus/AZ129/pIJRham*.

Extraction and Isolation of clone-specific metabolites: Compound **4** was isolated from cultures of *S. albus/AZ129* grown (200 rpm at 30 °C) in 125 ml baffled flasks containing 50 ml of R5A media. Seven days old cultures were extracted with ethyl acetate (3:1 v/v). The resulting extract was subjected to isocratic (65% aqueous, 35% acetonitrile containing 0.1% trifluoroacetic acid) reversed-phase HPLC (C_{18}, 10 mm × 250 mm, 3.5 mL/min) to afford 0.5 mg of **4** per 100 ml of culture broth. The structure of **4** was determined by HRESIMS (369.0976 $[M-H]^-$; 369.0974 calcd for $C_{20}H_{17}O_7$) and 1D and 2D NMR data (see supplementary table 3) to be the previously described steffimycin biosynthetic intermediate 8-demethoxy-10-deoxysteffimycinone.[7]

Arimetamycins A (**1**), B (**2**) and C (**3**) were isolated from cultures of *S. albus/AZ129/pIJRham*. 125 ml baffled flasks containing 50 ml of R5A were inoculated with *S. albus/AZ129/pIJRham* spore stocks. After seven days of shaking (200 rpm at 30 °C), cultures were extracted with ethyl acetate (EtOAc, 3:1 v/v) and concentrated in vacuo. Arimetamycins A (**1**), B (**2**) and C (**3**) were isolated from the EtOAc extract using two rounds of reversed-phase HPLC (C_{18} column, 10 mm × 250 mm, 3.5 mL/min). The first round of HPLC using 35% acetonitrile with 0.1% trifluoroacetic acid yielded three crude samples, which were re-purified by the second round of HPLC using 60%, 70% and 80% methanol with 0.1% trifluoroacetic acid to yield arimetamycins A (**1**, 2.5 mg/L), B (**2**, 0.5 mg/L) and C (**3**, 3 mg/L), respectively.

Arimetamycin A (1): red gum; $[\alpha]^{25}_D$ + 15 (*c* 0.2, MeOH); UV (MeOH) λ_{max} 230, 270, 289, 445 nm; IR (neat) ν_{max} 3200, 3073, 2990, 2852, 2118, 1673, 1625, 1568 cm^{-1}; ^1H and ^{13}C NMR, COSY, and HMBC data, see supplementary Table 1; HRESIMS m/z 757.3553 $[M+H]^+$ (cacld for $C_{39}H_{52}N_2O_{13}$, 757.3548).

Arimetamycin B (2): red powder; $[\alpha]^{25}_D$ + 37 (*c* 0.1, MeOH); UV (MeOH) λ_{max} 229, 270, 290, 446 nm; IR (neat) ν_{max} 3404, 3092, 2929, 2056, 1678, 1610 cm^{-1}; ^1H and ^{13}C NMR, see supplementary Table 2; HRESIMS m/z 543.1479 $[M-H]^-$ (cacld for $C_{27}H_{27}O_{12}$, 543.1503).

Arimetamycin C (3): red powder; $[\alpha]^{25}_D$ + 83 (*c* 0.2, MeOH); UV (MeOH) λ_{max} 229, 270, 290, 446 nm; IR (neat) ν_{max} 3272, 3084, 2932, 2854, 2109, 1707, 1677, 1623, 1565 cm^{-1}; ^1H and ^{13}C NMR, COSY, and HMBC data, see supplementary Table 2; HRESIMS m/z 557.1665 $[M-H]^-$ (cacld for $C_{28}H_{29}O_{12}$, 557.1659).

TAR assembly and analysis of overlapping eDNA clones: An eDNA clone overlapping with clone AZ129 (cosmid

AZ493) was identified in the initial min-PKS PCR screen and subsequently recovered from the AZ library as outlined above. AZ493 was fully sequenced and found to overlap with approximately 8.5 kb of AZ129. The eDNA captured on these two cosmids was reassembled into a single continuous fragment using transformation-associated recombination (TAR) in yeast and the pTARa bacterial artificial chromosome (BAC) shuttle vector.

The required pathway-specific TAR capture vector was constructed using InFusion cloning methodology (Clontech). One-kilobase upstream and downstream homology arms were amplified from AZ493 and AZ129, respectively, using the following primers:

AZ493UPS_FW (5'-CTATCGATCTCGAGGGTACGAAGCGATGAGCTTT-3'),
AZ493UPS_RV (5'-GTGACCCTT**GTTAAC**CTTTTACGAGCACTTCTCC-3'),
AZ129DWS_FW (5'-**GTTAAC**AAGGGTCACCCAACAGAACC-3'),
AZ129DWS_RV (5'-CCCTGCAGGAGCTCGCCTGATTGCCATTCGTC-3').

AZ493UPS_FW and AZ129DWS_RV were designed to include 15 bp of overlapping sequence (underlined) with either end of the linearized pTARa capture vector (BmtI/SphI). AZ493UPS_RV was designed to contain a 15 bp overlap (underlined) with the AZ129DWS_FW primer including an HpaI site (bold), which was added to facilitate the subsequent linearization of the pathway specific pTARa capture vector. Gel purified amplicons (Qiagen) and BmtI/SphI linearized pTARa were combined in a standard InFusion cloning reaction (Clontech) to yield the AZ493/AZ129 cosmid clone specific capture vector.

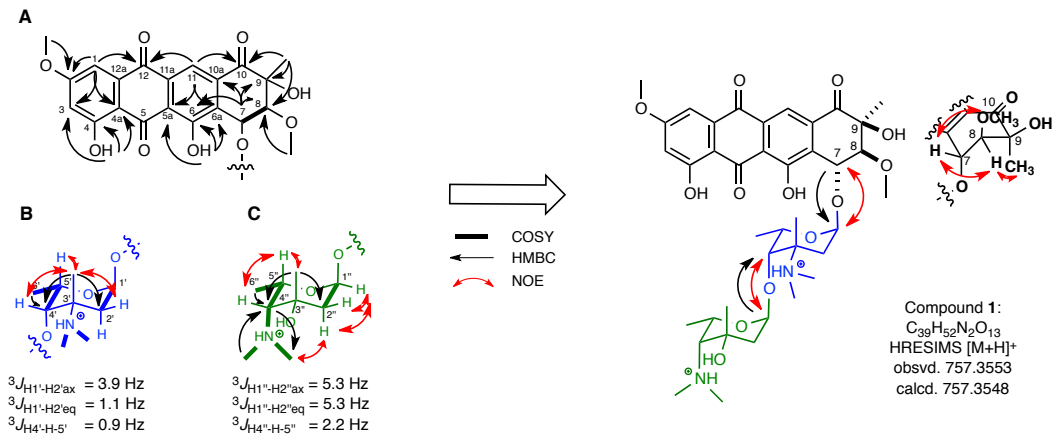
The AZ493/AZ129 TAR reconstruction reaction was carried out as follows. Two hundred nanogram of DraI-digested AZ493 and AZ129 were mixed with HpaI-cut AZ493/AZ129 specific capture vector and then transformed into 200 μl of *Saccharomyces cerevisiae* spheroplasts according to published methods.[8] Transformed speroplasts were mixed with the synthetic complete (SC) top agar (1 M sorbitol, 1.92 g/L synthetic complete uracil dropout supplement, 6.7 g/L yeast nitrogen base, 2% glucose and 2.5% agar) and overlaid onto SC uracil dropout plates. Plates were incubated at 30 °C for 72 hrs. DNA was prepared from 12 yeast colonies using a zymolyase lysis protocol (Zymoresearch) and screened by PCR with primers designed to recognize sequences from both AZ493 and AZ129. A BAC clone that produced amplicons from both AZ493 and AZ129 was electroporated into *E. coli* EPI300. DNA isolated from the resulting *E. coli* transformant was restriction mapped and sequenced to confirm the correct reassembly of the two cosmids into a 59 kb insert BAC clone (BAC*AZ493AZ129*).

For heterologous expression studies, BAC*AZ493AZ129* was transformed into *E. coli* S17.1 and conjugated into both *S. albus* and *S. albus/pIJrham* to generate *S. albus/BACAZ493AZ129* and *S. albus/BACAZ493AZ129/pIJRham*, respectively. These strains were cultured in R5A medium for 7 days (200 rpm at 30 °C) and extracted with ethyl acetate (3:1 v/v). The resulting ethyl acetate extracts were compared by HPLC with extracts from similarly grown cultures of *S.albus/AZ129* and *S.albus/AZ129/pIJRham* (5% to 100% aqueous: acetonitrile gradient (over 30 min) containing 0.1% trifluoro acetic acid, C_{18}, 4.6 mm x 150 mm, 1 ml/min).

Antiproliferative assay: The cytotoxicities of arimetamycins A (**1**), B (**2**) and C (**3**) were evaluated using four human cancer cell lines: HCT-116 colon carcinoma (ATCC; CCL-247),[9] WiDr colorectal adenocarcinoma (ATCC; CCL-218),[10] MDR-MB-231 breast adenocarcinoma (ATCC; HTB-26)[11] and H69AR multidrug resistant lung carcinoma (ATCC; CRL-11351)[12] cell lines. HCT-116 and H69AR cells were grown in McCoy's 5A Modified Medium (Gibco) supplemented with 10% (v/v) FBS and RPMI-1640 (ATCC) supplemented with 20% (v/v) FBS, respectively. WiDr and MDA-MB-231 cells were grown in DMEM Medium (Gibco) supplemented with 10% (v/v) FBS. All the cell lines were incubated at 37

°C with 5% CO_2. Cells in log phase growth were harvested by trypsinization. Cells were seeded into 96-well plates (HCT-116: 1,000 cells/well, WiDr: 3,000 cells/well, MDA-MB-231: 8,000 cells/well, H69AR: 25,000 cells/well) and incubated overnight at 37 °C with 5% CO_2. Compounds **1** – **3** (in DMSO) were sequentially diluted (3-fold starting at 50 µg/ml) across a 96-well plate and 100 µL was transferred to the appropriate wells in the assay plates. Doxorubicin (Sigma-Aldrich) and daunorubicin (Sigma-Aldrich) were used as positive controls and DMSO was used as a negative control. The plates were incubated at 37 °C for 3-6 days depending on the growth rate and then evaluated for viability using either an MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) (MDR-MB-231 and H69AR) or crystal violet (HCT-116 and WiDr) based colorimetric assay.[13] Cell viability is recorded based on the percent stain present in each well relative to no drug DMSO control wells. Assays were run in triplicate.

Supplementary Discussion 1. Structure determination of arimetamycin A (**1**)



Supplementary figure 1. 2D NMR correlations used to determine the structure of arimetamycin A (**1**).

The structure of **1** was determined using a combination of HRESIMS and 1D and 2D NMR data. The HRESIMS spectrum of **1** displayed a molecular ion peak at *m/z* 757.3553 ([M+H]$^+$), corresponding to a predicted molecular formula of $C_{39}H_{52}N_2O_{13}$. The ^{1}H NMR spectrum exhibited an overall signal distribution typical of glycosylated aromatic compounds including signals corresponding to phenolic hydroxyl, aromatic and sugar anomeric protons. The appearance of two anomeric proton signals (δ_H 5.69 and 5.26) and carbon signals (δ_C 99.6 and 98.2) in the ^{1}H, ^{13}C and HMQC NMR spectra suggested that compound **1** contained two sugar moieties.

