

UC San Diego

UC San Diego Previously Published Works

Title

Searching for Small Molecules with an Atomic Sort

Permalink

<https://escholarship.org/uc/item/1q05m9m2>

Journal

Angewandte Chemie, 132(3)

ISSN

0044-8249

Authors

Duggan, Brendan M

Cullum, Reiko

Fenical, William

et al.

Publication Date

2020-01-13

DOI

10.1002/ange.201911862

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Molecular Search Engine

International Edition: DOI: 10.1002/anie.201911862

German Edition: DOI: 10.1002/ange.201911862

Searching for Small Molecules with an Atomic Sort

Brendan M. Duggan,* Reiko Cullum, William Fenical, Luis A. Amador, Abimael D. Rodríguez, and James J. La Clair*

Abstract: The discovery of biologically active small molecules requires sifting through large amounts of data to identify unique or unusual arrangements of atoms. Here, we develop, test and evaluate an atom-based sort to identify novel features of secondary metabolites and demonstrate its use to evaluate novelty in marine microbial and sponge extracts. This study outlines an important ongoing advance towards the translation of autonomous systems to identify, and ultimately elucidate, atomic novelty within a complex mixture of small molecules.

One of the most critical aspects in the discovery of biologically active small molecules is the elucidation of small molecular motifs with unique three-dimensional displays. The combination of this process with detailed target-based mode of action research^[1] lies at the foundation of drug lead^[2] discovery. While automation,^[3] miniaturization,^[4] digital networking^[5] and machine learning-guided high-throughput screening^[6] have produced active leads, the bulk of screening efforts still follow a central approach that begins with a molecular ensemble, either an extract containing natural products or a smart library of synthetic compounds.^[7] Although both synthetic and natural approaches appear different, they typically apply a combination of molecular, cellular, or phenotypic screens. While effective, such approaches are often cluttered by the discovery of redundant structural features and motifs. This strategy has prevailed, in part, due to our inability to search for structural novelty.

Mass spectrometry (MS) methods, and associated profiling systems, provide an excellent means to characterize molecules, but are typically limited to databased compounds

with effective molecular ionization. For the elucidation of molecular motifs NMR spectroscopy is required, typically achieved by collecting a series of 2D spectra, which a trained user interprets to unequivocally identify every atom in the sample. Recently, computational systems to evaluate NMR data have been developed.^[8] Salient examples, such as the SMART system, allow one to rapidly identify the structural family of a purified compound.^[8b] However, these tools use NMR data to evaluate molecular species. While these approaches enable rapid clustering of a new compound with structural neighbours, they are not able to identify novel material, which for natural product lead discovery requires one to return to the classical methods of compound isolation via dereplication with repeated purification and characterization steps.

One can view a mixture of compounds simply as a collection of atoms. Since NMR reports atomic data, it will provide information on all the atoms in all the compounds present in a mixture. Since the most common atoms in molecules of pharmaceutical interest are hydrogen (H) and carbon (C), a ¹H-¹³C HSQC NMR spectrum provides a near complete map of the molecular frameworks within a screening collection. While some structural features are missed, this spectrum is routinely used by experts to classify and identify compounds, and for this reason we selected the ¹H-¹³C HSQC experiment as the data source for developing an objective method that could be readily automated for scoring and sorting atomic novelty.

We began by constructing a ¹H-¹³C HSQC peak database, where each peak describes a hydrogen attached to a carbon atom (Figure 1c, Supporting Figure S3) and is diagnostic of the atom's immediate atomic environment (Supporting Figures S1–S4). Using publicly available data, we abstracted peaks from spectra in the Human Metabolome Database^[9] and the BioMagResBank,^[10] and constructed peak lists from a tabulation of the chemical shifts of common solvents.^[11] The public data was supplemented with spectra of standards and natural products collected in house or obtained from the literature. The total number of peaks was 10308 obtained from 1207 spectra.

We then collected a ¹H-¹³C HSQC spectrum on a 50 µg sample of a model natural product, bromophycolide A^[12] (Figure 1a, Supporting Figure S4, Supporting Table S1). Automated digital peak picking of this spectrum produced a peak list (Figure 1b), which was compared against the database (Figure 1c) to provide a profiled spectrum (Figure 1d, Supporting Figure S5). A distance score for each peak in the profiled spectrum (Figure 1d) was determined by calculating the Euclidean distance to the closest peak in the database (Supporting Table S1). Since the ranges of the ¹H and ¹³C dimensions of a HSQC experiment differ by a factor

[*] Dr. B. M. Duggan

Skaggs School of Pharmacy and Pharmaceutical Sciences
University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093 (USA)
E-mail: bmduggan@ucsd.edu

R. Cullum, Prof. W. Fenical


Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California, San Diego
La Jolla, CA 92093-0204 (USA)

L. A. Amador, Prof. A. D. Rodríguez

Molecular Sciences Research Center, University of Puerto Rico
1390 Ponce de León Avenue, San Juan, 00926 (Puerto Rico)

Dr. J. J. La Clair

Department of Chemistry and Biochemistry, University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093 (USA)
E-mail: jlaclair@ucsd.edu

 Supporting information (including experimental procedures, copies of NMR spectra, lead characterization data, as well as access to raw FID data) and the ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/anie.201911862>.

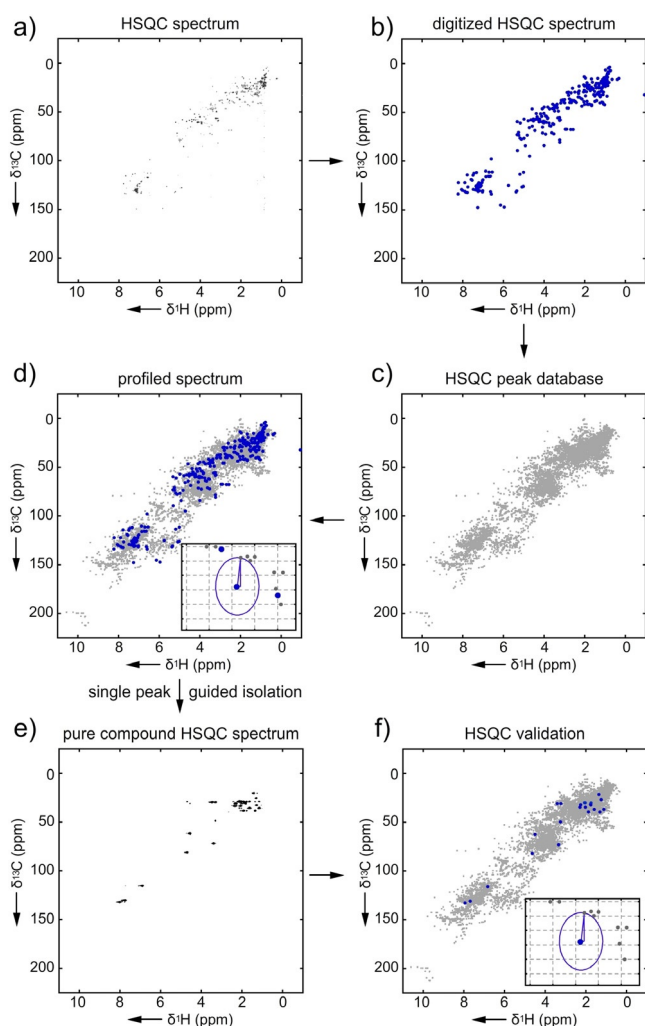


Figure 1. Searching atomic novelty in bromophycolide A. Workflow: a) The process begins by collection of a ^1H - ^{13}C HSQC spectrum of bromophycolide A. b) Once collected, the peaks are then digitally abstracted to provide the observed peak list (a 10 kB text file). The observed peak list is then compared one peak at a time against the HSQC database (a 850 kB text file) using the Euclidian distance algorithm. c) To illustrate this graphically, we rendered the database as a HSQC spectrum. Here, each peak within the database is represented by a grey dot. d) In a graphical representation, the peaks to be profiled (blue) are then compared to the database peaks (grey). An inset shows graphically the distance calculation made between one of the peaks to be profiled (blue) and its closest database peak (grey). Other than collecting the NMR spectrum, the steps are conducted computationally with peak lists. The graphical representation in (c), (d) is provided to visualize this process. This procedure can be used to guide isolation by following the most novel peaks. e) This results in a pure compound whose HSQC data can be f) re-subjected to the atomic novelty prioritization.

of more than 10, equal weighting of the dimensions was achieved by dividing the distances in each dimension by the total range of the chemical shifts in the database for that dimension. For the database used here, the ^1H chemical shift range was 10.25 ppm and the ^{13}C range 211.6 ppm. The distance score was then expressed as a percentage, which can be thought of as the fraction of the total chemical shift range. Sorting the scores identified the most unusual chemical shifts,

an indicator of a structurally unique proton and/or carbon atom.

The four peaks A1–A4 (Figure 2) with the largest distance scores are shown in Figure 2c–f. Pleasingly, these peaks corresponded to positions of structural novelty, namely the macrocyclic aryl system at C3 and C16, halogenation at C22 and a uniquely substituted macroaromatic lactone at C14.

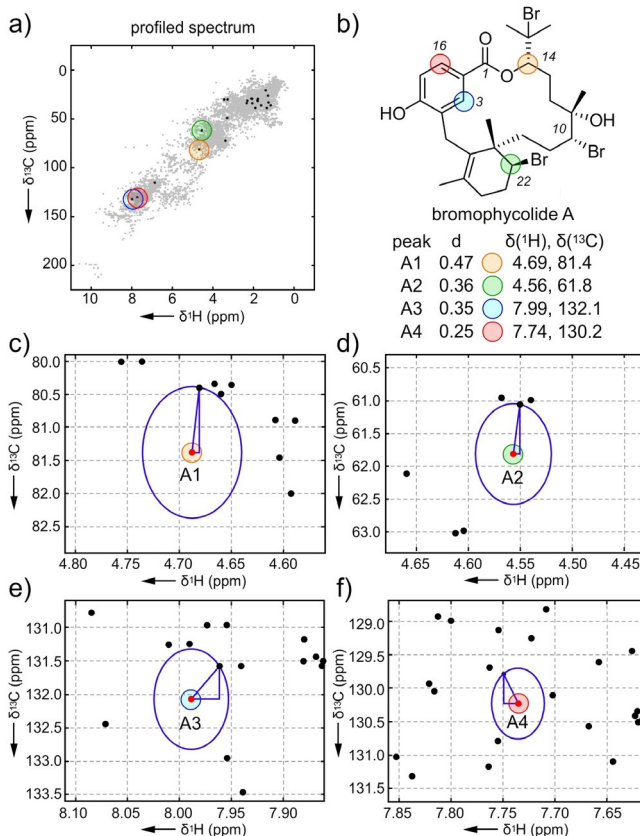


Figure 2. An Euclidian distance algorithm for atomic novelty scoring. a) The profiled spectrum of bromophycolide A with the top peaks A1–A4. b) The structure of bromophycolide A with A1–A4 highlighted. c–f) Graphical rendering of the Euclidian distance evaluations between the bromophycolide A peak (red point) and its closest peak in the database (black points).

Similarly, testing the algorithm with strychnine (Supporting Table S2 and Figures S6,S7), brusatol (Supporting Table S3 and Figures S8,S9), and paclitaxel (Supporting Table S4 and Figures S10,S11) also identified patterns of proton and carbon atom novelty. Like algorithms used for automated MS assignment^[13] or NMR protein structure elucidation,^[14] this scoring system provides critical information with regards to structural assignment as well as identifying regions of novelty.

After testing with pure compounds, we then evaluated the approach as a method to prioritize natural product isolation. Here, the goal was to use this method to rapidly identify extracts that may have natural products with unique structural features. As illustrated in Figure 1, the peaks with the highest novelty (Figure 1d) can be used to guide the isolation of pure materials from the extract (Figure 1e). ^1H - ^{13}C HSQC data collected on the pure material (Figure 1e) can be

reapplied to the atomic prioritization as a means of validation (Figure 1 f).

We began by exploring extracts from a marine microbial strain. We prepared EtOAc extracts from the cultivation of a marine-derived *Streptomyces* sp., strain CNB-982, and collected ^1H - ^{13}C HSQC data on 52 μg of crude material (Supporting Figure S12). The peaks from this spectrum were extracted digitally, and compared against the database (Figure 3 b, Supporting Table S5, Supporting Figure S13). Of the 289 peaks detected, 20 had distance scores greater than or equal to 1%.

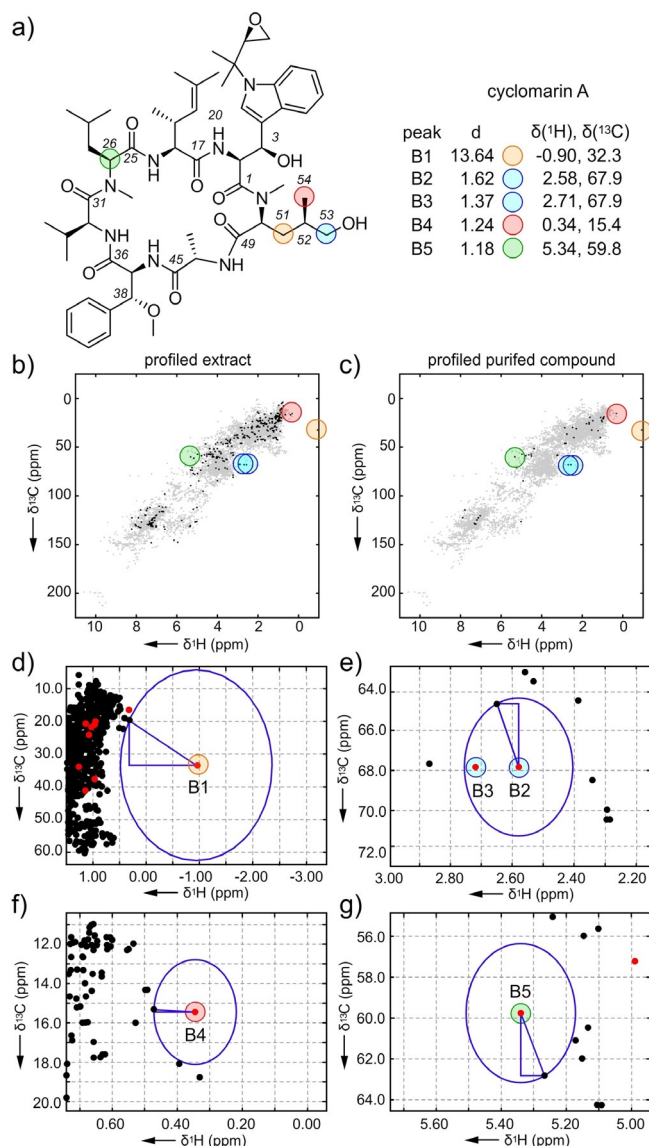


Figure 3. Exploration of the atomic novelty search on the marine microbial extract CNB-982 identifies the cyclic peptide cyclomarins A. a) Structure of cyclomarins A along with the chemical shifts of the top scoring peaks, B1–B5, from its ^1H - ^{13}C HSQC spectrum in CD_3OD . b) Profiled ^1H - ^{13}C HSQC of the CNB-982 extract. c) Profiled ^1H - ^{13}C HSQC of the purified cyclomarins A. d–g) Graphical rendering of the Euclidian distance evaluations between peaks B1–B5 (red points) and closest peaks in the database (black points). Supporting Figure S1 provides chemical structures of the closest database peaks to B1–B5.

We focused our attention on the top peak in the CNB-982 extract, B1 ($\delta_{\text{H}} = -0.91$, $\delta_{\text{C}} = 32.4$, distance = 13.74%, Figure 3, Supporting Table S5). Using peak B1 to guide selection of chromatography fractions, we isolated 1.8 mg of a pure material. The combination of the exact mass for $[M - \text{H}_2\text{O} < M + \text{H}]^+ m/z$ 1025.6062 observed by FAB mass spectrometry along with an expanded NMR data set (Supporting Figures S16–S19 and Supporting Table S7), enabled the pure material to be identified as cyclomarins A ($\text{C}_{56}\text{H}_{80}\text{O}_{11}\text{N}_8$ calcd 1025.6057) (see an expanded discussion in Supporting Figure S3).^[15] By profiling the pure cyclomarins A against the database (Figure 3 c, Supporting Table S6 and Supporting Figure S13) we found the top five scoring peaks (Figure 3 a) were observed in the top 16 found in the crude mixture, indicating that much of the novelty came from one compound, therein validating the approach as shown in Figure 1 f.

Inspection of peaks B1–B5 (Figure 3) was not only useful to guide isolation but provided structural information. As shown in Figure 3 d, we were able to determine that the top peak B1 (orange) was one of the protons at C51, which had an unusual proton chemical shift at -0.90 ppm. Peaks B2 and B3 arose from the geminal protons C53 (blue, Figure 3 e) with uncharacteristic proton shifts at 2.58 and 2.71 ppm. The fourth peak (B4, red, Figure 3 f) arose from the methyl group at C54. Peak B5 (green, Figure 3 g) was assigned to the alpha carbon (C26) of the leucine residue. Interestingly, all of the top four peaks arose from (2*S*,4*R*)-2-amino-5-hydroxy-4-methyl-pentanoic acid, a rare amino acid, therein confirming the method's aptness to find structural novelty. Here, we demonstrated how this tool can be used to prioritize the isolation of materials from a crude extract. In this example, the top peak, B1, provided a clear spectroscopic beacon to guide the isolation process, and comparison with the proximal peaks in the database provided preliminary structural information (Supporting Figure S8).

Next, we tested the system with extracts from a multicellular model using a marine sponge. A ^1H - ^{13}C HSQC spectrum was recorded using a 32 μg sample of a $\text{CH}_2\text{Cl}_2/\text{MeOH}$ extract from the marine sponge *Plakortis halichondrioides* (specimen code IM06-19) (Supporting Figure S20). Digital data extraction followed by profiling against the database (Figure 4 b, Supporting Figure S21) identified several peaks for prioritization (Supporting Table S9).

Selecting peak C1 as an NMR guide (Figure 4 a), we isolated 1.2 mg of pure material (Figure 4 c, Supporting Figures S22, S23 and Table S9) from 1.9 g of crude extract with a HRMS sodiated molecular ion with m/z 371.2193 suggesting a formula of $\text{C}_{21}\text{H}_{32}\text{O}_4\text{Na}$. Collection and evaluation of 1D and 2D NMR data (Supporting Figures S24–S27 and Supporting Table S10) identified the material as a new compound, gracilioether L (Figure 4 a, see an expanded discussion in Supporting Figure S4), a two-carbon chain elongated homologue of gracilioether B.^[16] As shown in Figure 4 d–f, the top three peaks C1–C3 were observed at the C11 *trans*-olefin (C1, Figure 4 d), furanyl-C4 olefin (C2, Figure 4 e) and terminal carbon C18 of an ethyl group at C6 (C3, Figure 4 f). The novelty search then identified the methyl ester C21 (peak C4, Figure 4 g) as the next unique chemical

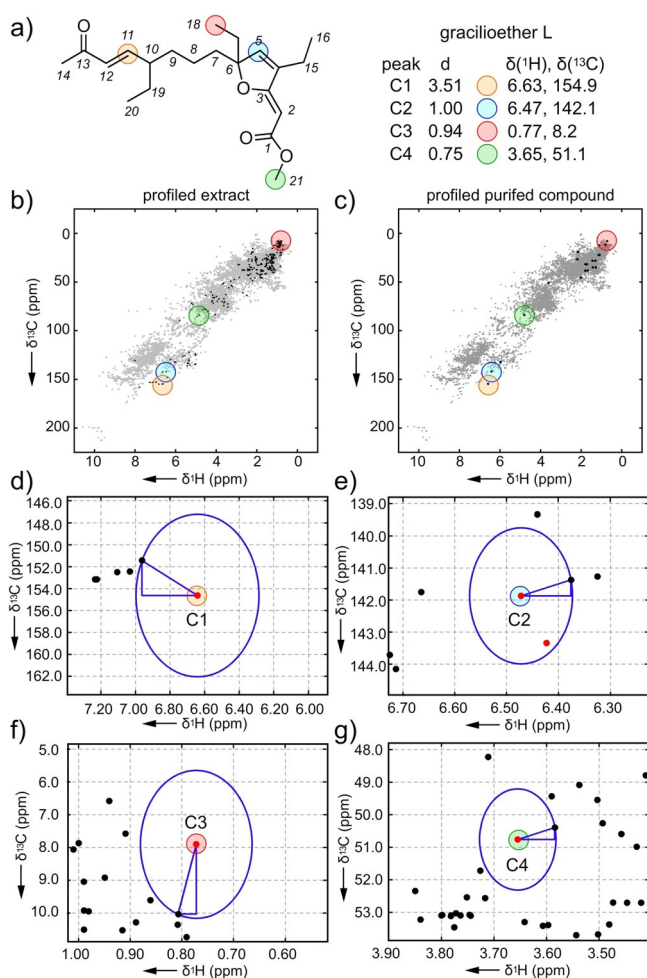


Figure 4. Exploration of the atomic novelty search on the *Plakortis halichondrioides* extract (IM06-19) identifies gracilioether L. a) Structure of gracilioether L along with the top peaks C1–C4 from ^1H - ^{13}C HSQC spectrum in CD_3OD . b) Profiled ^1H - ^{13}C HSQC of the IM06-19 extract. c) Profiled ^1H - ^{13}C HSQC of isolated gracilioether L. d)–g) Graphical rendering of the Euclidian distance evaluations between peaks C1–C4 observed in the spectrum of purified gracilioether L (red points) and its closest peaks in the database (black points). Supporting Figure S2 provides chemical structures of closest database peaks to C1–C4. The stereochemistry at C6 and C10 was not determined.

shift. Interestingly, these atoms are at positions where modifications are commonly observed within this large family of congeners. Further inspection of the entire prioritization (Supporting Table S9) additionally supports this conclusion, suggesting that this tool may also have applications to identify novelty within families of related natural products. Like cyclomarin A, comparing peaks C1–C4 against their proximal database peaks provided a direct structural correlation and was particularly useful in validating peak assignments (Supporting Figures S2, S4).

Overall, we have demonstrated the application of this tool to autonomously evaluate atomic novelty. These studies demonstrate how one can integrate NMR profiling as a means to identify novelty within an extract and isolate unique materials from these extracts using only micrograms of material. While one often views NMR analysis as expensive

and slow, the speed and material requirements, < 6 min using a 35 μL sample, rival those conventionally used by mass spectrometry. The fact that less than 5 min is required for data collection and 1 min for data abstraction and algorithmic processing, clearly demonstrates the practical nature of the process. Here, we found that the prioritizing of peaks was dependent on the quality of the peak list. While noise peaks rarely interfered, we found that two-bond correlations, cross peaks between atoms connected by two bonds instead of one, often scored highly. To flag these peaks we identified 99.99% confidence limits on the database peaks. Peaks falling outside these limits were checked for shared coordinates with two other peaks. Efforts are now underway to expand this tool to flag noise and artifacts, and exclude two-bond correlations.

We foresee this tool expanding to become a multicomponent algorithm that not only incorporates Euclidian distance scoring of ^1H - ^{13}C HSQC spectra but also includes data obtained from such spectra as ^1H - ^1H -COSY, ^1H - ^1H -TOCSY, ^1H - ^1H -NOESY, ^1H - ^1H -ROESY, ^1H - ^{13}C -HMBC, ^1H - ^{15}N -HSQC and ^1H - ^{15}N -HMBC, collected in a single interleaved experiment.^[17] Direct digital extraction of chemical shifts from the raw data^[18] will push the protocol closer to native computational interrogation. This in turn will expedite the growth of the database, which we anticipate will eventually be able to assign the structural features of each carbon and proton within a given molecule. While structural assignment is a key facet of the drug discovery process, the ability to search through molecular data one atom at a time offers a new perspective that can enable this system to operate through a conventional online portal, one that ideally will be united with biosynthetic genome mining tools^[19] such as antiSMASH,^[20] as well as proteomic^[21] and transcriptomic^[22] data. Ultimately, it will be interesting to understand the correlation between structural novelty and its role in inducing novel biological activity, perhaps done best one atom at a time.

Acknowledgements

This program was supported in part by funding from NIH Grant 1SC1GM086271-01A1 awarded to A.D.R., the Xenobe Research Institute to J.J.L. and NIH grant CA044848 to W.F.

Conflict of interest

The authors declare no conflict of interest.

Keywords: algorithms · drug discovery · molecular search engine · natural products · NMR spectroscopy

How to cite: *Angew. Chem. Int. Ed.* **2020**, *59*, 1144–1148
Angew. Chem. **2020**, *132*, 1160–1164

[1] a) S. T. Cardona, C. Selin, A. S. Gislason, *Crit. Rev. Microbiol.* **2015**, *41*, 465–472; b) J. J. La Clair, *Nat. Prod. Rep.* **2010**, *27*, 969–995.

