Highlight

Brief Overview of Recently Reported Misassigned Natural Products and Their *in Silico* Revisions Enabled by DU8ML, a Machine Learning-Augmented DFT Computational NMR Method.

Received 00th January 20xx,
Accepted 00th January 20xx

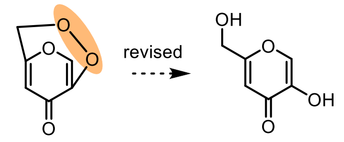
DOI: 10.1039/x0xx00000x

Ivan M. Novitskiy[a] and Andrei G. Kutateladze*[a]

This *Highlight* article describes a personal selection of recent misassigned structures of natural products and their revision with the aid of DU8ML, a machine learning-augmented DFT computational method for fast and accurate calculations of solution NMR chemical shifts and spin-spin coupling constants.

It is self-evident for the broader readership of this journal that natural product research is extraordinarily consequential. Among other things, it provides one of the most prolific pipelines for drug discovery. In this context it is critically important to determine the structures of NPs with utmost accuracy. Solution structure elucidation principally relies on NMR as the most informative method for non-crystallizable materials. There has been spectacular progress with new solution NMR techniques.[1] Yet, structure misassignments remain ubiquitous, perhaps excessively so. This brief overview of recent literature is intended to highlight such misassignments, offer revised structures, and further stimulate community discussions about pitfalls and common errors in structure elucidation. For this work we rely on our recently developed machine learning-augmented DFT computational NMR method, DU8ML.[2] Hopefully, such overviews will become a recurring feature in *Nat. Prod. Rep*.

At times, misassignments are grotesque and easy to spot. Recently reported new pyran derivative with antioxidant and anticancer properties, plantarone (**1**),[3] was flagged by the "Hot off the Press" reviewers Hill and Sutherland,[4] who called it "impossibly strained" and requiring revision. DU8ML analysis pointed to the known kojic acid (**2**), although even without computations the comparison of published *experimental* ^{13}C chemical shifts for plantarone and kojic acid, albeit in different solvents, produced rmsd$_{\text{[exp-exp]}}$ of 0.67ppm.

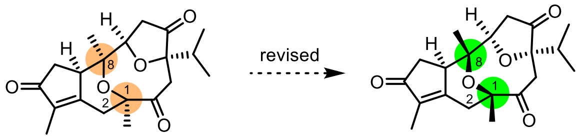


1
plantarone

2
(kojic acid)
rmsd(δ_{13C})$_{\text{exp-exp}}$ = 0.67ppm

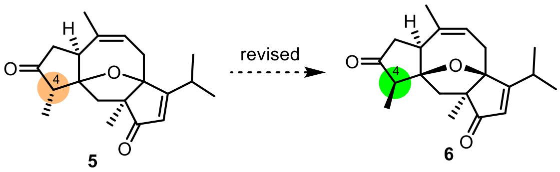
In contrast, it is by far more difficult to detect structure misassignments in cases when imperceptible errors involve a small number of stereogenic centers. Such subtle cases of misassignment are most frequent. Computational methods – including DU8ML – could expeditiously identify problems and help guide the process of structure revision.

For example, of ten new fusicoccane diterpenes recently isolated from the roots of *H. forsskaolii*,[5] at least three required revisions of stereoconfiguration for one or two stereogenic centers: compound **3** needed revision to diastereomer **4**; compound **5** was revised to 4-epimer **6**; and compound **7** was revised to 4-epimer **8** with stereoconfiguration at C6 – which was not defined in the original publication – clarified as shown. Notice that in all revisions *relative* stereochemistry is implied.



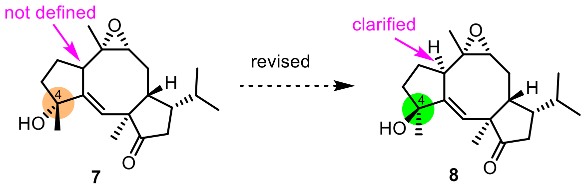
3
rmsd(δ_{13C}) = 3.94ppm

4
rmsd(δ_{13C}) = 1.03ppm



5
rmsd(δ_{13C}) = 2.16ppm

6
rmsd(δ_{13C}) = 0.85ppm



7
rmsd(δ_{13C}) = 1.64ppm

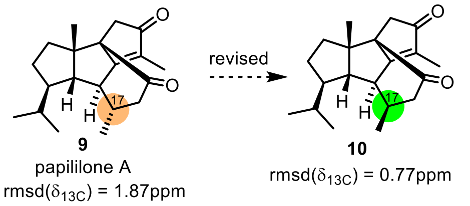
8
rmsd(δ_{13C}) = 1.15ppm

---

[a] *Department of Chemistry and Biochemistry, University of Denver, Denver, Colorado, USA 80210.*