Analysis of the HMBC, COSY and NOE spectra initially established three substructures designated as A (tetracyclic aromatic core), B (sugar) and C (sugar) (Supplementary Figure 1). Based on empirical chemical shift data and HMBC correlations, substructure A was determined to be the tetracyclic structure steffimycinone, which is an aglycone common to the steffimycin class of anthracyclines.[14] The small coupling constant (2.57 Hz) observed between H-1 (δ_H 7.28) and H-3 (δ_H 6.96) indicated a *meta* relationship between these two protons. HMBC correlations from H-1 to C-2 (δ_C 166.7), C-3 (δ_C 106.8), C-4a (δ_C 110.4) and C-12 (δ_C 180.5) and from 4-OH (δ_H 12.00) to C-3, C-4 (δ_C 164.8) and C-4a, together with an HMBC correlation from 2-OCH$_3$ (δ_H 3.97) to C-2 (δ_C 166.7) suggested the presence of an aromatic ring substituted with a methoxy at C-2, a hydroxyl at C-4 and a ketone at C-12a (δ_C 133.3). The remaining singlet aromatic proton signal (H-11, δ_H 8.06) was indicative of the presence of an additional penta-substituted aromatic ring. An HMBC correlation from H-11 to C-12 indicated that these two aromatic rings were bridged by the C12 ketone. HMBC correlations from H-11 to C-5a (δ_C 118.8) and C-6a (δ_C 132.8) as well as HMBC correlations from 6-OH (δ_H 12.80) to C-5a, C-6 (δ_C 161.3) and C-6a placed the hydroxyl on C-6. This second aromatic ring was expanded further by HMBC correlations from H-11 to C-10 (δ_C 198.6), from 9-CH$_3$ (δ_H 1.39) to C-8 (δ_C 86.3), C-9 (δ_C 75.9) and C-10 (δ_C 198.6), from H-7 (δ_H 5.04) to C-6 (δ_C 161.3), C-8, C-9 and C-10a (δ_C 135.2) and from 8-OCH$_3$ (δ_H 3.45) to C-8. Together these correlations indicated the presence of an aliphatic ring functionalized with a ketone at C-10, a methyl and hydroxyl at C-9, a methoxy at C-8 and an alkoxy at C-7. No direct HMBC correlation was observed to connect the two quaternary carbons C-4a and C-5a. However after completing substructures B and C (as described below) only two atoms (a carbon and oxygen) remain unaccounted for the structure based on the predicted molecular formula. The highly up-fielded carbon chemical shift of the remaining ketone (C-5, δ_C 189.9) strongly suggested that carbons C-4a and C-5a were bridged by this ketone to give substructure A as steffimycinone. This substructure is further supported by the nearly perfect correlation between ^{1}H and ^{13}C chemical shift data from substructure A and data reported for this aglycone substructure in known

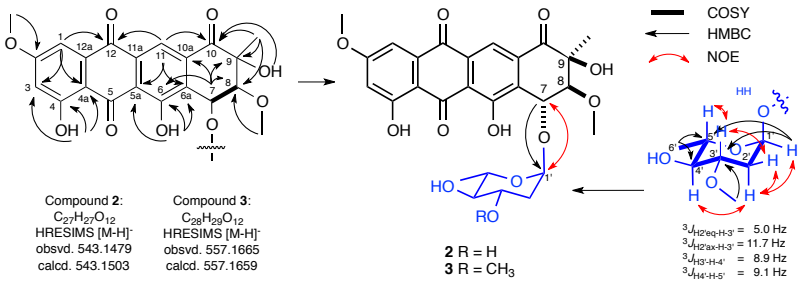
steffimycins.[15] The identical chemical shifts together with NOE correlations suggested that the relative configurations within the final ring in substructure A are the same as those reported for steffimycinone. An NOE correlation observed between 9-CH_3 and H-8 indicated a "*syn*" relationship between these protons, and thus between 9-OH and 8-OCH_3 as well. The proton H-7 showed NOE correlations with 8-OCH_3 and H-8 indicating an equatorial position for H-8, thus determining the relative configuration as "*anti*".

Substructure B was elucidated based on empirical chemical shift data as well as COSY and HMBC data. COSY correlations between H-1'/H_2-2', between H-4'/H-5'/H_3-6' and between 3'-NH and $N(CH_3)_2$ identified three spin systems. These three spin systems were connected by HMBC correlations from 3'-CH_3 (δ_H 1.36) to C-2' (δ_C 32.4), C-3' (δ_C 63.6) and C-4' (δ_C 75.9) and from $N(CH_3)_2$ (δ_H 2.70 and 2.77) to C-3'. Thus, defining substructure B as 2,6-dideoxy-4-hydroxy-3-methyl-3-*N,N*-dimethylamino-hexose. The relative configuration within this dideoxy sugar moiety was deduced by NOE correlations and coupling constants analysis. NOE correlations between H-5' and 3'-CH_3 and between H-4' and 3'-CH, together with the small coupling constant (0.9 Hz) observed between H-5' and H-4' placed 3'-CH_3 and 4'-O-R in axial positions and 3'-$N(CH_3)_2$ and CH_3-6' in equatorial positions. The small coupling constants observed between H-1' and both H-2'ax and H-2'eq indicated this sugar moiety was a β-anomer.

The remaining substructure C was determined in a similar fashion to that of substructure B. Two COSY correlation networks including H-1''/H_2-2" and H_3-6"/H-5"/H-4"/4"-NH/4"-$N(CH_3)_2$ established two spin systems. These two spin systems were connected through the quaternary carbon C-3" by HMBC correlations from 3''-CH_3 (δ_H 1.49) to C-2'' (δ_C 40.1), C-3'' (δ_C 68.3) and C-4'' (δ_C 68.1). The C-3" carbon chemical shift (δ_C 68.3) indicated the presence of a hydroxyl group at this position, thus defining substructure C as a 2,6-dideoxy-3-hydroxyl-3-methyl-4-*N,N*-dimethylamino-hexose. The relative configuration within this sugar moiety was also determined by NOE correlations and coupling constants. An NOE correlation between H-5'' and 3''-CH_3 as well as the small $^3J_{HH}$ (2.2 Hz) coupling observed between H-5'' and H-4'' placed both 3"-CH_3 and 4"-$N(CH_3)_2$ in axial positions and 3"-OH and CH_3-6" in an equatorial position. In the ^{1}H spectrum of **1**, H-1" appears as a triplet indicating the same coupling constant (5.3 Hz) between H-1" and both H-2"ax and H-2"eq, thus we inferred this sugar moiety to be a β-anomer.

The three substructures (A, B and C) were assembled into a final structure through additional HMBC correlations. An HMBC correlation from H-7 (δ_H 5.04) to C-1' (δ_C 99.6) indicated that the first sugar (substructure B) was connected to C-7 of the steffimycinone aglycone (substructure A) by an *O*-glycosidic linkage. The complete disaccharide arimetamycin A (**1**) was then defined by an HMBC correlation from H-1'' (δ_H 5.26) to C-4' (δ_C 75.9), which connects the second sugar moiety (substructure C) to C-4' of the first sugar through an *O*-glycosidic linkage.

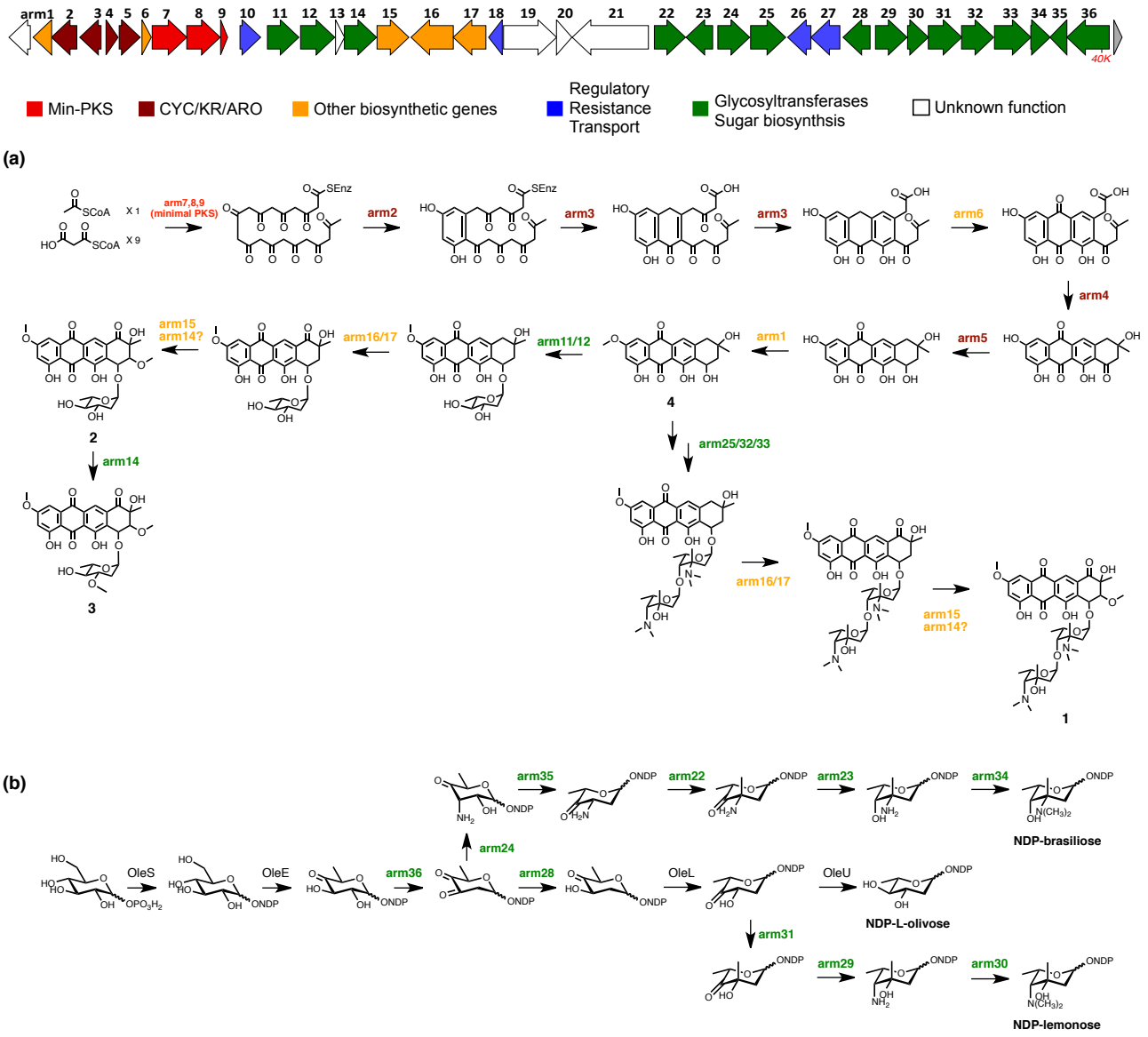
Supplementary Discussion 2. Structure determination of arimetamycins B (**2**) and C (**3**)



Supplementary figure 2. 2D NMR correlations used to determine the structures of arimetamycins B (**2**) and C (**3**).

HRESIMS analysis (*m/z* 543.1479 [M-H]$^-$ for **2**, *m/z* 557.1665 [M-H]$^-$ for **3**) predicted molecular formulas of $C_{27}H_{27}O_{12}$ and $C_{28}H_{29}O_{12}$ for **2** and **3**, respectively. The overall signal distribution patterns observed in the ^{1}H and ^{13}C NMR spectra were nearly identical for **2** and **3** except for the presence of one additional methoxy group (δ_H 3.25 and δ_C 56.6) in **3**. Analysis of 2D NMR spectra including COSY and HMBC indicated that compounds **2** and **3** shared the same aglycone structure (steffimycinone) as that of **1**, but differed in their sugar moieties. The presence of only one anomeric proton (δ_H 5.48 for **2** and 5.53 for **3**) and carbon (δ_C 100.9 for **2** and 100.7 for **3**) signals in the spectra for each of these compounds indicated that **2** and **3** were both mono-glycosylated. The sequential COSY correlations between H-1'/H_2-2'/H-3'/H-4'/H-5'/H_3-6' suggested the presence of a 2,6-dideoxy hexose sugar moiety. NOE correlations observed between H-5' (δ_H 3.71 and 3.74) and H-3' (δ_H 3.47 and 3.19), together with large coupling constants between H-2'ax and H-3' ($^3J_{HH}$ = 11.7 Hz), between H-3' and H-4' ($^3J_{HH}$ = 8.9 Hz), and between H-4' and H-5' ($^3J_{HH}$ = 9.1 Hz) defined the sugar moieties in **2** and **3** as olivose and 3-*O*Me-olivose, respectively. The connection between C-7 of the steffimycinone aglycone and C-1' of these respective sugar moieties through an *O*-glycosidic linkage was verified by an HMBC correlation from H-7 (δ_H 5.04 and 5.05) to C-1' (δ_C 100.9 and 100.7), thus completing the structure determinations of **2** and **3**.