- [2] J. Eder, R. Sedrani, C. Wiesmann, *Nat. Rev. Drug Discovery* **2014**, *13*, 577–587.
- [3] D. E. Leahy, V. Sykora, *Drug Discovery Today Technol.* **2013**, *10*, e437–e441.
- [4] P. Neuzi, S. Giselbrecht, K. Länge, T. J. Huang, A. Manz, *Nat. Rev. Drug Discovery* **2012**, *11*, 620–632.
- [5] R. A. Quinn, L. F. Nothias, O. Vining, M. Meehan, E. Esquenazi, P. C. Dorrestein, *Trends Pharmacol. Sci.* **2017**, *38*, 143–154.
- [6] C. Scheeder, F. Heigwer, M. Boutros, *Curr. Opin. Syst. Biol.* **2018**, *10*, 43–52.
- [7] K. L. Spear, S. P. Brown, *Drug Discovery Today Technol.* **2017**, *23*, 61–66.
- [8] a) W. H. Gerwick, *J. Nat. Prod.* **2017**, *80*, 2583–2588; b) C. Zhang, Y. Idelbayev, N. Roberts, Y. Tao, Y. Nannapaneni, B. M. Duggan, J. Min, E. C. Lin, E. C. Gerwick, G. W. Cottrell, W. H. Gerwick, *Sci. Rep.* **2017**, *7*, 14243; c) F. Olivon, P. M. Allard, A. Koval, D. Righi, G. Genta-Jouve, J. Neyts, C. Apel, C. Pannecouque, L. F. Nothias, X. Cachet, L. Marcourt, F. Roussi, V. L. Katanaev, D. Touboul, J. L. Wolfender, M. Litaudon, *ACS Chem. Biol.* **2017**, *12*, 2644–2651.
- [9] D. S. Wishart, Y. D. D. Feunang, A. Marcu, A. C. C. Guo, K. Liang, R. Vázquez-Fresno, T. Sajed, D. Johnson, C. Li, N. Karu, Z. Sayeeda, E. Lo, N. Assempour, M. Berjanskii, S. Singhal, D. Arndt, Y. Liang, H. Badran, J. Grant, A. Serra-Cayuela, Y. Liu, R. Mandal, V. Neveu, A. Pon, C. Knox, M. Wilson, C. Manach, A. Scalbert, *Nucleic Acids Res.* **2018**, *46*, D1.
- [10] E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. K. Wenger, H. Yao, J. L. Markley, *Nucleic Acids Res.* **2008**, *36*, D402–D408.
- [11] H. E. E. Gottlieb, V. Kotlyar, A. J. Nudelman, *J. Org. Chem.* **1997**, *62*, 7512–7515.
- [12] J. Kubanek, A. C. Prusak, T. W. Snell, R. A. Giese, K. I. Hardcastle, C. R. Fairchild, W. Aalbersberg, C. Raventos-Suarez, M. E. Hay, *Org. Lett.* **2005**, *7*, 5261–5264.
- [13] a) T. R. Sandrin, P. A. Demirev, *Mass Spectrom. Rev.* **2018**, *37*, 321–349; b) T. Kind, H. Tsugawa, T. Cajka, Y. Ma, Z. Lai, S. S. Mehta, G. Wohlgemuth, D. K. Barupal, M. R. Showalter, M. Arita, O. Fiehn, *Mass Spectrom. Rev.* **2018**, *37*, 513–532.
- [14] a) J. M. Würz, S. Kazemi, E. Schmidt, A. Bagaria, P. Güntert, *Arch. Biochem. Biophys.* **2017**, *628*, 24–32; b) L. B. Andreas, T. Le Marchand, K. Jaudzems, G. Pintacuda, *J. Magn. Reson.* **2015**, *253*, 36–49; c) P. Güntert, *Eur. Biophys. J.* **2009**, *38*, 129–143; d) D. E. Zimmerman, G. T. Montelione, *Curr. Opin. Struct. Biol.* **1995**, *5*, 664–673.
- [15] M. K. Renner, Y.-C. Shen, X.-C. Cheng, P. R. Jensen, W. Frankmolle, C. A. Kauffman, W. Fenical, E. Lobkovsky, J. Clardy, *J. Am. Chem. Soc.* **1999**, *121*, 11273–11276.
- [16] a) M. D. Norris, M. V. Perkins, E. J. Sorensen, *Org. Lett.* **2015**, *17*, 668–671; b) C. Festa, G. Lauro, S. De Marino, M. V. D'Auria, M. Monti, C. Casapullo, A. D'Amore, B. Renga, A. Mencarelli, S. Petek, G. Bifulco, S. Fiorucci, A. Zampella, *J. Med. Chem.* **2012**, *55*, 8303–8317; c) R. Ueoka, Y. Nakao, S. Kawatsu, J. Yaegashi, Y. Matsumoto, S. Matsunaga, K. Furihata, R. W. van Soest, N. Fusetani, *J. Org. Chem.* **2009**, *74*, 4203–4207; d) F. Carmen, C. D'Amore, B. Renga, L. Gianlugi, S. De Marino, M. V. D'Auria, G. Bifulco, *Mar. Drugs* **2013**, *11*, 2314–2328.
- [17] E. Kupče, T. D. W. Claridge, *Angew. Chem. Int. Ed.* **2017**, *56*, 11779–11783; *Angew. Chem.* **2017**, *129*, 11941–11945.
- [18] a) K. Krishnamurthy, *Magn. Reson. Chem.* **2013**, *51*, 821–829; b) K. Krishnamurthy, A. M. Sefler, D. Russell, *Magn. Reson. Chem.* **2017**, *55*, 224–232.
- [19] a) P. N. Tran, M. R. Yen, C. Y. Chiang, H. C. Lin, P. Y. Chen, *Appl. Microbiol. Biotechnol.* **2019**, *103*, 3277–3287; b) L. Foulston, *Curr. Opin. Microbiol.* **2019**, *51*, 1–8; c) A. Alonso-Betanzos, V. Bolón-Canedo, *Adv. Exp. Med. Biol.* **2018**, *1065*, 607–626; d) G. Seghal Kiran, P. Ramasamy, S. Sekar, M. Ramu, S. Hassan, A. S. Ninawe, J. Selvin, *Int. J. Biol. Macromol.* **2018**, *112*, 1278–1288.
- [20] K. Blin, T. Wolf, M. G. Chevrette, X. Lu, C. J. Schwalen, S. A. Kautsar, H. G. Suarez Duran, E. L. C. de Los Santos, H. U. Kim, M. Nave, J. S. Dickschat, D. A. Mitchell, E. Shelest, R. Breitling, E. Takano, S. Y. Lee, T. Weber, M. H. Medema, *Nucleic Acids Res.* **2017**, *45*, W36–W41.
- [21] a) J. Griss, *Proteomics* **2016**, *16*, 729–740; b) K. G. Calderón-González, J. Hernández-Monge, M. E. Herrera-Aguirre, J. P. Luna-Arias, *Adv. Exp. Med. Biol.* **2016**, *919*, 281–341; c) B. Turriziani, A. von Kriegsheim, S. R. Pennington, *Adv. Exp. Med. Biol.* **2016**, *919*, 383–396.
- [22] a) G. C. A. Amos, T. Awakawa, R. N. Tuttle, A. C. Letzel, M. C. Kim, Y. Kudo, W. Fenical, B. S. Moore, P. R. Jensen, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E11121–E11130; b) J. S. Bauer, S. Fillinger, K. Förstner, A. Herbig, A. C. Jones, K. Flinspach, C. Sharma, H. Gross, K. Nieselt, A. K. Apel, *RNA Biol.* **2017**, *14*, 1617–1626.

Manuscript received: September 17, 2019

Revised manuscript received: October 24, 2019

Accepted manuscript online: November 6, 2019

Version of record online: December 2, 2019

Supporting Information

Searching for Small Molecules with an Atomic Sort

Brendan M. Duggan, Reiko Cullum, William Fenical, Luis A. Amador, Abimael D. Rodríguez,
and James J. La Clair**

anie_201911862_sm_miscellaneous_information.pdf

Contents:	Page
Materials, methods and experimental procedures	S3-S4
Table S1. Atomic novelty scores for bromophycolide A	S5
Table S2. Atomic novelty scores for strychnine	S6
Table S3. Atomic novelty scores for brusatol	S7
Table S4. Atomic novelty scores for paclitaxel	S8
Table S5. Atomic novelty scores for the CNB-982 extract	S9-15
Table S6. Atomic novelty scores for cyclomarin A	S16-17
Table S7. NMR spectral data for cyclomarin A in CD ₃ OD	S18-19
Table S8. Atomic novelty scores for the IM06-19 extract	S20-24
Table S9. Atomic novelty scores for gracilioether L	S25
Table S10. NMR spectral data for gracilioether L in CD ₃ OD	S26
Figure S1. Peak comparative analyses for cyclomarin A	S27
Figure S2. Peak comparative analyses for gracilioether L	S28
Figure S3. Peak shift analyses for cyclomarin A	S29
Figure S4. Structure elucidation of gracilioether L	S30
Figure S5. ¹ H- ¹³ C HSQC database (Fig. 1c)	S31
Figure S6. ¹ H- ¹³ C HSQC spectrum of bromophycolide A in CD ₃ OD (Fig. 1a)	S32
Figure S7. Profiled spectrum of bromophycolide A (Fig. 1d)	S33
Figure S8. ¹ H- ¹³ C HSQC spectrum of strychnine in CDCl ₃	S34
Figure S9. Profiled spectrum of strychnine	S35
Figure S10. ¹ H- ¹³ C HSQC spectrum of brusatol in CD ₃ OD	S36
Figure S11. Profiled spectrum of brusatol	S37
Figure S12. ¹ H- ¹³ C HSQC spectrum of paclitaxel in CD ₃ OD	S38
Figure S13. Profiled spectrum of paclitaxel	S39
Figure S14. ¹ H- ¹³ C HSQC spectrum of the CNB-982 extract	S40
Figure S15. Profiled spectrum of the CNB-982 extract	S41
Figure S16. ¹ H- ¹³ C HSQC spectrum of cyclomarin A in CD ₃ OD	S42
Figure S17. Profiled spectrum of cyclomarin A	S43
Figure S18. ¹ H NMR spectrum of cyclomarin A in CD ₃ OD	S44
Figure S19. ¹ H- ¹ H gCOSY spectrum of cyclomarin A in CD ₃ OD	S45
Figure S20. ¹ H- ¹³ C HMBC spectrum of cyclomarin A in CD ₃ OD	S46
Figure S21. ¹ H- ¹³ C H2BC spectrum of cyclomarin A in CD ₃ OD	S47
Figure S22. ¹ H NMR spectrum of cyclomarin A in CDCl ₃	S48
Figure S23. ¹ H- ¹³ C HSQC spectrum of the IM06-19 extract	S49
Figure S24. Profiled spectrum of the IM06-19 extract	S50
Figure S25. ¹ H- ¹³ C HSQC spectrum of gracilioether L	S51
Figure S26. Profiled spectrum of gracilioether L	S52
Figure S27. ¹ H NMR spectrum of gracilioether L in CD ₃ OD	S53
Figure S28. ¹ H- ¹ H gCOSY spectrum of gracilioether L in CD ₃ OD	S54
Figure S29. ¹ H- ¹ H NOESY spectrum of gracilioether L in CD ₃ OD	S55
Figure S30. ¹ H- ¹³ C HMBC spectrum of gracilioether L in CD ₃ OD	S56

A. General Experimental Procedures. Chemical reagents were purchased from Acros, Fluka, Sigma-Aldrich, or TCI. Deuterated NMR solvents were purchased from Cambridge Isotope Laboratories. Analytical Thin Layer Chromatography (TLC) was performed on Silica Gel 60 F254 precoated glass plates (EM Sciences). Preparative TLC (pTLC) was conducted on Silica Gel 60 plates (EM Sciences). Visualization was achieved with UV light and/or an appropriate stain (I_2 on SiO_2 , $KMnO_4$, bromocresol green, dinitrophenylhydrazine, ninhydrin, and ceric ammonium molybdate). Flash chromatography was carried out on Geduran Silica Gel 60 (40-63 mesh) from Analtech or EM Biosciences. Analytical HPLC was carried out using a HarmonySecure RP18 (250 x 4.6 mm i.d., 5 μ m) column and was performed on an Agilent 1260 series system controller provided with an Agilent 1260 G1315D photodiode array detector with ChemStation software (Agilent). Optical rotations were measured in $CHCl_3$ with an Autopol IV automatic Polarimeter using a 10 mm microcell. Infrared and UV spectra were recorded using a Bruker Tensor 27 FTIR spectrometer and a Shimadzu UV-2401 PC UV-VIS recording spectrophotometer, respectively. NMR data were acquired with a Bruker DRX-500 spectrometer, Varian VS500 spectrometer, Varian VX500 spectrometer equipped with a Xsens Cold probe or a Bruker Avance III 600 equipped with a 1.7mm cryoprobe. Chemical shifts were referenced using the corresponding solvent signals (δ_H 7.26 and δ_C 77.00 for $CDCl_3$, δ_H 3.31 and δ_C 49.0 for CD_3OD). The NMR spectra were processed using Mestrenova (Mnova 11.0 Mestrelab Research) or TopSpin 3.0 (Bruker Biospin) software.

B. Microbial culturing. Strain CNB-382 was initially cultured in a 1 L volume using a seawater based A1 medium composed of 6 g of starch, 4 g of yeast, 2 g of glucose, 2 g of peptone, and 1 L seawater. After 7 days of cultivation, the broth was extracted with 1 L EtOAc, and the solvent was removed under vacuum to yield 30 mg of organic extract.

C. Purification of cyclomarin A. The organic extract (30 mg) was subjected to size exclusion chromatography (Sephadex LH-20) eluting with MeOH. Fractions containing cyclomarin A, as determined by NMR data, were then collected and fractionated by C-18 reversed-phase semi-prep HPLC (Phenomenex Luna C-18 column, 250 x 10 mm column, 5 μ m; 3 mL/min; 35% for 10 min and increase to 100% MeCN/ H_2O over 60 min; UV detection at 210 nm) to yield cyclomarin A (1.8 mg).

Cyclomarin A. Colorless oil; $[\alpha]_D^{20} = -51.7^\circ$ (c 0.48, $CHCl_3$); UV (MeOH) λ_{max} 222 (22,900), 287 (1000), and 293 (11,200); IR ν (neat) 3400-3300, 3030, 2962, 2928, 2871, 1644, 1512, 1453, and 748 cm^{-1} ; 1D and 2D-NMR data provided in Table S7; HR-FABMS obsd. $[M]^+ m/z$ 1025.6062, calcd. 1025.6057 for $C_{56}H_{80}O_{10}N_9$; EI-MS (% relative intensity) 814 (1), 731 (1), 459 (2), 368 (3), 313 (4), 282 (16), 229 (35), 186 (21), 144 (69), 121 (100), and 116 (33).

D. Animal Material. The sponge *Plakortis halichondrioides* was collected in June 2006 during an underwater expedition near Mona Island, Puerto Rico. The sponge was frozen at $-20^\circ C$, and then lyophilized. A voucher specimen (IM06-19) is stored at the Molecular Sciences Research Center, University of Puerto Rico.

E. Purification of gracilioether L. The dry sponge *Plakortis halichondrioides* (32.4 g dry weight) was carefully cut into small chunks and blended in 1:1 $CHCl_3$ /MeOH (2 x 100 L) at rt. After filtration, the crude extract was *concentrated in vacuo* to yield a brown thick paste (1.9 g). This extract was presented to flash column (38 mm ID x 200 mm height) washing with six rows of 15 test tubes with 150 mL of solvent (10 mL/tube) as given by row 1 (hexanes), row 2 (8:1 hexanes/EtOAc), row3 (4:1 hexanes/EtOAc), row 4 (2:1 hexanes/EtOAc), row 5 (1:1 hexanes/EtOAc) and row 6 (EtOAc). Each tube was explored by TLC analysis and tube 10 in row 5 displayed the targeted peaks. This was then subjected twice to pTLC first with 1:1 hexanes/EtOAc then 1:1 hexanes/acetone to deliver 1.2 mg of gracilioether L.

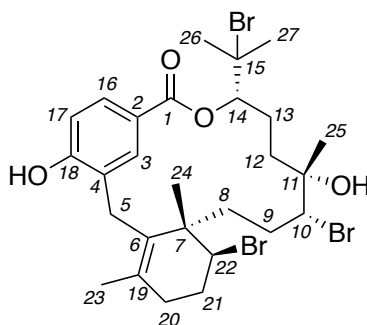
Gracilioether L. Colorless oil; $[\alpha]_D^{20} = -89.2$ (c 0.05, CH₃OH); UV (CH₃OH) λ_{\max} (log ϵ) 286 (2.74), 243 (2.17), 205 (2.42) nm; 1D and 2D-NMR data provided in Table S11; HR-ESI-MS m/z calcd. for C₂₁H₃₂O₄ [M+Na]⁺: m/z 371.2198, found 371.2193.

F. HSQC database. Publicly available ¹H-¹³C HSQC spectra were downloaded from HMDB (www.hmdb.ca) and BMRB (bmrw.wisc.edu/metabolomics). Chemical shift referencing of the spectra was not modified after downloading. Chemical shifts were extracted from the peak lists in the downloaded data using perl and shell scripts and used without further modification. Precision of the chemical shifts varies depending on the data source. Plotting the chemical shift data enabled visual identification of outliers, which were inspected manually using TopSpin. Peaks picked on noise, artifacts, or two bond correlations were removed. Chemical shifts for common solvents were obtained from Tables 1 and 2 in Gottlieb, H. E.; Kotlyar, V.W.; Nudelman, A. *J. Org. Chem.* **1997**, *62*, 7512-7515. In-house ¹H-¹³C HSQC spectra of standards cholesteryl acetate, sucrose, o-dichlorobenzene, ethylbenzene, strychnine and streptomycin and from recent projects (see references: (a) Mehrotra S.; Duggan, B. M. Tello-Aburto, R.; Newar, T. D.; Gerwick, W. H.; Murray, T. F.; Maio, W. A. *J. Nat. Prod.* **2014**, *77*, 2553-2260; (b) Kleigrew, K.; Almaliti, J.; Tian, I. Y.; Kinnel, R. B.; Korobeynikov, A.; Monroe, E. A.; Duggan, B. M.; Di Marzo, V.; Sherman, D. H.; Dorrestein, P. C.; Gerwick, L.; Gerwick, W. H. *J. Nat. Prod.* **2015**, *78*, 1671-1682; (c) Wang, X.; Duggan, B. M., Molinski, T. F. *J. Am. Chem. Soc.* **2015**, *137*, 12343-12351; (d) White, A. R.; Duggan, B. M.; Tsai, S. C.; Vanderwal, C. D. *Org. Lett.* **2016**, *18*, 1124-1127; (e) Carling, C. J.; Olejniczak, J.; Foucault-Collet, A.; Collet, G.; Viger, M. L.; Huu, V. A.; Duggan, B. M.; Almutairi, A. *Chem. Sci.* **2016**, *7*, 2392-2398; (f) Li, Z. R.; Li, J.; Gu, J. P.; Lai, J. Y.; Duggan, B. M.; Zhang, W. P.; Li, Z. L.; Li, Y. X.; Tong, R. B.; Xu, Y.; Lin, D. H.; Moore, B. S.; Qian, P. Y. *Nat. Chem. Biol.* **2016**, *12*, 773-775) were processed with TopSpin 3.6. Spectra were referenced to internal TMS or residual solvent. Peaks were picked automatically at a high threshold with the interpolation type set to “parabolic”. The threshold was then reduced and additional weak peaks were added. Noise, artifacts and two bond correlations were removed. Peaks picked on individual lines in multiplets were combined to a single peak. Peaks from all four data sources were combined to give the final database. A subset of the database that excluded aldehydes was fitted to a straight line to obtain the equation $\delta_C = 16.9989 \delta_H + 1.06803$. Including aldehydes gave a poor fit. Taking every point in the database the difference between the experimental δ_C and that calculated from the fitted line was determined and the standard deviation of these differences calculated. In the entire database only six points were found to fall more than four standard deviations (99.99% confidence limit) from the line.

G. Chemical shift extraction and distance scoring. Samples for profiling were dissolved in 50 μ l of methanol-d₄ and transferred to 1.7mm NMR tubes. ¹H-¹³C HSQC spectra were collected using a 1.7mm microcryoprobe and processed using TopSpin 3.6. Peaks were picked as described above for the database spectra. Peaks more than four standard deviations from the line fitted to the database were flagged. The distance score was calculated as

$$\text{distance score} = \left\{ \left[(\delta_{H,\text{query}} - \delta_{H,\text{db}}) / \text{range } \delta_H \right]^2 + \left[(\delta_{C,\text{query}} - \delta_{C,\text{db}}) / \text{range } \delta_C \right]^2 \right\}^{1/2}$$

where δ_H is the proton chemical shift, δ_C is the carbon chemical shift, the subscript “query” indicates the query peak and “db” a peak in the database, and range is the difference between the minimum and maximum chemical shifts of the indicated nucleus observed in the database. For each query peak the distance score was calculated for all peaks in the database and the minimum value reported.

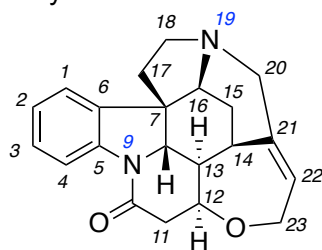
Table S1. Atomic novelty scores for bromophycolide A

peak ¹	position ²	δ_H	δ_C	distance
14	14	4.69	81.4	0.47%
16	22	4.56	61.8	0.36%
1	3	7.99	132.1	0.35%
12	16	7.74	130.2	0.25%
18	5	3.47	30.0	0.25%
19	20	2.07	34.0	0.23%
5	12	1.17	36.1	0.20%
13	17	6.88	115.4	0.20%
3	8	1.93	38.8	0.19%
21	21	2.32	31.4	0.15%
4	9	1.80	29.5	0.14%
23	9	2.08	29.4	0.13%
17	5	3.30	30.0	0.12%
11	24	1.30	26.2	0.12%
7	27	1.81	31.7	0.12%
8	23	1.42	20.9	0.09%
2	12	1.64	36.1	0.08%
10	26	1.83	31.0	0.07%
6	8	1.36	38.8	0.06%
15	10	3.40	72.3	0.03%
9	25	1.28	33.2	0.02%
20	20	2.35	34.0	0.02%
22	13	2.12	29.3	0.02%

¹ Peak is a number that identifies the order that the peaks were abstracted from the raw data.

² Position identifies the atom number given by Kubanek, J.; Prusak, A. C.; Snell, T. W.; Giese, R. A.; Hardcastle, K. I.; Fairchild, C. R.; Aalbersberg, W.; Raventos-Suarez, C.; Hay, M. E. *Org. Lett.* **2005**, *7*, 5261-5264.

Table S2. Atomic novelty scores for strychnine



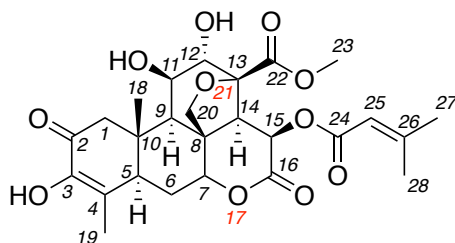
peak ¹	position ²	δ_H	δ_C	distance ³
2	4	8.10	116.3	1.89%
14	13	1.27	48.2	1.01%
3	22	5.90	127.4	0.60%
6	20	2.72	52.6	0.49%
7	20	3.70	52.6	0.43%
19	12	4.28	77.6	0.42%
12	18	3.20	50.4	0.37%
16	17	1.89	42.8	0.37%
21	16	3.95	60.2	0.30%
15	14	3.14	31.6	0.29%
9	15	2.37	27.0	0.27%
20	8	3.85	60.1	0.26%
13	18	2.87	50.4	0.26%
8	11	2.65	42.5	0.17%
17	23	4.07	64.5	0.15%
11	11	3.11	42.4	0.15%
18	23	4.14	64.6	0.13%
1	3	7.26	128.5	0.12%
10	15	1.47	26.8	0.08%
5	1	7.16	122.4	0.07%
4	2	7.10	124.2	0.03%

¹ Peak is a number that identifies the order that the peaks were abstracted from the raw data.

² Position identifies the atom number as given by Verpoorte R. *J. Pharm Sci.* **1980**, 69, 865-866.

³ The strychnine data was removed from the database before calculating these distance scores.

Table S3. Atomic novelty scores for brusatol

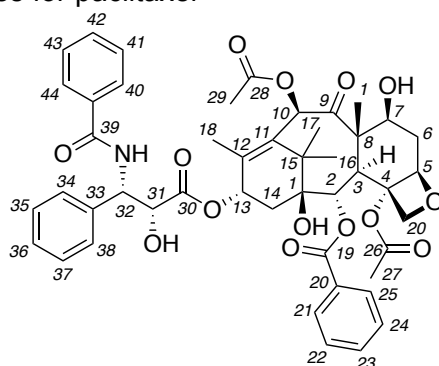


peak ¹	position ²	δ_H	δ_C	distance
2	7	4.91	84.5	0.91%
11	19	1.84	13.1	0.54%
9	28	2.17	20.2	0.53%
16	1	2.84	49.8	0.35%
17	1	2.53	49.8	0.33%
18	6	2.30	29.7	0.30%
15	5	2.97	42.9	0.26%
5	14	3.80	52.8	0.21%
4	11,15	4.17	72.6	0.16%
10	27	1.94	27.2	0.12%
13	20	4.69	74.2	0.08%
14	20	3.71	74.2	0.07%
8	6	1.87	29.8	0.07%
12	18	1.37	15.3	0.07%
3	25	5.68	115.6	0.05%
1	12	4.20	76.2	0.05%
6	23	3.72	52.7	0.03%
19	9	2.21	42.1	0.02%

¹ Peak is a number that identifies the order that the peaks were abstracted from the raw data.