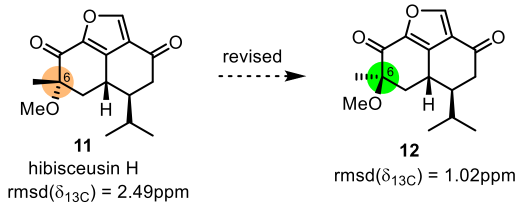
In another recent study, secondary metabolites from *Papiliomyces* sp were investigated and the first tetracyclic diterpenoid with an unprecedented 5/5/5/6 ring system, papililone A (**9**), was reported.[6] While the authors augmented its structure elucidation by both HOSE and Neural Net methods for ^{13}C chemical shift prediction, our DU8ML analysis still necessitated revision to 17-epimer **10**.

Eight new cadinene sesquiterpenoids, hibisceusins A-H, were isolated from infected stems of the semi-mangrove plant, *Hibiscus tiiaceus*.[7] Structures of five of them were determined by X-ray crystallographic analysis. Yet according to DU8ML, hibisceusin H (**11**), characterized only by solution NMR, required revision to its 6-epimer **12**. The new spirocyclic lactone, terreinlactone C (**13**),[8] isolated from the fungus *Aspergillus* terreus, also needed revision to 4-epimer **14**.



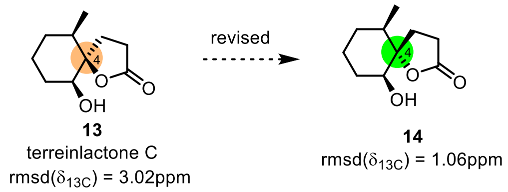
9
papililone A
rmsd(δ_{13C}) = 1.87ppm

10
rmsd(δ_{13C}) = 0.77ppm



11
hibisceusin H
rmsd(δ_{13C}) = 2.49ppm

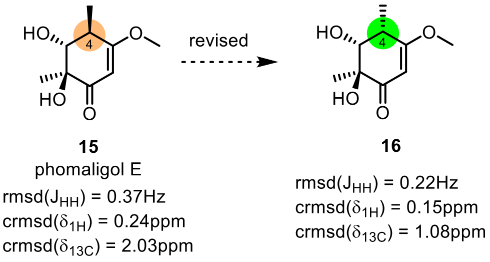
12
rmsd(δ_{13C}) = 1.02ppm



13
terreinlactone C
rmsd(δ_{13C}) = 3.02ppm

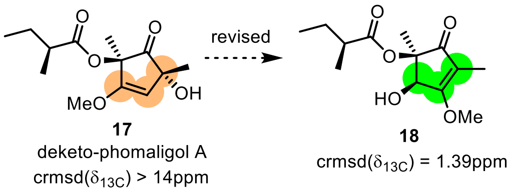
14
rmsd(δ_{13C}) = 1.06ppm

Culture broth from marine-derived fungus, *Aspergillus flocculosus*, produced two new compounds, phomaligol E (**15**), and deketo-phomaligol A (**17**).[9] Our analysis showed that phomaligol E (**15**), needs revision to 4-epimer (**16**), while deketo-phomaligol A (**17**) required a more extensive revision into α,β-unsaturated cyclopentanone **18**.