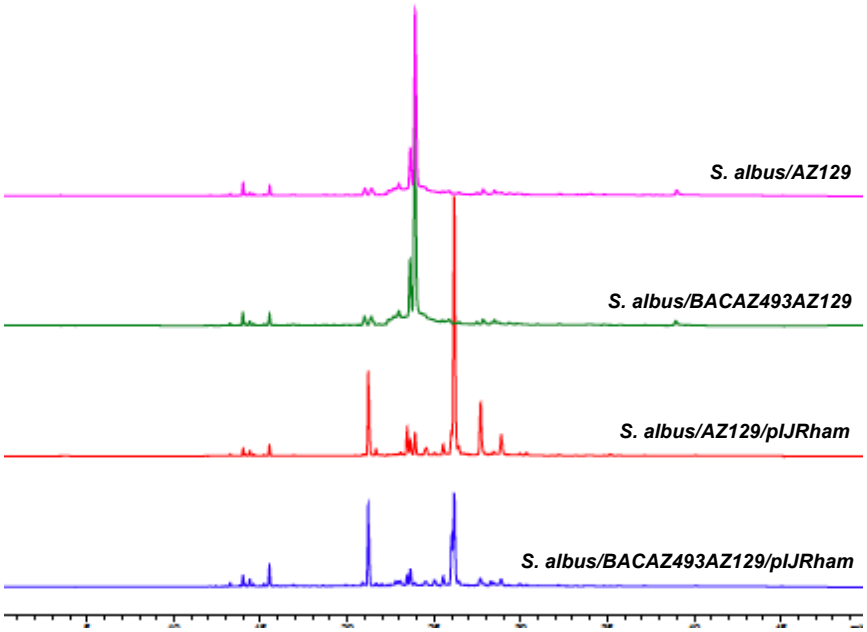
Supplementary Discussion 3. The proposed biosynthesis of arimetamycins A – C (**1** – **3**)



Supplmentary figure 3. (**a**) The proposed biosynthetic pathway for the dissacharide anthracycline arimetamycin A (**1**) and the monosaccharide anthracyclines arimetamycins B (**2**) and C (**3**). The exact order of the terminal steps (post compound **4**) is not clear. (**b**) Proposed biosynthetic pathway for the biosynthesis of the three deoxy sugars found in arimetamycins A, B and C. Arm32/33 and arm11/12 are predicted to be P450-oxidoreductase/glycosyltransferase two component catalytic systems.[16]

The glycocysotranferase arm12, which shows high sequence identity (82%) to the stfG glycocysotranferase from steffimycin biosynthesis, is predicted to transfer the NDP-L-olivose to the common biosynthetic intermediate **4** to generate arimetamycin B (**2**) and eventually arimetamycin C (**3**) after methylation by arm14.[7] The two remaining glycosyltransferases found in the *arm* cluster, arm25 and arm33, do not show significant sequence similarity to any known glycosyltransferases and are predicted to incorporate the two aminodeoxy sugars (NDP-brasiliose and NDP-lemonose) seen in arimetamycin A (**1**). All genes required for the synthesis of these two aminodeoxy sugars, as well as L-olivose, are present in the *arm* cluster with the exception of the four sugar biosynthesis genes, NDP-D-glucose synthase (oleS), 4,6-dehydratase (oleE), 3,5-epimerase (oleL), and 4-ketoreductase (oleU), which were added in trans.

Supplementary Discussion 4. TAR assembly of overlapping clones



Supplmentary figure 4. Transformation-associated recombination (TAR) assembly of the overlapping clones AZ493 and AZ129 into a 59 kb insert containing BAC clone. HPLC comparison of ethyl acetate extracts from cultures of *S. albus* transformed with either AZ129 or TAR-assembled BAC-AZ493AZ129 with and without pIJRham co-transformation, indicate that AZ129 and BAC-AZ493AZ129 confer the production of the same set of metabolites.

To confirm that clone AZ129 contained the full complement of biosynthetic genes comprising this eDNA-derived anthracycline biosynthetic gene cluster, we extended the AZ129 eDNA insert by using an overlapping cosmid clone (AZ493) and transformation-associated recombination[8, 17] to produce a bacterial artificial chromosome consisting of eDNA captured on both cosmids (BAC-AZ493AZ129). Following conjugation into *S. albus*/pIJRham, *S. albus*/BAC-AZ493AZ129/pIJRham was found to produce the same set of metabolites as *S. albus*/AZ129/pIJRham, indicating that clone AZ129 likely contains a complete anthracycline biosynthetic gene cluster.

Supplementary table 1: NMR spectroscopic data of arimetamycin A (**1**) in DMSO-d_6

No.	Arimetamycin A (**1**)				
	δ_C[b]	δ_H[a]	*J* in Hz	HMBC	NOE
1	108.2	7.28	d (2.57)	2, 3, 4a, 12	2-OCH_3, 4-OH
2	166.7				
3	106.8	6.96	d (2.57)	1, 2, 4, 4a	2-OCH_3, 4-OH
4	164.8				
4a	110.4				
5	189.9				
5a	118.8				
6	161.3				
6a	132.8				
7	71.8	5.04	d (2.1)	1', 6, 6a, 8, 9, 10a	8, 8OCH_3, 1'
8	86.3	3.55	d (2.1)	6a, 7, 9, 10, 9CH_3, 8OCH_3	7, 9CH_3, 5', 6'
9	75.9				
10	198.6				
10a	135.2				
11	115.3	8.06	s	5a, 6a, 10, 12	
11a	135.1				
12	180.5				
12a	133.3				
9CH_3	23.5	1.39	s	8, 9, 10	
2OCH_3	56.6	3.97	s	2	
8OCH_3	59.7	3.45	s	8	7, 8, 6’
4-OH		12.00	s	3, 4, 4a	1, 3
6-OH		12.80	s	5a, 6, 6a	7, 1’
1'	99.6	5.69	dd (3.5, 1.2)	3'	7, 2'
2'	32.4	2.00	brm	3', 3'CH_3	1’
3'	63.6				
4'	75.9	4.05	m		3'CH_3, 3'NCH_3, 6', 1"
5'	64.8	4.13	qd (7.1, 0.9)	1', 4', 6'	8, 9CH_3, 6', 3'CH_3
6'	17.8	1.35	d (7.3)	5', 4'	4’, 5’, 8, 1’’, 2’’
3'NCH_3	37.8	2.70	d (4.4)	3'	2’, 4', 3'CH_3, 5", 6"
3'NCH_3	36.2	2.77	d (4.7)		
3'NH		8.15	brs		2’, 3’$N(CH_3)_2$, 5’’
3'CH_3	14.5	1.36	s	2', 3', 4'	2’, 3’NH, 3’$N(CH_3)_2$
1"	98.2	5.26	t (5.3)	4', 3", 5"	4', 5', 2", 4"$N(CH_3)_2$, 6’’
2" (ax)	40.1	1.96	dd (14.5, 5.3)	3", 4", 3"CH_3	6', 1", 5", 3"CH_3
2" (eq)		2.16	dd (14.5, 5.3)		6', 1", 3’’CH_3, 4"NCH_3
3"	68.3				
4"	68.1	3.37	brd (2.2)	3", $N(CH_3)_2$	5", 6", 3"CH_3, 4"$N(CH_3)_2$
5"	64.7	4.67	qd (7.1, 2.2)	6"	2', 3'NCH_3, 4", 6", 3"CH_3
6"	17.1	1.45	d (7.1)	5", 4"	1", 4", 5", 4"$N(CH_3)_2$
3"CH_3	30.5	1.49	s	2", 4"	2", 4", 5", 4"$N(CH_3)_2$
4"NCH_3	42.9	2.99	brs	4"	1", 2"b, 4", 5", 6", 3"CH_3
4"NCH_3	45.1	2.99	brs		
4"NH		8.99	brs		

[a]recorded at 600 MHz, [b]recorded at 150 MHz, [a,b]signals were referenced to the DMSO-d_6 solvent signals (δ_H 2.50 and δ_H 39.51)

Supplementary table 2: NMR spectroscopic data of arimetamycins B (**2**) and C (**3**) in DMSO-d_6

No.	Arimetamycin B (**2**)			Arimetamycin C (**3**)				
	δ_C[b]	δ_H[a]	*J* in Hz	δ_C[b]	δ_H[a]	*J* in Hz	HMBC	NOE
1	108.1	7.27	s	108.1	7.25	s	2, 3, 4a, 12	2OCH$_3$
2	166.6			166.6				
3	106.8	6.95	s	106.8	6.93	s	1, 2, 4, 4a	2OCH$_3$
4	164.7			164.7				
4a	110.4			110.4				
5	189.9			189.9				
5a	118.8			118.8				
6	161.3			161.3				
6a	133.1			133.1				
7	70.8	5.04	brs	70.9	5.05	brs	6, 6a, 8, 9, 10a, 1’	8, 8OCH$_3$, 1’
8	85.9	3.61	brs	85.9	3.61	brs	6a, 7, 9, 10, 9CH$_3$, 8OCH$_3$	7, 9CH$_3$, 5’
9	76.3			76.3				
10	198.9			198.8				
10a	135.3			135.3				
11	115.2	8.06	s	115.3	8.05	s	5, 5a, 6, 10, 12, 12a	
11a	135.0			135.0				
12	180.6			180.5				
12a	133.7			133.4				
9CH$_3$	23.3	1.40	s	23.3	1.42	s	8, 9, 10	8, 5’
2OCH$_3$	56.6	3.97	s	56.6	3.96	s	2	
8OCH$_3$	59.7	3.43	s	59.7	3.43	s	8	7
4-OH		12.00	s		11.98	s	3, 4, 4a	
6-OH		12.76	s		12.76	s	5a, 6, 6a	
9-OH		5.65	s		5.66	s	8, 9, 10, 9CH$_3$	
1'	100.9	5.48	brd (3.8)	100.7	5.53	brd (3.2)	3’, 5’	7, 2’ax, 2’eq
2'ax	38.8	1.58	td (13.8, 3.8)	34.5	1.50	m	1’, 3’, 4’,	1, 2’eq, 4’
2'eq		1.94	dd (13.8, 2.9)		2.17	dd (12.7, 3.2)		1, 2’ax, 3’
3'	67.5	3.47	m	77.6	3.19	m	4’, 3’OCH$_3$	2’eq, 5’
4'	76.9	2.87	m	75.3	2.99	td (9.1, 4.2)		2’ax, 6’, 4’OH
5'	69.6	3.71	m	69.5	3.74	dq (9.3, 6.1)	4’, 6’	8, 9CH$_3$, 3’, 4’, 6’
6'	17.9	1.26	d (5.8)	17.8	1.27	d (6.1)	4’, 5’	4’, 5’
3’OH or 3'OCH$_3$		4.77	d (3.9)	56.6	3.25	s	3’	2’ eq
4'OH		5.03	d (5.9)		5.17	d (4.2)		

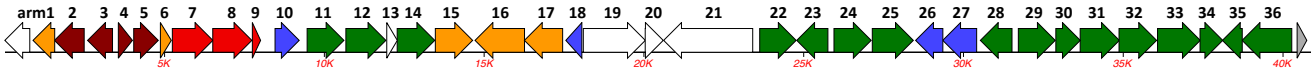
[a]recorded at 600 MHz, [b]recorded at 150 MHz, [a,b]signals were referenced to the DMSO-d_6 solvent signals (δ_H 2.50 and δ_H 39.51)

Supplementary table 3: NMR spectroscopic data of 8-demethoxy-10-deoxysteffimycinone (**4**) in DMSO-d_6

No.	8-demethoxy-10-deoxysteffimycinone (**4**)			
	δ_C [b]	δ_H [a]	*J* in Hz	HMBC
1	107.1	7.20	d, *J* = 2.4	2, 3, 4a
2	165.8			
3	107.0	6.85	d, *J* = 2.4	1, 2, 4, 4a
4	165.0			
4a	110.9			
5	NA[c]			
5a	113.6			
6	NA[c]			
6a	135.7			
7	63.5	5.05	t, *J* = 6.6	6a, 9, 10a
8	44.3	1.74 2.08	dd, *J* = 13.6, 7.1 ddd, J = 13.6, 6.7, 1.6	6a, 7, 9, 10
9	68.0			
10	44.6	2.72 2.89	d, *J* = 16.8	6a, 10a, 9, 9CH_3, 11
10a	145.4			
11	119.4	7.31	s	5a, 6a, 10, 12
11a	NA[c]			
12	181.9			
12a	NA[c]			
9CH_3	30.1	2.40	s	8, 9, 10
2OCH_3	56.6	3.92	s	2
4-OH		NA[c]		
6-OH		NA[c]		
9-OH		4.50	s	8, 9, 10

[a]recorded at 600 MHz, [b]assinged from HMQC and HMBC, [a,b]signals were referenced to the DMSO-d_6 solvent signals (δ_H 2.50 and δ_H 39.51), [c]not assigned.