² Position identifies the atom number as given by Hagigaya, Y.; Konda, Y.; Iguchi, M.; Onda, M.; Li, X.; Wu, L.; Li, S.; Sun, X. *J. Nat. Prod.* **1989**, 52, 740-748.

Table S4. Atomic novelty scores for paclitaxel



peak ¹	position ²	δ_H	δ_C	distance
14	10	6.45	76.5	1.76%
12	32	5.64	57.5	1.73%
11	13	6.16	71.9	1.12%
27	5	5.00	85.6	0.65%
19	29	2.17	20.5	0.48%
26	2	5.65	75.9	0.46%
20	27	2.36	22.9	0.37%
28	3	3.83	47.6	0.36%
21	19	1.66	10.1	0.31%
8	21,25	8.11	130.9	0.24%
1	23	7.67	134.3	0.19%
7	34,38	7.49	128.2	0.18%
5	41,43	7.47	129.3	0.18%
3	42	7.55	132.6	0.17%
2	36	7.29	128.7	0.16%
15	27	4.19	77.2	0.15%
22	6	2.47	37.2	0.14%
16	16	1.15	22.0	0.14%
6	35, 37	7.42	129.4	0.13%
9	40, 44	7.86	128.2	0.13%
4	22, 24	7.58	129.4	0.10%
13	31	4.74	74.6	0.07%
18	18	1.91	14.4	0.06%
24	6	1.81	37.2	0.03%
17	17	1.16	26.6	0.03%
25	14	1.96	36.2	0.02%
23	14	2.23	36.3	0.01%
10	7	4.32	72.0	0.01%

¹ Peak is a number that identifies the order that the peaks were abstracted from the raw data.

² Position identifies the atom number as given by Chmurny, G. N.; Hilton, B. D.; Brobst, S.; Look, S. A.; Witherup, K. M.; Beutler, J. A. *J. Nat. Prod.* **1982**, *55*, 414-423.

Table S5. Atomic novelty scores for the CNB-982 extract.¹

peak²	δ_H	δ_C	distance
289	-0.91	32.4	13.74%
95	6.15	145.4	2.53%
94	5.97	147.4	2.36%
2	7.27	147.9	2.01%
270	0.80	3.9	1.91%
192	2.58	67.9	1.60%
146	5.33	89.5	1.51%
191	2.70	67.9	1.42%
7	8.23	134.0	1.36%
6	8.22	122.4	1.28%
259	0.75	7.6	1.28%
60	0.34	15.4	1.28%
194	2.97	67.6	1.15%
274	1.05	42.5	1.14%
260	0.85	5.0	1.13%
287	5.33	59.9	1.13%
271	0.77	7.8	1.11%
137	5.11	126.6	1.10%
203	2.62	25.4	1.00%
58	4.93	84.4	1.00%
288	5.03	126.2	0.98%
275	1.10	42.0	0.97%
122	7.79	114.3	0.95%
9	6.61	98.0	0.90%
280	4.72	95.1	0.89%
149	4.59	64.5	0.73%
40	5.00	57.3	0.71%
139	5.21	113.7	0.71%
181	4.01	45.0	0.71%
151	4.54	57.1	0.69%
43	4.74	66.0	0.69%
145	5.26	70.9	0.66%
272	0.42	16.7	0.66%
169	3.25	78.5	0.64%
134	6.88	110.6	0.63%
91	5.12	77.9	0.60%
190	3.60	46.1	0.59%
193	2.91	67.6	0.56%
143	5.20	76.7	0.56%
142	5.26	73.7	0.55%
153	4.35	54.9	0.54%
170	3.24	80.9	0.50%

183	4.02	51.4	0.50%
97	6.64	131.6	0.49%
261	1.00	6.9	0.48%
158	3.94	54.1	0.47%
176	3.39	43.2	0.47%
35	4.53	59.1	0.47%
140	5.00	76.8	0.46%
157	3.91	60.5	0.46%
215	1.68	43.4	0.45%
232	2.01	23.3	0.45%
22	4.40	51.5	0.43%
132	6.71	111.5	0.43%
217	1.77	41.9	0.43%
279	5.02	111.7	0.43%
163	4.31	57.0	0.43%
133	6.56	111.4	0.42%
67	2.03	23.0	0.41%
104	7.04	131.9	0.41%
286	1.10	29.5	0.40%
167	4.01	84.3	0.40%
99	7.78	132.5	0.39%
29	2.04	13.8	0.39%
150	4.46	68.8	0.39%
196	2.82	45.9	0.38%
79	2.50	29.6	0.37%
189	3.42	44.3	0.36%
184	3.50	48.9	0.36%
101	8.02	130.4	0.36%
236	1.19	25.7	0.35%
266	0.92	33.7	0.35%
156	4.00	59.8	0.34%
36	3.37	45.6	0.33%
144	5.34	73.2	0.33%
177	3.37	41.9	0.33%
100	7.85	132.2	0.33%
27	2.08	22.2	0.33%
126	7.46	112.7	0.32%
221	1.27	40.5	0.32%
218	1.87	42.1	0.32%
209	2.32	31.0	0.31%
206	2.81	32.8	0.31%
257	0.96	17.6	0.31%
123	7.58	119.5	0.31%
20	1.17	28.8	0.31%

108	7.21	123.1	0.30%
49	4.06	59.8	0.29%
242	0.96	21.9	0.29%
70	2.32	29.2	0.29%
175	4.18	51.8	0.29%
72	3.54	45.8	0.28%
45	1.98	15.1	0.28%
119	7.35	126.1	0.27%
141	4.84	73.7	0.27%
103	7.91	131.3	0.27%
131	6.66	116.0	0.27%
222	1.12	40.4	0.26%
44	3.25	58.5	0.26%
246	1.14	17.3	0.26%
121	7.90	122.3	0.26%
201	2.49	39.0	0.25%
148	4.79	58.9	0.25%
211	2.06	39.3	0.25%
179	3.74	46.7	0.25%
262	0.67	16.6	0.25%
224	1.25	34.3	0.24%
216	1.68	42.6	0.24%
164	4.42	56.9	0.24%
31	1.13	24.5	0.24%
83	4.26	60.0	0.24%
52	4.60	46.7	0.23%
17	5.35	130.6	0.23%
278	2.03	42.2	0.23%
250	1.07	15.2	0.23%
225	1.23	32.9	0.23%
200	2.55	38.2	0.23%
178	3.51	43.2	0.23%
248	1.18	18.8	0.22%
159	3.90	54.5	0.22%
74	4.19	59.7	0.22%
16	7.28	130.1	0.22%
63	1.93	23.0	0.22%
208	2.30	28.8	0.22%
65	4.44	57.4	0.22%
256	1.01	15.4	0.21%
118	7.46	129.2	0.21%
82	2.88	29.9	0.21%
284	1.19	28.3	0.21%
10	7.13	122.4	0.21%

76	0.90	19.7	0.21%
152	4.65	57.9	0.21%
155	4.26	59.0	0.20%
239	1.06	22.3	0.20%
102	8.06	129.4	0.20%
129	6.75	115.7	0.20%
13	7.24	129.1	0.20%
277	1.43	38.2	0.20%
75	4.03	61.2	0.19%
172	3.54	62.7	0.19%
138	5.38	131.2	0.19%
96	6.60	134.1	0.19%
42	3.25	55.7	0.19%
204	2.77	36.5	0.19%
213	2.35	29.8	0.19%
77	1.64	19.4	0.19%
90	3.34	56.0	0.19%
173	3.59	62.1	0.19%
105	7.03	130.6	0.19%
240	1.00	20.8	0.18%
214	2.29	23.5	0.18%
212	2.05	28.8	0.18%
258	1.02	21.9	0.18%
241	0.97	20.1	0.18%
66	1.11	19.0	0.18%
1	7.22	131.3	0.18%
41	1.74	28.1	0.17%
86	4.07	61.0	0.17%
46	1.12	18.5	0.17%
81	1.68	24.7	0.17%
11	7.29	127.6	0.17%
54	2.01	28.8	0.17%
220	1.57	32.4	0.17%
26	4.36	57.6	0.17%
254	1.00	18.1	0.17%
182	4.10	46.6	0.17%
116	7.36	128.0	0.16%
180	3.83	45.0	0.16%
68	3.64	71.2	0.16%
264	1.69	16.1	0.16%
273	2.16	44.0	0.16%
263	0.86	9.9	0.16%
276	1.91	39.1	0.16%
38	1.17	25.6	0.15%

78	1.81	25.9	0.15%
223	1.18	40.0	0.15%
135	7.04	120.3	0.15%
154	4.23	57.2	0.15%
186	3.03	37.4	0.15%
28	1.91	19.9	0.14%
47	3.04	39.9	0.14%
19	1.60	26.6	0.14%
53	2.19	36.3	0.14%
109	7.23	124.1	0.14%
269	0.86	10.9	0.14%
249	1.17	19.9	0.14%
197	2.67	40.3	0.14%
107	7.01	119.7	0.14%
185	3.09	37.4	0.14%
219	1.82	42.1	0.14%
265	1.12	32.1	0.14%
5	7.10	125.3	0.14%
252	1.00	23.1	0.14%
238	1.03	23.8	0.14%
188	3.19	40.7	0.13%
136	5.77	125.1	0.13%
244	1.09	18.6	0.13%
187	2.99	40.6	0.13%
117	7.40	129.6	0.13%
127	7.33	112.0	0.13%
199	2.71	39.1	0.13%
8	7.30	128.2	0.13%
243	0.96	23.0	0.13%
147	5.18	61.2	0.13%
165	4.09	71.4	0.13%
71	1.12	19.6	0.12%
24	1.31	20.7	0.12%
115	7.18	127.0	0.12%
62	0.89	22.8	0.12%
233	1.74	25.8	0.12%
106	7.09	122.3	0.12%
21	1.35	20.7	0.12%
285	1.53	28.9	0.12%
64	1.52	39.1	0.11%
268	0.91	11.0	0.11%
33	1.44	25.1	0.11%
15	7.23	127.6	0.11%
3	7.24	130.0	0.11%

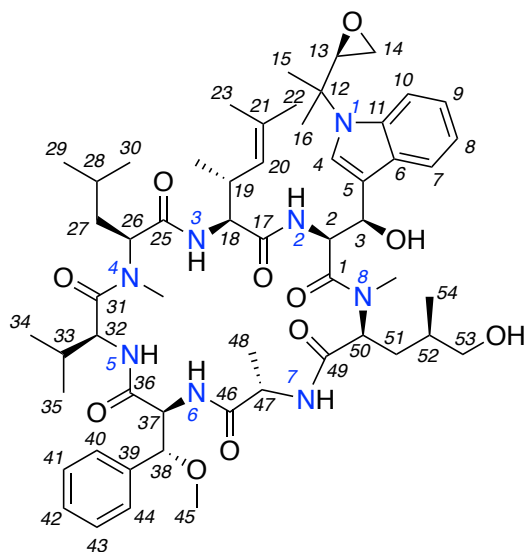
112	7.28	130.9	0.11%
229	1.68	26.2	0.11%
255	1.03	15.3	0.11%
80	4.13	54.3	0.11%
125	7.50	114.7	0.11%
227	1.52	33.3	0.11%
51	1.14	18.9	0.11%
230	1.88	25.5	0.11%
128	6.86	116.1	0.10%
210	1.98	33.3	0.10%
61	2.68	30.3	0.10%
231	1.86	25.0	0.10%
4	7.35	129.2	0.10%
195	2.88	45.9	0.10%
228	1.35	30.4	0.10%
89	3.31	49.0	0.10%
120	7.26	124.8	0.10%
162	3.78	61.2	0.10%
92	3.59	41.9	0.10%
32	1.91	22.3	0.09%
88	3.51	46.1	0.09%
198	2.75	40.1	0.09%
205	2.81	29.7	0.09%
48	2.67	37.8	0.09%
160	3.84	56.0	0.09%
237	1.14	20.9	0.09%
124	7.54	120.0	0.08%
171	3.68	64.4	0.08%
59	0.96	19.1	0.08%
37	2.03	27.9	0.08%
207	2.32	34.7	0.08%
114	7.18	131.0	0.08%
34	1.91	23.3	0.08%
18	1.32	25.1	0.08%
130	6.71	115.9	0.08%
57	1.15	18.4	0.07%
55	2.16	36.7	0.07%
30	2.09	29.1	0.07%
84	1.50	22.8	0.07%
235	1.45	20.6	0.07%
87	0.93	12.3	0.06%
12	7.26	128.9	0.06%
282	1.33	30.0	0.06%
39	1.32	23.4	0.06%

161	3.85	56.7	0.06%
110	7.32	128.1	0.06%
234	1.38	15.4	0.06%
247	1.05	19.0	0.06%
98	7.76	126.7	0.06%
202	2.43	40.1	0.05%
25	2.27	34.8	0.05%
281	4.09	65.4	0.05%
168	3.53	72.3	0.05%
251	0.93	16.4	0.05%
283	1.21	29.1	0.05%
56	1.93	29.2	0.05%
166	3.94	72.6	0.04%
245	1.07	17.8	0.04%
93	1.29	30.5	0.04%
226	1.41	31.5	0.04%
113	7.24	130.8	0.03%
253	0.87	19.9	0.03%
14	7.21	124.6	0.03%
50	1.29	32.8	0.03%
69	1.80	22.5	0.03%
85	3.22	58.1	0.03%
73	3.16	37.9	0.02%
174	3.74	64.0	0.02%
267	0.89	14.3	0.01%
111	7.29	129.2	0.01%
23	1.60	25.9	0.01%

¹ Green shading denotes the top five peaks observed in the pure cyclomarin A spectrum (Table S6). These peaks were identified after the pure material was obtained to emphasize that many of the most highly scoring peaks in the mixture came from a single compound, namely cyclomarin A.

² Peak is a number that identifies the order that the peaks were abstracted from the raw data.

Table S6. Atomic novelty scores for cyclomarlin A



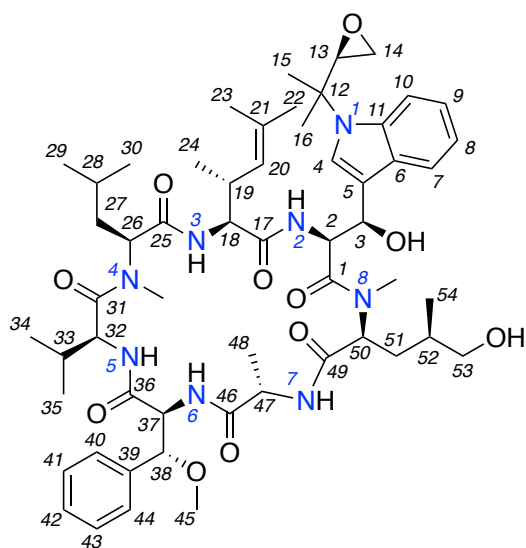
peak ¹	position ²	δ_H	δ_C	distance
41	51 β	-0.92	32.4	13.64%
7	53 β	2.58	67.9	1.62%
34	53 α	2.71	67.9	1.37%
12	54	0.35	15.6	1.24%
29	26	5.34	61.7	1.18%
27	20	5.03	126.3	0.99%
4	38	4.97	84.5	0.99%
25	7	7.79	114.3	0.91%
32	32	4.54	57.1	0.69%
28	3	5.26	70.8	0.64%
6	37	4.99	57.4	0.62%
31	18	4.52	59.1	0.54%
33	47	4.39	51.4	0.43%
20	NMe-8	2.50	27.7	0.39%
36	33	2.31	31.1	0.36%
13	14 β	2.83	46.1	0.36%
24	10	7.91	122.5	0.29%
15	13	3.25	58.5	0.26%
11	29	1.07	22.4	0.23%
30	50	4.79	58.0	0.22%
40	52	1.23	33.1	0.20%
18	NMe-4	2.88	28.1	0.20%
5	2	4.65	58.0	0.19%
1	8	7.13	122.3	0.19%
45	23	1.61	22.7	0.18%
22	15	1.69	24.8	0.17%

2	9	7.04	120.2	0.17%
16	34	0.98	20.1	0.15%
19	22	1.81	26.0	0.15%
38	48	1.00	20.8	0.15%
37	35	1.12	19.5	0.13%
14	19	2.77	36.6	0.13%
17	30	1.04	23.9	0.11%
23	16	1.50	22.7	0.11%
43	41,43	7.29	129.3	0.09%
9	27 β	1.14	40.1	0.08%
26	40,44	7.32	128.0	0.07%
10	27 α	2.44	40.1	0.07%
35	14 α	2.89	46.1	0.06%
8	51 α	1.57	32.4	0.06%
44	28	1.77	24.9	0.05%
39	24	0.96	19.1	0.04%
21	45	3.22	58.1	0.04%
3	4	7.21	124.8	0.03%
42	42	7.26	129.1	0.02%

¹ Peak is a number that identifies the order that the peaks were abstracted from the raw data.

² Position identifies the atom number as given in Renner, M. K.; Shen, Y.-C.; Cheng, X.-C.; Jensen, P. R.; Frankmole, W.; Kaufmann, C. A.; Fenical, W.; Lobkovsky, E.; Clardy, J. *J. Am. Chem. Soc.* **1999**, *121*, 11273-11276.

Table S7. NMR spectral data for cyclomarin A in CD₃OD



cyclomarin A

Position	δ_C , Type	δ_H , mult (<i>J</i> in Hz)	¹ H- ¹ H COSY	¹ H- ¹³ C HMBC ¹
1	171.8, C	-		
2	58.0, CH	4.65, d (9.3)	3	1,3,5,17
3	70.8, CH	5.26, d (9.3)	2	2,4,5,6
4	124.8, CH	7.21 s		2,3,5,6,10w,11
5	114.3, C	-		
6	127.5, C	-		
7	122.5, CH	7.91 d (7.9)	8	6,9
8	122.3, CH	7.13, t (7.5)	7,9	7w,10,11
9	120.2, CH	7.04, t (7.6)	8,10	6,7
10	114.3, CH	7.79, d (8.5)	9	7,8,11
11	135.2, C	-		
12	57.4, C	-		
13	58.5, CH	3.25 dd (2.6, 4.1)	14 α ,14 β	14
14 α	46.1, CH ₂	2.89 m	13,14 β	13
14 β		2.83 dd (2.7, 4.7)	13,14 α	
15	24.8, CH ₃	1.69 s	16	
16	22.7, CH ₃	1.50 s	15	
17	172.3, C	-		
18	59.1, CH	4.52 d (10.0)	19	17,19,20,25
19	36.6, CH	2.77 dt (6.5, 10.2)	18,20,24	17,18,20,21
20	126.3, CH	5.03 d (9.8)	19,22,23	18,19,22,23
21	134.5, C	-		
22	26.0, CH ₃	1.81 d (0.4)	20	20,21,23
23	22.7, CH ₃	1.61 d (0.3)	20	20,21,22
24	19.1, CH ₃	0.96 d (6.7)	19	18,19,21
25	170.0, C	-		

26	61.7,CH	5.34 dd (3.0,11.3)	27 α ,27 β	23,25,27 α ,27 β ,31,NMe-4
27 α	40.1,CH ₂	2.44 ddd (4.0, 11.3, 13.1)	26,27 β ,28	26,28,29,30
27 β		1.14 m	26,27 α ,28	26,28,29,30
28	24.9,CH	1.77 m	27 α ,27 β ,29,30	26,27 α ,27 β ,29,30
29	22.4,CH ₃	1.07 d (6.6)	28	23,27,30
30	23.9,CH ₃	1.04 d (6.6)	28	23,27,29
31	173.2,C	-		
32	57.1,CH	4.54 d (9.5)	33	31,33,34,36
33	31.1,CH	2.31 ddt (13.3, 10.1, 6.7)	32,34,35	32,35
34	20.1,CH ₃	0.98 (d, 6.7)	33	32,33,35
35	19.5,CH ₃	1.12 (d, 6.6)	33	32,33,34
36	171.6,C	-		
37	57.4,CH	4.99 d (3.4)	38	36,38,45 ^w ,46
38	84.5,CH	4.97 d (3.6)	37	36,37,39,40,44
39	137.5	-		
40	128.0,CH	7.32 d (6.5)	41	38,44
41	129.3,CH	7.29 t (6.6)	40,42	39,40
42	129.1,CH	7.26 t (6.8)	41,43	39,40,44
43	129.3,CH	7.29 t (6.6)	42,44	39,44
44	128.0,CH	7.32 d (6.5)	43	38,40
45	58.1, CH ₃	3.22 s		38
46	172.7,C	-		
47	51.4, CH	4.39 q (7.2)	48	46,48,49
48	20.8, CH ₃	1.00 d (7.2)	47	46,47
49	170.7,C	-		
50	58.0,CH	4.79 dd (2.8, 12.4)	51 α ,51 β	1,49,52,NMe-8
51 α	32.4,CH ₂	1.57 m	50,51 β ,52	49,50
51 β		-0.92 m	50,51 α ,52	49,50,NMe-8
52	33.1,CH	1.23 m	51 α ,51 β ,54,53	54
53 α	67.9,CH ₂	2.71 dd (6.1, 10.6)	52,53 β	52,54
53 β		2.58 dd (7.3, 10.6)	52,53 α	52,54
54	15.6,CH ₃	0.35 d (6.7)	52	51,52,54
NMe-4	28.1,CH ₃	2.88 s		26,31
NMe-8	27.7,CH ₃	2.50 s		1,3 ^w ,50

¹ w denotes a weak cross peak

Table S8. Tabulation of the atomic novelty scores for the IM06-19 extract¹

peak²	δ_H	δ_C	distance
159	6.43	148.3	3.30%
1	6.68	154.9	3.13%
153	5.03	134.5	1.62%
3	5.08	134.7	1.19%
152	6.48	142.2	1.13%
161	5.55	125.7	1.09%
146	0.80	8.1	0.93%
169	5.32	135.2	0.87%
86	0.84	29.5	0.85%
165	5.36	133.6	0.79%
13	3.66	51.1	0.75%
141	0.85	8.0	0.66%
5	4.58	78.1	0.63%
163	5.04	125.4	0.63%
4	5.42	133.7	0.61%
151	7.13	153.2	0.53%
68	1.75	39.1	0.47%
95	0.83	9.0	0.43%
8	4.85	83.9	0.42%
104	0.96	8.1	0.40%
160	6.24	129.8	0.39%
157	5.16	73.0	0.38%
66	0.87	33.9	0.38%
154	4.51	78.7	0.38%
113	1.18	21.9	0.37%
166	5.35	130.9	0.36%
176	1.86	45.6	0.35%
122	1.23	34.6	0.34%
188	1.63	34.1	0.34%
175	1.75	45.8	0.33%
186	1.89	43.8	0.33%
124	1.57	33.9	0.33%
144	1.18	12.6	0.32%
100	0.96	13.1	0.31%
14	2.13	45.9	0.31%
69	0.96	10.5	0.30%
106	1.09	12.3	0.30%
2	6.06	132.4	0.29%
136	2.27	19.5	0.29%
21	1.26	35.8	0.29%
61	1.27	36.7	0.28%
36	0.96	9.7	0.28%

35	1.03	12.2	0.28%
116	0.88	9.2	0.27%
102	1.22	40.1	0.26%
96	0.94	26.2	0.26%
182	1.54	41.3	0.26%
184	1.83	42.0	0.26%
173	3.70	52.2	0.26%
183	1.62	41.9	0.25%
67	0.83	22.1	0.23%
105	2.08	24.3	0.23%
115	1.21	22.3	0.23%
50	0.82	9.7	0.23%
17	2.00	37.4	0.23%
148	0.98	14.7	0.23%
180	2.48	38.3	0.22%
125	1.19	26.2	0.22%
46	1.12	38.2	0.22%
158	4.53	63.7	0.22%
37	1.69	37.9	0.22%
19	0.97	20.6	0.22%
133	1.19	29.5	0.21%
30	1.58	31.9	0.21%
71	1.24	36.3	0.21%
89	0.86	13.1	0.20%
83	1.53	25.7	0.20%
81	1.29	35.0	0.20%
70	1.01	10.3	0.20%
88	1.45	37.3	0.20%
15	1.80	45.8	0.20%
84	1.54	32.9	0.19%
187	1.69	38.8	0.19%
177	2.31	41.7	0.19%
123	2.92	32.3	0.19%
192	1.60	30.8	0.18%
194	1.38	22.0	0.18%
57	2.33	35.1	0.18%
7	4.07	83.5	0.18%
185	1.84	43.1	0.17%
49	1.03	19.9	0.17%
193	1.26	25.2	0.17%
178	2.24	37.3	0.17%
120	0.99	13.7	0.17%
12	2.89	45.4	0.17%
168	5.37	131.3	0.16%

