Discovery and Characterization of the TB Drug Lead Ecumicin

Wei Gao,[†,‡] Jin-Yong Kim,[§,∥] Shao-Nong Chen,[†] Sang-Hyun Cho,[‡] Jongkeun Choi,[⊥] Birgit U. Jaki,[†,‡] Ying-Yu Jin,[∥] David C. Lankin,[†] Ji-Ean Lee,[∥] Sun-Young Lee,[§,∥] James B. McAlpine,[†,‡] José G. Napolitano,[†] Scott G. Franzblau,[‡] Joo-Won Suh,[§,∥,*] Guido F. Pauli[†,‡,*]

[†]Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60612, U.S.A. [‡]Institute for Tuberculosis Research, College of Pharmacy, University of Illinois at Chicago, 833 S. Wood St. Chicago, Illinois 60612 U.S.A. [§]Division of Bioscience and Bioinformatics, College of Natural Science, Myongji University, Cheoin-gu, Yongin, Gyeonggi-Do 449-728, South Korea. [∥]Center for Nutraceutical and Pharmaceutical Materials, Myongji University, Cheoin-gu, Yongin, Gyeonggi-Do 449-728, South Korea.[⊥] Department of Cosmetic Science, Chungwoon University, Hongseong Chungnam 350-701, South Korea.

Supporting Information

Table of Contents

I. General Procedures

1D and 2D NMR spectra were obtained on a Bruker DPX 400 MHz with a 5 mm ATM CPTCI Z-gradient probe, Bruker Avance 600 MHz NMR spectrometer with a 5 mm CPTXI Z-gradient probe and a Bruker Avance II 900MHz NMR spectrometer with a 5 mm ATM CPTCI Z-gradient probe. ^{1}H and ^{13}C NMR chemical shifts were referenced to the methanol-d_4 solvent signals (δ_H 3.310 and δ_C 49.00, respectively). A mixing time of 60 ms was set for the TOCSY experiment and 200 ms for the ROESY experiment. The HMBC spectra were recorded with the $^{3}J_{C\text{-}H}$ set to 8 Hz, and the HSQC spectra were collected with the $^{1}J_{C\text{-}H}$ set to140 Hz. High-resolution ESI mass spectra were obtained using a Shimadzu IT-TOF LC mass spectrometer. Tandem mass analysis was performed using a Micromass Q-TOF LC mass spectrometer.

Marfey's analysis

The analysis was carried out as described.[1] 0.3 mg peptide was mixed with 0.5 ml 6N HCl solution. The mixture was heated at 110^{o}C for 16 hours, and the resultant solution was dried under vacuum. 100 µg hydrolyte, L-valine (Sigma-Aldrich), D-valine (Sigma-Aldrich), *N*-methyl-L-valine (Sigma-Aldrich) was mixed with 20 µL 1M $NaHCO_3$ and 10 µL 10% L-FDAA (Sigma-Aldrich) in acetone. The solution was heated to 80^{o}C for 3 min, and then cooled down to room temperature. 20 µL 1M HCl was then added to neutralize the solution. The reprivatized sample was then diluted to approximately 1 µg/mL with acetonitrile. 5 µL of diluted sample was injected into a RP-C18 column (Xterra® MS C18 3.5 µm 2.1x150 mm), and a bi-phase gradient method starting from 0.1% FA acetonitrile solution: 0.1% FA water solution = 20:80 (v/v) to 50:50 (v/v) in 30 min was applied. MS detector was set to scan mode with *m/z* from 200 to 600, and SIM mode with *m/z* of 370, and 384.

References

(1) Bhushan, R.; Bruckner, H. *J. Chromatogr. B,* **2011**, *879*, 3148.

II. Supplementary Results

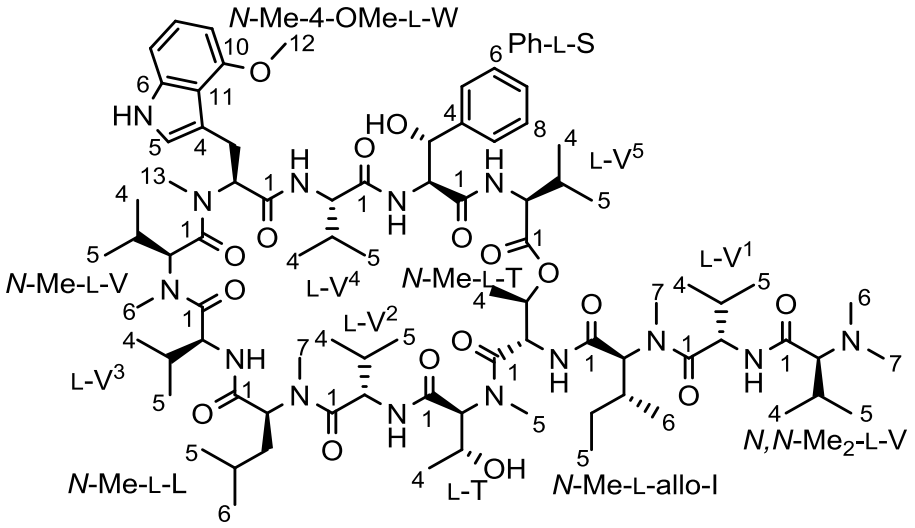
Purity of Ecumicin (1) in Sample H14 Assessed by 100% qHNMR

The ^{1}H NMR spectrum of fraction H14 (900 MHz, CD_3OD, Figure S6) was acquired under quantitative conditions [90 degree excitation pulse (10.25 μs, acquisition time 4.0 s; relaxation delay 60.0 s] using a Bruker Avance II 900 instrument. Processing, used zero filling (2x), Lorentzian-Gaussian apodization (LB=-0.3, GB =0.05), and polynomial baseline correction.

The aromatic singlet (δ_H 6.698) of **1** was selected as the most pure signal, and used to quantitate **1** by applying the following integral-based 100% qHNMR method. The integral of the singlet was set to 1.000, representing the integral of one proton (1H). All other proton signals between δ_H 0.0 to 7.5 in the spectrum, except for the solvent signals (HOD and CHD_2OD), were integrated, and the sum of these integrals was determined to be 137.36.

A total of 125 non-exchangeable protons are present in **1**. Considering that all resonances not belonging to **1** (i.e., the impurity signals) were typical peptide resonances, and assuming that they have the same molecular weight as **1** (i.e., isomeric impurities), the purity of **1** in fraction H14 was calculated as:

Purity = 125×1.00/137.36×100 % = 91 %



1

Detailed 2D NMR Assignments of Ecumicin (1)

Extensive analysis of the 2D NMR spectra of **1**, particularly based on COSY, TOCSY, HSQC HMBC and semi-selective HMBC experiments (Figure S7 to Figure S12) resulted in the elucidation of 15 discrete ^{1}H, ^{1}H spin systems: *N,N*-Me_2-V, V^1, *N*-Me-I, T, *N*-Me-T, V^2, *N*-Me-L, V^3, *N*-Me-V, *N*-Me-4-OMe-W^a, *N*-Me-4-OMe-W^b, V^4, Ph-S^a, Ph-S^b and V^5.

The ^{13}C and ^{1}H signals of aliphatic methyl groups were in very crowded regions, where 14 carbon signals representing 14 carbons were presented within a δ_C 0.94 window (δ_C 19.10 to 20.04, Figure S6). The resolution of normal HSQC and HMBC experiments was not high enough to establish connectivity for these signals, so semi-selective HMBC experiment was used. The direct H-C connectivity was built by using ^{13}C-^{1}H single bond correlations extracted from semi-selective HMBC spectrum (Figure S11). Such correlations were observed with a coupling constant around 126 Hz along F2 direction. The methyl groups were connected to their spin systems respectively by using the COSY correlations between methyl proton signals and β-proton signals. Such connectivities were confirmed by ^{1}H,^{13}C long range couplings between methyl proton signals and β-carbon signals, or between methyl carbon signals and β-proton signals.

Connectivity inside the *N*-Me-4-OMe-W and Ph-S units was established by using ^{1}H,^{13}C long-range correlations extracted from HMBC experiment (Figure S10). Correlations were observed for *N*-Me-4-OMe-W H^4/ *N*-Me-4-OMe-W C^{β}, Ph-S $H^{2'}$/ Ph-S C^{β}, and Ph-S $H^{5'}$/ Ph-S C^{β}.

The position of the methoxyl group of *N*-Me-4-OMe-W was also determined by analyzing the HMBC spectrum. HMBC correlations were observed for *N*-Me-4-OMe-W H^{12}/ *N*-Me-4-OMe-W C^{10}, *N*-Me-4-OMe-W H^8/ *N*-Me-4-OMe-W C^6, *N*-Me-4-OMe-W H^7/ *N*-Me-4-OMe-W C^{11}, *N*-Me-4-OMe-W H^9/ *N*-Me-4-OMe-W C^{11}.

The positions of *N*-methyl and methoxyl groups were determined by analyzing HMBC correlations: (i, i) H^{NMe}/C^{α} (^3J) and/or (i, i+1) $H^{NMe}/C^{C=O}$ (^3J), where "i" indicates the location of individual amino acid residue from the N-terminus. Such correlations were observed for *N,N*-Me_2-V $H^{6(7)}$/ *N,N*-Me_2-V C^{α}, *N*-Me-I H^7/V^1 $C^{C=O}$, *N*-Me-I $H^{7/}$ *N*-Me-I C^{α}, *N*-Me-T H^5/T $C^{C=O}$, *N*-Me-T H^5/ *N*-Me-T C^{α}, *N*-Me-L H^7/V^2 $C^{C=O}$, *N*-Me-L H^7/ *N*-Me-L C^{α}, *N*-Me-V H^6/V^3 $C^{C=O}$, *N*-Me-V H^6/ *N*-Me-V C^{α}, *N*-Me-4-OMe-W H^{13}/ *N*-Me-4-V $C^{C=O}$, *N*-Me-4-OMe-W H^{13}/ *N*-Me-4-OMe-W C^{α}.