Supplementary table 4: Gene annotation table of the eDNA clone AZ129



Gene	Size (bp)	Proposed function	Homologous gene	Origin	Identity/ Similarity (%)	Accession NO.
ORF1	780		putative hydroxylase	*Streptomyces nogalater*	50/65	AF187532.1
arm1	678	O-methyl transferase	stfMI	*Streptomyces steffisburgensis*	72/84	CAJ42328.1
arm2	930	Aromatase	stfQ	*Streptomyces steffisburgensis*	77/84	CAJ42327.1
arm3	780	Cyclase	stfY	*Streptomyces steffisburgensis*	80/89	CAJ42324.1
arm4	441	Cyclase	stfX	*Streptomyces steffisburgensis*	79/86	CAJ42323.1
arm5	768	Ketoreductase	stfT	*Streptomyces steffisburgensis*	71/82	CAJ42322.1
arm6	342	Monooygenase	stfOI	*Streptomyces steffisburgensis*	77/86	CAJ42321.1
arm7	1269	KSα	stfP	*Streptomyces steffisburgensis*	85/92	CAJ42320.1
arm8	1224	KSβ (CLF)	stfK	*Streptomyces steffisburgensis*	83/89	CAJ42319.1
arm9	258	ACP	stfS	*Streptomyces steffisburgensis*	67/79	CAJ42318.1
arm10	756	SARP family transcriptional regulator	AckI	*Streptomyces galilaeus*	64/75	BAB72052.1
arm11	1152	P450-oxidoreductase	stfPII	*Streptomyces steffisburgensis*	57/66	CAJ42339.1
arm12	1251	Glycosyltransferase	stfG	*Streptomyces steffisburgensis*	82/87	CAJ42338.1
arm13	330	unknown	stfF	*Streptomyces steffisburgensis*	81/83	CAJ42337.1
arm14	1176	O-methyltransferase	stfMII	*Streptomyces steffisburgensis*	74/86	CAJ42336.1
arm15	1167	Oxygenase	stfOII	*Streptomyces steffisburgensis*	76/84	CAJ42335.1
arm16	1548	Dehydrogenase	stfE	*Streptomyces steffisburgensis*	77/88	CAJ42334.1
arm17	1197	P450-oxidoreductase	stfPI	*Streptomyces steffisburgensis*	75/84	CAJ42333.1
arm18	501	MarR family transcriptional regulator	stfRIII	*Streptomyces steffisburgensis*	75/80	CAJ42332.1
arm19	1941		stfD	*Streptomyces steffisburgensis*	78/84	CAJ42330.1
arm20	564		stfC	*Streptomyces steffisburgensis*	68/79	CAJ42329.1
arm21	2787		hypothetical protein SCAB_77201	Streptomyces scabiei 87.22	38/49	YP_003493221.1
arm22	1140	NDP-hexose-C3-methyltransferase	DacS3	*Dactylosoporangium* sp. SC14051	70/79	AFU65915.1
arm23	975	NDP-4-keto-6-deoxyhexose reductase	hedI	*Streptomyces griseoruber*	59/66	AAP85355.1
arm24	1164	NDP-hexose-3-aminotransferase	Med-20	*Streptomyces* sp. AM-7161	76/83	BAC79028.1
arm25	1284	Glycosyltransferase	AknK	*Streptomyces galilaeus*	51/67	AAF70102.1
arm26	852	ABC transporter	AcrW	*Streptomyces galilaeus*	53/71	BAB72060.1
arm27	1062	ANC transporter ATP-binding protein	AcrV	*Streptomyces galilaeus*	69/81	BAB72059.1
arm28	990	NDP-hexose-3-ketoreductase	AknQ	*Streptomyces galilaeus*	61/77	AAF73453.1
arm29	1155	NDP-hexose-4-aminotransferase	NocS6	*Nocardia* sp. ATCC202099	69/80	ADR01074.1
arm30	765	NDP-hexose-N,N-dimethyltransferase	SnogA	*Streptomyces nogalater*	58/73	AAF01819.1
arm31	1227	NDP-hexose-C3-methyltransferase	ApoM2	*Nocardiopsis* sp. FU40	72/83	AEP40909.1
arm32	1176	P450-oxidoreductase	dnrQ	*Streptomyces peucetius*	40/55	AAD15266.1
arm33	1338	Glycosyltransferase	RdmH	*Streptomyces pururascens*	52/68	AAL24450.1
arm34	711	NDP-hexose-N,N-dimethyltransferase	hedH	*Streptomyces griseoruber*	56/69	AAP85356.1
arm35	621	NDP-hexose-3,5-epimerase	staE	*Streptomyces* sp. TP-A0274	60/70	BAC55217.1
arm36	1554	NDP-hexose-2,3-dehydratase	SnogH	*Streptomyces nogalater*	63/74	CAA12009.1
ORF38			transposase	*Rhodococcus jostii* RHA1	58/67	YP_703194.1

References for supplementary material:

[1] S. F. Brady, *Nat. Protoc.* **2007**, *2*, 1297-1305.

[2] K. T. Seow, G. Meurer, M. Gerlitz, E. Wendt-Pienkowski, C. R. Hutchinson, J. Davies, *J. Bacteriol.* **1997**, *179*, 7360-7368.

[3] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, *Mol. Biol. Evol.* **2011**, *28*, 2731-2739.

[4] W. Zhu, A. Lomsadze, M. Borodovsky, *Nucleic Acids Res.* **2010**, *38*, e132.

[5] M. Bierman, R. Logan, K. O'Brien, E. T. Seno, R. N. Rao, B. E. Schoner, *Gene* **1992**, *116*, 43-49.

[6] L. Rodriguez, C. Oelkers, I. Aguirrezabalaga, A. F. Brana, J. Rohr, C. Mendez, J. A. Salas, *J. Mol. Microbiol. Biotechnol.* **2000**, *2*, 271-276.

[7] S. Gullon, C. Olano, M. S. Abdelfattah, A. F. Brana, J. Rohr, C. Mendez, J. A. Salas, *Appl. Environ. Microbiol.* **2006**, *72*, 4172-4183.

[8] D. Kallifidas, S. F. Brady, *Methods Enzymol.* **2012**, *517*, 225-239.

[9] M. G. Brattain, W. D. Fine, F. M. Khaled, J. Thompson, D. E. Brattain, *Cancer Res.* **1981**, *41*, 1751-1756.

[10] P. Noguchi, R. Wallace, J. Johnson, E. M. Earley, S. O'Brien, S. Ferrone, M. A. Pellegrino, J. Milstien, C. Needy, W. Browne, J. Petricciani, *In Vitro* **1979**, *15*, 401-408.

[11] B. R. Brinkley, P. T. Beall, L. J. Wible, M. L. Mace, D. S. Turner, R. M. Cailleau, *Cancer Res.* **1980**, *40*, 3118-3129.

[12] S. E. Mirski, J. H. Gerlach, S. P. Cole, *Cancer Res.* **1987**, *47*, 2594-2598.

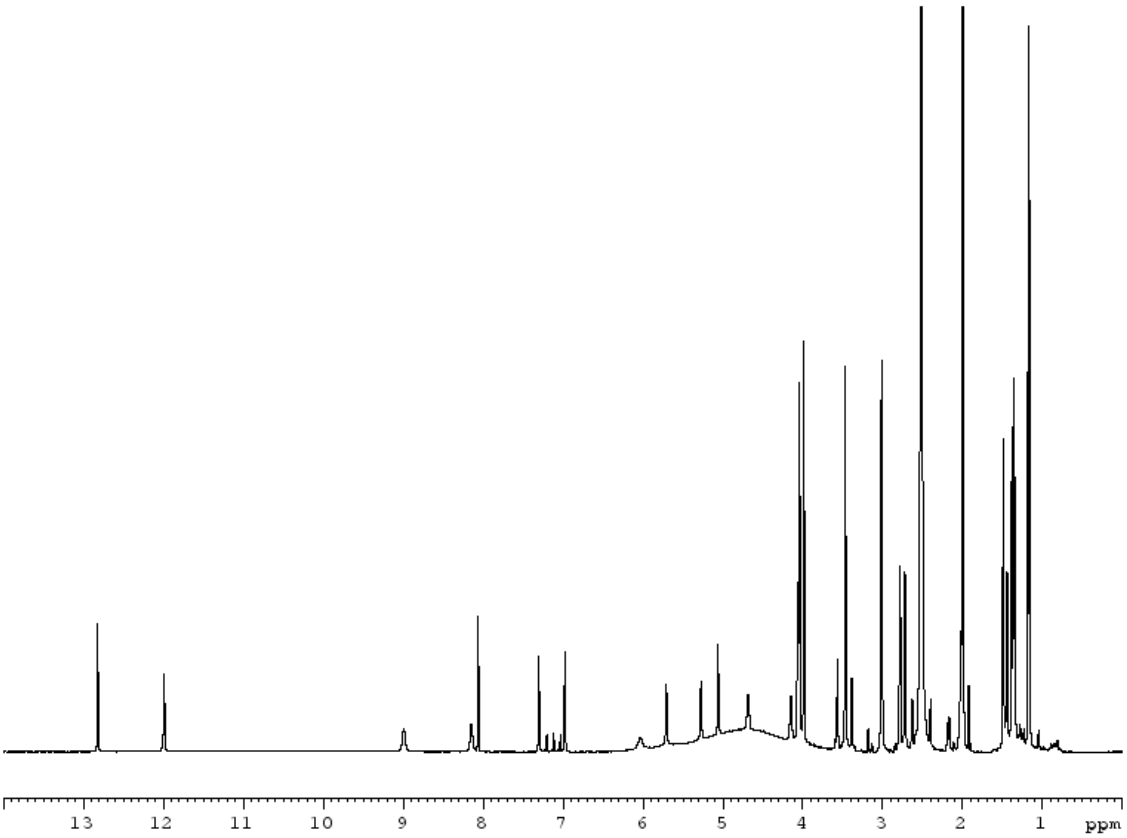
[13] a) R. Supino, *Methods Mol. Biol.* **1995**, *43*, 137-149; b) D. Zivadinovic, B. Gametchu, C. S. Watson, *Breast Cancer Res.* **2005**, *7*, R101-112.

[14] H. Laatsch, S. Fotso, *Top. Curr. Chem.* **2008**, *282*, 3-74.

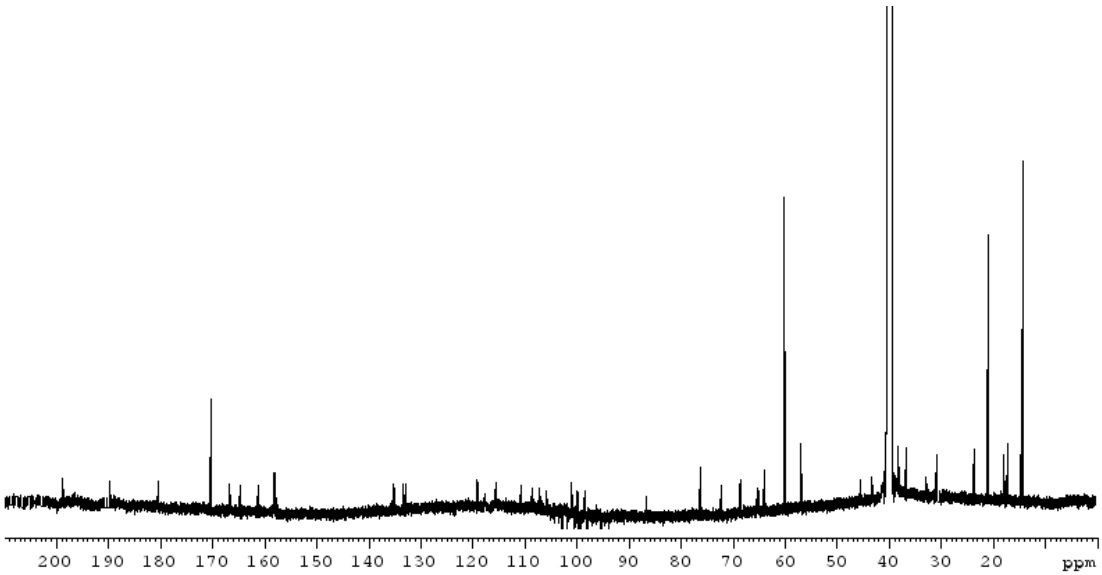
[15] R. C. Kelly, I. Schletter, J. M. Koert, F. A. MacKellar, P. F. Wiley, *J. Org. Chem.* **1977**, *42*, 3591-3596.

[16] W. Lu, C. Leimkuhler, G. J. Gatto, Jr., R. G. Kruger, M. Oberthur, D. Kahne, C. T. Walsh, *Chem. Biol.* **2005**, *12*, 527-534.