80	2.30	21.2	0.16%
107	1.21	24.1	0.15%
98	1.80	39.0	0.15%
9	4.42	80.2	0.15%
72	0.88	21.4	0.15%
47	1.30	23.2	0.15%
78	1.84	38.8	0.15%
170	3.64	67.4	0.15%
20	1.51	26.6	0.15%
118	0.87	20.2	0.14%
42	1.47	35.6	0.14%
145	2.02	26.7	0.14%
56	1.00	9.1	0.14%
138	2.26	26.9	0.14%
55	0.89	22.1	0.14%
65	1.49	27.4	0.14%
121	1.34	36.7	0.14%
44	1.21	23.5	0.13%
16	1.50	36.1	0.13%
11	3.23	54.7	0.13%
156	4.24	70.3	0.13%
147	0.94	11.4	0.13%
140	0.88	23.0	0.13%
60	2.09	28.4	0.13%
26	0.74	12.6	0.13%
94	1.25	40.1	0.13%
93	1.81	37.8	0.12%
128	1.26	30.1	0.11%
181	2.69	37.4	0.11%
63	1.17	40.4	0.11%
45	0.83	13.0	0.11%
23	1.48	29.3	0.11%
129	1.81	31.1	0.11%
73	1.93	32.9	0.11%
58	2.04	37.5	0.11%
103	1.38	36.8	0.11%
126	1.26	26.2	0.11%
162	5.05	123.9	0.11%
39	1.68	26.1	0.11%
111	1.38	28.4	0.11%
77	1.00	17.2	0.10%
76	1.33	21.3	0.10%
134	1.27	33.9	0.10%
92	1.78	38.1	0.10%

40	1.02	11.3	0.10%
191	1.74	31.0	0.10%
32	1.05	11.3	0.10%
38	1.34	35.1	0.10%
41	1.44	26.7	0.10%
79	1.26	29.1	0.10%
6	4.21	69.4	0.10%
196	0.91	12.2	0.10%
108	0.93	13.3	0.10%
33	0.81	19.7	0.10%
85	1.14	22.4	0.10%
53	2.09	27.5	0.10%
52	1.54	30.8	0.09%
132	1.31	27.0	0.09%
135	1.38	29.5	0.09%
31	1.19	28.8	0.09%
127	1.79	32.0	0.08%
117	1.31	28.3	0.08%
64	1.30	38.3	0.08%
28	1.21	25.2	0.08%
190	2.03	25.8	0.08%
109	1.46	30.7	0.08%
43	1.25	32.9	0.08%
143	0.91	14.5	0.08%
59	1.30	21.7	0.08%
22	1.53	35.5	0.07%
27	1.10	25.3	0.07%
179	2.39	38.1	0.07%
155	4.17	71.7	0.07%
142	0.86	11.2	0.07%
195	1.04	13.1	0.07%
90	2.04	28.2	0.07%
139	2.21	19.5	0.06%
74	1.74	27.5	0.06%
164	5.44	136.0	0.06%
167	5.36	130.1	0.06%
114	1.24	22.0	0.06%
34	2.47	36.6	0.06%
112	1.60	26.1	0.06%
131	1.84	32.0	0.05%
110	1.01	8.1	0.05%
119	1.20	36.9	0.05%
25	2.41	37.3	0.05%
62	1.31	25.2	0.05%

48	1.19	30.5	0.05%
91	1.29	33.1	0.05%
97	1.16	36.4	0.05%
130	1.30	24.2	0.05%
82	0.95	9.0	0.04%
137	2.16	35.8	0.04%
149	0.84	12.2	0.03%
171	3.46	72.6	0.03%
29	1.03	21.5	0.03%
189	2.40	32.3	0.03%
172	4.27	60.5	0.03%
87	0.89	10.3	0.03%
10	3.13	47.5	0.03%
24	1.26	32.0	0.03%
174	2.47	42.9	0.03%
150	1.30	30.8	0.02%
18	0.81	10.3	0.02%
54	1.38	35.6	0.02%
75	1.73	38.2	0.02%
101	1.54	28.4	0.01%
51	1.32	22.3	0.01%
99	2.28	35.0	0.01%

¹ Green shading denotes the top four peaks observed in the pure gracilioether L spectrum (Table S9). These peaks were identified after the pure material was obtained to emphasize that many of the high scoring peaks in the mixture came from a single compound, namely gracilioether L (Table S9).

² Peak is a number that identifies the order that the peaks were abstracted from the raw data.

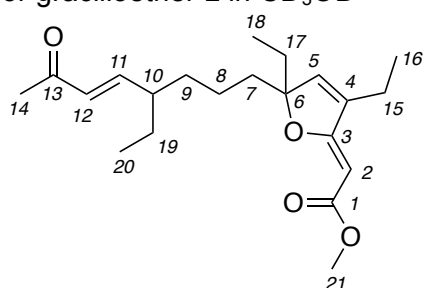
Table S9. Atomic novelty scores for gracilioether L

peak ¹	position ²	δ_H	δ_C	distance
16	11	6.63	154.9	3.51%
1	5	6.47	142.1	1.00%
21	18	0.77	8.2	0.94%
3	21	3.65	51.1	0.75%
18	16	1.16	12.2	0.47%
2	2	4.85	84.0	0.40%
9	10	2.07	45.9	0.36%
4	9 α	1.46	35.2	0.31%
14	8 α	1.18	22.4	0.22%
12	14	2.24	26.9	0.18%
10	17 β	1.78	32.0	0.17%
20	15	2.20	19.5	0.17%
19	9 β	1.31	35.2	0.16%
5	19 α	1.50	28.3	0.14%
6	7 β	1.73	38.5	0.14%
15	8 β	1.15	22.4	0.14%
11	20	0.85	12.0	0.13%
17	12	6.02	132.2	0.09%
13	19 β	1.35	28.3	0.09%
8	17 α	1.85	32.0	0.08%
7	7 α	1.83	38.5	0.04%

¹ Peak is a number that identifies the order that the peaks were abstracted from the raw data.

² Position identifies the atom number based on those given in Ueoka, R.; Nakao, Y.; Kawatsu, S.; Yaegashi, J.; Matsumoto, Y.; Matsunaga, S.; Furihata, K.; van Soest, R. W. M.; Fuestani, N. *J. Org. Chem.* **2009**, *74*, 4203-4207.

Table S10. NMR spectral data for gracilioether L in CD₃OD



gracilioether L

Position	δ_c , Type	δ_H , mult (<i>J</i> in Hz)	¹ H- ¹ H COSY	¹ H- ¹³ C HMBC
1	169.3, C	-		
2	84.0, CH	4.85, s	5w	3,4
3	174.2, C	-		
4	141.4, C	-		
5	142.1, CH	6.47, t (1.7)	2w, 15	1w, 3,4,6,15
6	99.7, C	-		
7 α	38.5, CH ₂	1.83, m	7 β , 8 β	5w, 6w, 8,9,17
7 β		1.73, m	7 α , 8 α	5w, 6w, 8,9,17w
8 α	22.4, CH ₂	1.18, m	7 α , 7 β , 8 β , 9 β	6,7,9,10
8 β		1.15, m	7 α , 7 β , 8 α , 9 α	6,7,9,10
9 α	35.2, CH ₂	1.46, m	8 β , 10	7,8,10,11,19w
9 β		1.31, m	8 α , 10	7,8,10,11,19w
10	45.9, CH	2.07, m	9 α , 9 β , 11, 19 α , 19 β	8,9,11,12,19,20
11	154.9, CH	6.63, dd (16.0, 9.2)	10,12	9,10,13,19
12	132.2, CH	6.02, d (0.8, 16.0)	11,14w	9w,10,13,14,19w
13	201.3, C	-		
14	26.9, CH ₃	2.24, s	11w	11,12,13
15	19.5, CH ₂	2.20, qt (7.4, 1.7)	5,16	3,4,5,16
16	12.2, CH ₃	1.16, t (7.4)	15	4,15
17 α	32.0, CH ₂	1.85, m	18	5,6,7,18
17 β		1.78, q (7.2)	18	5,6,7w,18
18	8.2, CH ₃	0.77, t (7.4)	17 α , 17 β	6,17
19 α	28.3, CH ₂	1.50, m	10, 19 β , 20	9,10,11,20
19 β		1.35, m	10, 19 α , 20	9,10,11,20
20	12.0, CH ₃	0.85, t (7.5)	19 α , 19 β	10,19
21	51.1, CH ₃	3.65, s		1

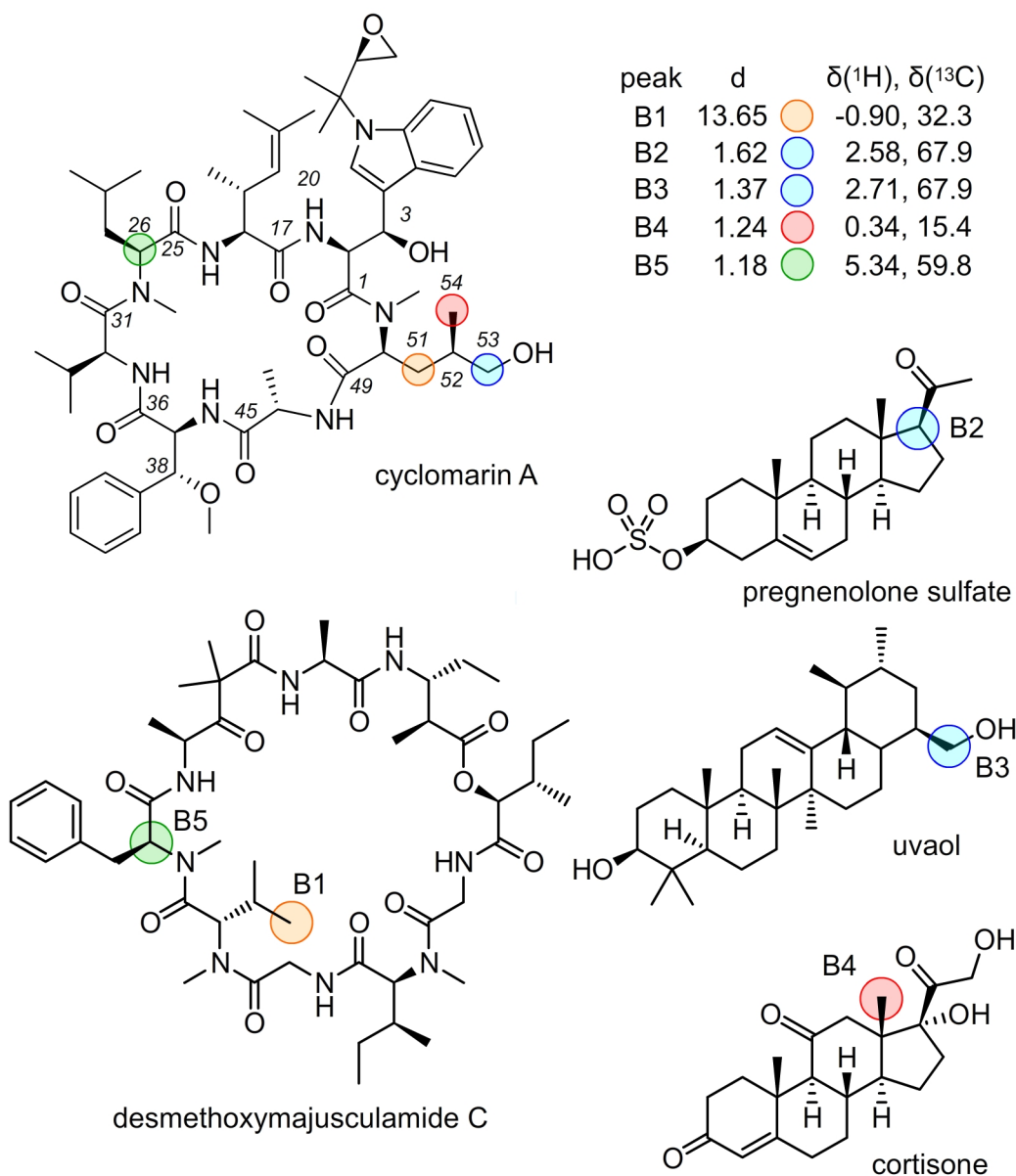


Figure S1. Peak comparative analyses for cyclomarins A. The structures producing the top prioritized peaks in cyclomarins A are compared with the compounds producing their closest peak within the database as shown in Fig. 2 of the manuscript. The closest peak to B1 was from a side chain methyl of an *N*-methylvaline residue in desmethoxymajusculamide C. Interestingly, both the methyl in desmethoxymajusculamide C and methylene within cyclomarins A were directly proximal to an *N*-methylated amide and shared a comparable environment. The second and third peaks B2 and B3 were close to peaks in pregnenolone sulfate and uvaol, respectively. While the former was not a good fit, the latter demonstrated a very similar chemical shift environment with the C53 in cyclomarins A sharing a common motif (blue circle) with uvaol. Peaks B4 and B5 showed also similar correlations with methyl groups (cyclomarins A *versus* cortisone) and α -protons on an amino acid within a cyclic peptide or despeptide. Remarkably, this small database was able to return structures for each peak with a reliable counterpart.

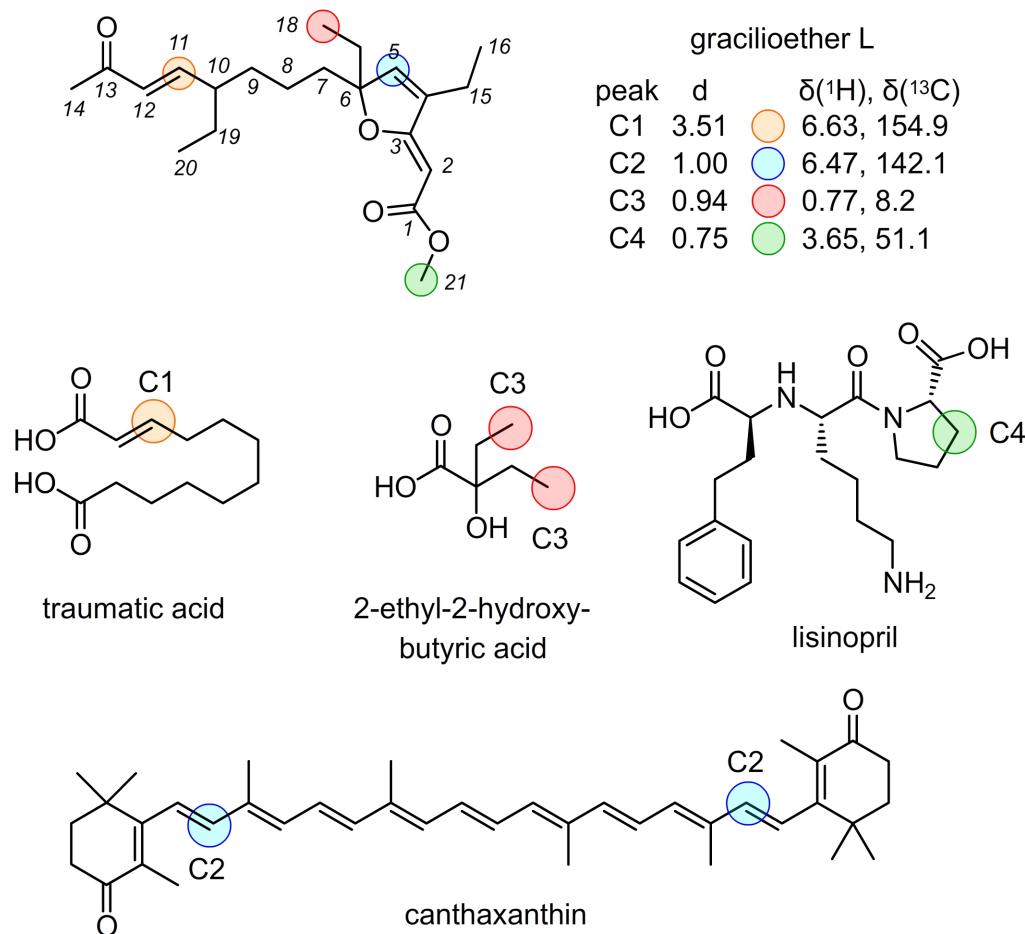


Figure S2. Peak comparative analyses for gracilioether L. The closest peak in the database to C1 was the β -position of traumatic acid, a functionality that directly correlated with the β -position of the *trans*-enone in gracilioether L. Peak C2 again shared remarkable similarity to its most proximal peak. Here, C2, a δ proton within an unconjugated ester, was proximal to an olefin peak that correlated to δ proton of an unsaturated ketone within canthaxanthin. The closest peak to C3 also contained a high degree of similarity being contained within an ethyl group proximal to hydrogen bond donating oxygen atom and a carbonyl. Peak C4 did not show comparable peaks due to a lack of methyl esters in our database. The latter point suggests the needs to develop an intelligent database, one that offers clear predictions for each peak and enables one to generate structural assignments. Eventually, this would need to be paired with correlation data from 2-bond and 3-bond couplings from spectra such as ^1H - ^{13}C HMBC spectra to automate *de novo* structural assignment.

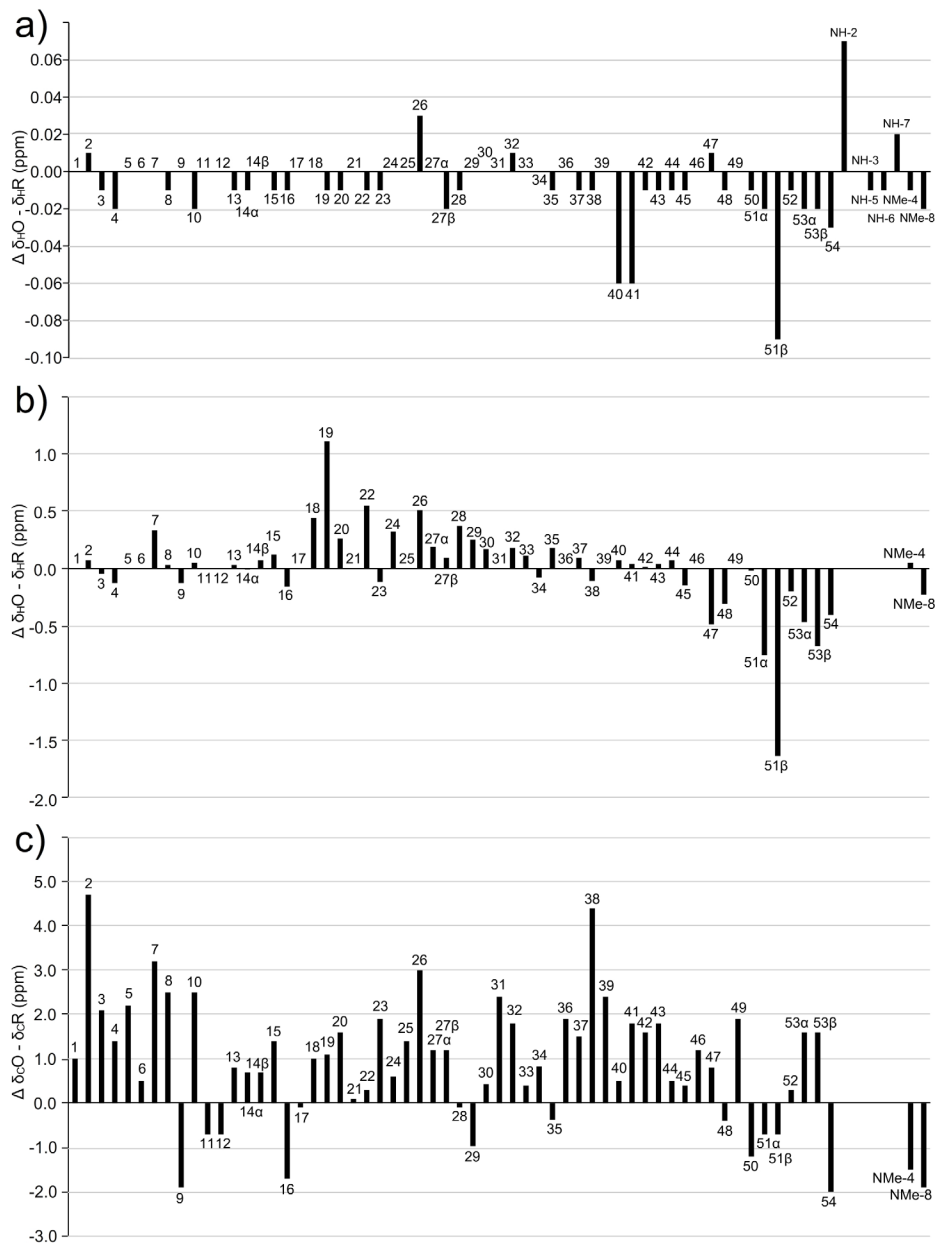


Figure S3. Peak shift analyses for cyclomar A. Our NMR studies were conducted in CD_3OD while the reported characterization of isolated cyclomar A [M. K. Renner, Y.-C. Shen, X.-C. Cheng, P. R. Jensen, W. Frankmoeller, C. A. Kauffman, W. Fenical, E. Lobkovsky, J. Clardy, *J. Am. Chem. Soc.* **1999**, *121*, 112736] (isolation data) and material produced *via* total synthesis [P. Barbie, U. Kazmaier, *Org. Lett.* **2016**, *18*, 204-7] was conducted in CDCl_3 . As noted in these publications the chemical shifts and coupling constants of cyclomar A modulate according to concentration and amounts of water present. **a)** Comparison of the proton chemical shifts from cyclomar A isolated herein in CDCl_3 against the isolation data in CDCl_3 . **b)** Comparison of the proton chemical shifts from cyclomar A isolated herein in CD_3OD against the isolation data in CDCl_3 . **c)** Comparison of the carbon chemical shifts from cyclomar A isolated herein in CD_3OD against the isolation data in CDCl_3 . Both proton in b) and carbon c) shift perturbations illustrate the complexities associated with the use of different solvents.

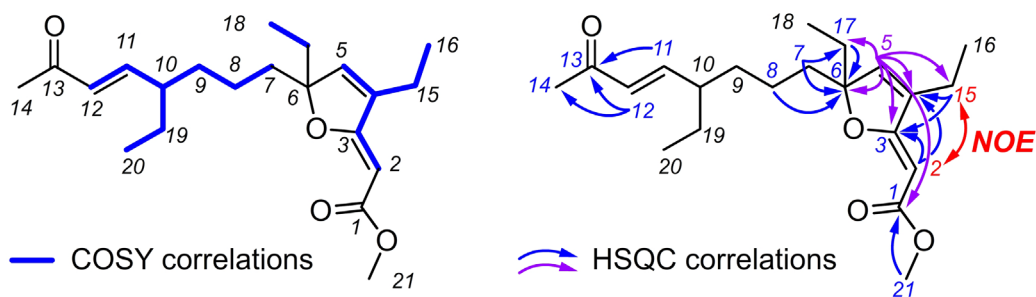


Figure S4. Structure elucidation of gracilioether L. Gracilioether L was isolated as a wax with molecular formula $C_{21}H_{32}O_4$ inferred by high resolution ESI HRMS ($M+Na^+$, m/z 371.2193 calcd. 371.2198). The 1H - ^{13}C HSQC indicated the presence of 3 ethyl groups, 3 contiguous methylenes, 1 acyl group, a carbomethoxyl, a sp^3 methine, an unusually high field sp^2 methine, a 1,2, disubstituted double bond, and a trisubstituted double bond. 1H - 1H COSY correlations defined a spin system consisting of the disubstituted double bond (H11,H12), the methine (H10), one of the ethyl groups (H19,H20) and the three adjoining methylenes (H9,H8,H7). 1H - ^{13}C HMBC correlations placed the acetyl group (H14 and C13) to the other side of the disubstituted double bond (H12). 1H - ^{13}C HMBC correlations to the trisubstituted double bond (H5) identified three quaternary carbons (C3,C4,C6) that were also linked to the remaining two ethyl groups and the highly upfield sp^2 methine. A literature search for similar chemical shifts suggested the trisubstituted cyclic ether framework. Initially, we envisioned C2 (δ_C 84.0) as an oxygenated methine but subsequently found it to be a part of a conspicuous furanylidene motif. The unusual chemical shifts of this group have been remarked upon several times [R. J. Capon, S. Singh, A. Sadaquat, S. Subramaniam, *Aust. J. Chem.* **2005**, *58*, 18-20 or D. B. Stierle, D. J. Faulkner, *J. Org. Chem.*, **1980**, *45*, 3396-3401]. HMBC correlations positioned the remaining ethyl groups on the five membered ring and located the methoxyl as part of a methyl ester attached to the Δ^2 trisubstituted double bond. The C11-C12 double bond was assigned the trans configuration based on the 16.0 Hz coupling between H11 and H12. NOEs from H2 to H15 and H16 defined the Z geometry about the C2-C3 double bond.

Note: The stereochemistry at C6 and C10 could not be determined by NMR methods and would likely require validation by chemical synthesis.