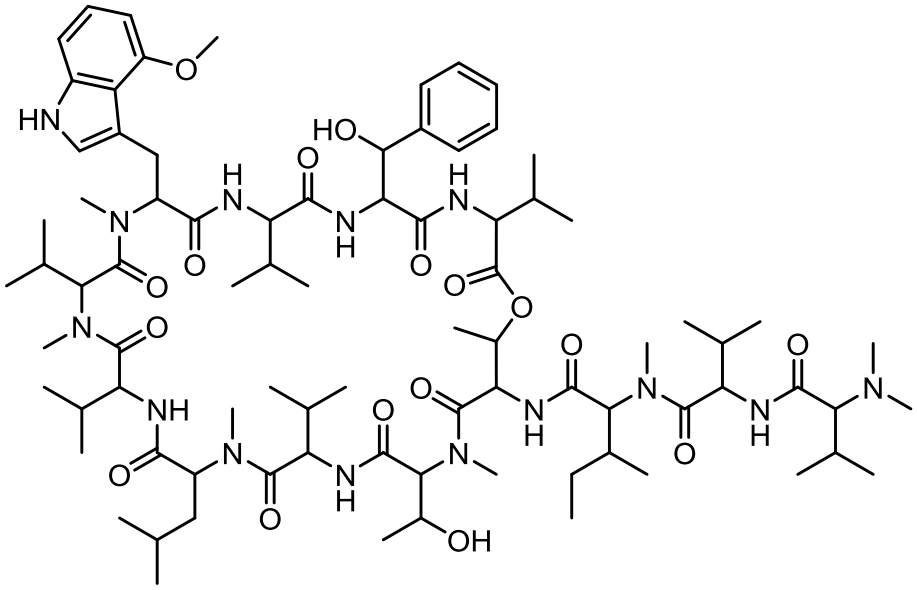
The individual amino acid residues were linked sequentially via correlations between the carbonyl carbon and the following protons: α-protons [(i, i) $H^{\alpha}/C^{C=O}$ (^2J)], β-protons [(i, i) $H^{\beta}/C^{C=O}$ (^3J)], neighboring α-protons [(i, i+1) $H^{\alpha}/C^{C=O}$ (^3J)], and/or neighboring *N*-methyl protons [(i, i+1) $H^{N\text{-}Me}/C^{C=O}$ (^3J)]. In the formula, "i" indicates the location of individual amino acid residue from the N-terminus. Another semi-selective HMBC spectrum was acquired, to resolve the crowded carbonyl region with 13 ^{13}C signals presented in a δ_C 4.23 window (δ_C 170.91 to 175.14, Figure S6). Correlations were observed for *N,N*-Me_2-V H^{α}/ *N,N*-Me_2-V $C^{C=O}$, *N,N*-Me_2-V H^{β}/ *N,N*-Me_2-V $C^{C=O}$, V^1 H^{α}/ V^1 $C^{C=O}$, V^1 H^{β}/ V^1 $C^{C=O}$, *N*-Me-I H^{α}/ *N*-Me-I $C^{C=O}$, *N*-Me-I H^{α}/ V^1 $C^{C=O}$, *N*-Me-I H^{β}/ *N*-Me-I $C^{C=O}$, T H^{α}/ T $C^{C=O}$, T H^{α}/ *N*-Me-I $C^{C=O}$, T H^{β}/ T $C^{C=O}$, *N*-Me-T H^{α}/ *N*-Me-T $C^{C=O}$, *N*-Me-T H^{α}/ T $C^{C=O}$, *N*-Me-T H^{β}/ *N*-Me-T $C^{C=O}$, V^2 H^{α}/ V^2 $C^{C=O}$, V^2 H^{α}/ *N*-Me-T $C^{C=O}$, V^2 H^{β}/ V^2 $C^{C=O}$, *N*-Me-L H^{α}/ *N*-Me-L $C^{C=O}$, *N*-Me-L H^{α}/ V^2 $C^{C=O}$, *N*-Me-L H^{β}/ *N*-Me-L $C^{C=O}$, V^3 H^{α}/ V^3 $C^{C=O}$, V^3 H^{β}/ V^3 $C^{C=O}$, *N*-Me-V H^{α}/ *N*-Me-V $C^{C=O}$, *N*-Me-V H^{α}/ V^3 $C^{C=O}$, *N*-Me-V H^{β}/ *N*-Me-V $C^{C=O}$, *N*-Me-4-OMe-W H^{α}/ *N*-Me-4-OMe-W $C^{C=O}$, *N*-Me-4-OMe-W H^{α}/ *N*-Me-V $C^{C=O}$, *N*-Me-4-OMe-W H^{β}/ *N*-Me-4-OMe-W $C^{C=O}$, V^4 H^{α}/ V^4 $C^{C=O}$, V^4 H^{α}/ *N*-Me-4-OMe-W $C^{C=O}$, V^4 H^{β}/ V^4 $C^{C=O}$, Ph-S H^{α}/ Ph-S $C^{C=O}$, Ph-S H^{α}/ V^4 $C^{C=O}$, Ph-S H^{β}/ Ph-S $C^{C=O}$, V^5 H^{α}/ V^5 $C^{C=O}$, V^5 H^{α}/ Ph-S $C^{C=O}$, V^5 H^{β}/ V^5 $C^{C=O}$.

An HMBC correlation between T H^{β} and V^5 $C^{C=O}$ was observed, suggesting **1** be a cyclic depsipeptide cyclized between the C-terminal carboxyl and the hydroxyl group of the T residue. That the *N,N*-Me_2-V and *N*-Me-L units as their

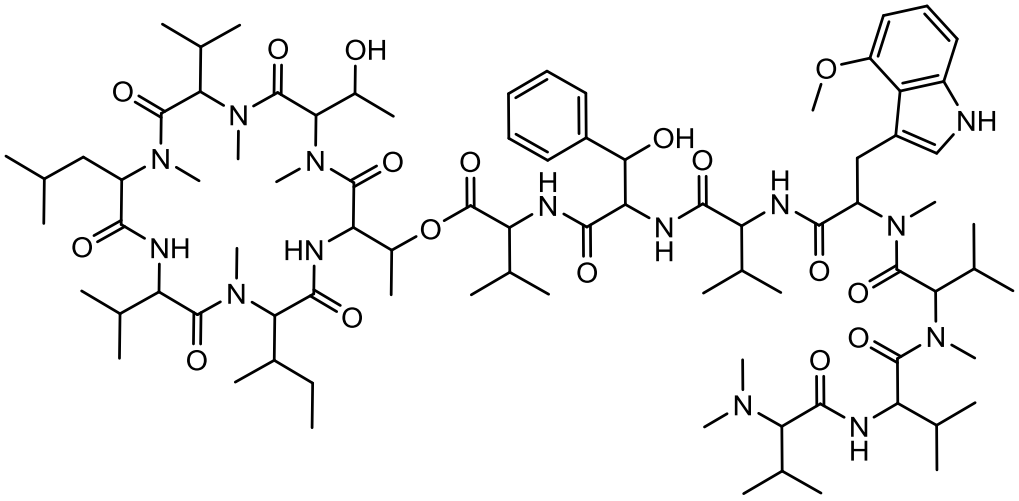
carbonyl carbons resonate at the same frequency leaves two structure possibilities for **1** (Figure S1).

Figure S1. The two possible structures of **1** (a) and (b) were constructed from extensive 1D and 2D NMR analyses and could not be resolved due to the isochronicity ($\Delta\delta$<0.01 ppm) of two carbonyl carbons.

(a)



(b)



III. IR, UV and CD Spectra of Ecumicin (1)

Figure S2. IR spectrum of **1**.