15
phomaligol E
rmsd(J_{HH}) = 0.37Hz
crmsd(δ_{1H}) = 0.24ppm
crmsd(δ_{13C}) = 2.03ppm

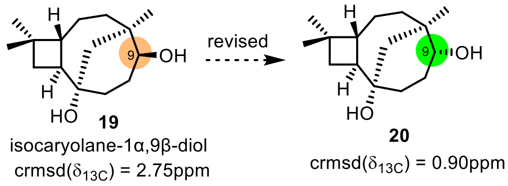
16
rmsd(J_{HH}) = 0.22Hz
crmsd(δ_{1H}) = 0.15ppm
crmsd(δ_{13C}) = 1.08ppm



17
deketo-phomaligol A
crmsd(δ_{13C}) > 14ppm

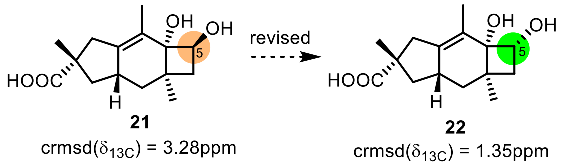
18
crmsd(δ_{13C}) = 1.39ppm

Carbon-13 NMR data for isocaryolane-1α,9β-diol (**19**), extracted from aerial parts of the plant *Perovskia artemisioides*[10] produced poor match with the computed chemical shifts. It was revised to 9-epimer **20**. A sterpurane derivative **21** isolated from a coculture of *P. orientoasiaticus* and *X. flaviporus* required similar single-stereogenic center revision to 5-epimer **22**.[11]



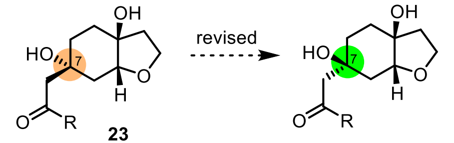
19
isocaryolane-1α,9β-diol
crmsd(δ_{13C}) = 2.75ppm

20
crmsd(δ_{13C}) = 0.90ppm



21
crmsd(δ_{13C}) = 3.28ppm

22
crmsd(δ_{13C}) = 1.35ppm



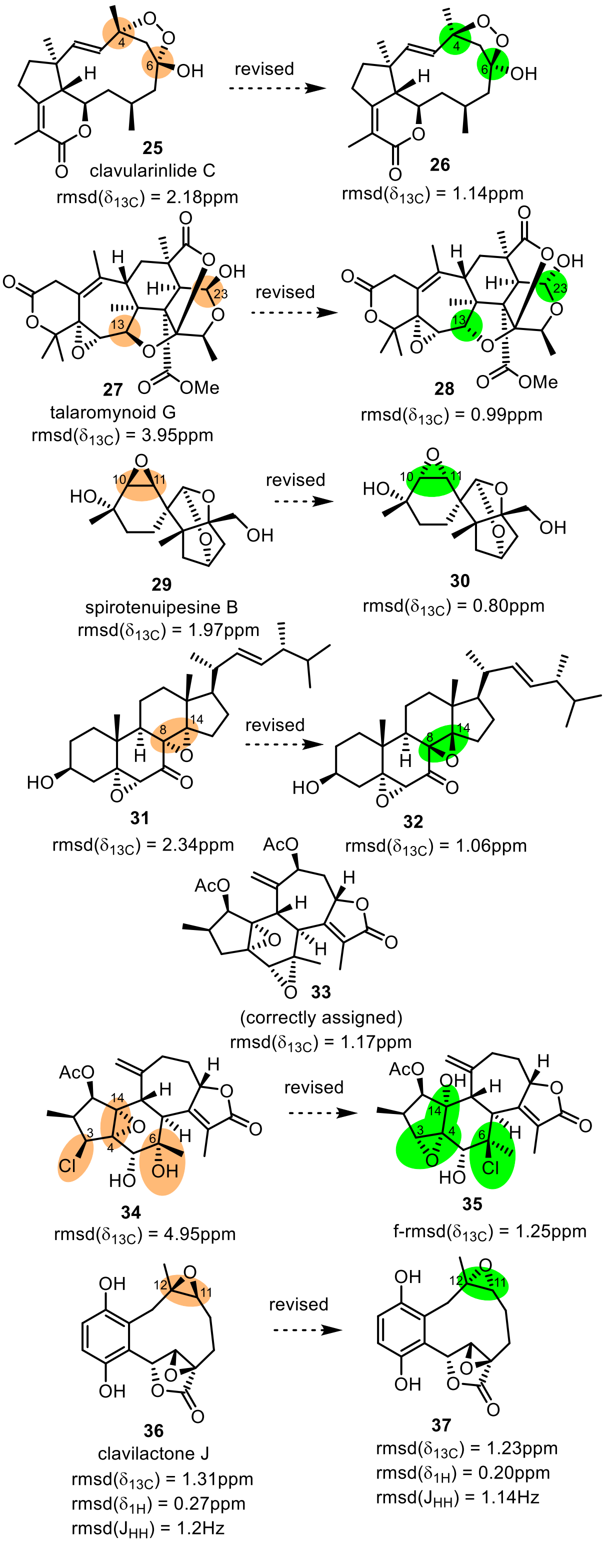
23
(a) delavatone A, R = CH=CMe$_2$
f-crmsd(δ_{13C}) = 1.82ppm
(b) delavatrone B, R = Me
rmsd(δ_{13C}) = 2.38ppm