Figure S5. ^1H - ^{13}C HSQC database

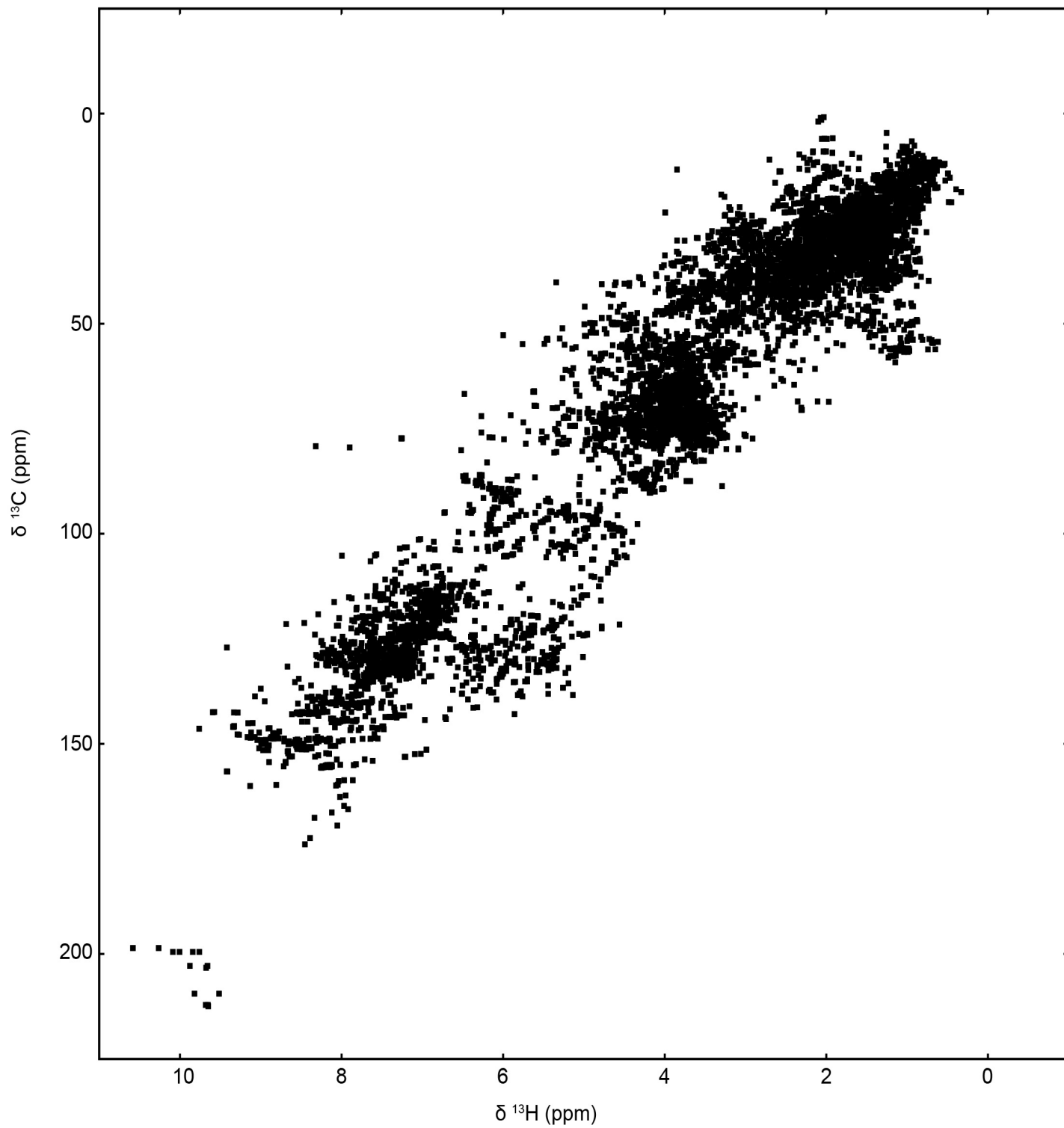


Figure S6. ^1H - ^{13}C HSQC (600 MHz) spectrum of bromophycolide A in CD_3OD

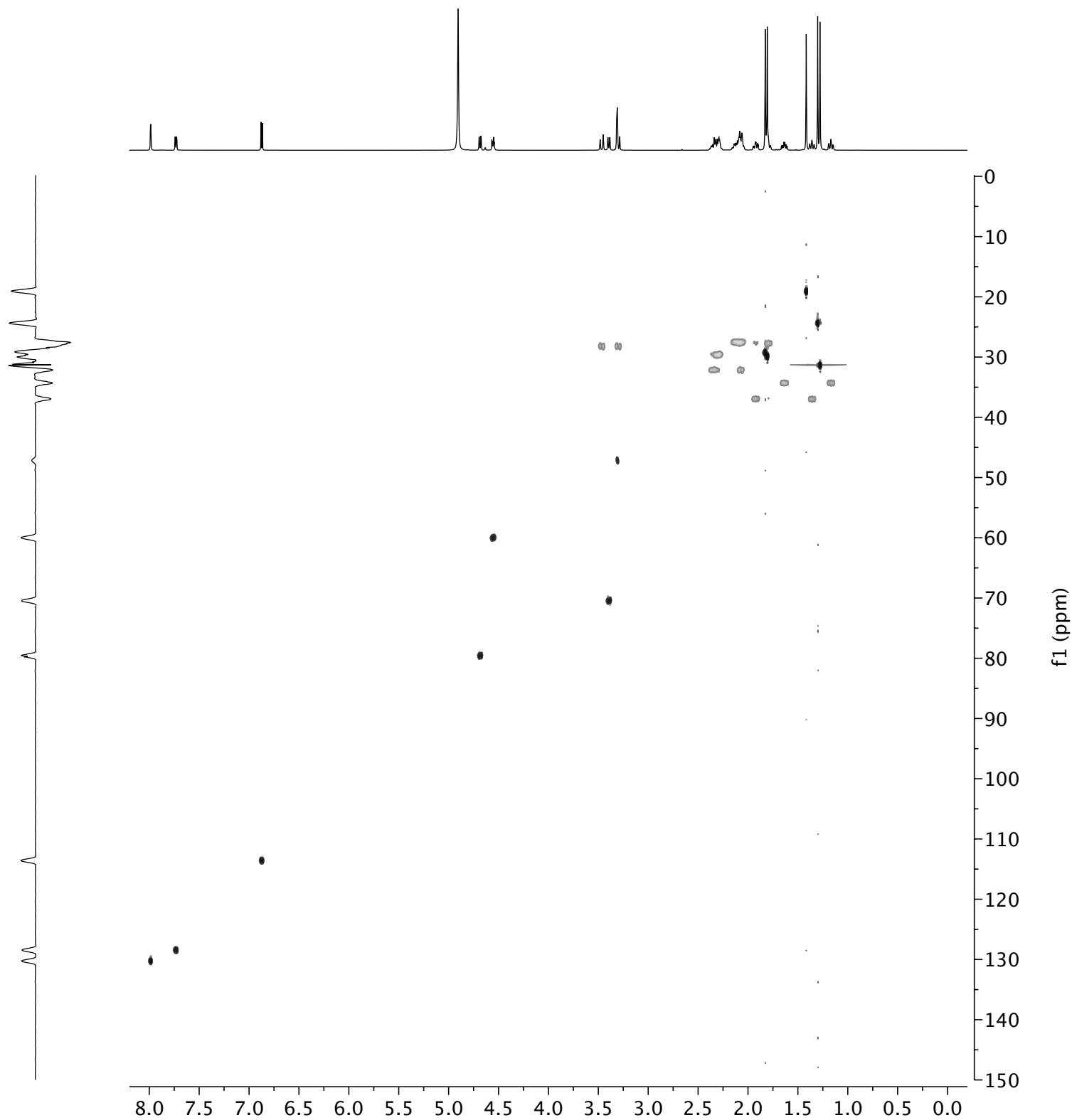


Figure S7. Profiled ^1H - ^{13}C HSQC spectrum of bromophycolide A

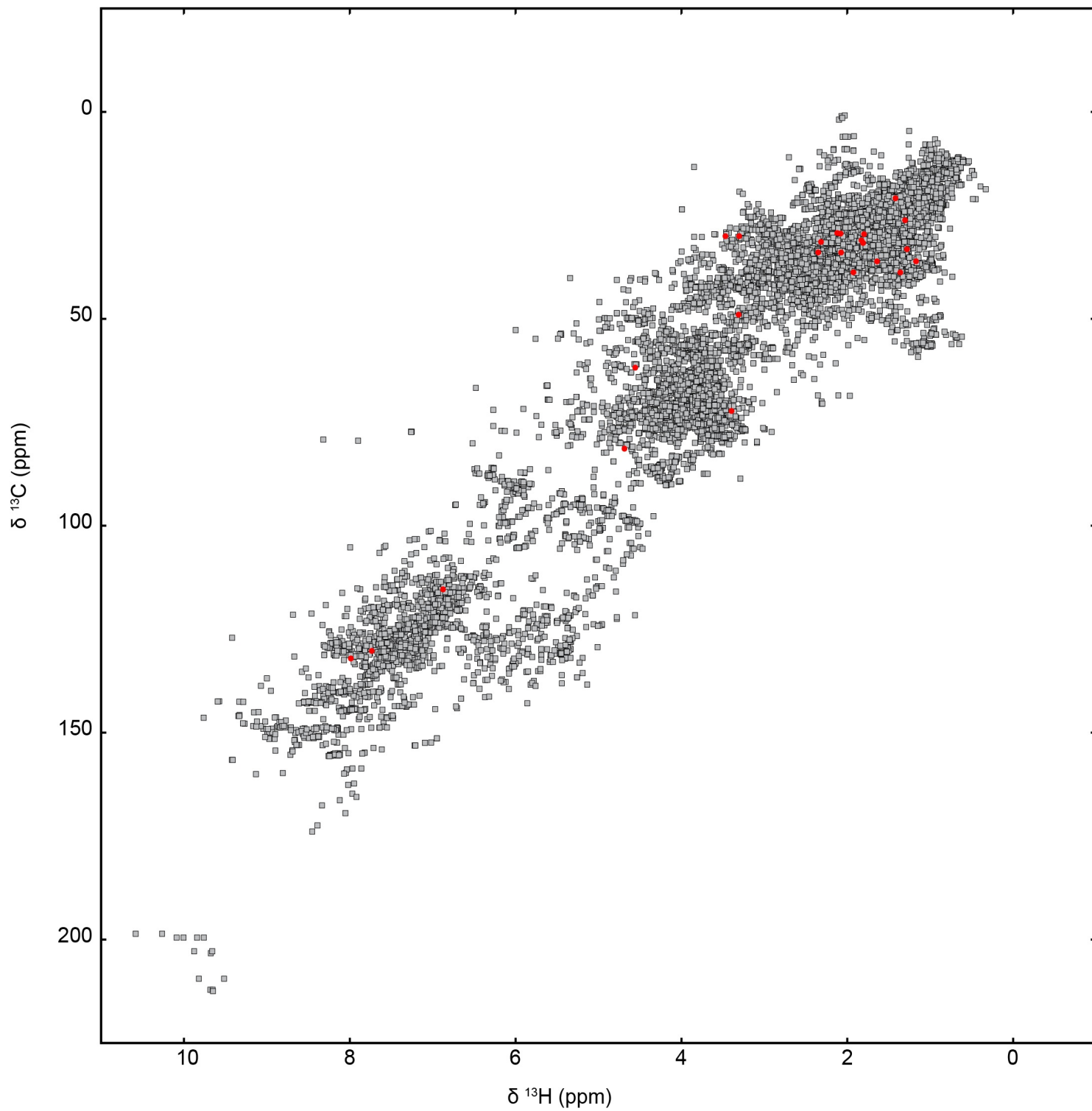


Figure S8. ^1H - ^{13}C HSQC(600 MHz) spectrum of strychnine in CD_3OD

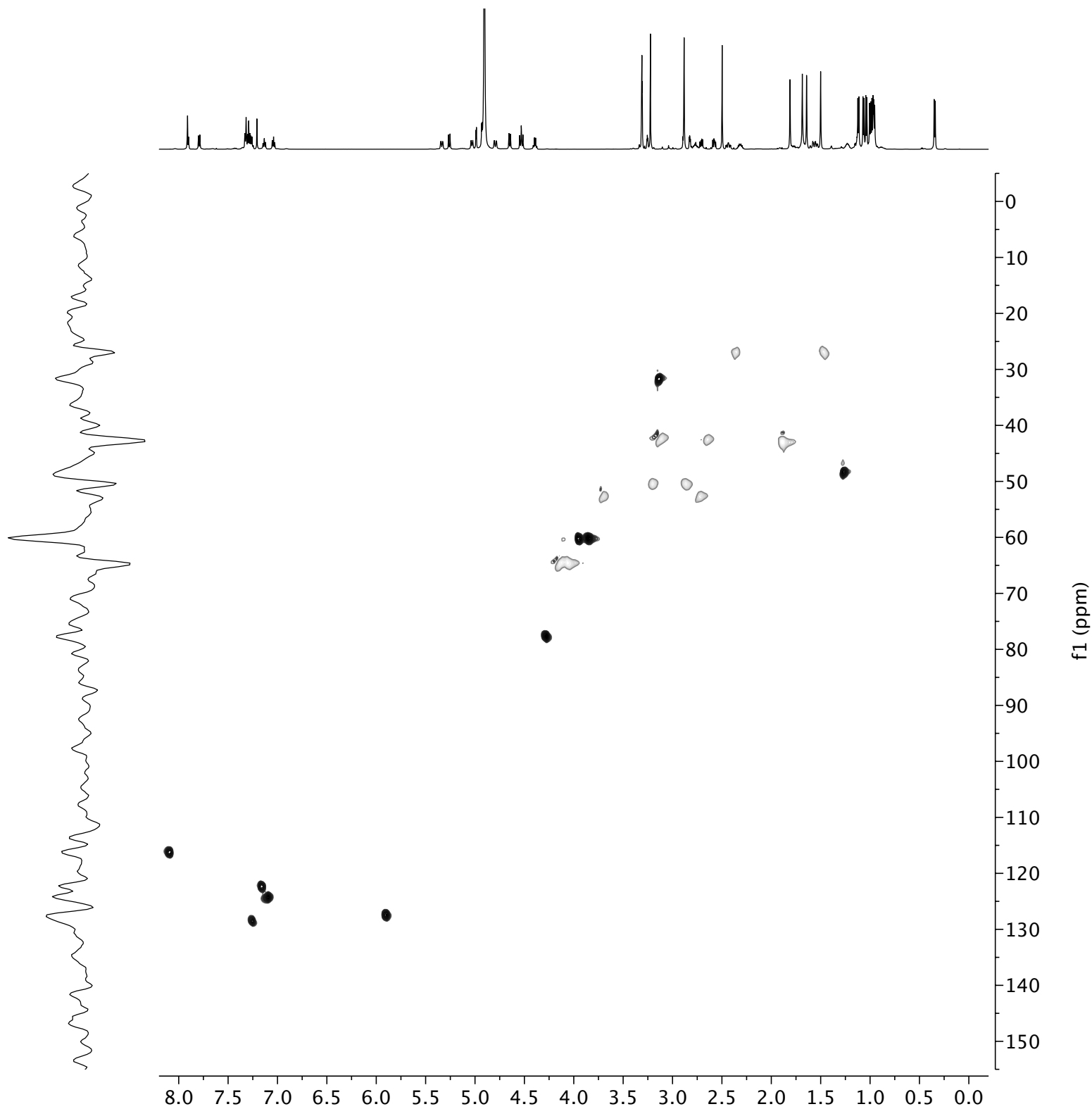


Figure S9. Profiled ^1H - ^{13}C HSQC spectrum of strychnine

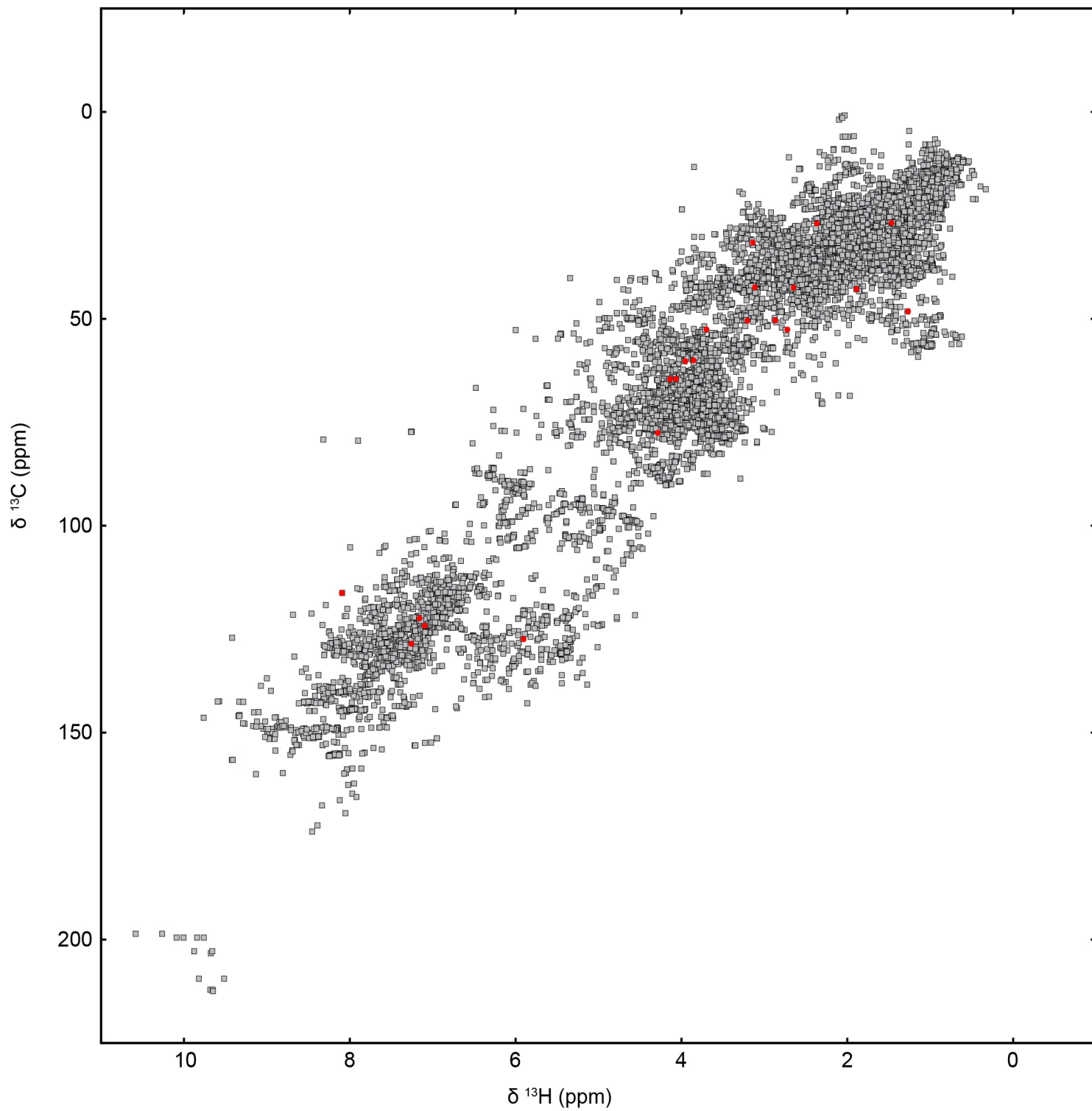


Figure S10. ^1H - ^{13}C HSQC (600 MHz) spectrum of brusatol in CD_3OD

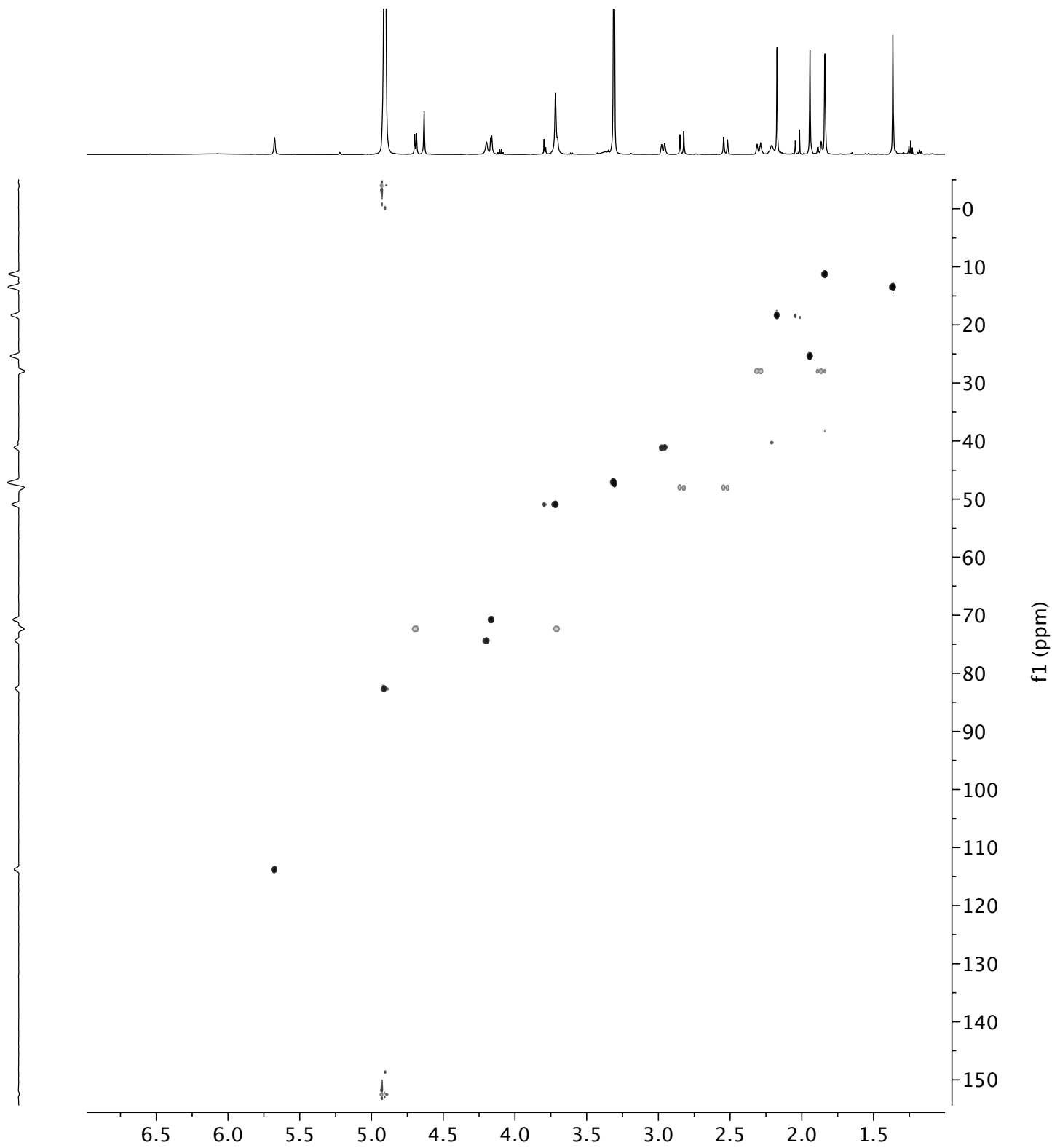


Figure S11. Profiled ^1H - ^{13}C HSQC spectrum of brusatol

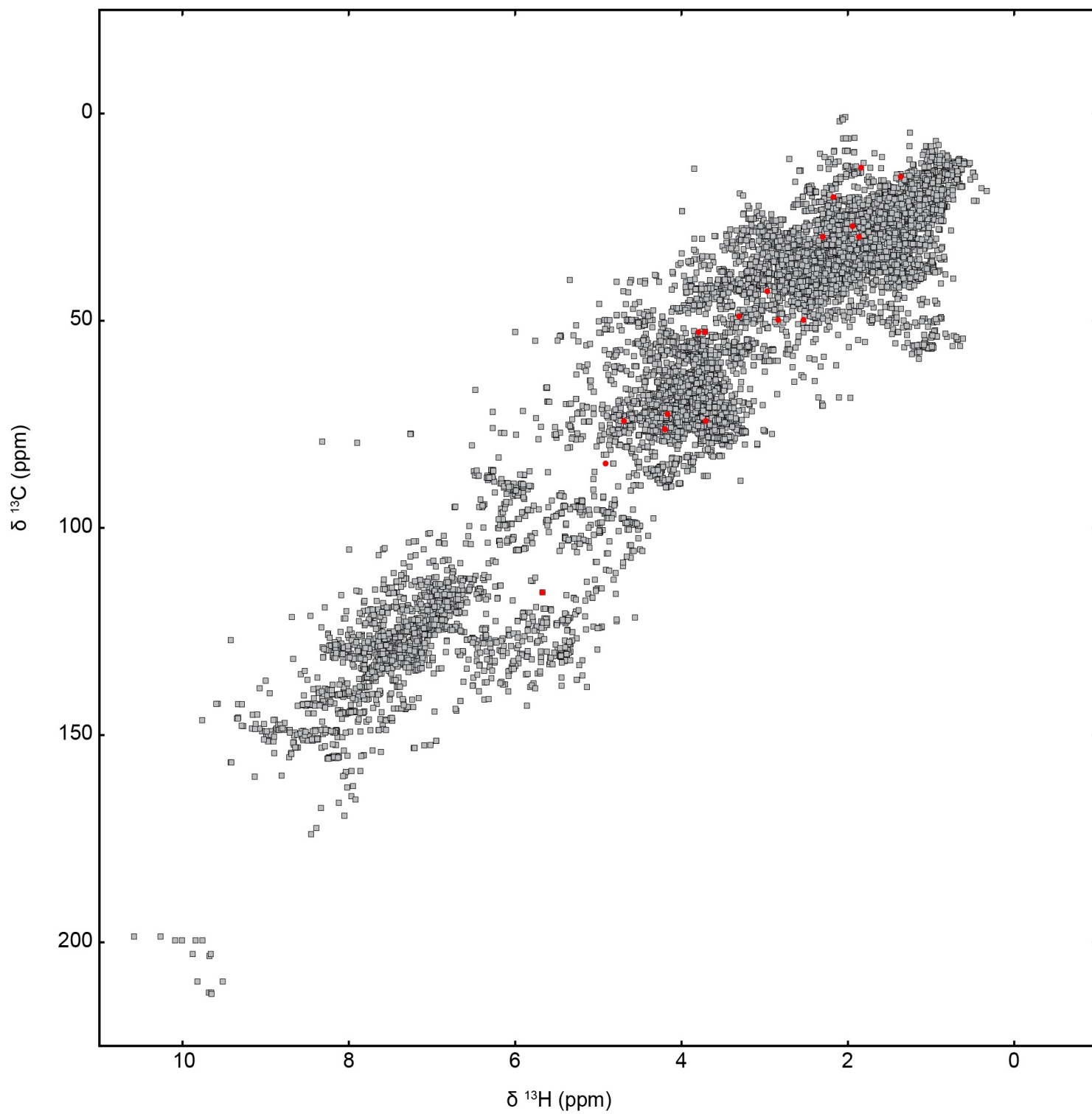


Figure S12. ^1H - ^{13}C HSQC (600 MHz) spectrum of paclitaxel in CD_3OD

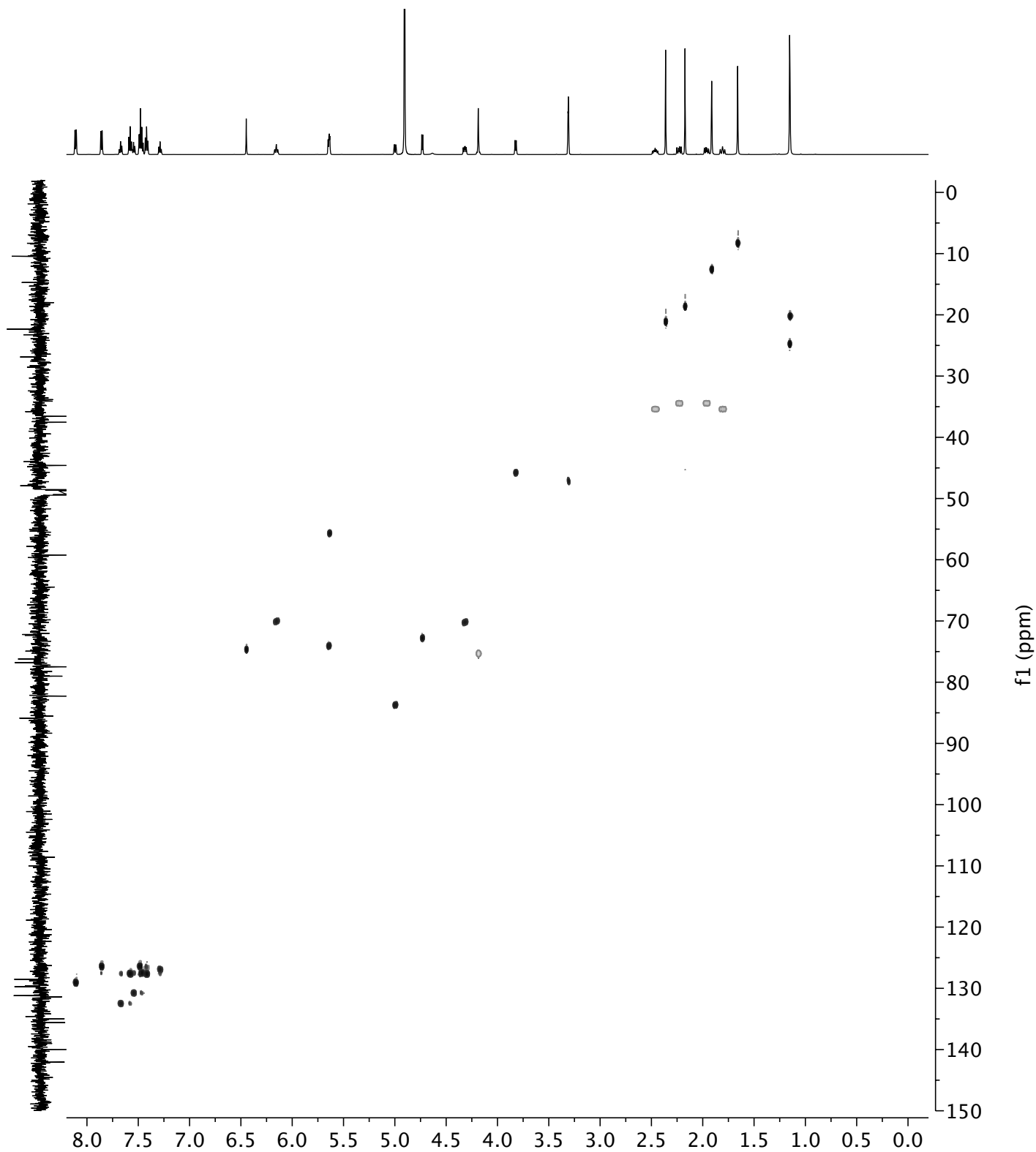


Figure S13. Profiled ^1H - ^{13}C HSQC spectrum of paclitaxel

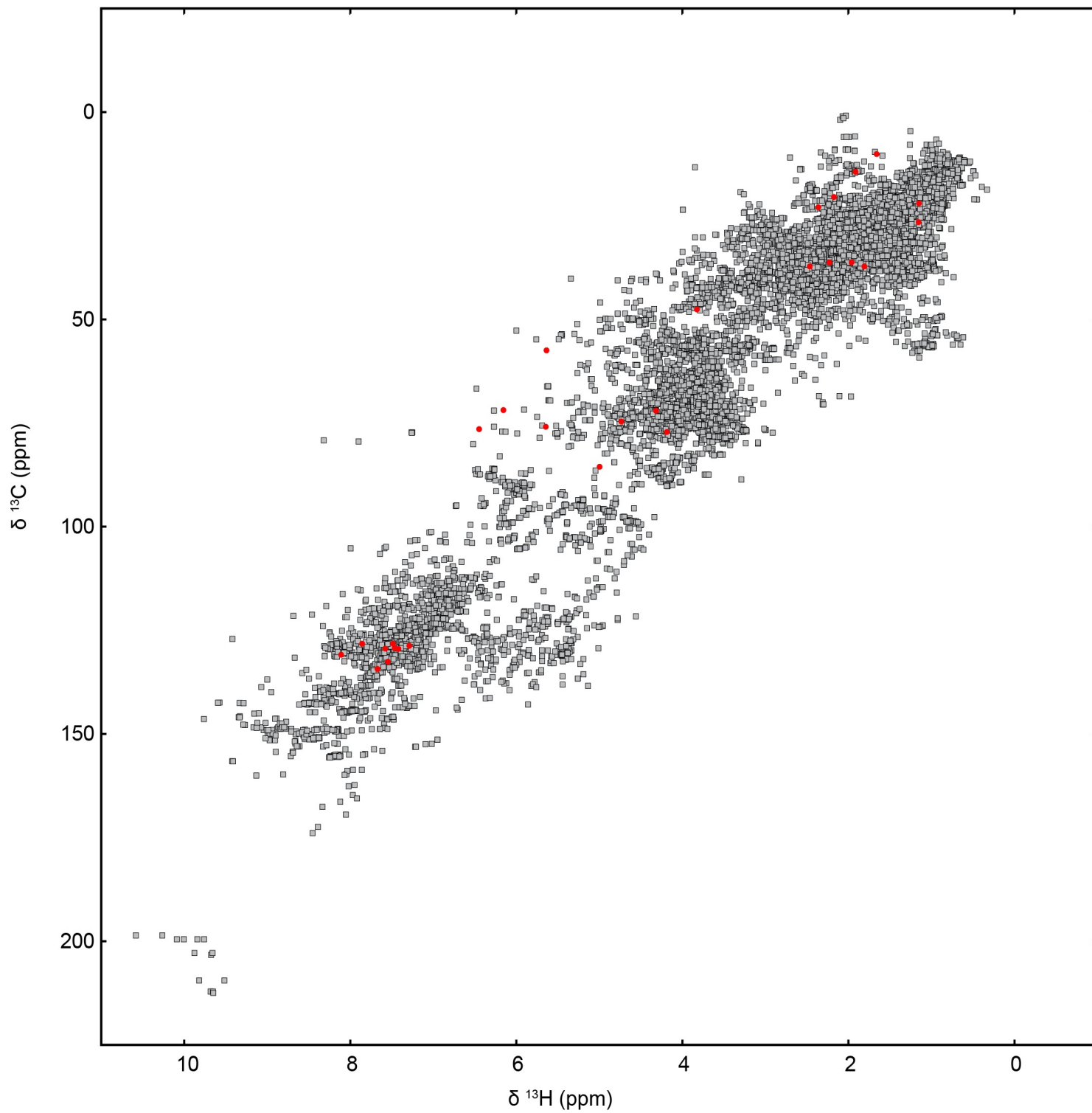


Figure S14. ^1H - ^{13}C HSQC (600 MHz) spectrum of the CNB-982 extract in CD_3OD