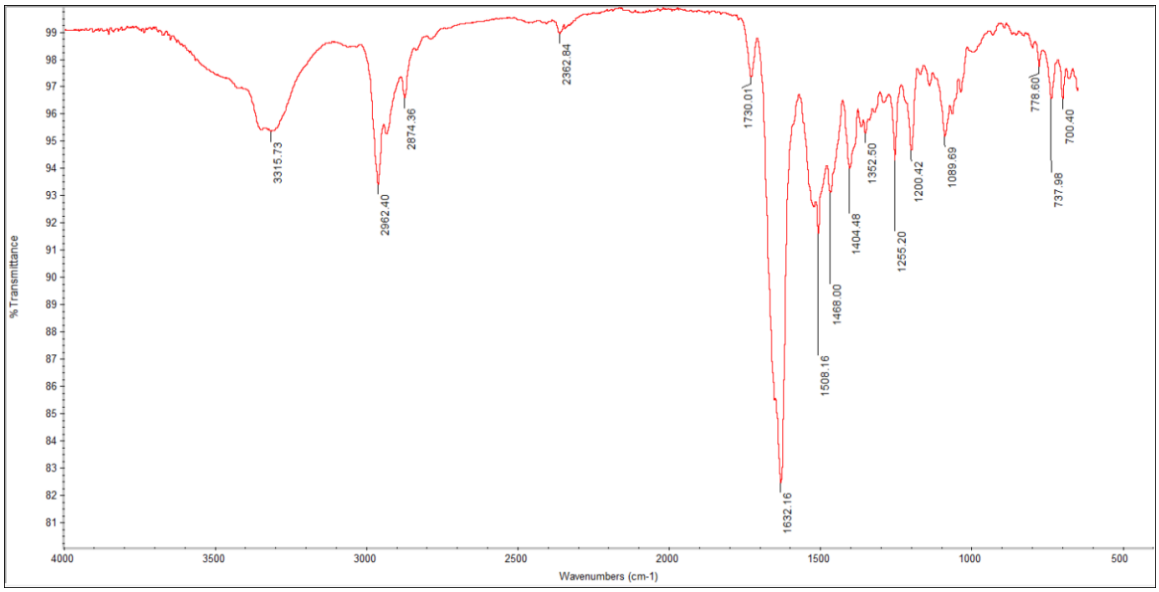
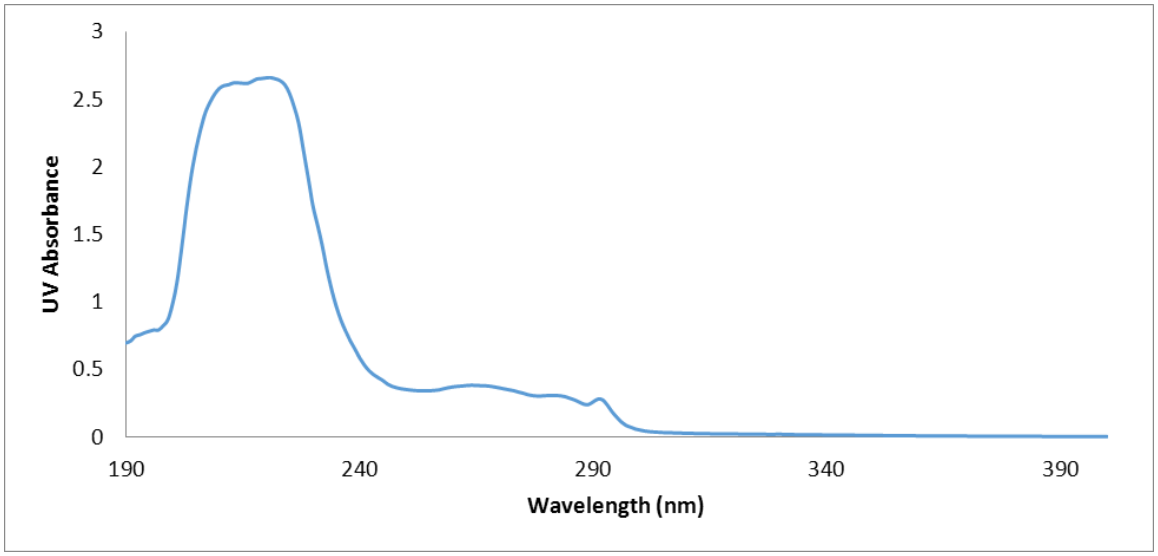


Figure S3. UV spectrum of **1**.



The spectrum was recorded with a methanol solution of **1** at 48.0 μM, in 1 cm quartz cuvette, 25°C (background subtracted).

ε = 54209 M^{-1}·cm^{-1} (211 nm)

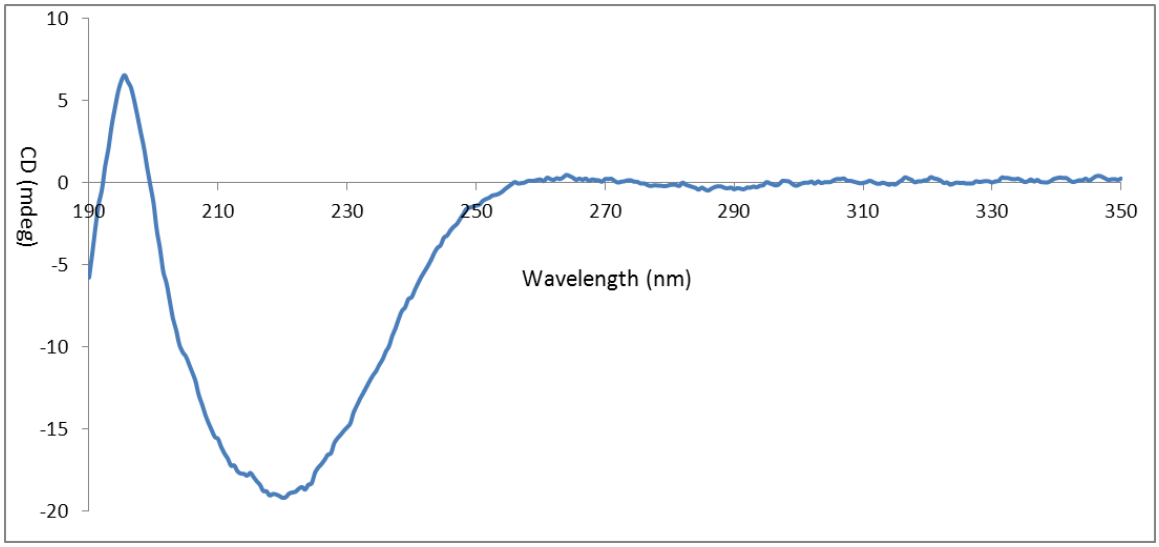
ε = 55266 M^{-1}·cm^{-1} (219 nm)

ε = 7925 M^{-1}·cm^{-1} (263 nm)

ε = 6465 M^{-1}·cm^{-1} (281 nm)

ε = 5839 M^{-1}·cm^{-1} (291 nm)

Figure S4. CD spectrum of **1**.



The spectrum was recorded with an acetonitrile solution of **1** at 10 µM, in 1 cm quartz cuvette, 25°C (background subtracted), on a Jasco J 715 polarimeter.

$[\theta]_{molar}$ = -1918530 deg·cm^2·mol^{-1} (220 nm)

$[\theta]_{molar}$ = 616094 deg·cm^2·mol^{-1} at 196 nm (25°C, in acetonitrile).

IV. NMR Spectra and Assignments of Ecumicin (1)

The 1D NMR spectra (FIDs) of ecumicin are available online as separate Supporting Information (ZIP file). The ^{1}H NMR raw data of ecumicin in CD_3OD (900 MHz, Bruker Avance II 900) and the ^{13}C NMR raw data of ecumicin in CD_3OH (100 MHz, Bruker DPX-400) can be found in 1D-NMR.zip, under the folder ^{1}H and ^{13}C, respectively.

Figure S5. ^{1}H NMR Spectra of **1** in (A) CD_3OD (900 MHz, Bruker Avance II 900), and (B) CD_3OH (600MHz, Bruker Avance 600, processed with water subtraction). Expansions of the 900 MHz spectrum in regards to (C) methyl/methylene region, (D) α-proton region, (E) aromatic region, in CD_3OD, and (G) amide region in CD_3OH (600 MHz), with further expansion of the most crowded methyl region (F) in CD_3OD.

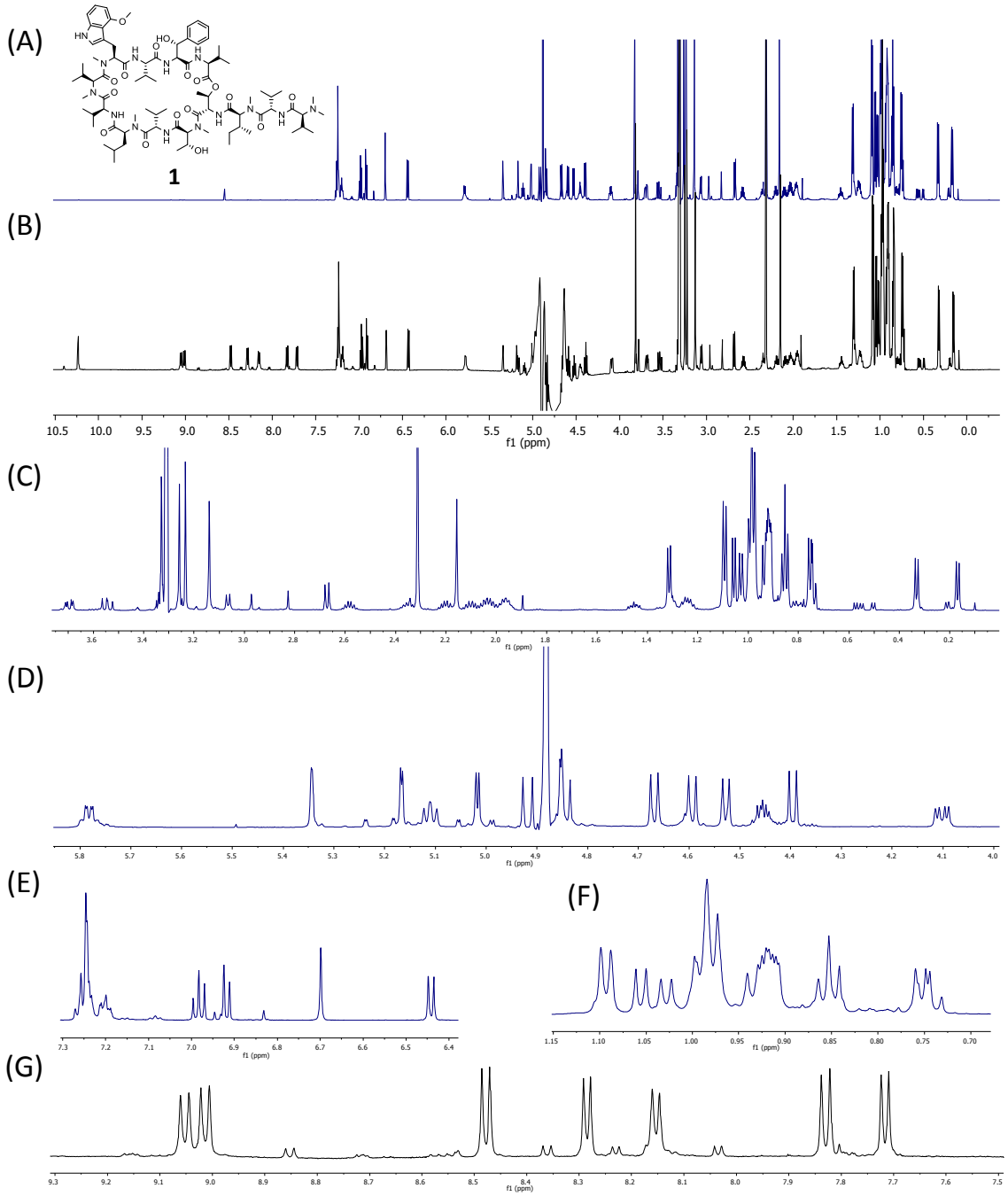


Figure S6. (A) ^{13}C and (B) DEPT-135 spectra of **1** in CD_3OH (100 MHz, Bruker DPX-400). Expansions of ^{13}C spectrum on (C) methyl/methylene region, (D) α-carbon region, (E) aromatic region, (F) carbonyl region and further expansion on the most crowded methyl region (G). (H) and (I) are expansions of DEPT-135 spectrum on the methylene carbons.