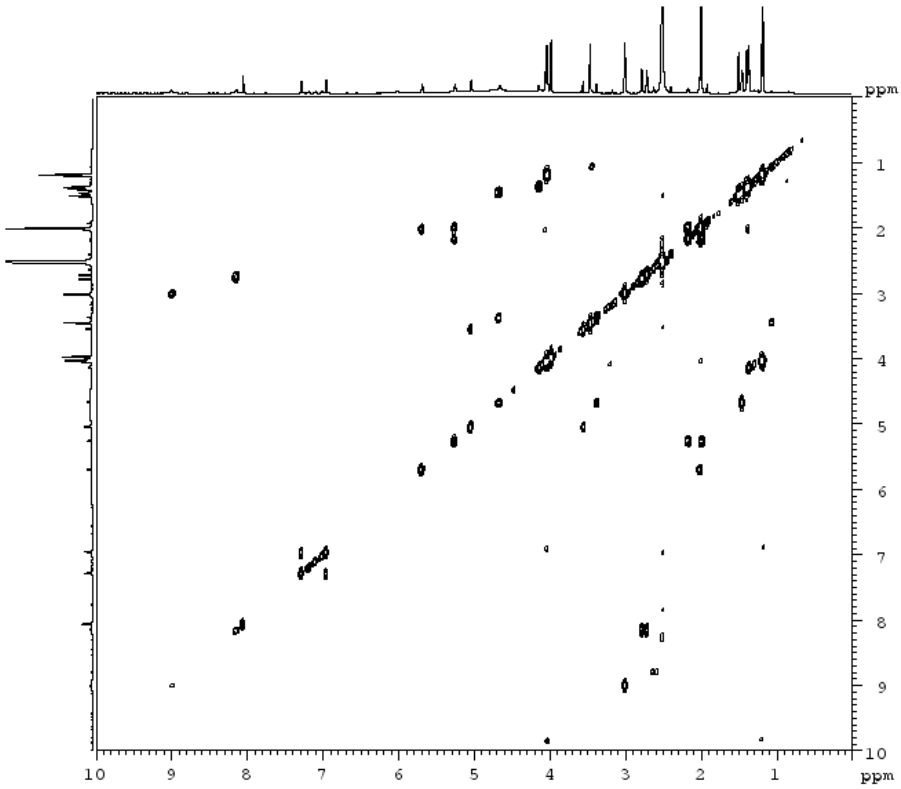
[17] J. H. Kim, Z. Feng, J. D. Bauer, D. Kallifidas, P. Y. Calle, S. F. Brady, *Biopolymers* **2010**, *93*, 833-844.



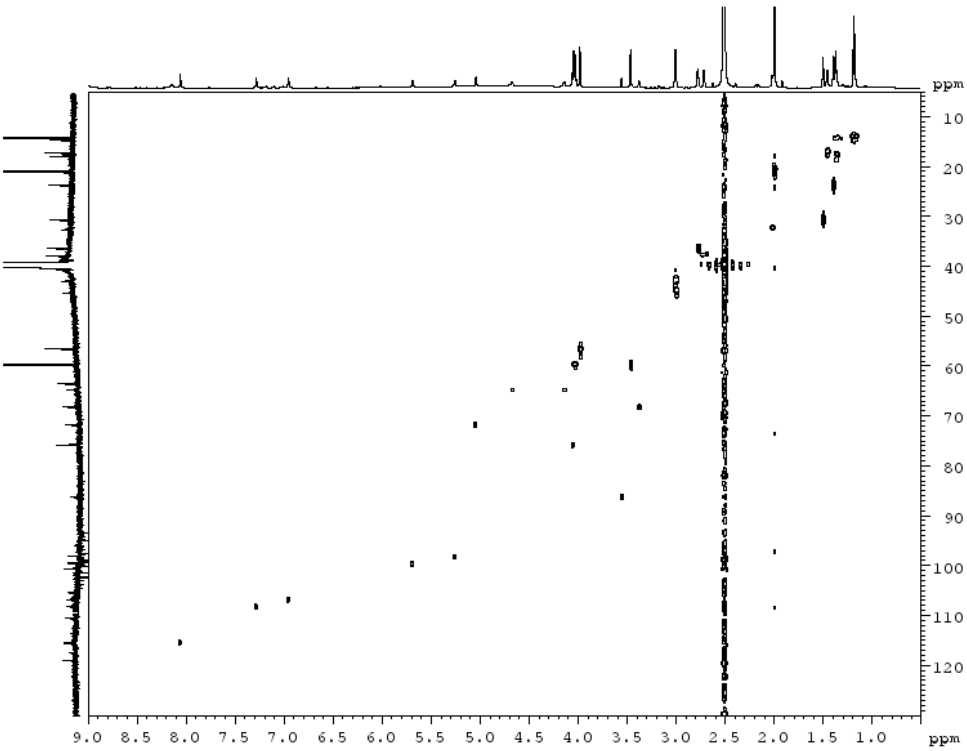
Supplementary figure 5. ^{1}H NMR spectrum (DMSO-d_6, 600 MHz) of arimetamycin A (**1**).



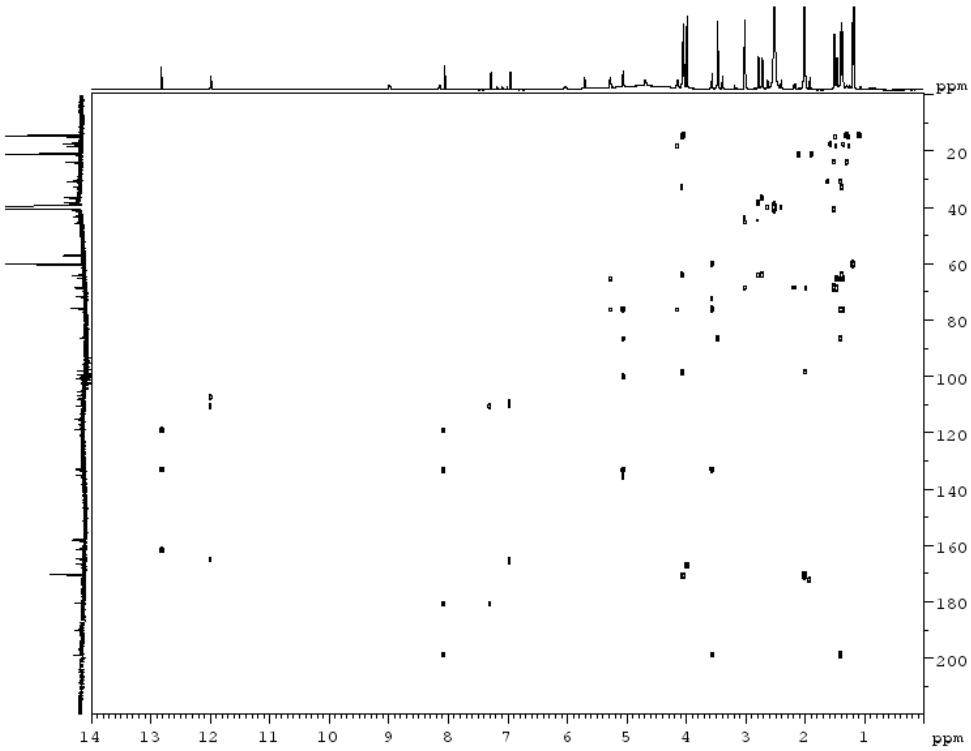
Supplementary figure 6. ^{13}C NMR spectrum (DMSO-d_6, 150 MHz) of arimetamycin A (**1**).



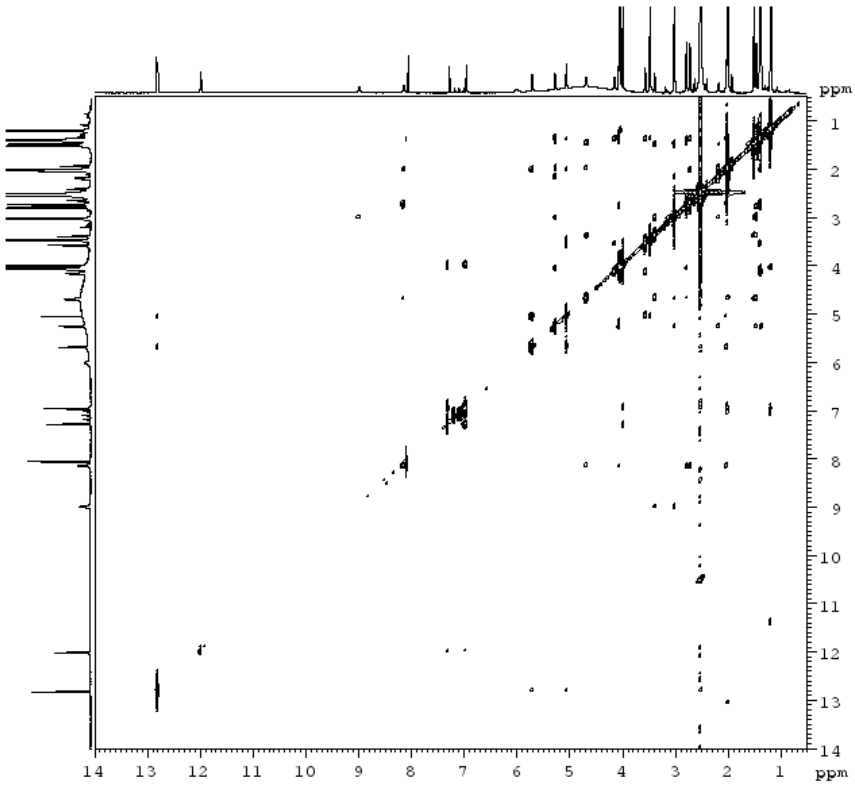
Supplementary figure 7. COSY spectrum (DMSO-d_6, 600 MHz) of arimetamycin A (**1**).



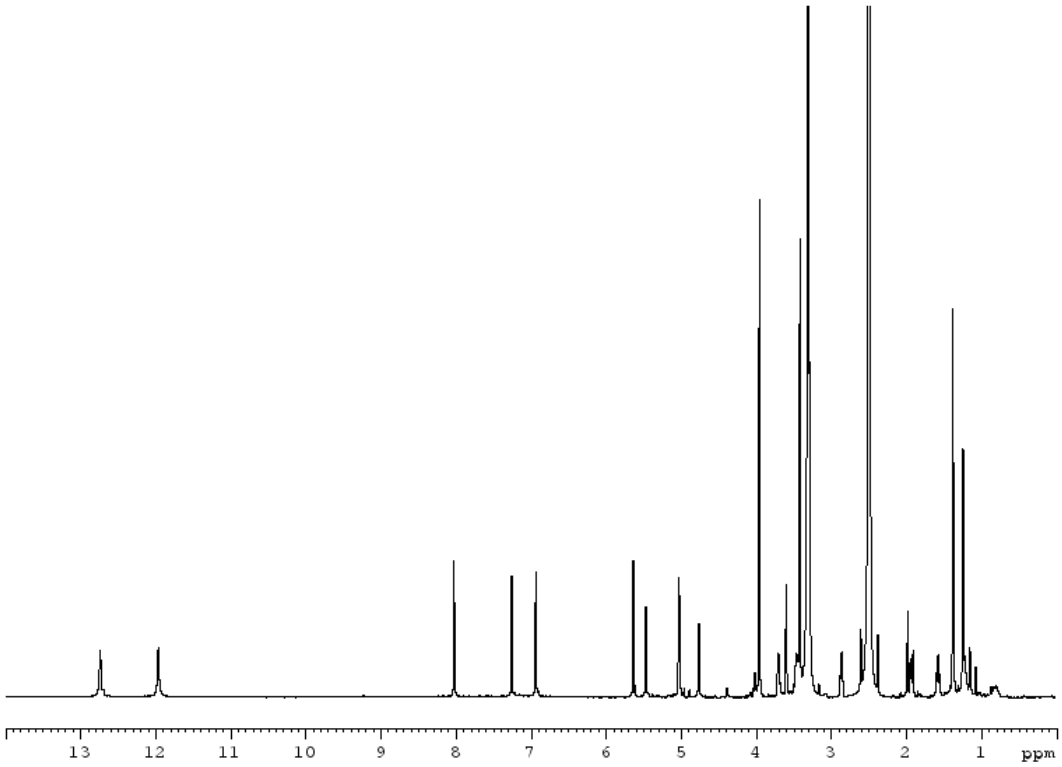
Supplementary figure 8. HMQC spectrum (DMSO-d_6, 600 MHz) of arimetamycin A (**1**).