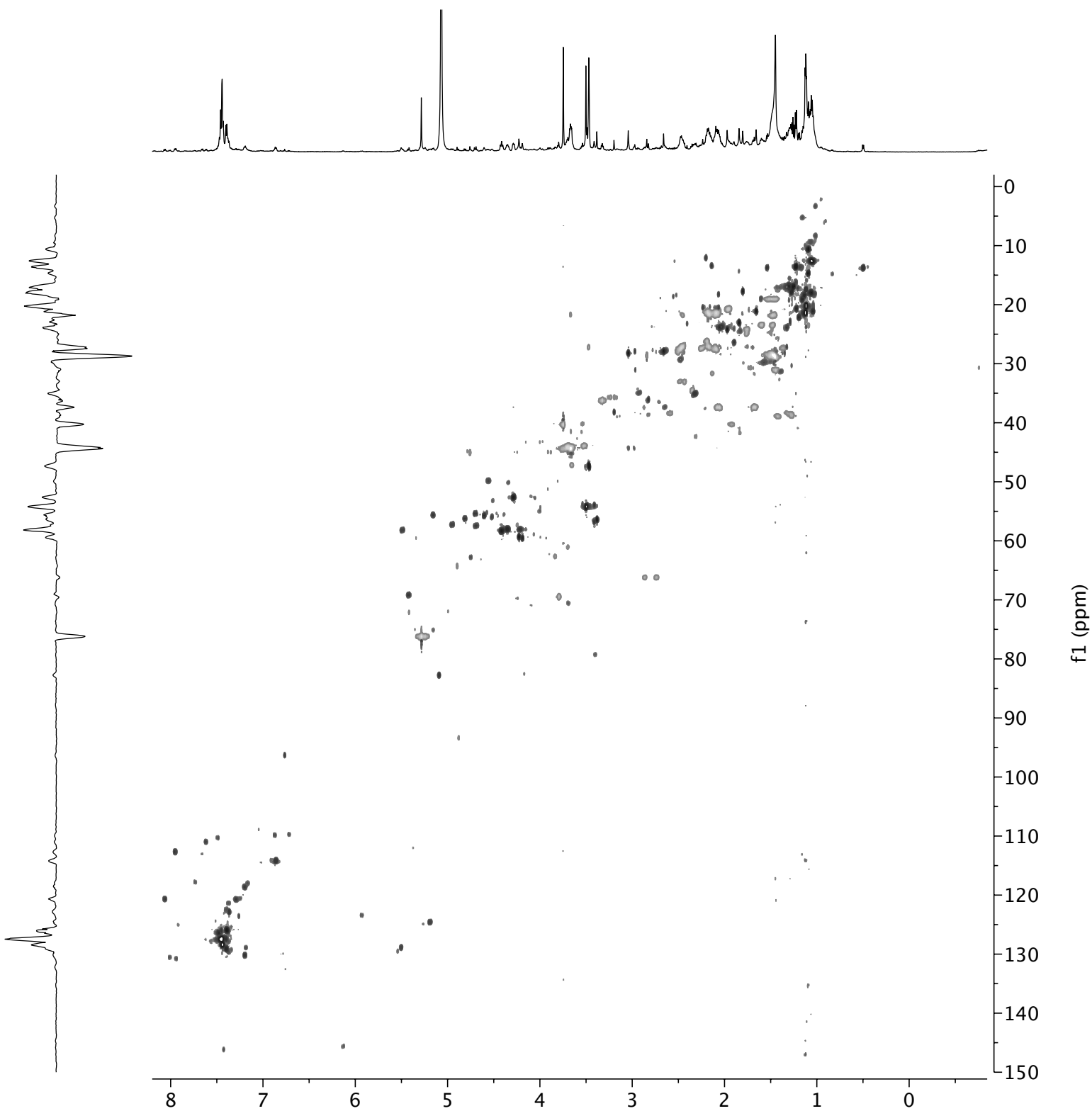


Figure S15. Profiled ^1H - ^{13}C HSQC spectrum of the CNB-982 extract

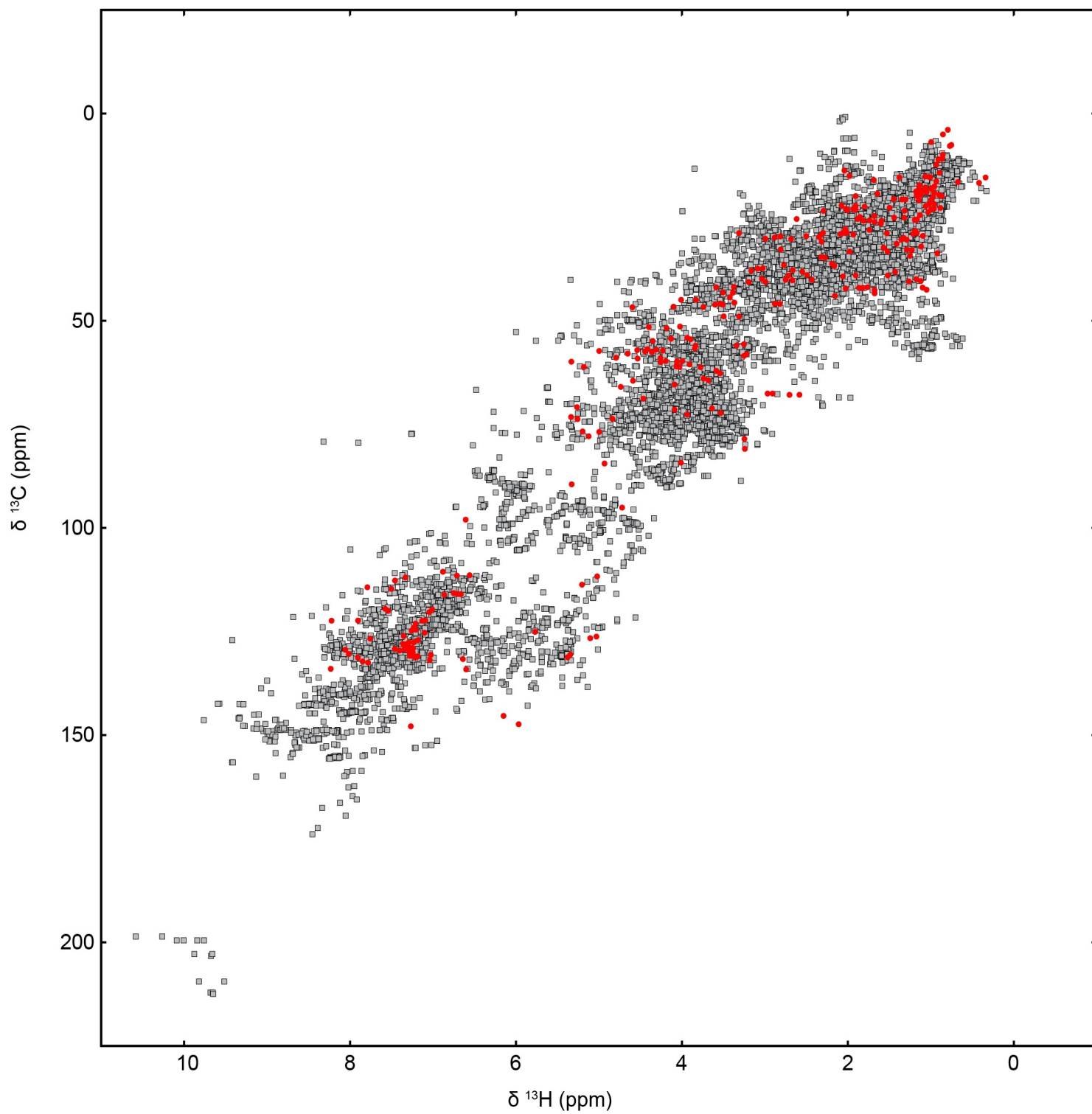


Figure S16. ^1H - ^{13}C HSQC (600 MHz) spectrum of cyclomarlin A in CD_3OD

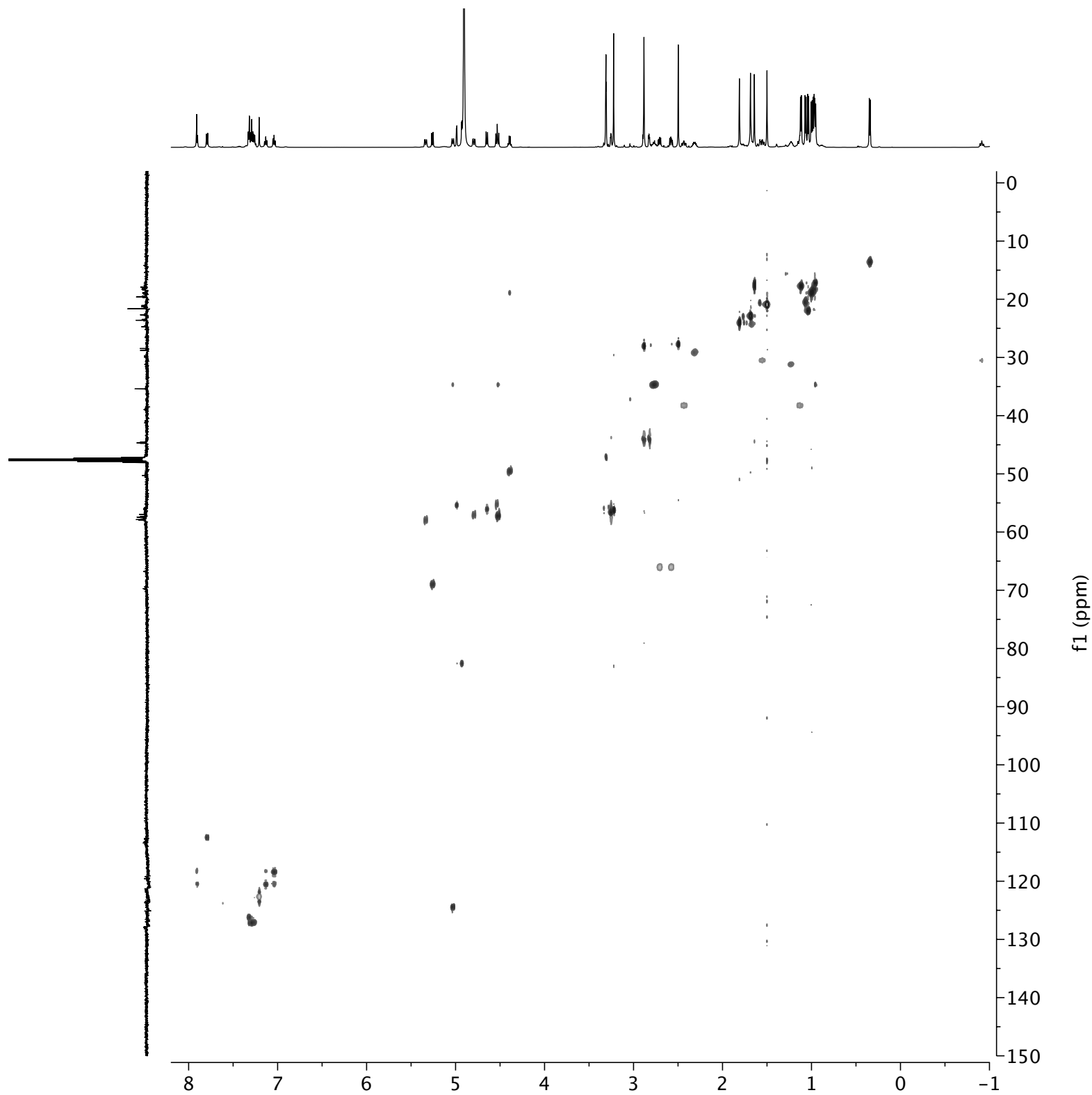


Figure S17. Profiled ^1H - ^{13}C HSQC spectrum of cyclomarin A

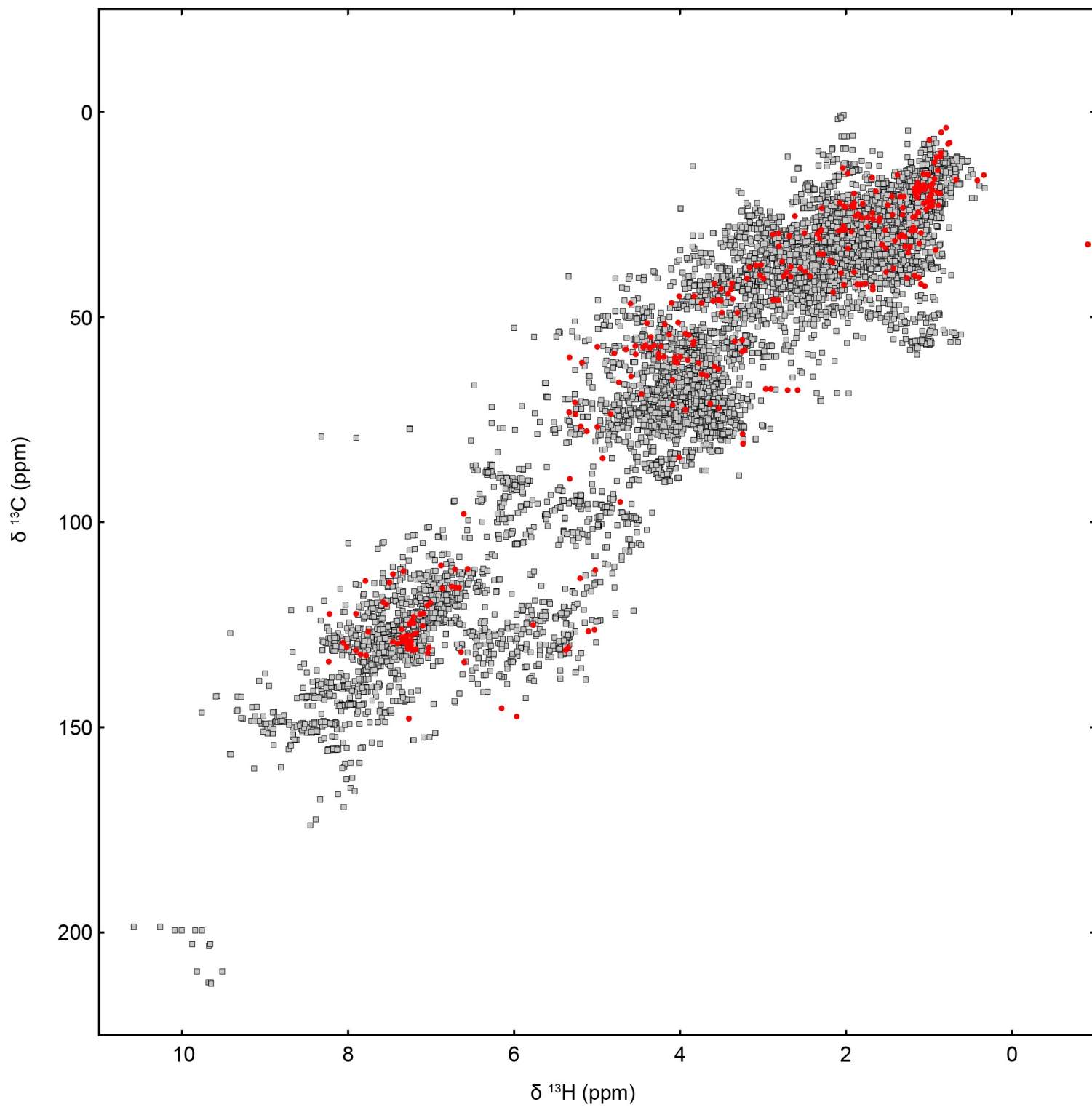


Figure S18. ^1H NMR (600 MHz) spectra of cyclomarin A in CD_3OD

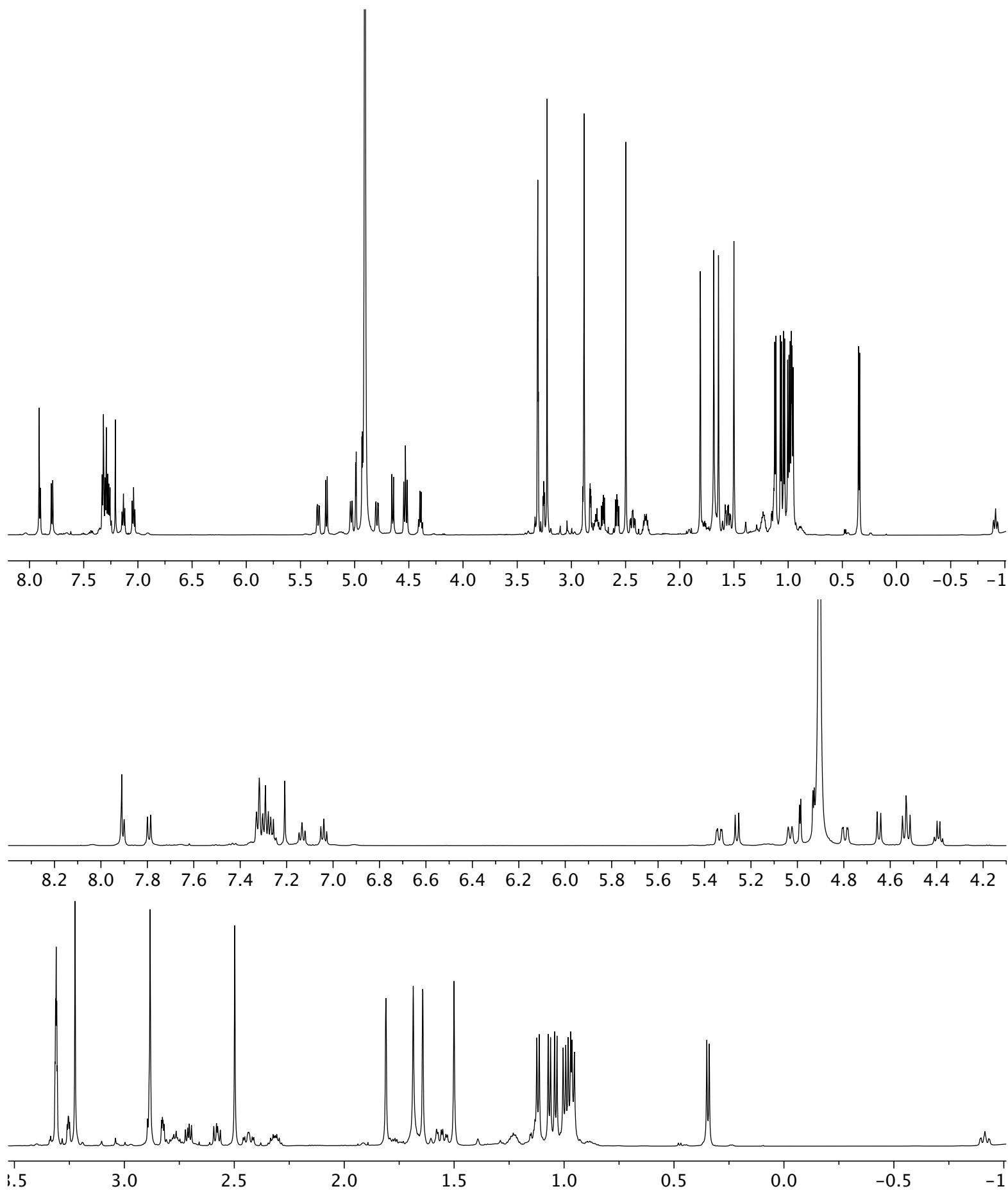


Figure S19. ^1H - ^1H gCOSY (600 MHz) spectrum of cyclomarlin A in CD_3OD

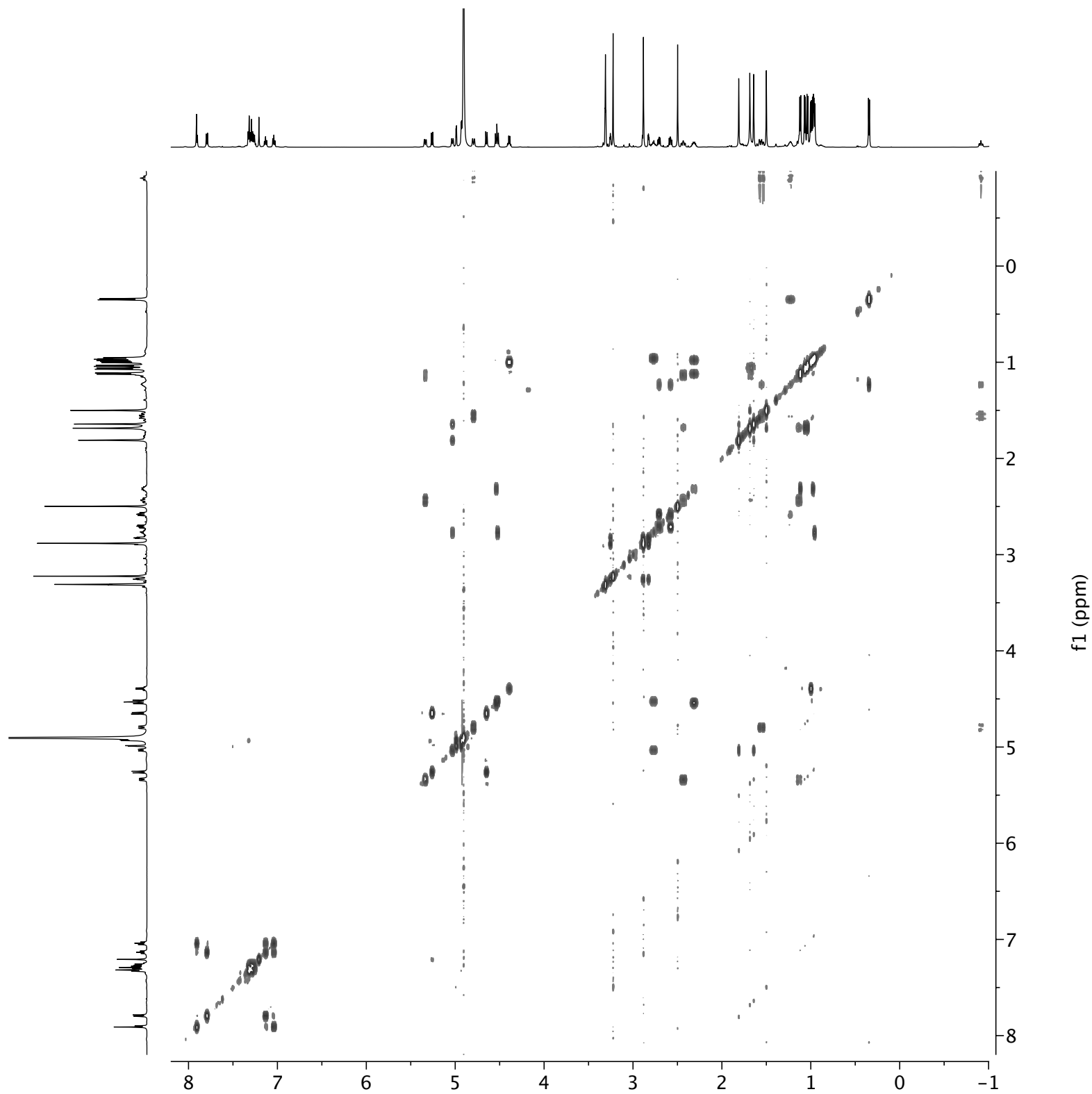