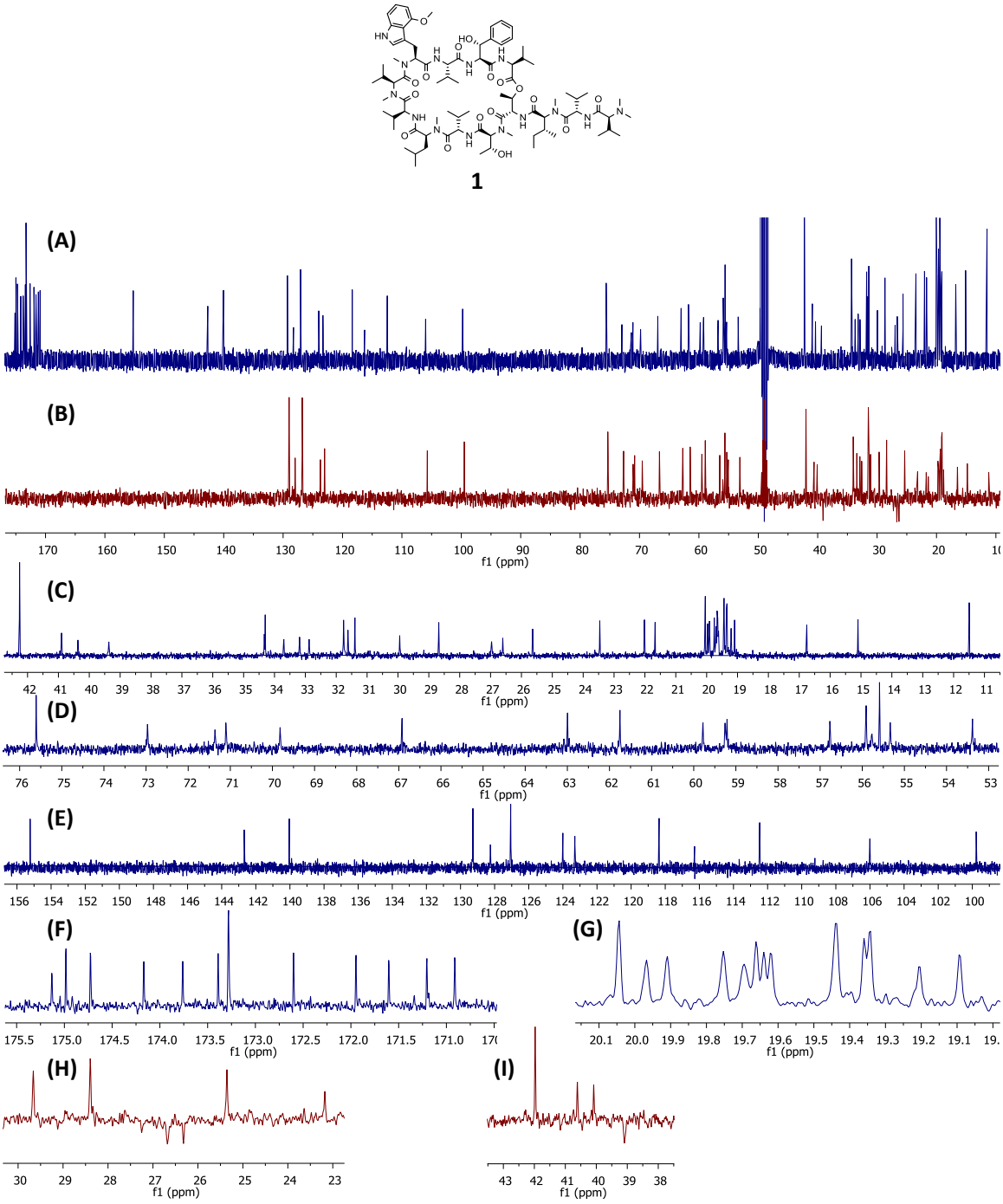


Figure S7. COSY spectrum of **1** in CD_3OD (600 MHz, Bruker Avance 600). The spectrum was acquired at (TD) 2048×256, in 32 scans. It was zero filled to 4k×4k, without linear prediction. The spectrum was not symmetrized during processing.

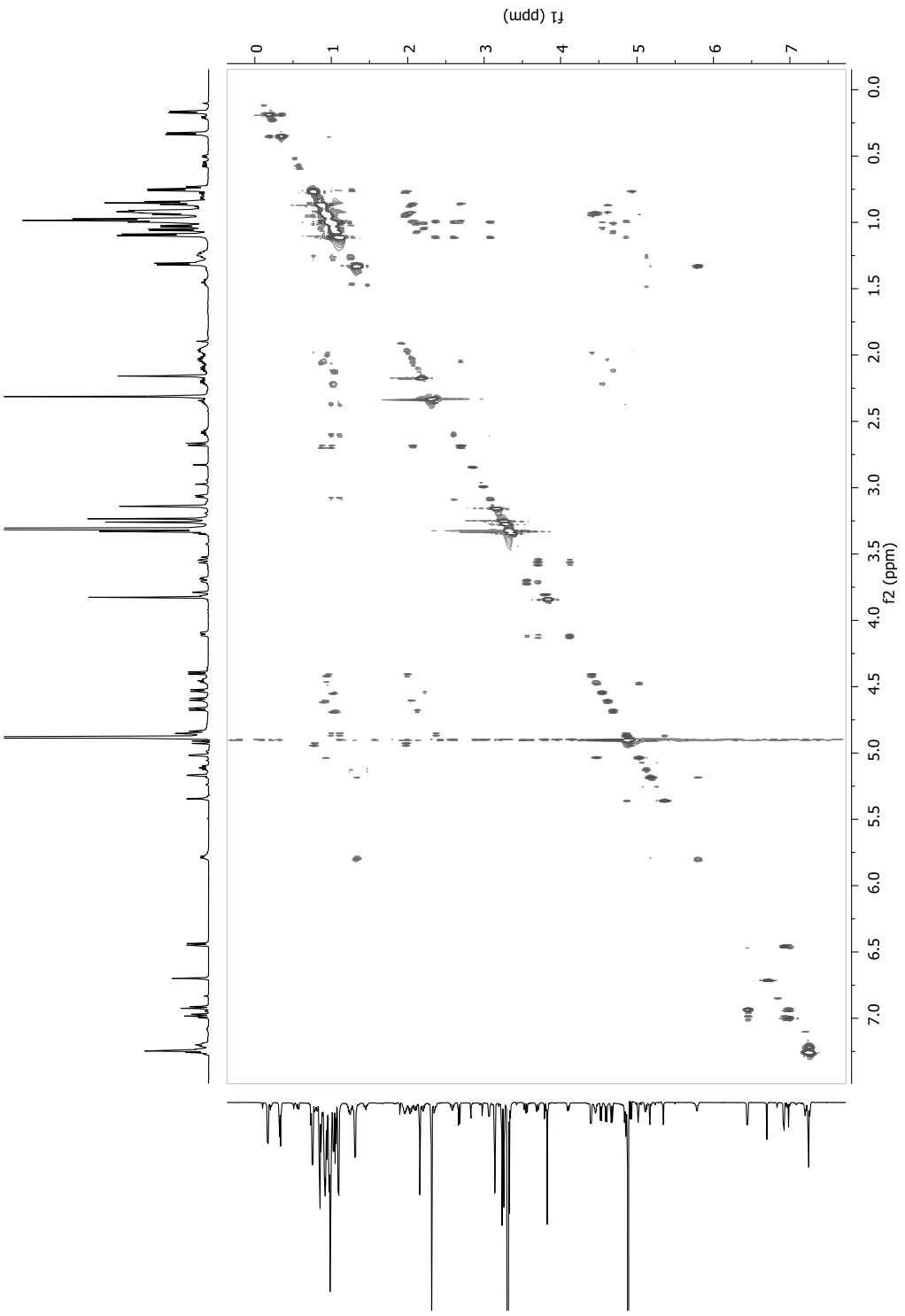


Figure S8. TOCSY spectrum of **1** in CD_3OD (600MHz, Bruker Avance 600). The spectrum was acquired at (TD) 4096×256, in 16 scans. It was zero filled to 4k×4k, without linear prediction. The spectrum was not symmetrized during processing.