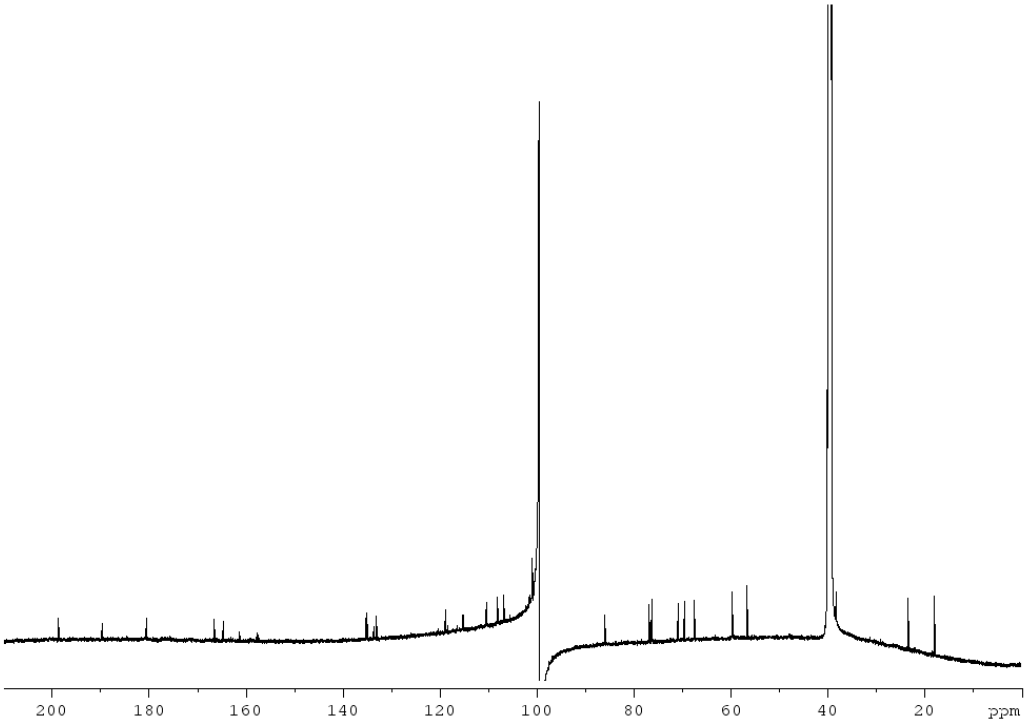
Supplementary figure 9. HMBC spectrum (DMSO-d_6, 600 MHz) of arimetamycin A (**1**).



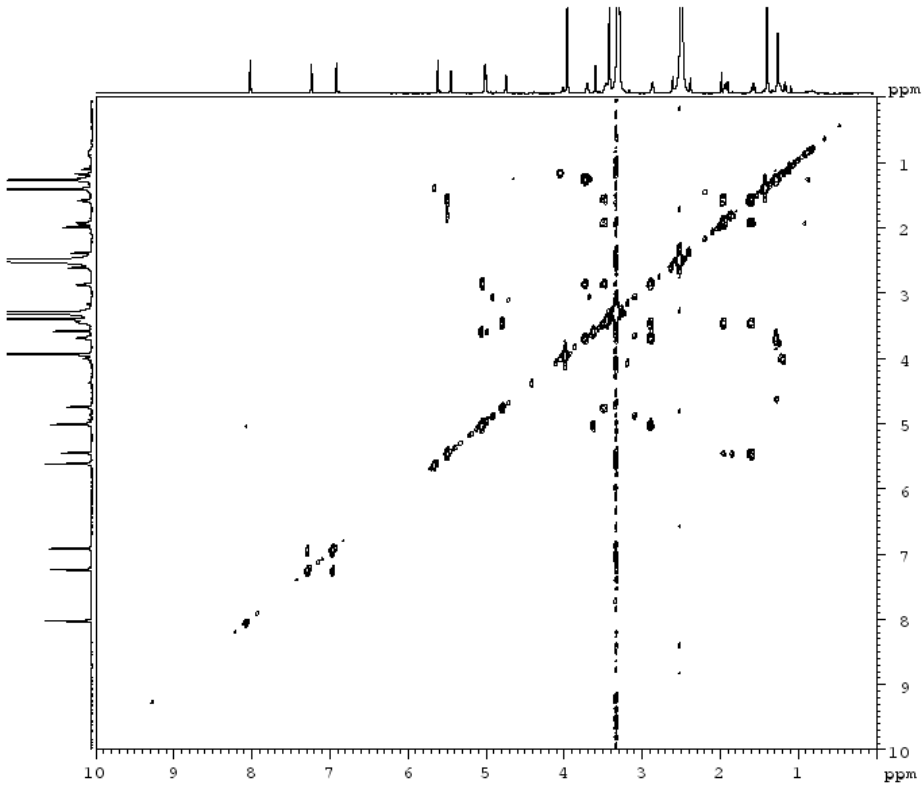
Supplementary figure 10. ROESY spectrum (DMSO-d_6, 600 MHz, mixing time: 250 ms) of arimetamycin A (**1**).



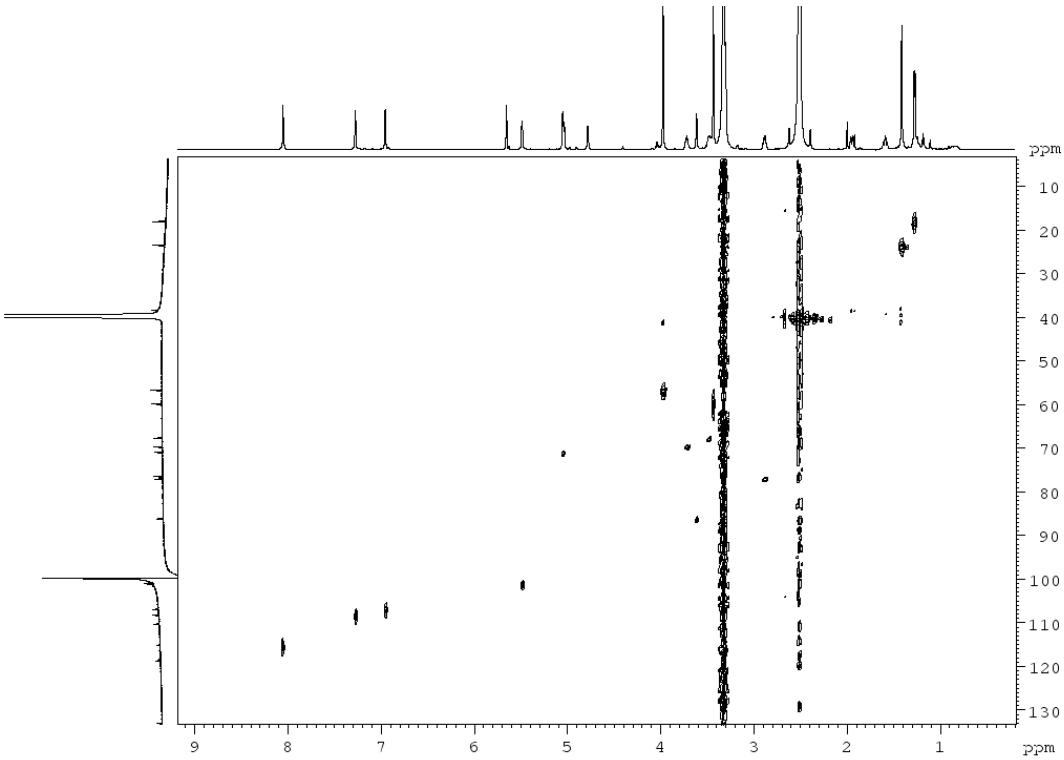
Supplementary figure 11. ^{1}H NMR spectrum (DMSO-d_6, 600 MHz) of arimetamycin B (**2**).



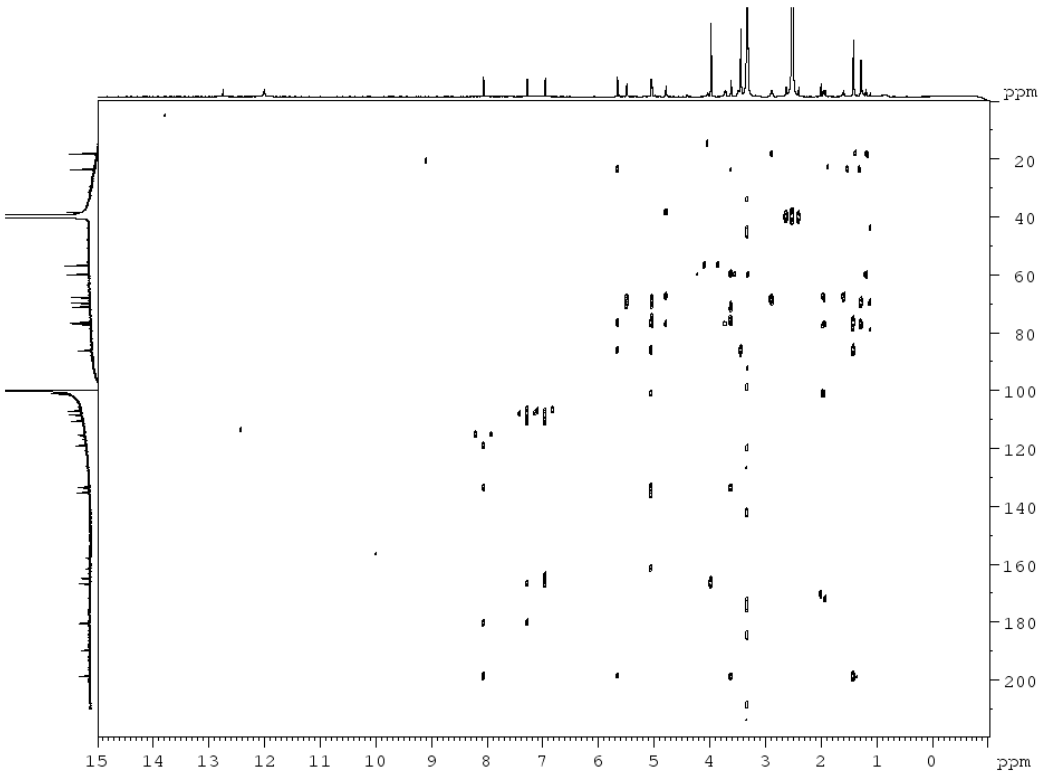
Supplementary figure 12. ^{13}C NMR spectrum (DMSO-d_6, 150 MHz) of arimetamycin B (**2**).



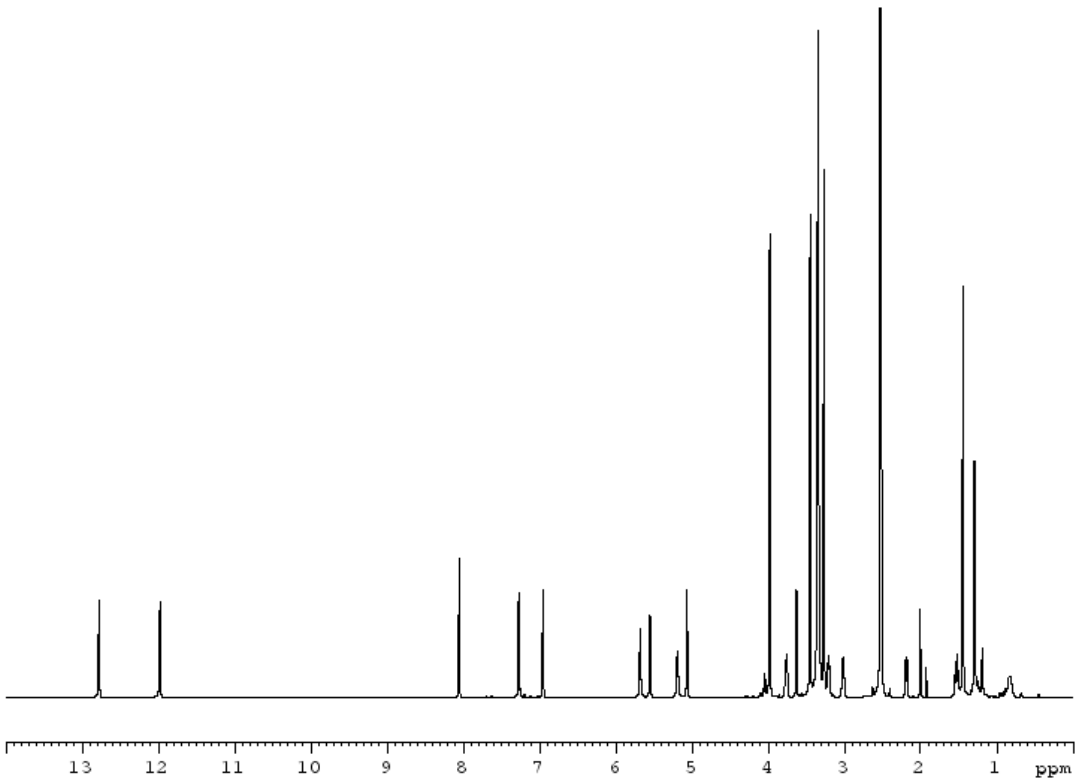
Supplementary figure 13. COSY spectrum (DMSO-d_6, 600 MHz) of arimetamycin B (**2**).