Figure S20. ^1H - ^{13}C HMBC (600 MHz) spectrum of cyclomarlin A in CD_3OD

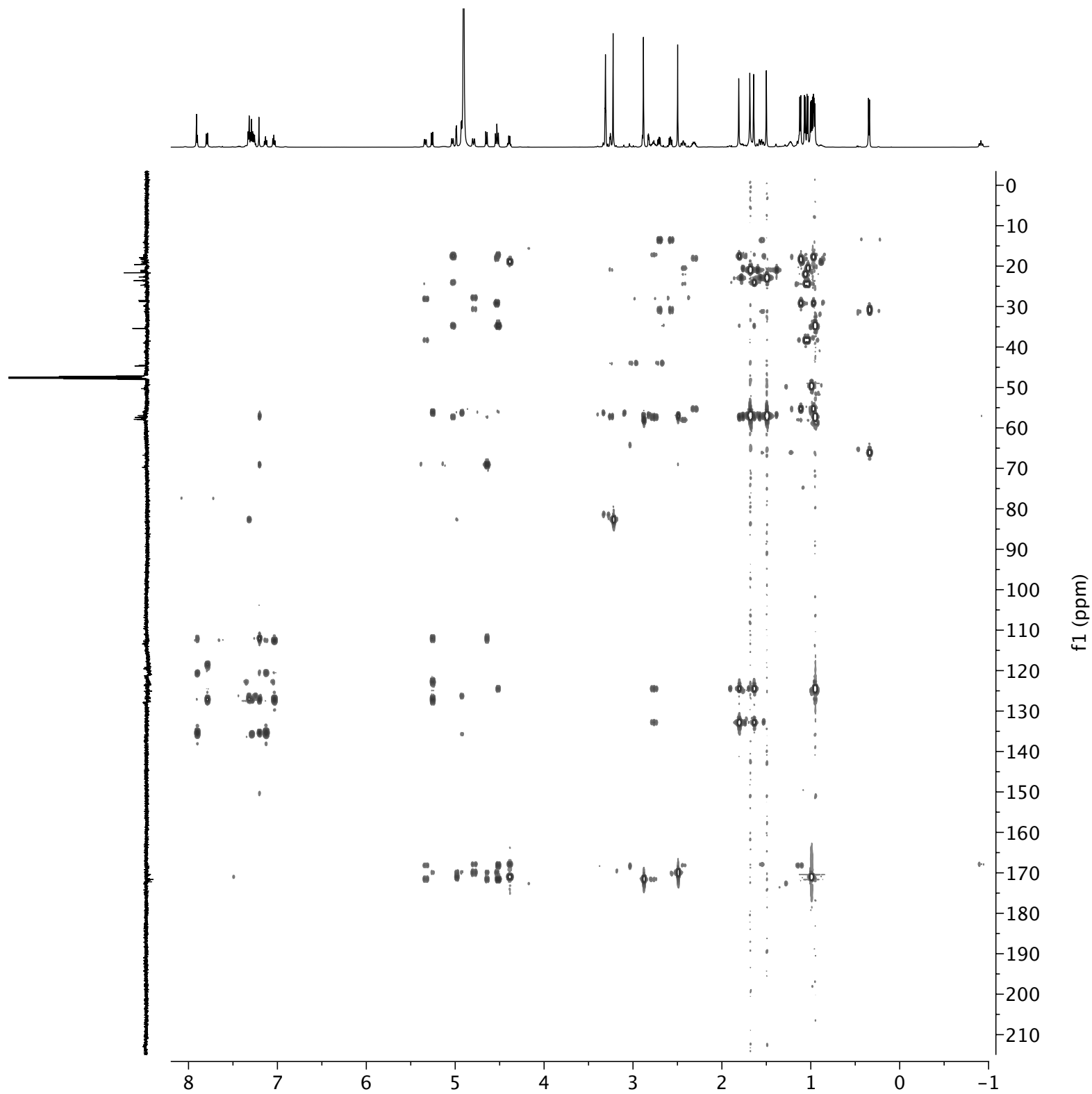


Figure S21. ^1H - ^{13}C H2BC (600 MHz) spectrum of cyclomarlin A in CD_3OD

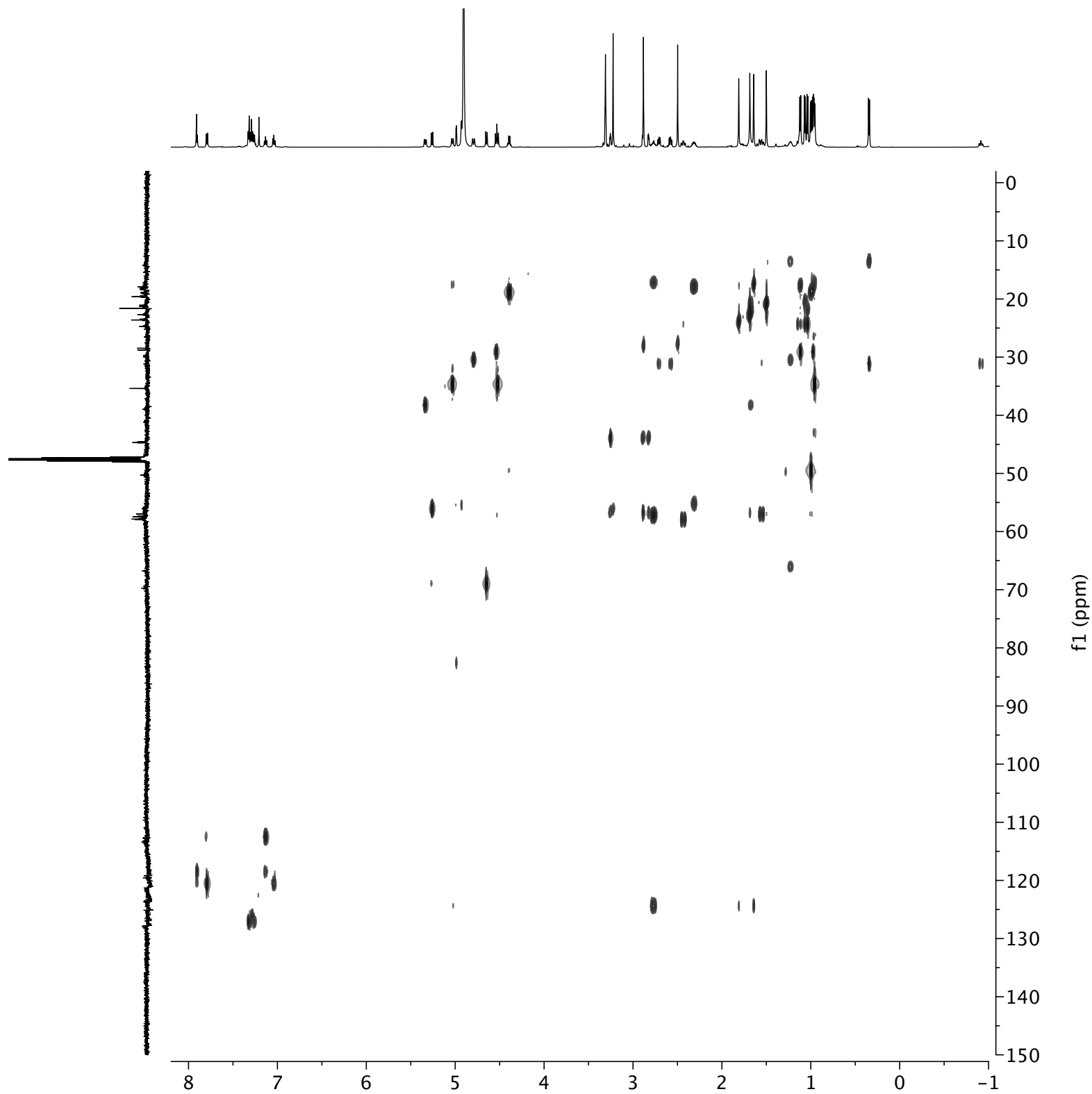


Figure S22. ^1H NMR (600 MHz) spectra of cyclomarlin A in CDCl_3

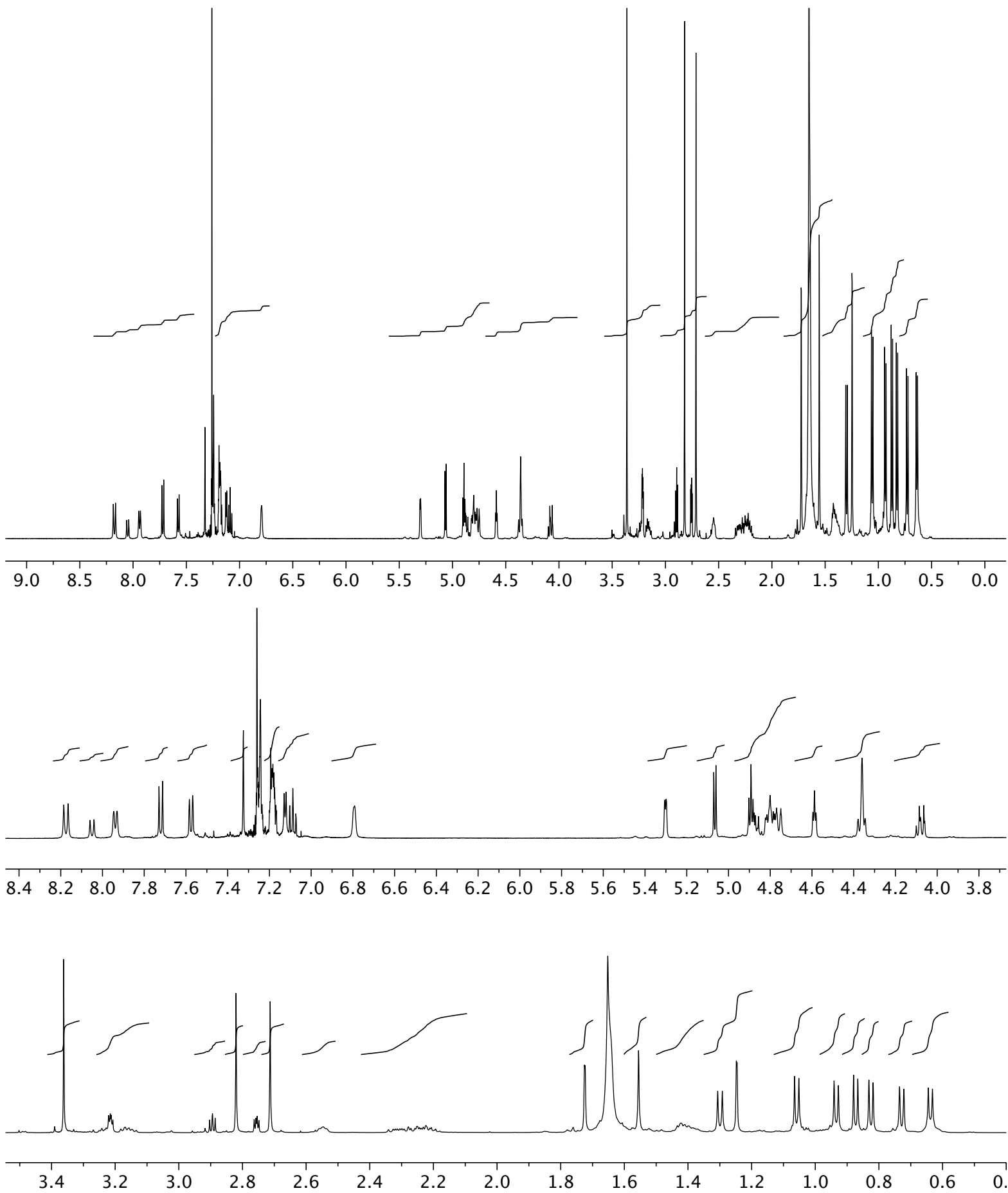


Figure S23. ^1H - ^{13}C HSQC (600 MHz) spectrum of IM06-19 extract in CD_3OD

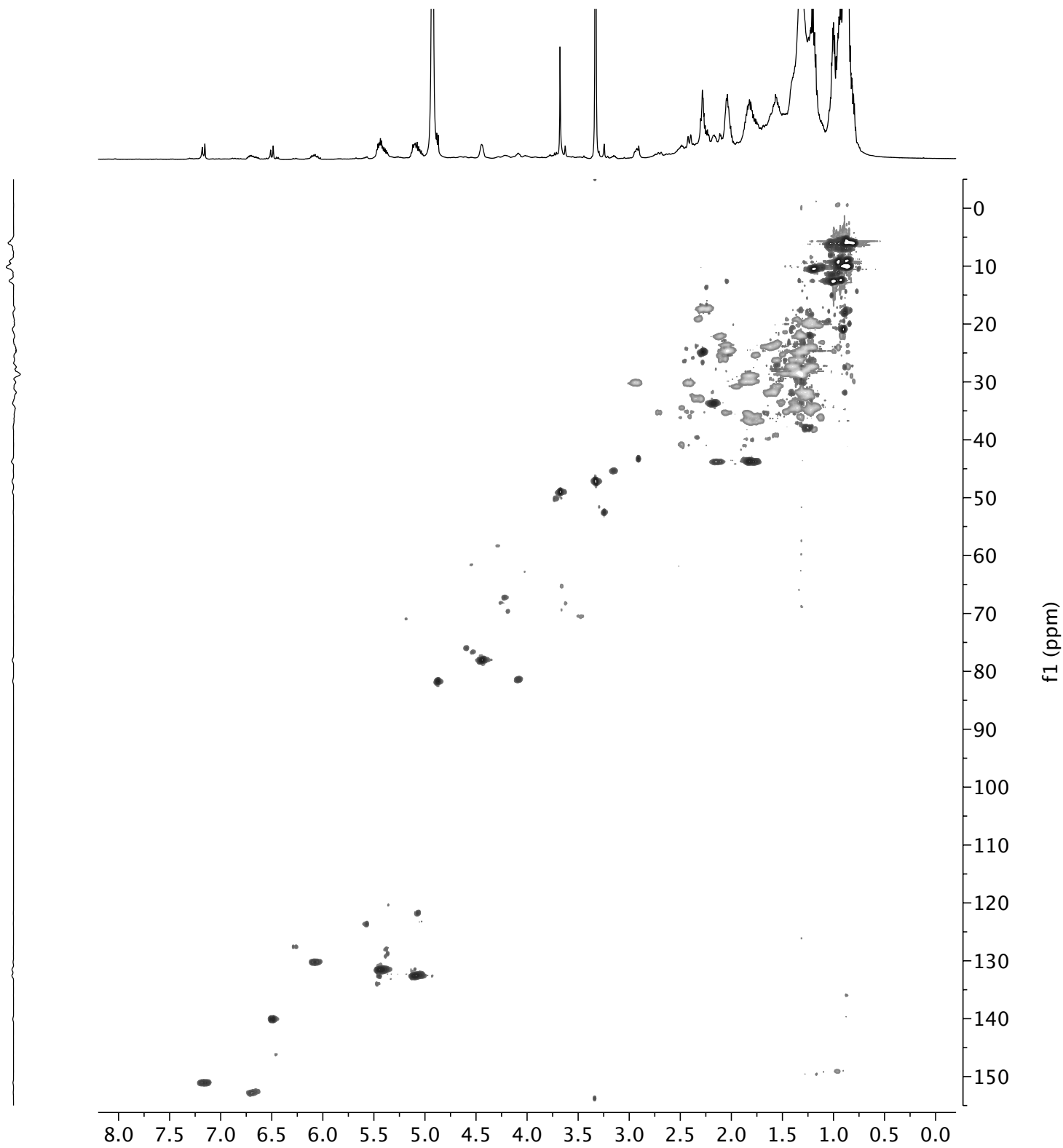


Figure S24. Profiled ^1H - ^{13}C HSQC spectrum of the IM06-19 extract

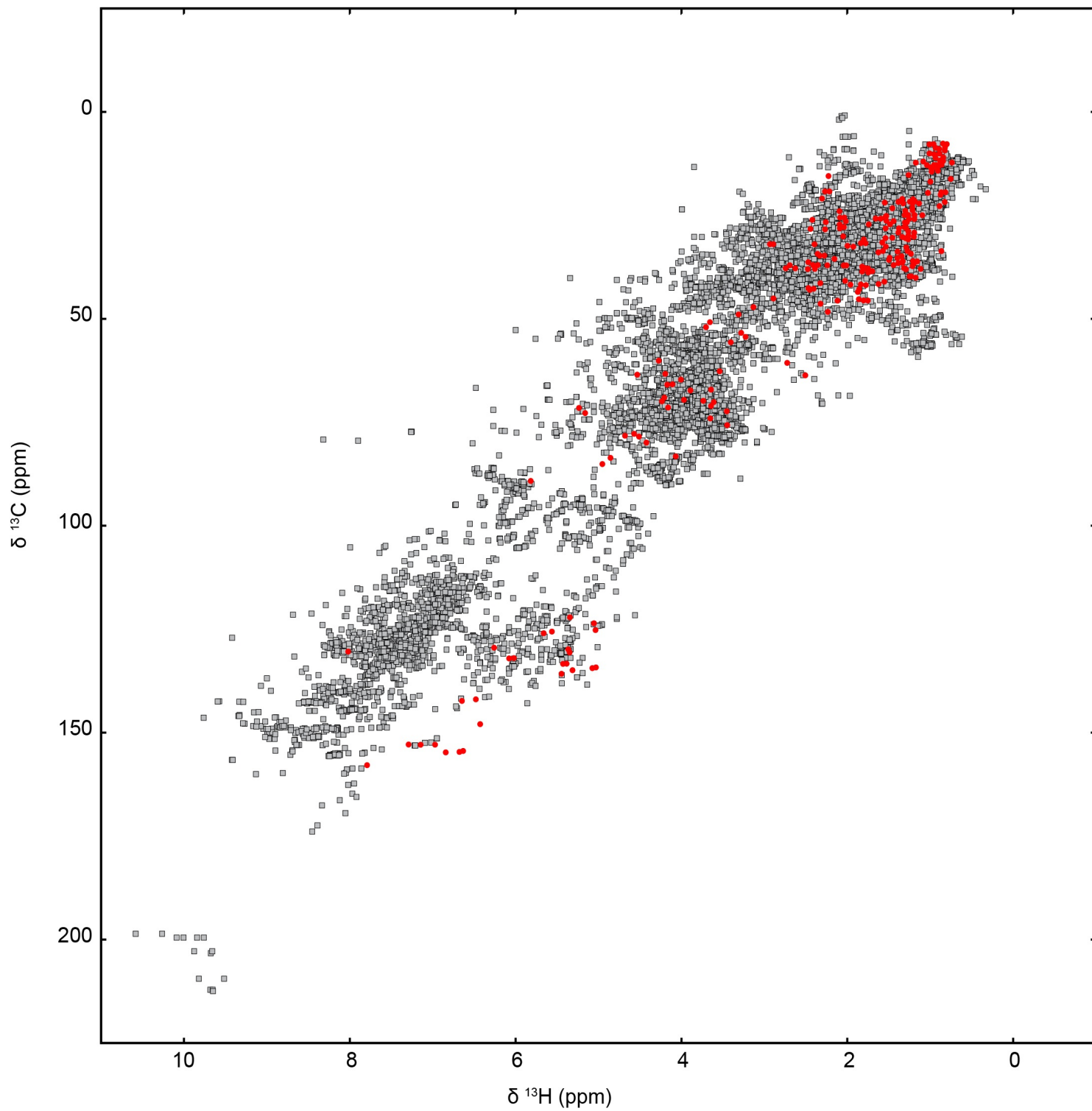


Figure S25. ^1H - ^{13}C HSQC (600 MHz) spectrum of gracilioether L in CD_3OD