24
(a) f-crmsd(δ_{13C}) = 1.04ppm
(b) rmsd(δ_{13C}) = 1.20ppm

Delavatone A (**23a**, R = CH=CMe$_2$) and its close relative, delavatrone B (**23b**, R = Me) were isolated from the ethanol extract of whole plant of *Incarvillea delavyi* and exhibited "remarkable suppression on NO production in LPS-induced BV-2 cells."[12] Their structures were elucidated by NMR "with the help of chemical shift calculations coupled with DP4+ probability analysis". Yet, DU8ML necessitated the revision of stereoconfiguration at C7 to structures **24** with the calculated chemical shifts matching the experimental data very well.

Clavularinlide C (**25**)[13] isolated from soft coral *Clavularia inflata* needed revision to diastereomer **26**. Nine new highly oxygenated 3,5-dimethylorsellinic acid-derived meroterpenoids, talaromynoids A–I, were isolated from the marine-derived fungus *Talaromyces purpureogenus* SCSIO 41517.[14] All of them possessed unprecedented polycyclic ring systems, with talaromynoid A characterized with X-ray crystallographic analysis. However, talaromynoid G (**27**), needed revision to diastereomer **28**.

Recent synthesis of spirotenuipesine A, a promoter of neurotrophic factor secretion from glial cells[15] prompted us to look at the original isolation work for spirotenuipesine A and its epoxide, spirotenuipesine B, which was assigned structure **29**.[16] DU8ML analysis necessitated revision of spirotenuipesine B to a stereoisomeric epoxide **30**.[17] As we reported in the past, complex epoxides often present significant challenge to accurate determination of their stereochemistry.[18] Another recent example of this is bis-epoxy ergostadienediol.[19] The authors reassigned an incorrect pentahydroxy ergostane structure to bis-epoxide **31**, because it matched the experimental chemical shifts for this bis-epoxide reported a decade ago.[20] The problem is that this revision was incomplete and, according to DU8ML, required additional correction of stereoconfiguration for the second oxirane moiety, i.e. to diastereomer **32**.



25 clavularinlide C rmsd(δ_{13C}) = 2.18ppm

26 rmsd(δ_{13C}) = 1.14ppm

27 talaromynoid G rmsd(δ_{13C}) = 3.95ppm

28 rmsd(δ_{13C}) = 0.99ppm

29 spirotenuipesine B rmsd(δ_{13C}) = 1.97ppm

30 rmsd(δ_{13C}) = 0.80ppm

31 rmsd(δ_{13C}) = 2.34ppm

32 rmsd(δ_{13C}) = 1.06ppm

33 (correctly assigned) rmsd(δ_{13C}) = 1.17ppm

34 rmsd(δ_{13C}) = 4.95ppm

35 f-rmsd(δ_{13C}) = 1.25ppm

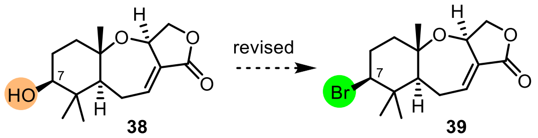
36 clavilactone J
rmsd(δ_{13C}) = 1.31ppm
rmsd(δ_{1H}) = 0.27ppm
rmsd(J_{HH}) = 1.2Hz

37
rmsd(δ_{13C}) = 1.23ppm
rmsd(δ_{1H}) = 0.20ppm
rmsd(J_{HH}) = 1.14Hz

Two new anti-HIV diterpenoids **33** and **34** were isolated from *C. megalocarpus*, with chloride **34** described as a potential artifact of extraction.[21] According to our analysis, diepoxide **33** is assigned correctly. However, chloride **34** needed significant revisions to **35**.