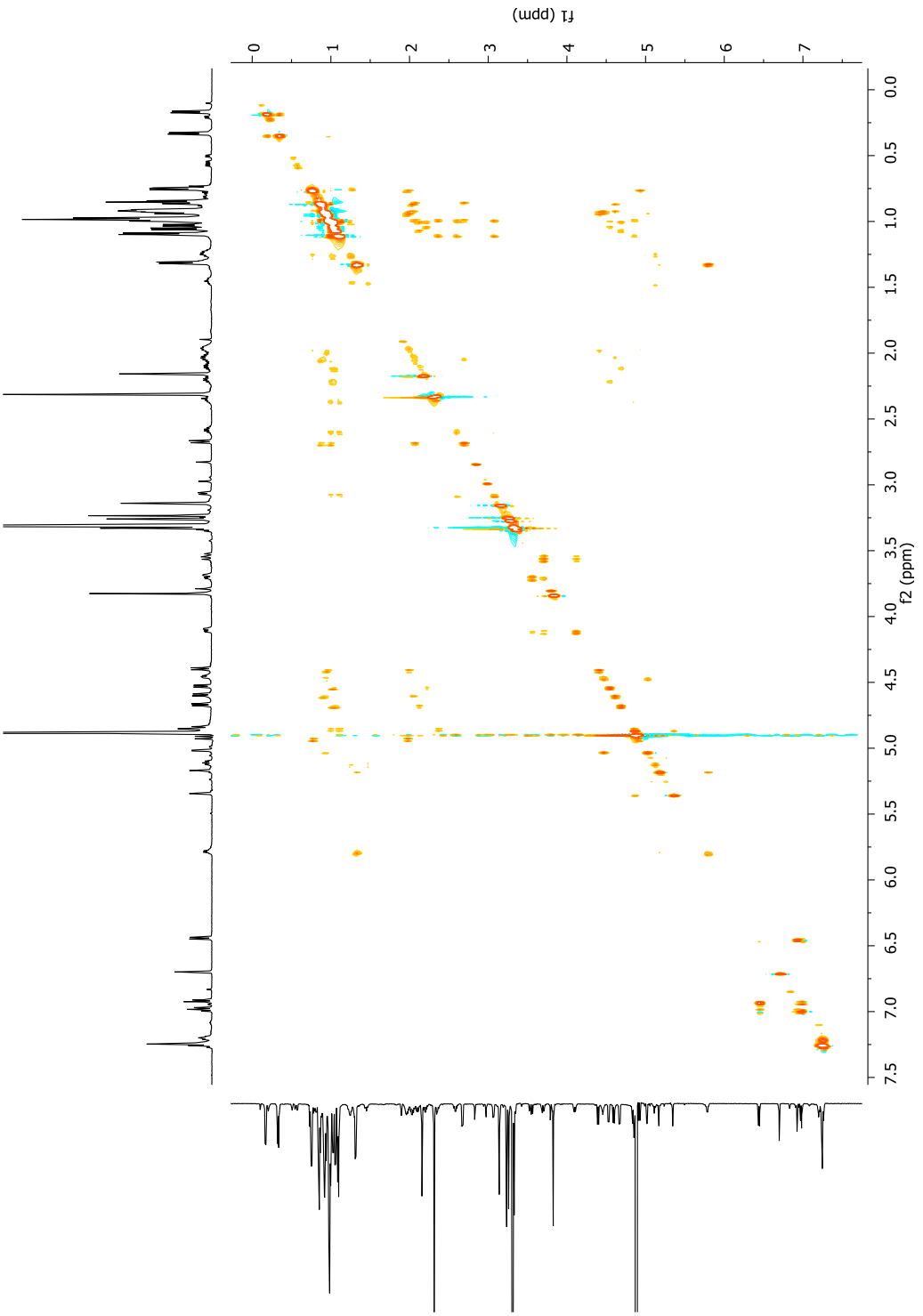


Figure S9. HSQC spectrum of **1** in CD_3OD (600 MHz, Bruker Avance 600). The spectrum was acquired at (TD) 4096×256, in 32 scans. It was zero filled to 4k×4k, with forward linear prediction on f1 direction.

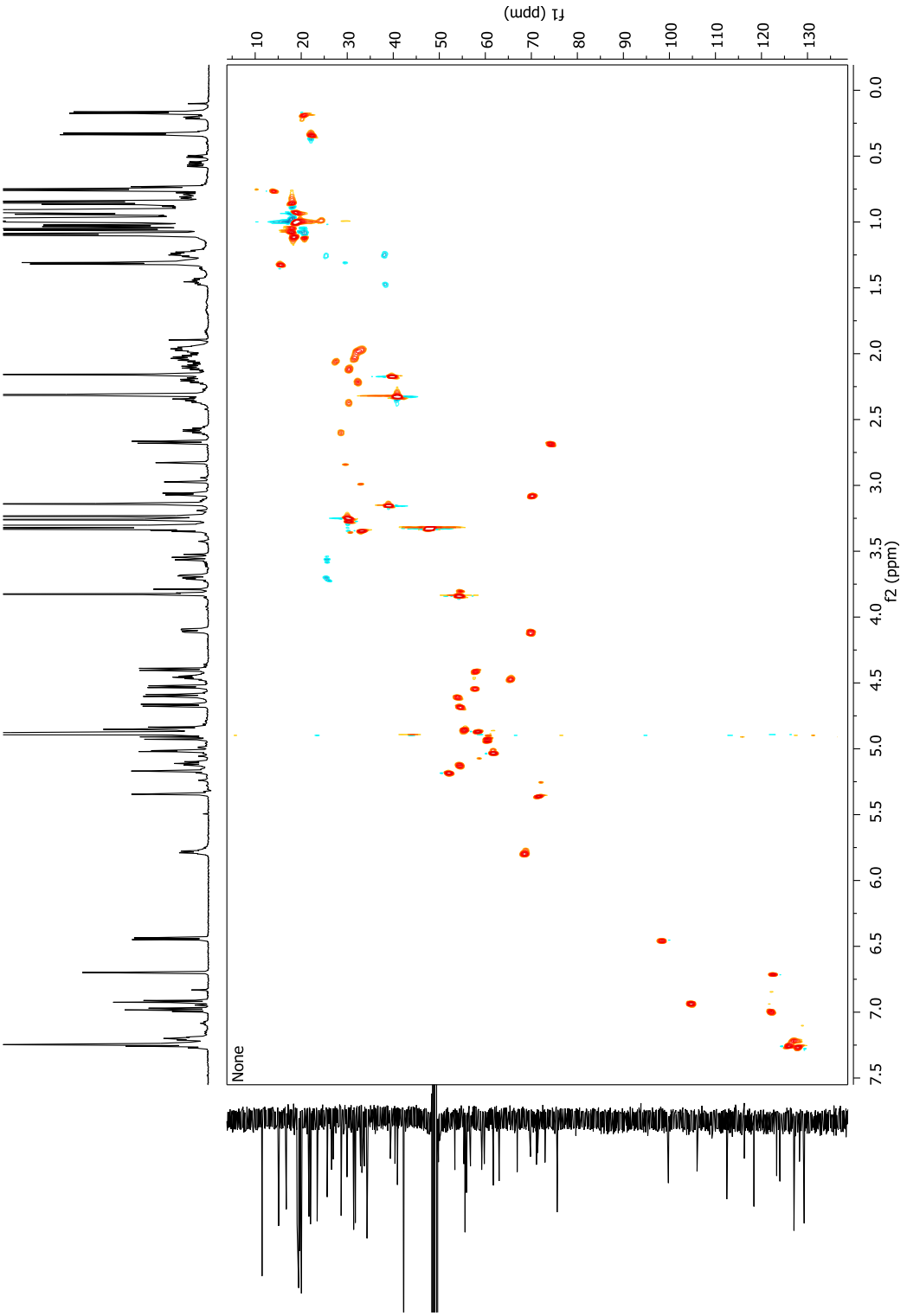


Figure S10. HMBC spectrum of **1** in CD_3OD (600 MHz, Bruker Avance 600). The spectrum was acquired at (TD) 4096×256, in 64 scans. It was zero filled to 4k×4k, with forward linear prediction on f1 direction.

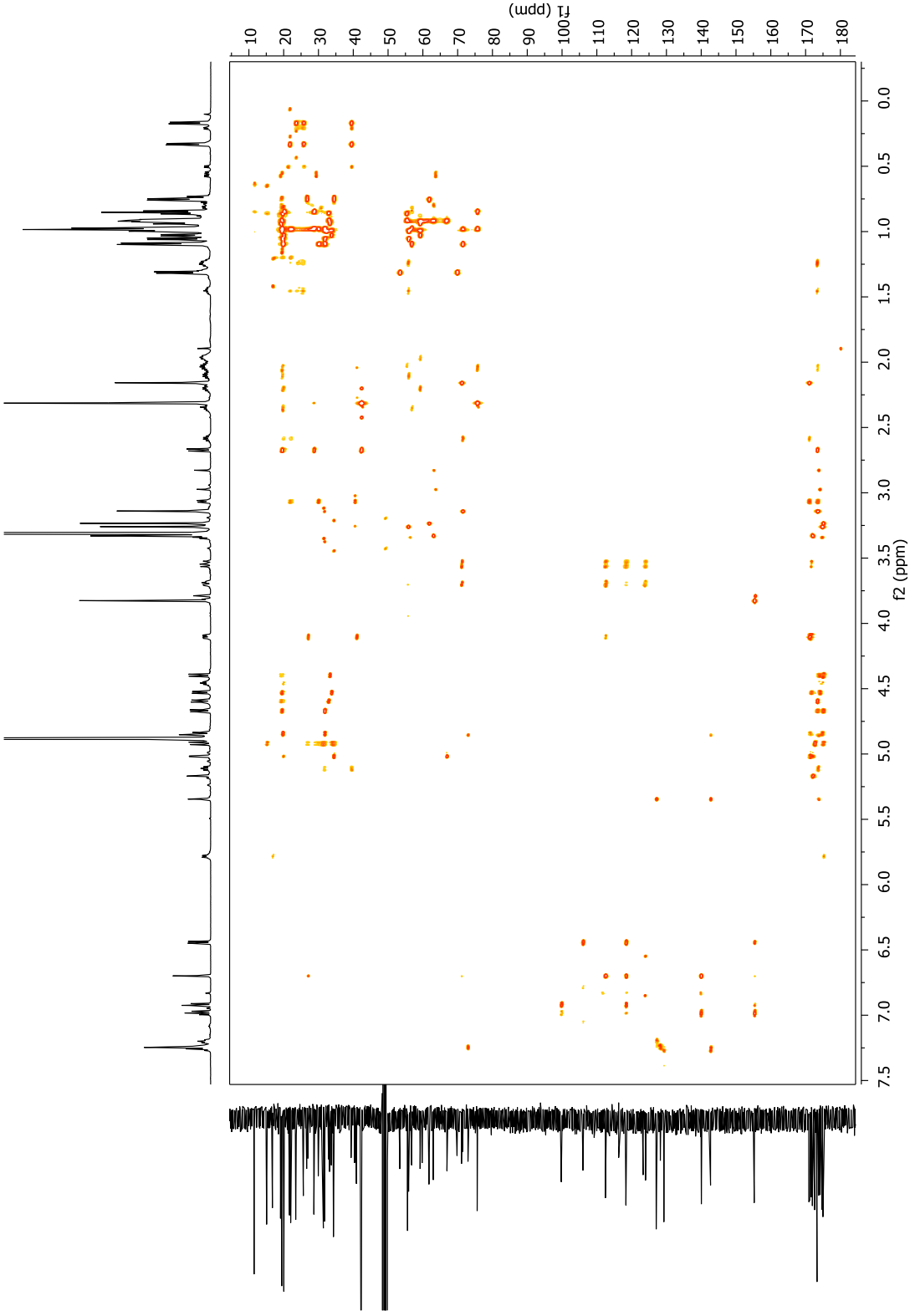


Figure S11. Semi-selective HMBC spectrum of **1** carbonyl region in CD_3OD (600 MHz, Bruker Avance 600). The spectrum was acquired at (TD) 4096×256, in 64 scans. It was zero filled to 4k×4k, with forward linear prediction on f1 direction.