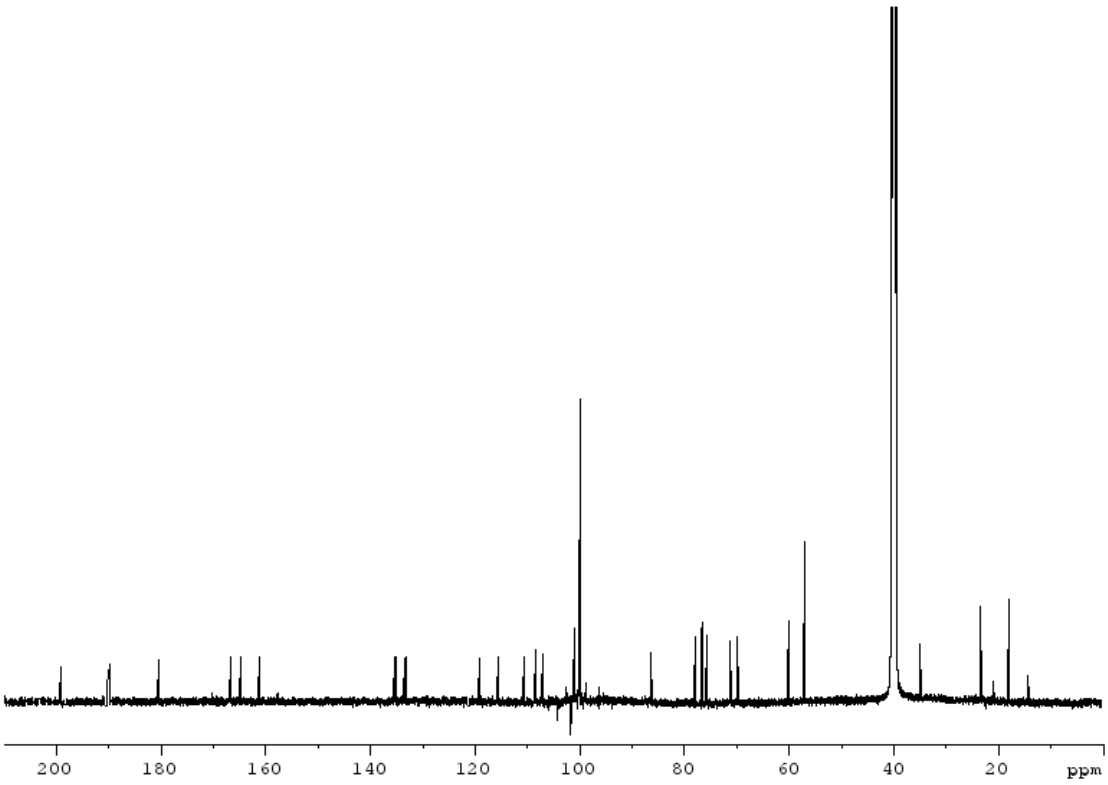
Supplementary figure 14. HMQC spectrum (DMSO-d_6, 600 MHz) of arimetamycin B (**2**).



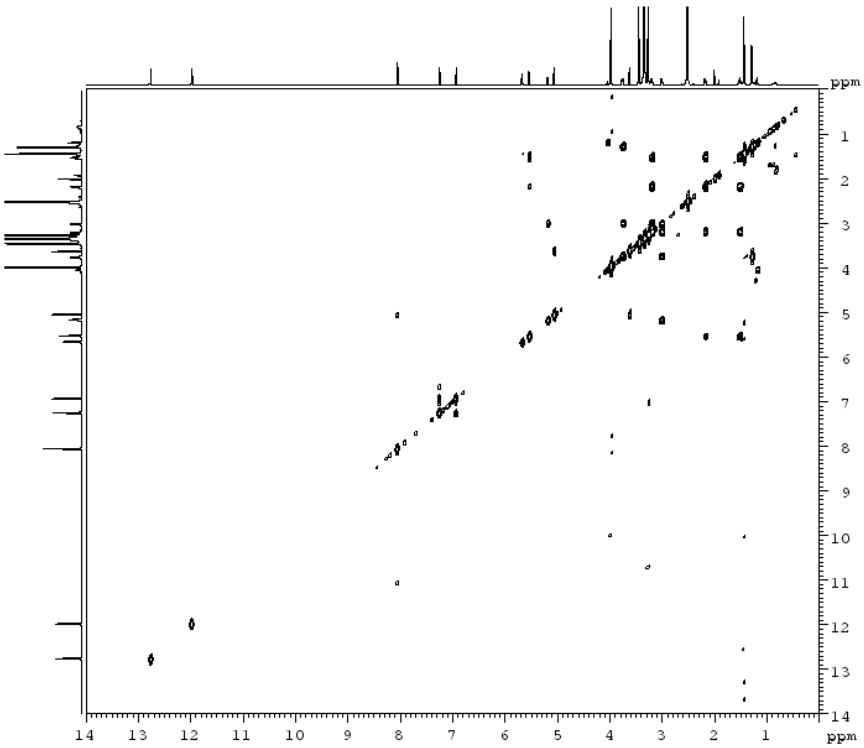
Supplementary figure 15. HMBC spectrum (DMSO-d_6, 600 MHz) of arimetamycin B (**2**).



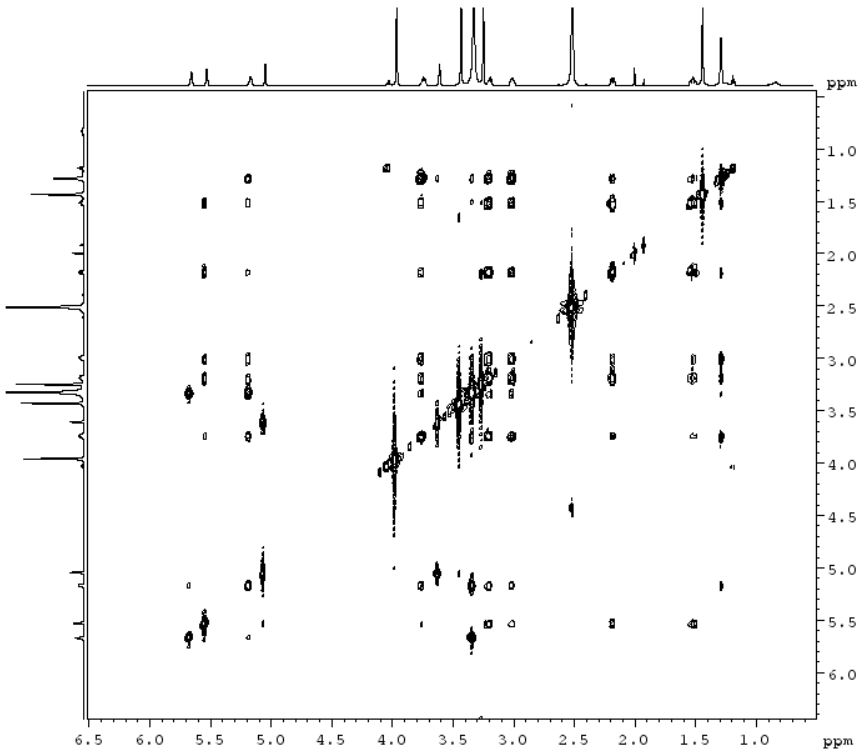
Supplementary figure 16. ^{1}H NMR spectrum (DMSO-d_6, 600 MHz) of arimetamycin C (**3**).



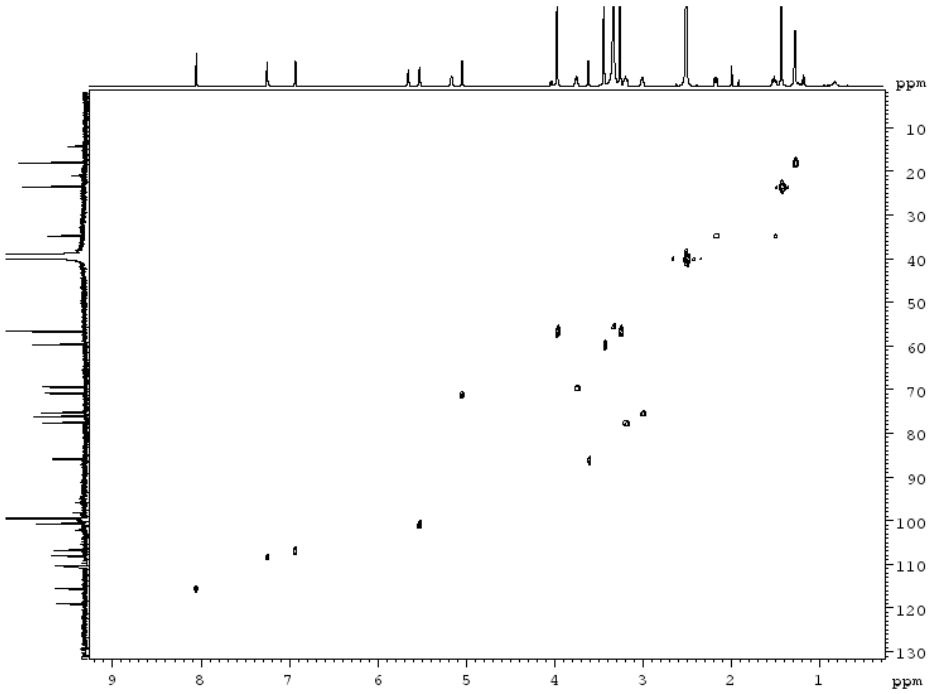
Supplementary figure 17. ^{13}C NMR spectrum (DMSO-d_6, 150 MHz) of arimetamycin C (**3**).



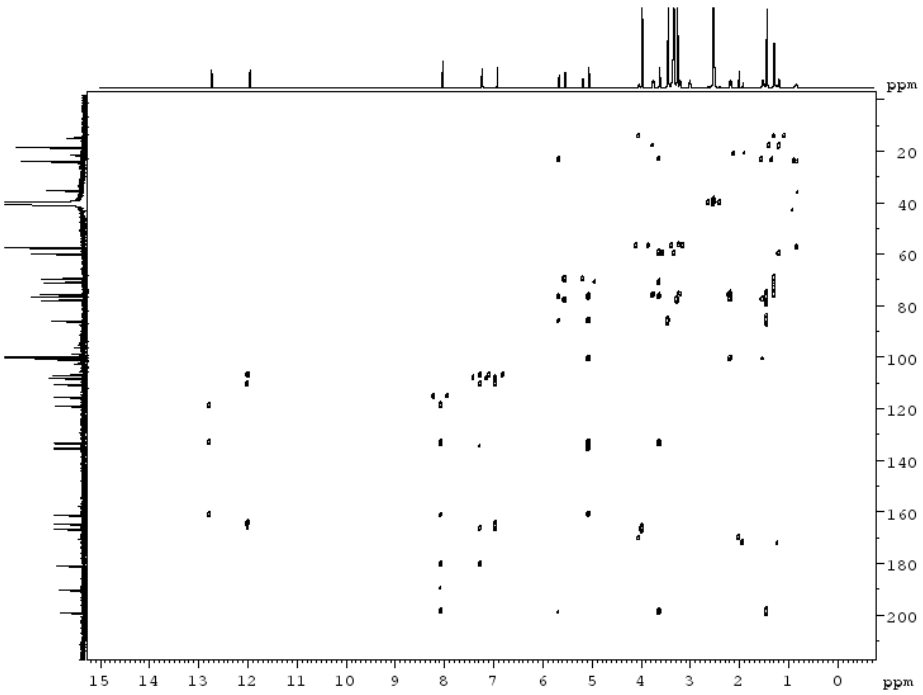
Supplementary figure 18. COSY spectrum (DMSO-d_6, 600 MHz) of arimetamycin C (**3**).