The case of meroterpenoid clavilactone J (**36**), recently isolated from the basidiomycete *Clitocybe clavipes*,[22] further underscores significant difficulty that conformationally flexible NPs possessing the oxirane moiety pose for structure elucidation. The authors calculated DP4+ all-data probabilities for the candidate structures, which pointed with high confidence (99.87%) to *11S,12R* diastereomer **36**, although the H data-based probability disagreed with C data. DU8ML analysis showed very small differences between the two diastereomers. However, all criteria that we used, i.e. ^{13}C and ^{1}H NMR chemical shifts, and proton spin-spin coupling constants matched the *11R,12S* diastereomer **37** slightly better. We therefore revise clavilactone J to structure **37**.

Another challenge for the *in silico* structure elucidations present the cases where computer-driven revisions are not in keeping with the published interpretation of mass-spectrometry data. One such example is a new sesquiterpene lactone panaxolide (**38**) which was recently isolated from the leaves of Vietnamese ginseng found in the Ngoc Linh mountain range.[23] While most of the calculated chemical shifts closely tracked the experimental values, the C7 carbon bearing the hydroxy group was singularly off by >10ppm, contributing to poor overall rmsd of 3.1ppm. Replacing the 7-OH group by bromine improved rmsd to 0.83ppm, indicating that panaxolide should be revised to **39**, which implies misinterpretation of the MS data. Luckily, bromide **39** is a known compound, aplysistatin, first isolated in 1977 from a sea hare, with structure established by X-ray crystallography,[24] imparting confidence that DU8ML has arrived at the correct revision. We are not certain how a brominated marine NP ended up in Vietnamese ginseng found at high elevation.

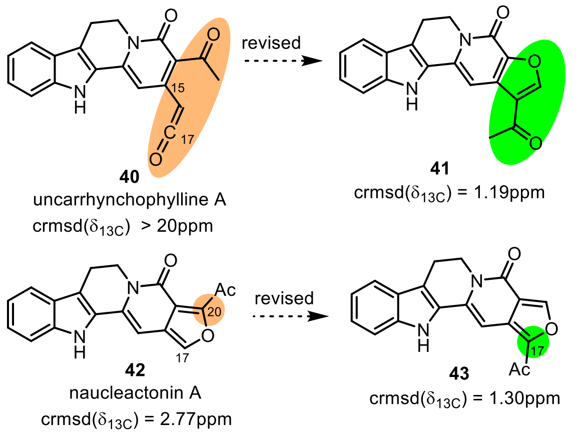


38 panaxolide crmsd(δ_{13C}) = 3.10ppm

39 aplysistatin crmsd(δ_{13C}) = 0.83ppm

Highly reactive moieties offer another red flag of structure misassignment. An unusual alkaloid, uncarrhynchophylline A (**40**), purportedly containing a ketene moiety was isolated from stems of *Uncaria rhynchophylla*.[25] DU8ML analysis showed irreconcilable problems with this structure, which we revised to acetylfuran **41**.

Revised uncarrhynchophylline A (**41**) is a new unreported structure. However, an alkaloid with a high degree of similarity, naucleactonin A (**42**),[26] was previously described. Instructively, according to DU8ML, naucleactonin A also required revision to **43**, i.e. relocation of the acetyl group from C20 to C17.