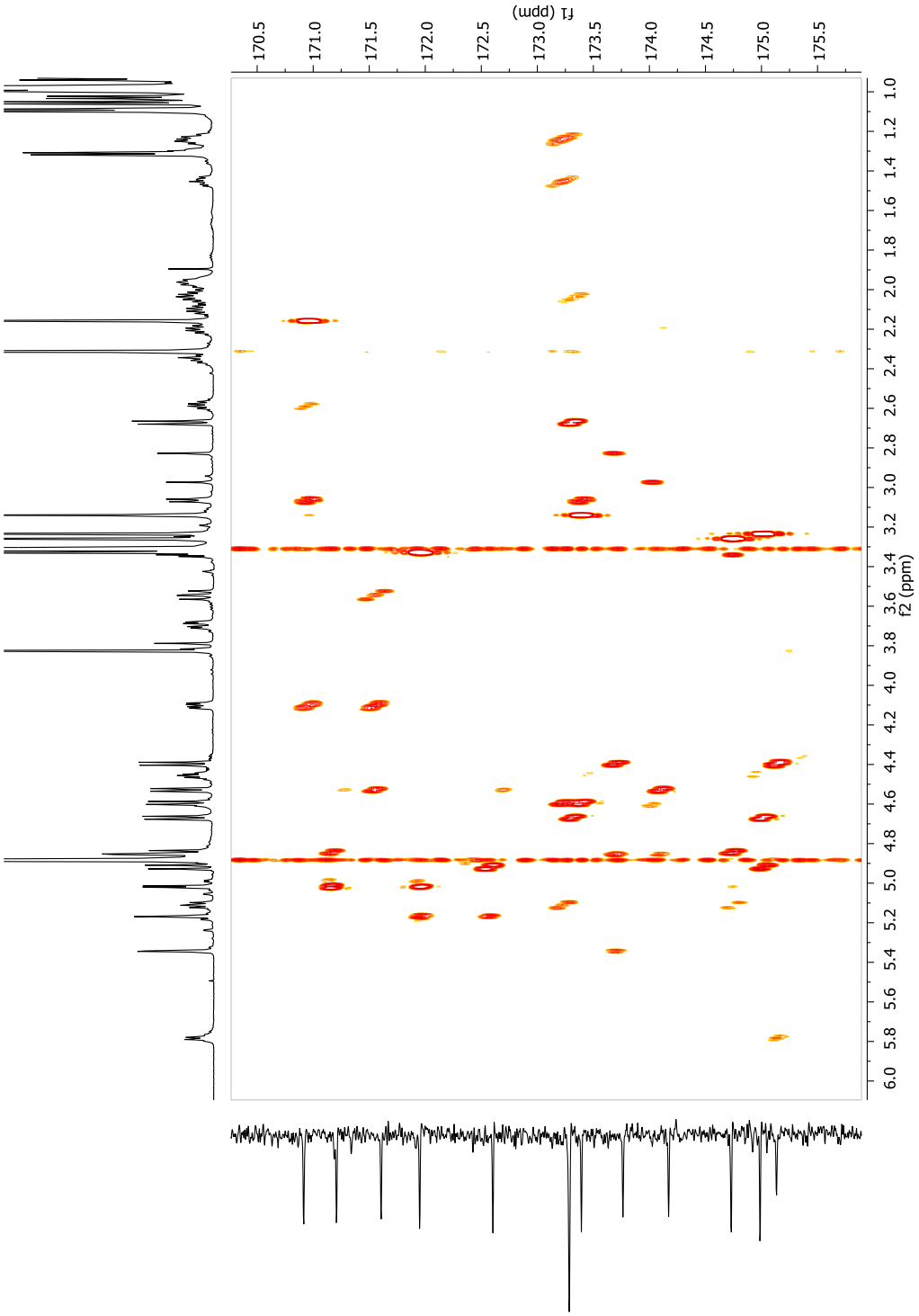


Figure S12. Semi-selective HMBC spectrum of **1** aliphatic methyl region, in CD_3OD (600 MHz, Bruker Avance 600). The spectrum was acquired at (TD) 4096×256, in 48 scans. It was zero filled to 4k×4k, with forward linear prediction on f1 direction.

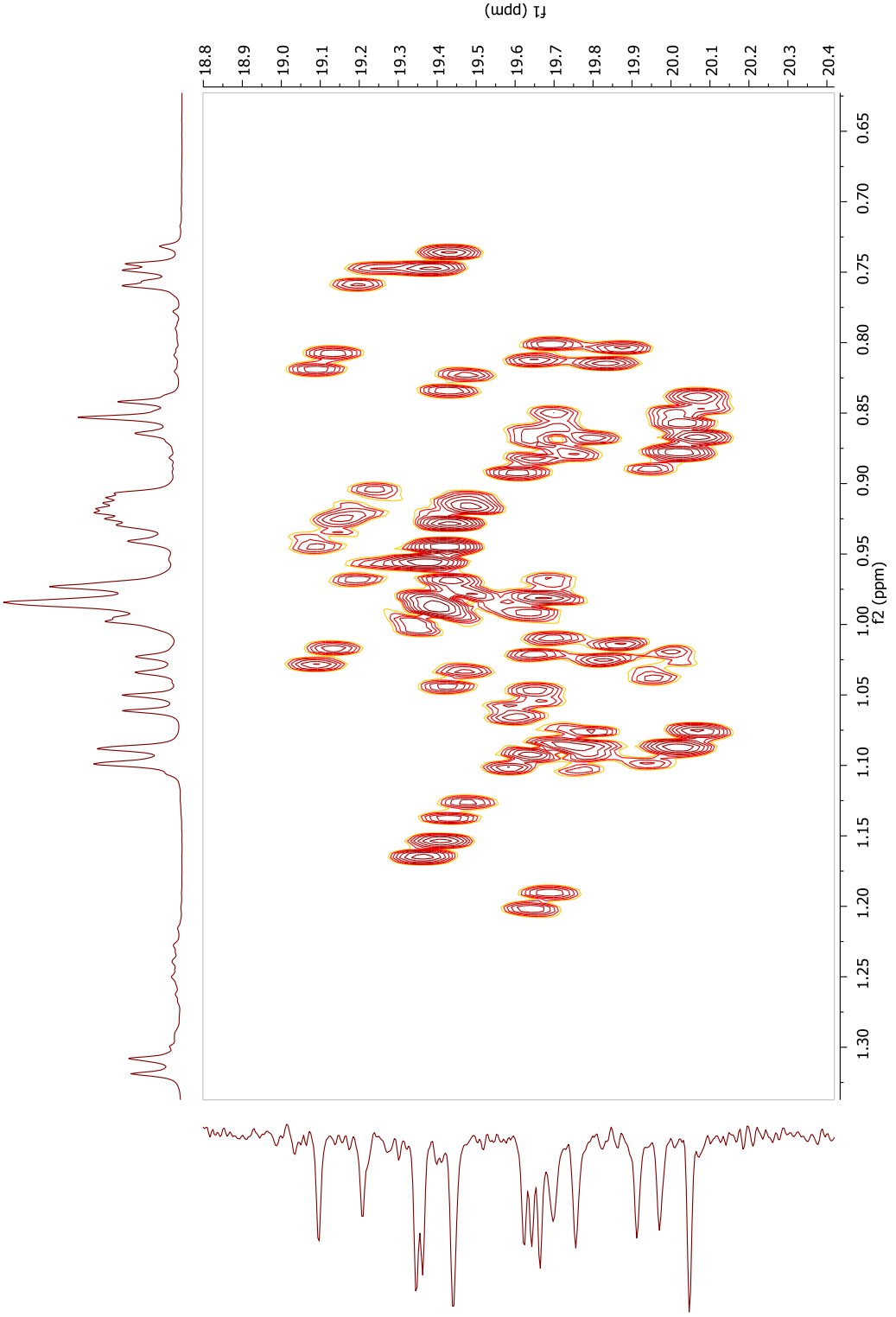


Figure S13. ROESY spectrum of **1** in CD_3OD (600 MHz, Bruker Avance 600). The spectrum was acquired at (TD) 2048×256, in 32 scans. It was zero filled to 4k×4k, without linear prediction. The spectrum was not symmetrized during processing.