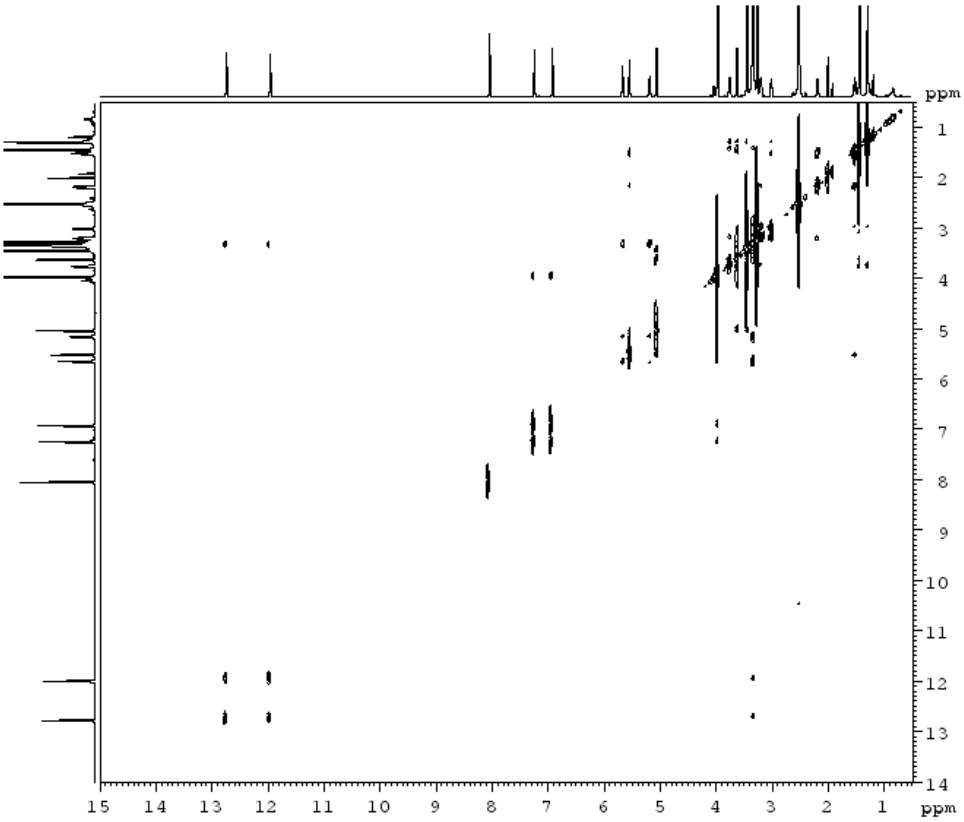
Supplementary figure 19. TOCSY spectrum (DMSO-d_6, 600 MHz) of arimetamycin C (**3**).



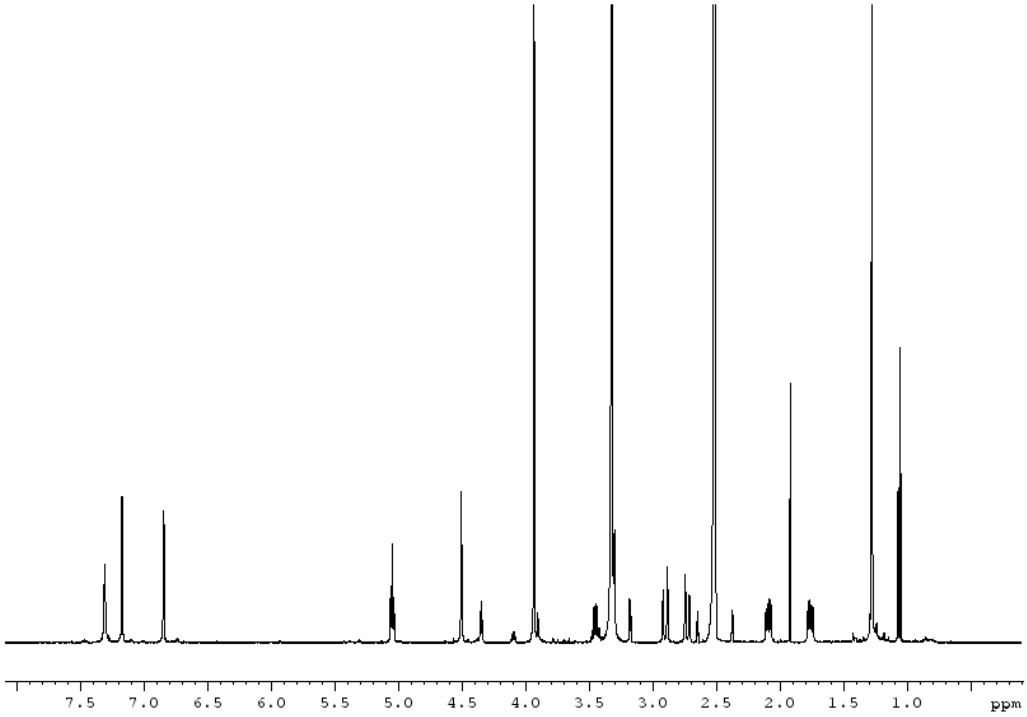
Supplementary figure 20. HMQC spectrum (DMSO-d_6, 600 MHz) of arimetamycin C (**3**).



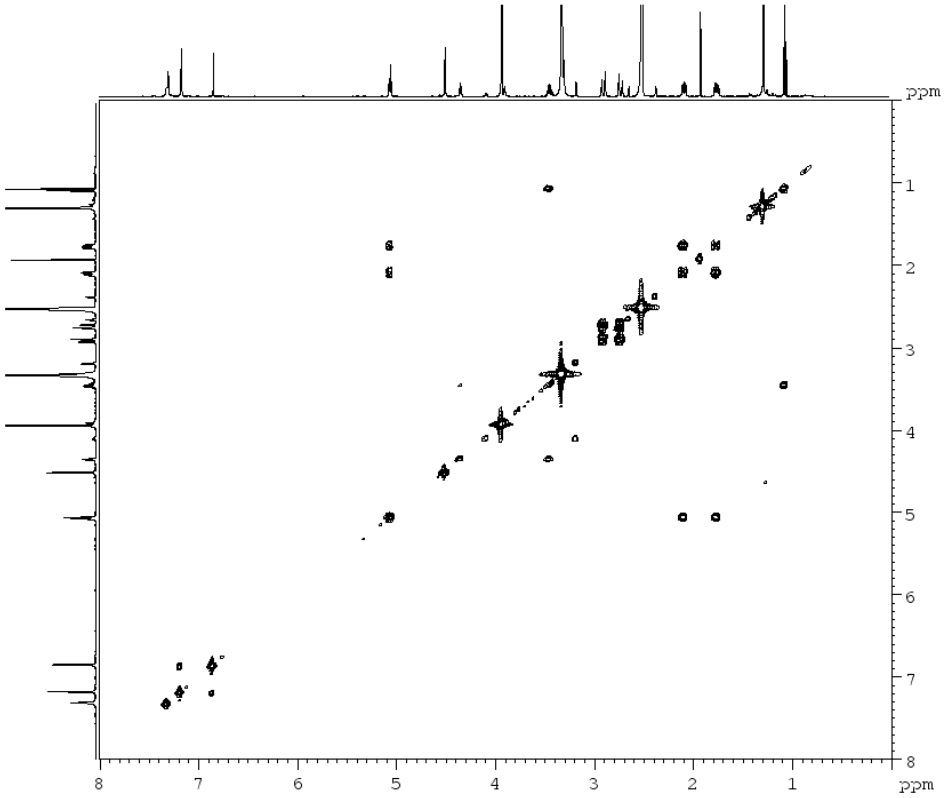
Supplementary figure 21. HMBC spectrum (DMSO-d_6, 600 MHz) of arimetamycin C (**3**).



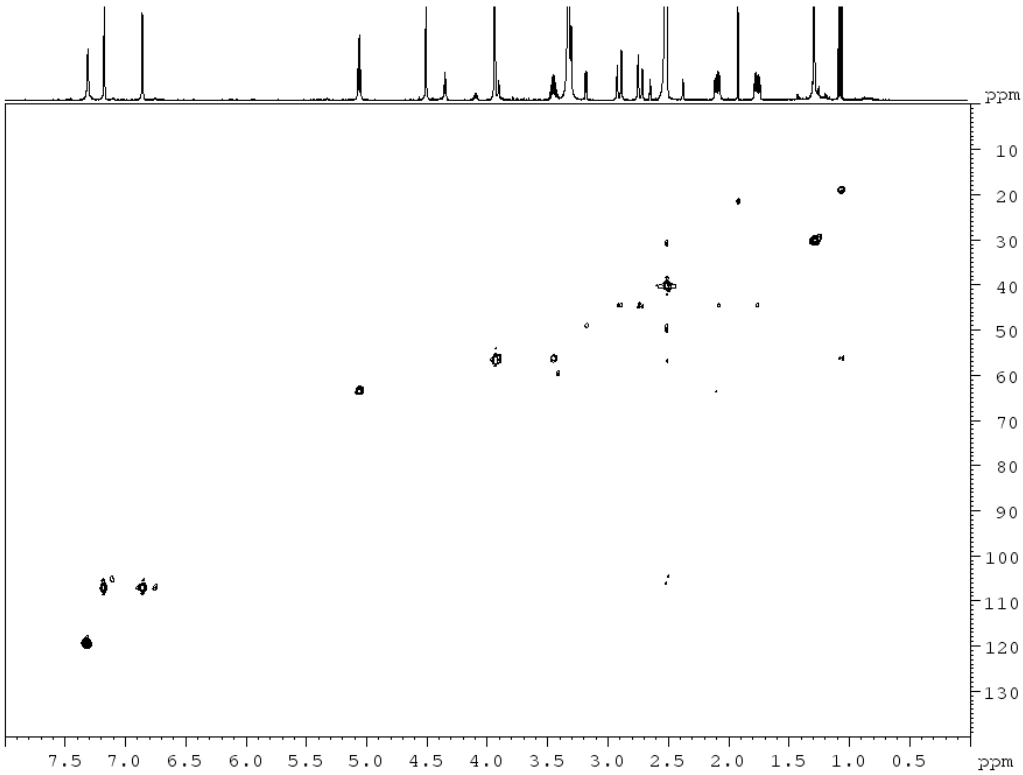
Supplementary figure 22. ROESY spectrum (DMSO-d_6, 600 MHz, mixing time: 250 ms) of arimetamycin C (**3**).



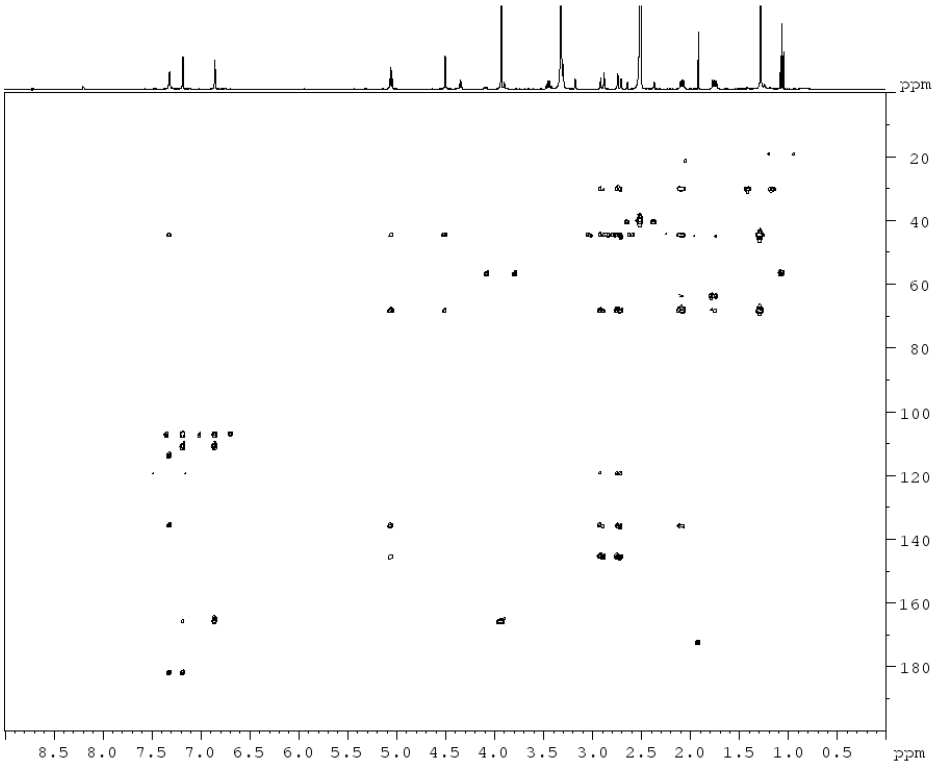
Supplementary figure 23. ^{1}H NMR spectrum (DMSO-d_6, 600 MHz) of compound **4**.



Supplementary figure 24. COSY spectrum (DMSO-d_6, 600 MHz) of compound **4**.



Supplementary figure 25. HMQC spectrum (DMSO-d_6, 600 MHz) of compound **4**.



Supplementary figure 26. HMBC spectrum (DMSO-d_6, 600 MHz) of compound **4**.