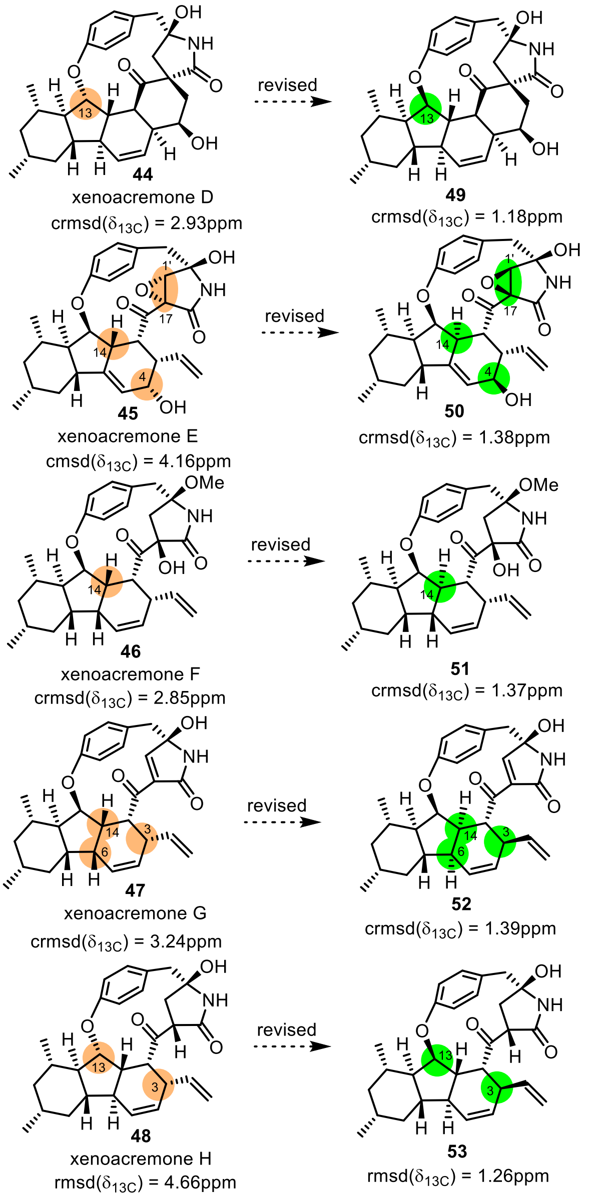
40
uncarrhynchophylline A
crmsd(δ_{13C}) > 20ppm

41
crmsd(δ_{13C}) = 1.19ppm

42
naucleactonin A
crmsd(δ_{13C}) = 2.77ppm

43
crmsd(δ_{13C}) = 1.30ppm

Finally, a group of bioactive tyrosine-decahydrofluorene analogues, xenoacremones D-H (**44**-**48**), were recently isolated from the plant-derived fungus *Xenoacremonium sinensis*.[27] While the structure connectivities in these complex NPs were determined correctly, stereoconfiguration of one or more stereogenic centers in each of these NPs required revision.



44
xenoacremone D
crmsd(δ_{13C}) = 2.93ppm

49
crmsd(δ_{13C}) = 1.18ppm

45
xenoacremone E
cmsd(δ_{13C}) = 4.16ppm

50
crmsd(δ_{13C}) = 1.38ppm

46
xenoacremone F
crmsd(δ_{13C}) = 2.85ppm

51
crmsd(δ_{13C}) = 1.37ppm

47
xenoacremone G
crmsd(δ_{13C}) = 3.24ppm

52
crmsd(δ_{13C}) = 1.39ppm

48
xenoacremone H
rmsd(δ_{13C}) = 4.66ppm

53
rmsd(δ_{13C}) = 1.26ppm

The case of xenoacremones is particularly challenging because of structure distortions related to the *p*-methylenephenoxy bridge linking positions 13 and 2'. The cyclic C2'-hemiaminal moiety further aggravated the situation. DU8ML-guided analysis helped us revise xenoacremone D (**44**) to 13-epimer **49**; xenoacremone E (**45**) – to diastereomer **50**; xenoacremone F (**46**) – to 14-epimer **51**; xenoacremone G (**47**) – to diastereomer **52**; and xenoacremone H (**48**) – to diastereomer **53**. It appears that stereoconfiguration of centers C7,C9,C11,C12, and C13 is "conserved," unlike the other stereogenic centers, e.g., C3,C6, C14, and C15.

Conclusions

DU8ML offers ample cost-benefit outcomes, allowing for fast and accurate computations of NMR parameters, and thus enabling *high-throughput* validation or revision of published structures on a massive scale. However, no method of structure elucidation is perfect. While most of the computational NMR methods these days are reasonably reliable in flagging potential misassignments, the proposed revisions are what they are – the most likely structures. This overview is an invitation to a discussion of how to make structure elucidation by NMR more reliable and learn to recognize most challenging aspects of assigning connectivity and stereochemistry in newly isolated natural products.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work is supported by the NSF, CHE-1955892.