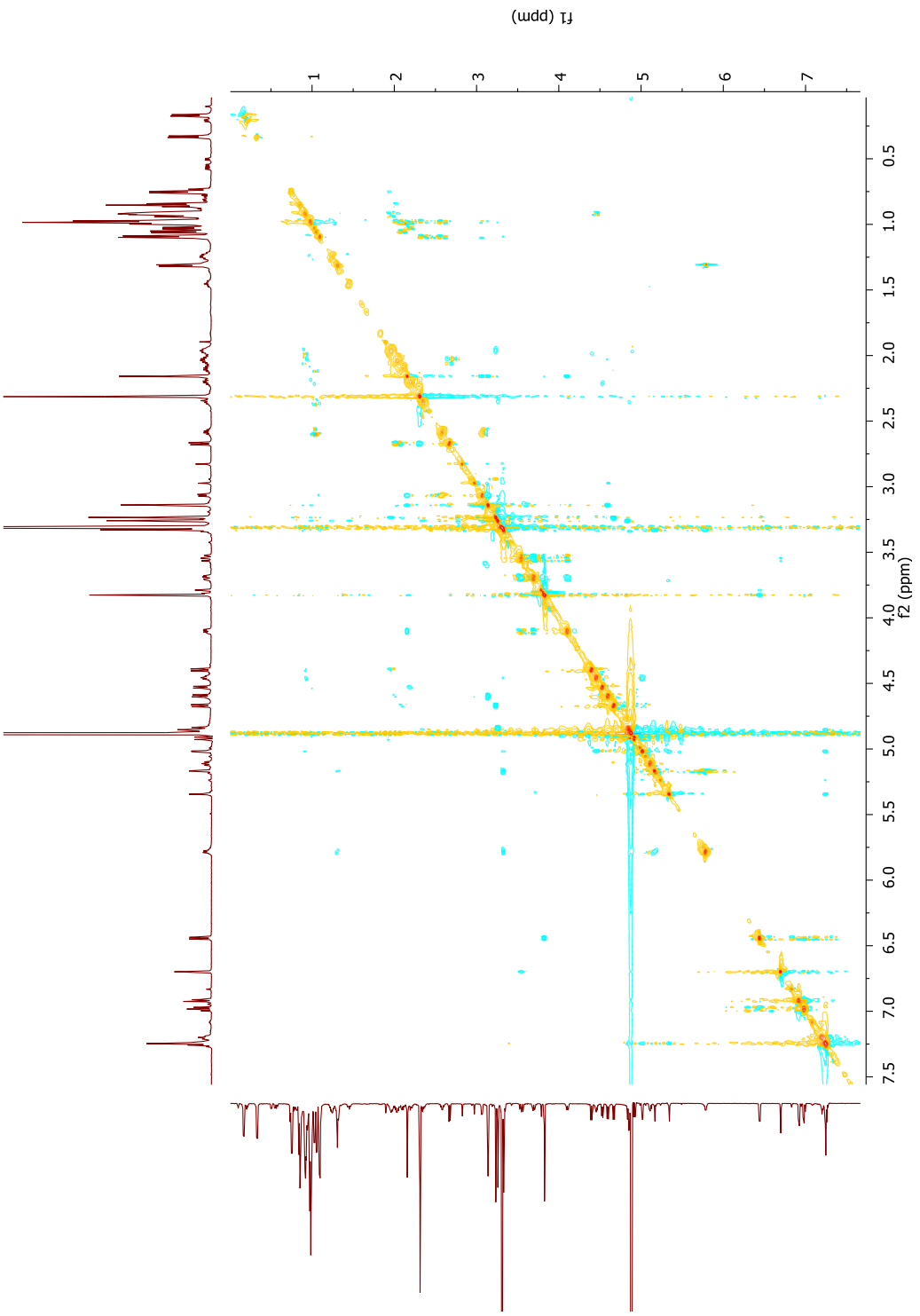
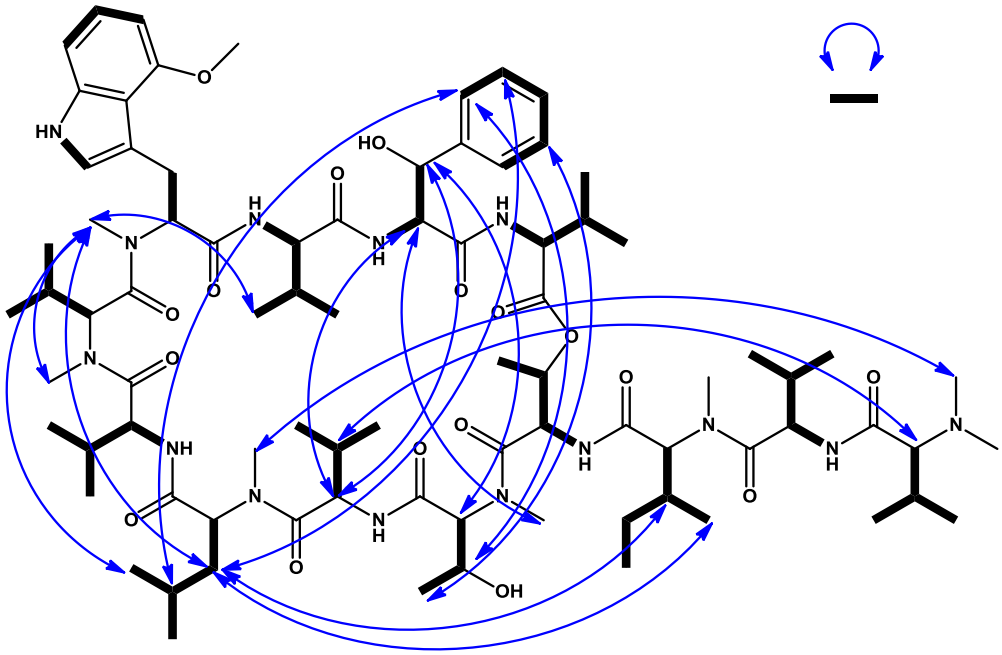


Figure S14. Observed ROESY correlations for **1**.



V. MS spectra and assignments.

Figure S15. LC-MS^2 spectrum of **1**.

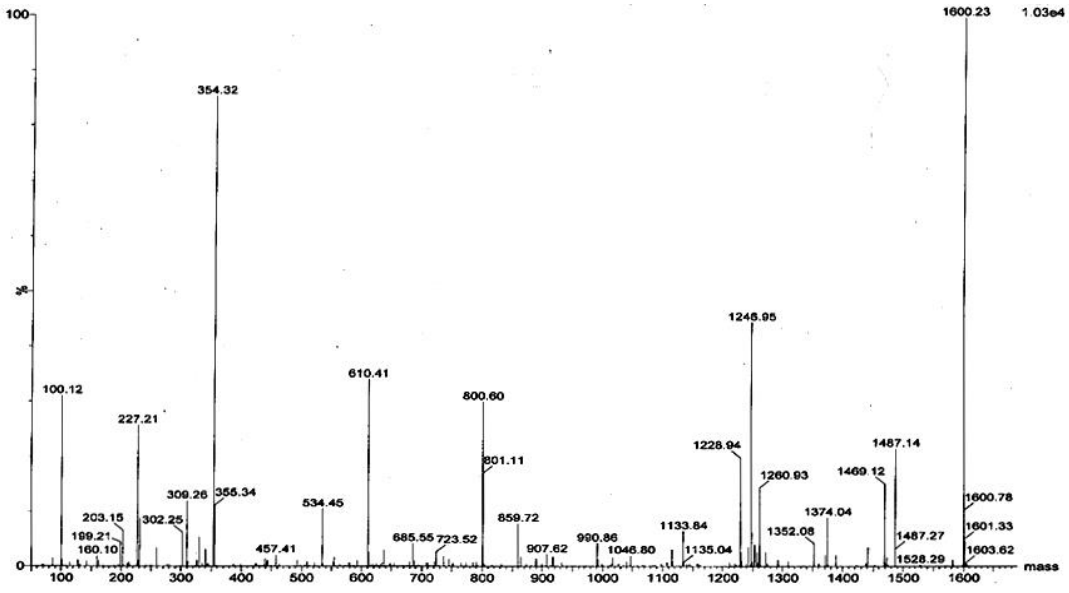


Figure S16. Fragmentation pattern in tandem mass spectroscopy provided conclusive evidence for the planar structure of **1**.