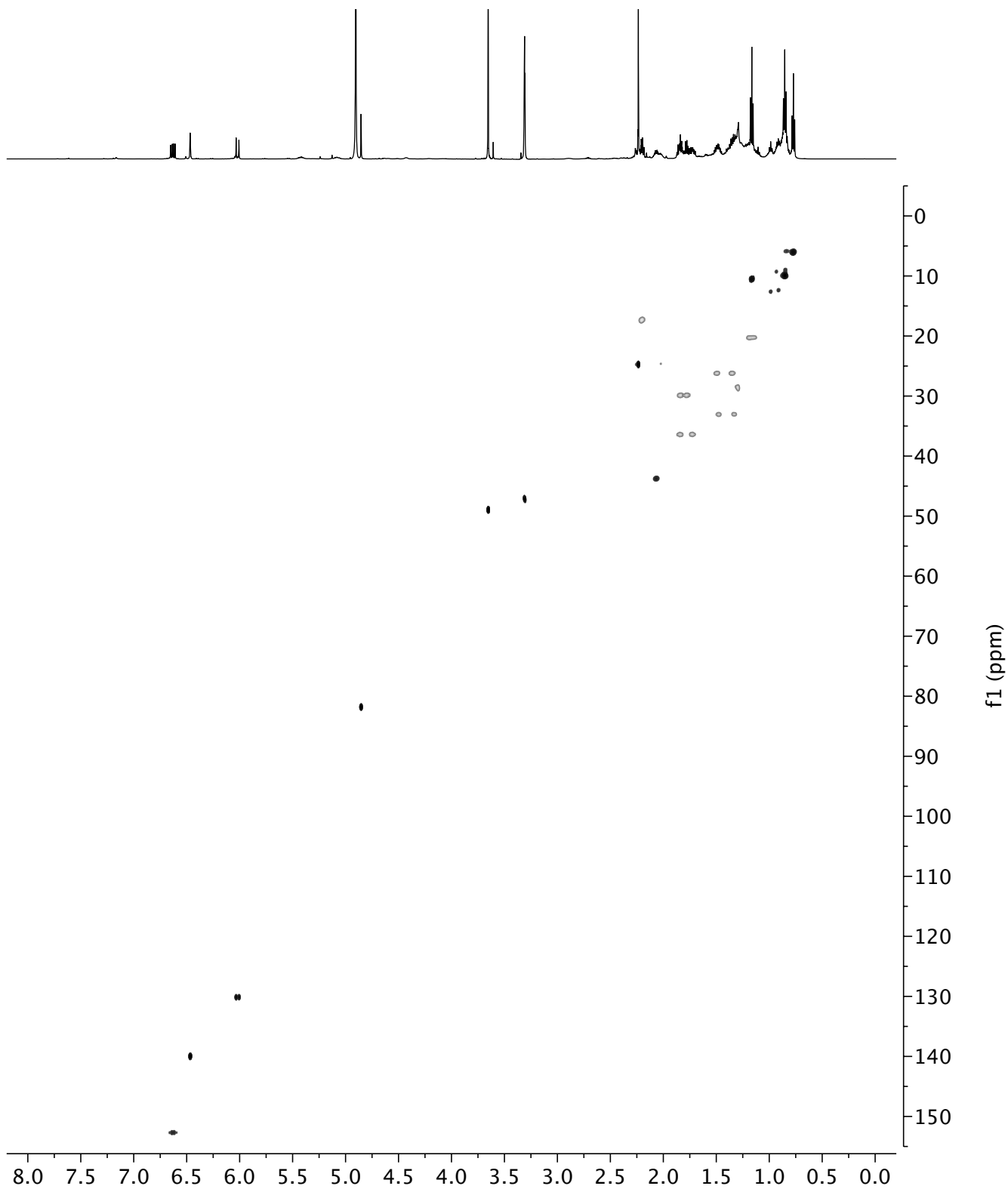


Figure S26. Profiled ^1H - ^{13}C HSQC spectrum of of gracilioether L

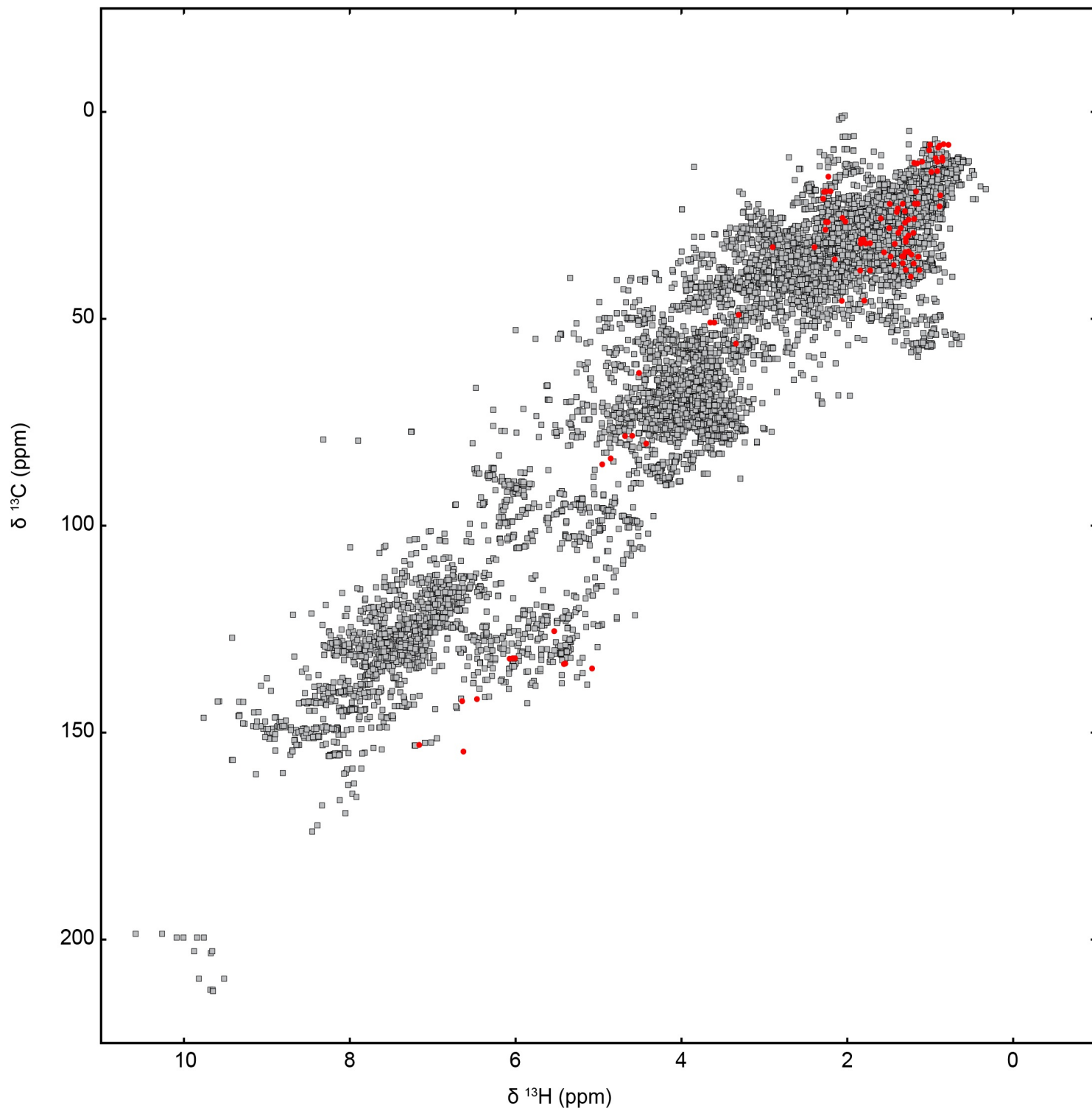


Figure S27. ^1H NMR (600 MHz) spectra of gracilioether L in CD_3OD

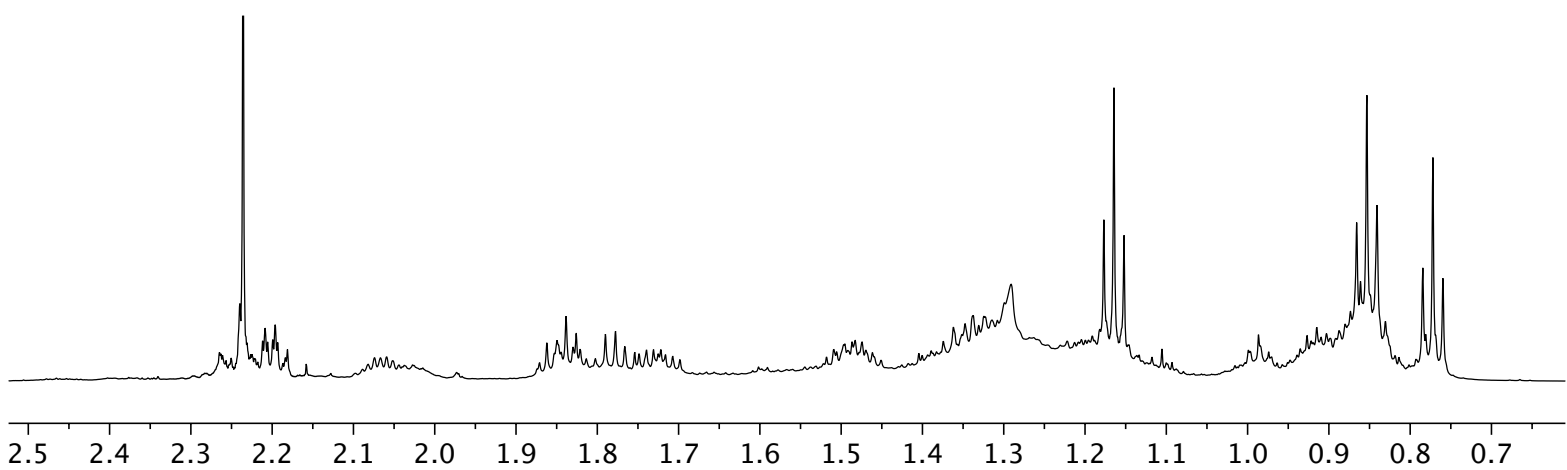
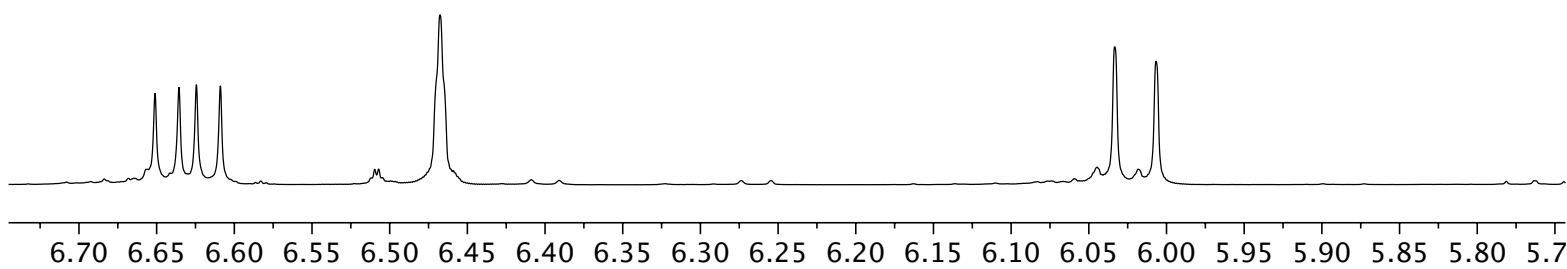
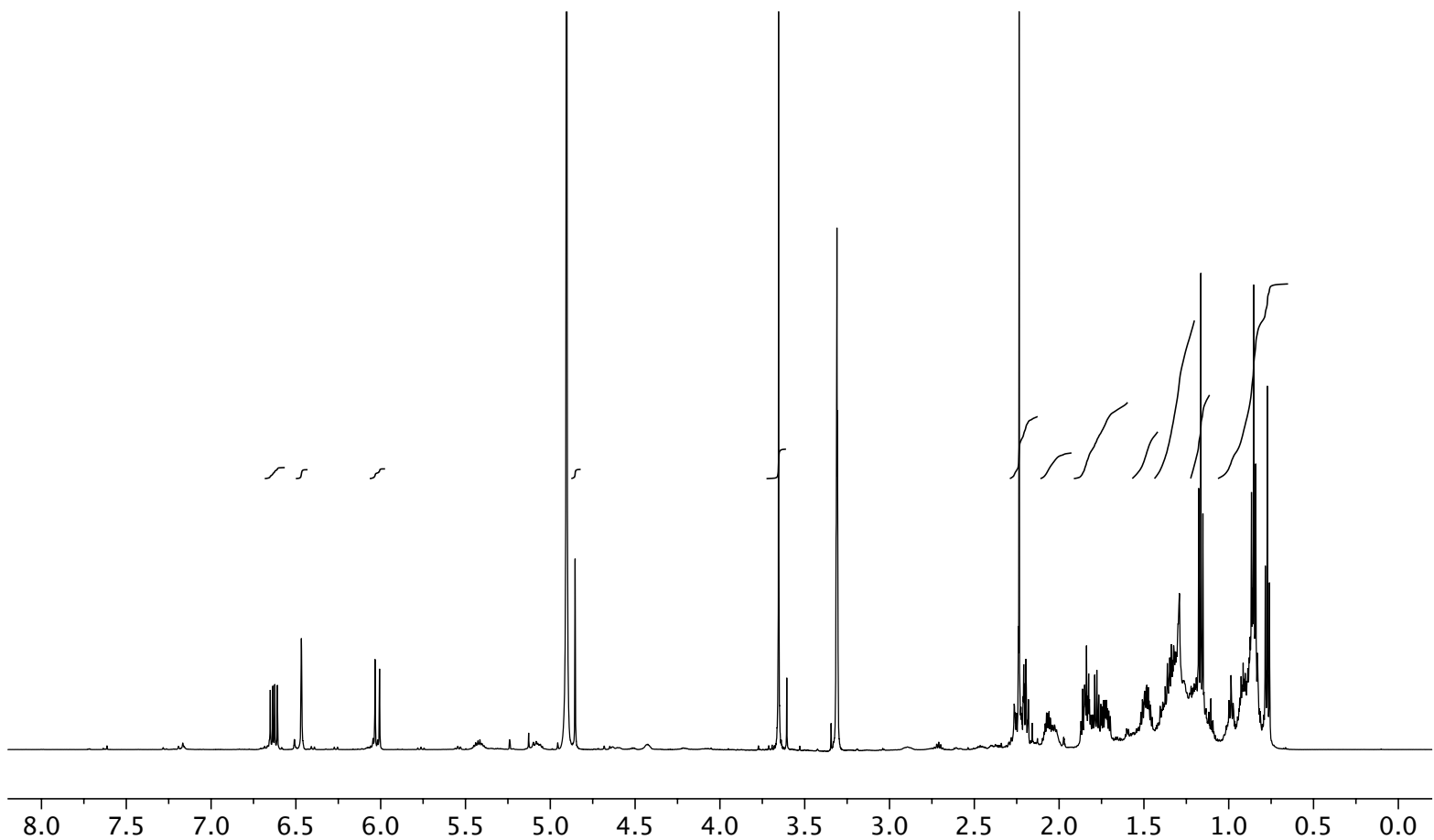


Figure S28. ^1H - ^1H gCOSY (600 MHz) spectrum of gracilioether L in CD_3OD

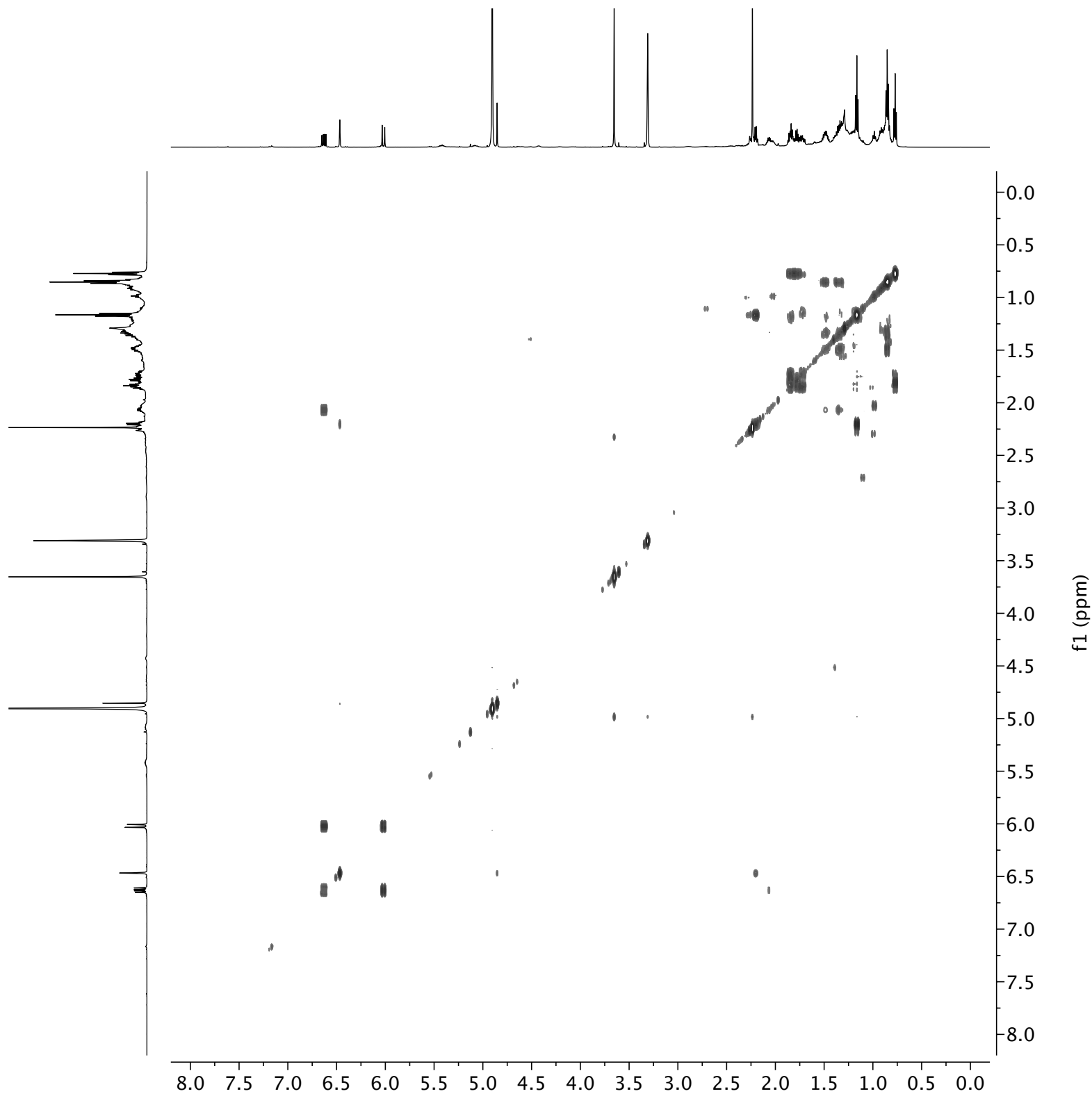


Figure S29. ^1H - ^1H NOESY (600 MHz) spectrum of gracilioether L in CD_3OD

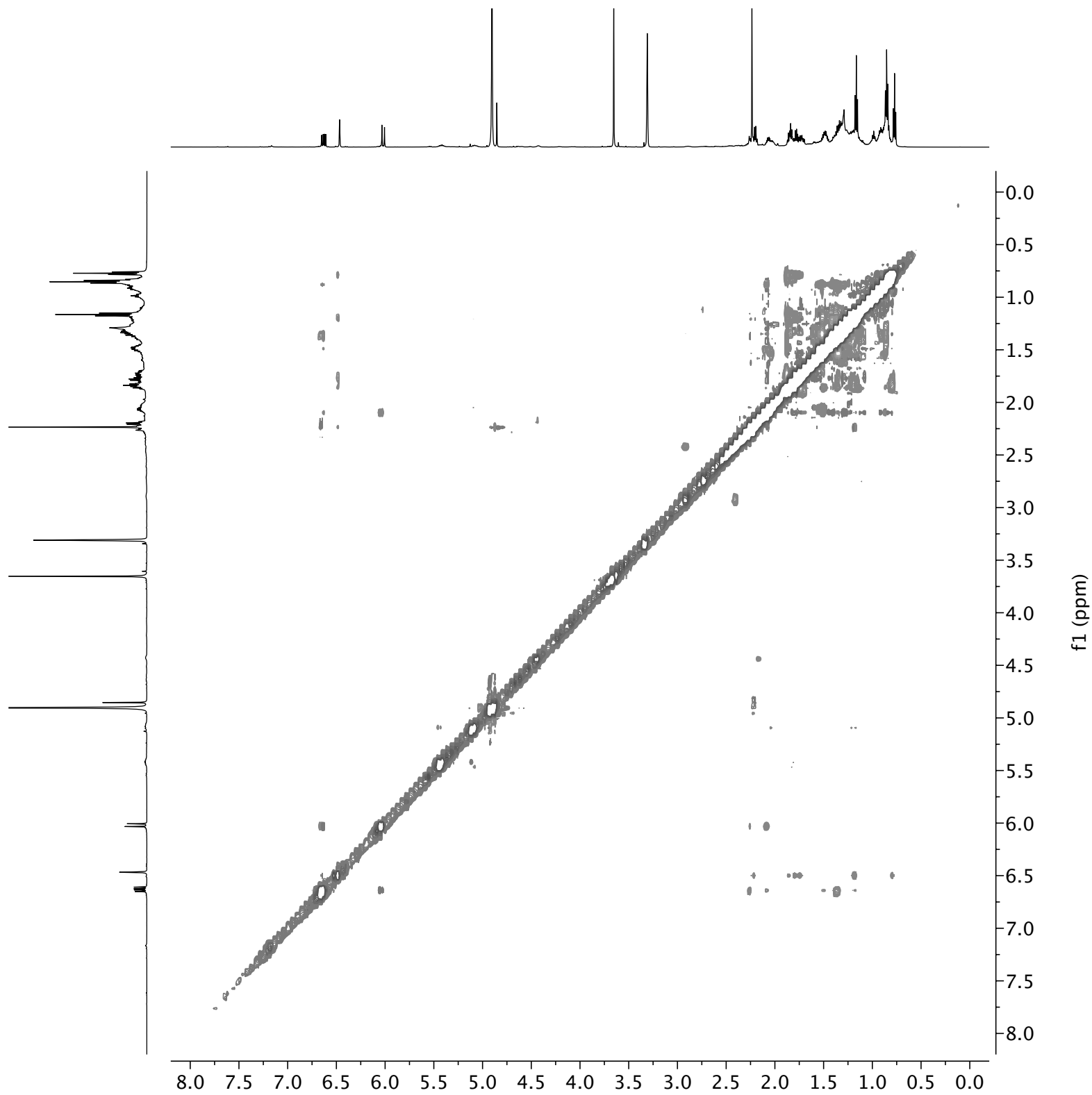


Figure S30. ^1H - ^{13}C HMBC (600 MHz) spectrum of gracilioether L in CD_3OD