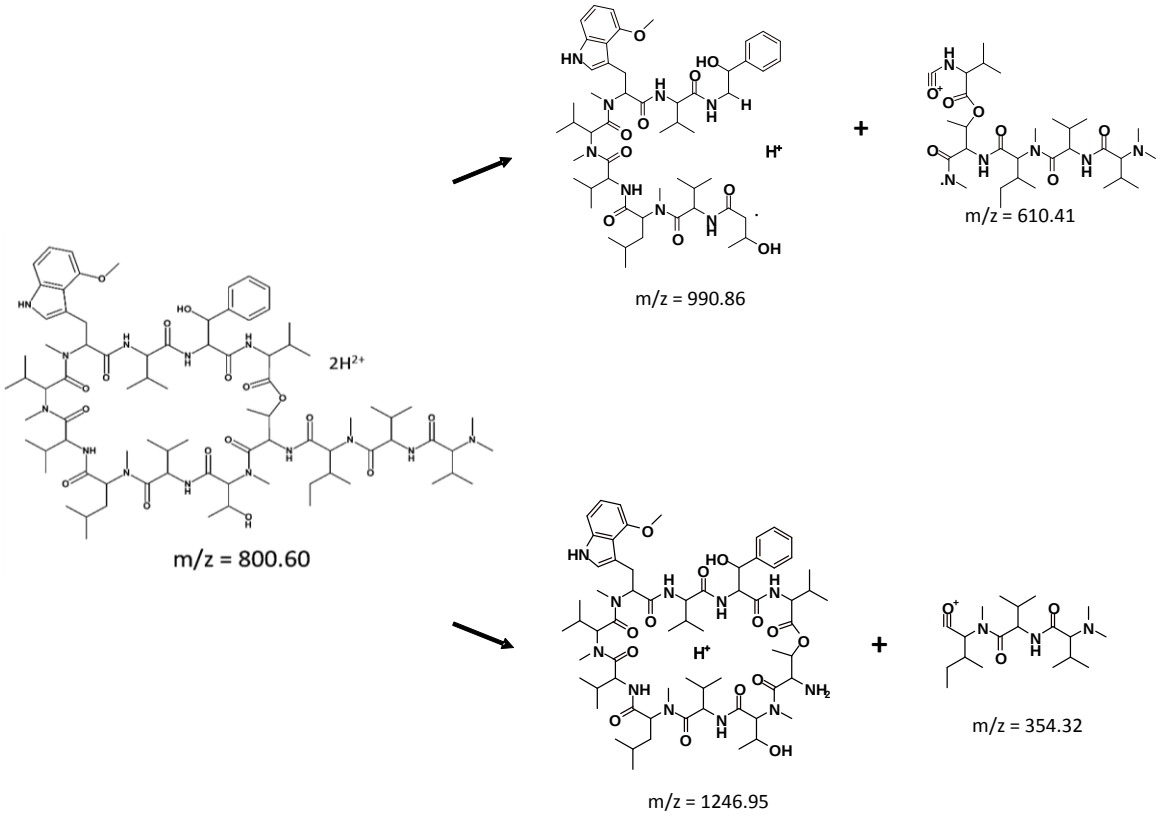
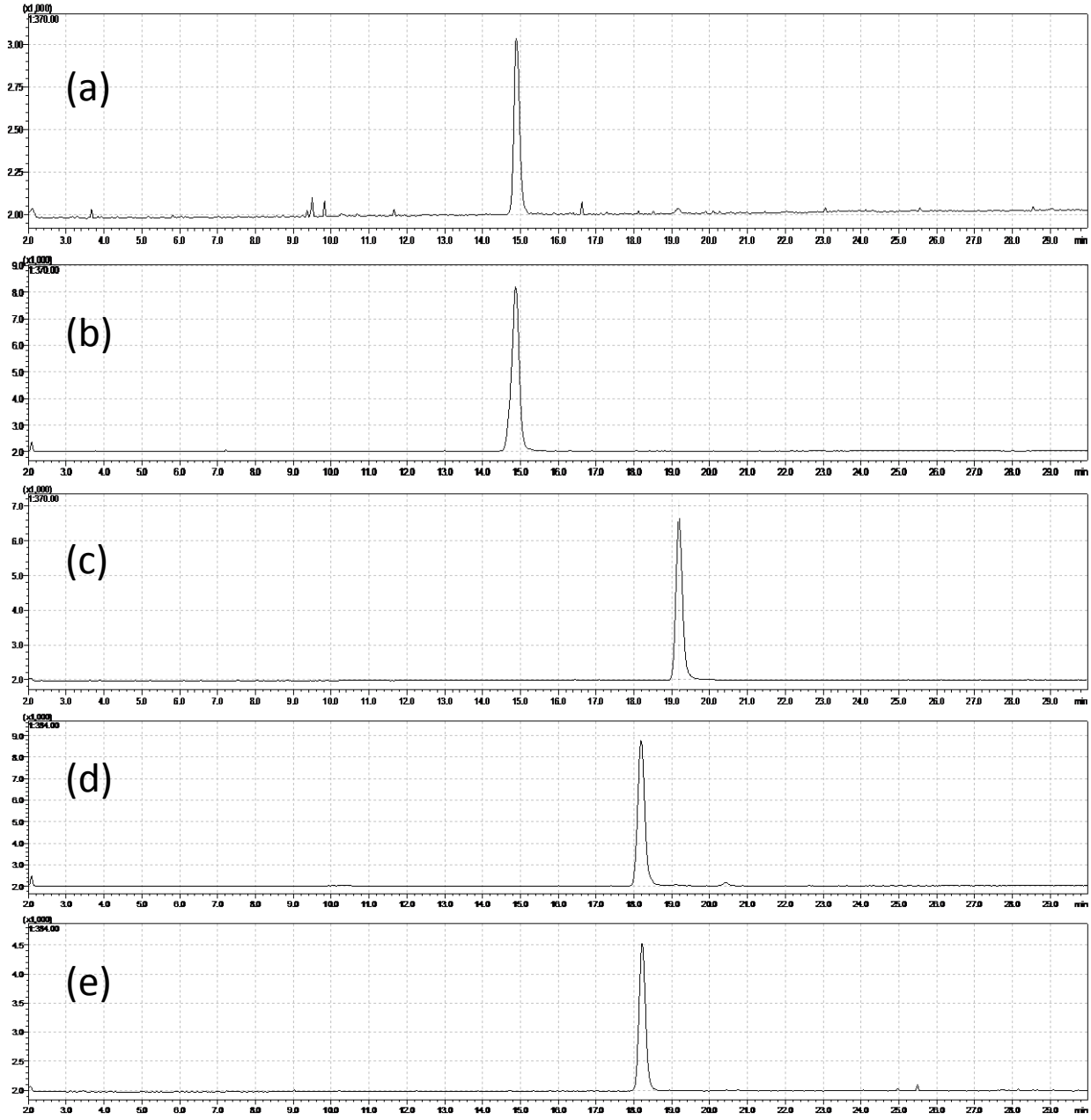


Figure S17. Result of the Marfey's experiment.



LC-ESI-MS (Shimadzu 2020) on a RP C18 column; peaks shown in (a), (b), and (c) were detected with SIM mode at *m/z* = 370.3; peaks shown in (d) and (e) were detected with SIM mode at *m/z* = 384.0.

(a) Hydrolate of **1** reacted with L-FDAA, r.t. = 14.87 min.
(b) L-ala-NH_2-DNP-L-valine, r.t. = 14.87 min.
(c) L-ala-NH_2-DNP-D-valine, r.t. = 19.19 min.
(d) Hydrolate of **1** reacted with L-FDAA, r.t. =18.22 min.
(e) L-ala-NH_2-DNP-*N*-methyl-L-valine, r.t. =18.22 min.

VI. X-ray crystallography data.

Figure S18. Oak Ridge Thermal-Ellipsoid Plot Program representation of the asymmetric unit of ecumicin, shown at 50% probability displacement ellipsoids. X-ray data were collected on a Bruker D8 discover X-ray system at room temperature and the crystal diffracted X-ray to 0.83 Å. Crystal data for ecumicin: $C_{83}H_{134}O_{17}N_{14}$, M = 1599. 01, space group $P2_12_12$ (no. 18), a = 71.64(7) Å, b = 11.434(11) Å, c = 12.697(13) Å, V= 10400.110 $Å^3$, Z = 4, F(000) = 3972, Dc = 1.172 g/cm^3, μ (MoKα) = 18.71 cm^{-1}, T = 293(2), 1142 variables refined with 10140 reflections with I > 4σ(I) to R1 = 0.1722. The structure was solved by direct methods using ShelxD program and refined by full-matrix least-squares using ShelxL.

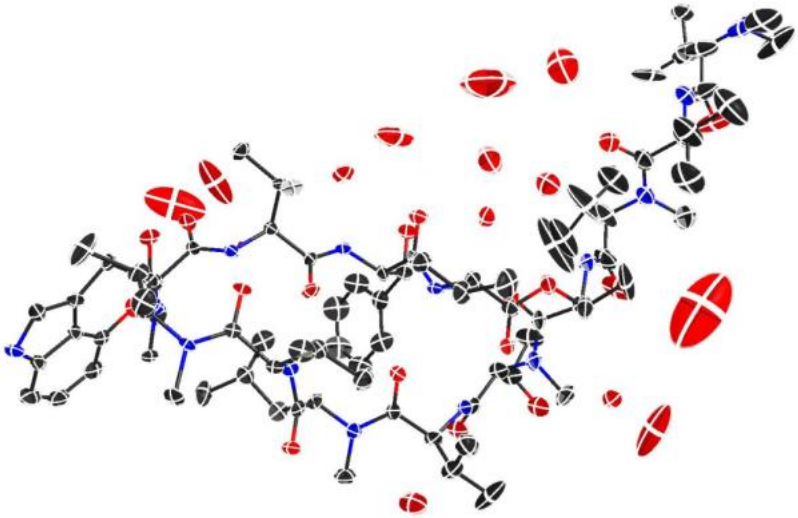


Figure S19. The *B*-factor putty tube representation of ecumicin shows that its structure is mostly ordered, with only the two *N*-terminal amino acids showing flexibility. The segments with the highest temperature factor are shown as thicker red cylinders.

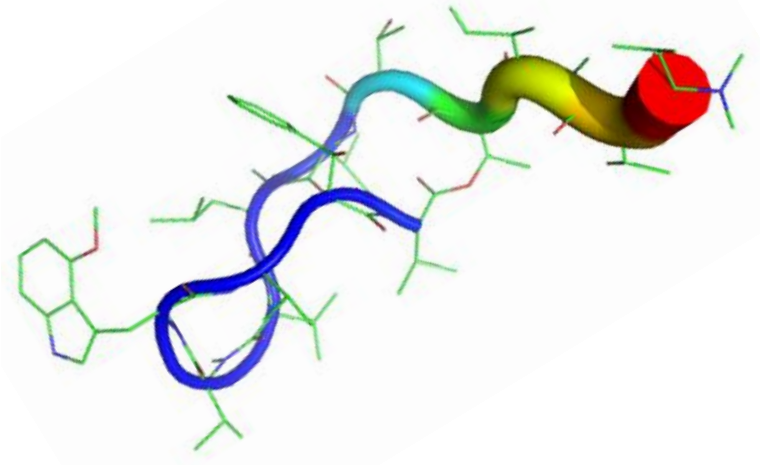


Table S1. Summary of carbonyl groups that may contribute to n→π* interactions in ecumicin (AA = amino acid residue).

AA (C=)O[1]	AA C(=O[2])	$d_{n\rightarrow\pi^*}$ (Å)	θ
V[3]	*N*-Me-V	2.647	97.6°
N-Me-V	*N*-Me-4-OMe-W	2.911	124.4°