Notes and References

---

1 Fisher, J. Modern NMR Techniques for Synthetic Chemistry, CRC Press, **2021**.

2 *DU8ML in a nutshell:* fast DFT calculations of chemical shifts and spin-spin coupling constants at the light basis sets are augmented by post-processing of the calculated chemical shift values with empirically derived corrections via SMARTS labeling of substructure fragments known to give large deviations from the experimental values. For details see: I. M. Novitskiy, A. G. Kutateladze. *J. Org. Chem.*, 2022, **87**, 4818 and references therein.

3 C. H. K. Momo, A. D. K. Mboussaah, N. F. Zambou, M. A. Shaiq. *Nat. Prod. Res.* 2022, **36**, 909.

4 R. A. Hill, A. Sutherland. *Nat. Prod. Rep.* 2022, https://doi-org.du.idm.oclc.org/10.1039/D2NP90011D

5 M. D'Ambola, L. Fiengo, M. G. Chini, R. Cotugno, A. Bader, G. Bifulco, A. Braca, N. De Tommasi, F. Dal Piaz. *J. Nat. Prod.* 2019, **82**, 539.

6 Y. Sun, C.-M. Yuan, S.-Y. Xu, Y. Li, X.-B. Yang, T.-C. Wen, K. Zhou. *Tetrahedron Lett.* 2022, **93**, 153680.

7 D.-L. Chen, M.-Y. Chen, Y. Hou, C.-H. Wang, Zh.-C. Sun, Y. Yang, H.-Q. Liang, G.-X. Ma, X.-D. X., J.-H. Wei. *J. Nat. Prod.* 2022*,* **85**, 127.

8 D.-Y. Li, P. Han, Zh.-D. Wu, Ch.-M. Chen, Q.-Y. Tong, H.-Ch. Zhu, J.-R. Guo, Y.-J. Lai, Y.-H. Zhang. *J. Asian Nat. Prod. Res*. 2021, **23**, 429.

9 B.-K. Choi, D.-Y. Cho, D.-K. Choi, P. T. H. Trinh, H. J. Shin. *Mar. Drugs* 2021, **19**, 65.

10 Z. Sadeghi, M. Masullo, A. Cerulli, F. Nazzaro, M. M. Farimani, S. Piacente. *J. Nat. Prod.* 2021, **84**, 26.

11 (a) H. T. Pham, T. P. Doan, H. W. Kim, T. W. Kim, S.-Y. Park, H. Kim, M. Lee, K. H. Kim, W. K. Oh, Y. W. Lim, K. B. Kang. *J. Nat. Prod.* 2022, **85**, 511. (b) notice that the authors named compound **21** 6-hydroxysterpuric acid, matching their abstract graphics; however, the figures in the text are clearly depicting 5-hydroxysterpuric acid, which it is.

12 J.-Y. Li, X.-P. Shen, H. Chen, S.-S. Li, W.-D. Zhang, Y.-H. Shen. *Phytochem. Lett.* 2020, **50**, 10.

13 X. Han, X. Luo, L. Xue, L. van Ofwegen, W. Zhang, K. Liu, Y. Zhang, X. Tang, P. Li, G. Li. *J. Nat. Prod.* 2022, **85**, 276.

14 Z.-H. Huang, X. Liang, Ch.-J. Li, Q. Gu, X. Ma, S.-H. Qi. *J. Nat. Prod.* 2021*,* **84**, 2727.

15 T. Yanagimoto, S. Yamada, Y. Kasai, H. Yamamoto, M. Kubo, Y. Fukuyama, H. Imagawa. *Tetrahedron Lett.* 2020, **64**, 152723.

16 H. Kikuchi, Y. Miyagawa, Y. Sahashi, S. Inatomi, A. Haganuma, N. Nakahata, Y. Oshima. *J. Org. Chem.* 2004, **69**, 352.

17 we thank Prof. Hiroshi Imagawa, Tokushima Bunri University, for a private communication that our computational revision of spirotenuipesine B is confirmed by x-ray crystallography in his lab.

18 Kutateladze, A.G.; Kuznetsov, D.M.; Beloglazkina, A. A.; Holt, T. *J. Org. Chem.*, 2018, **83**, 8341.

19 Y. Yaoita, K. Machida. *Nat. Prod. Comm*. 2020, **15**, 1.

20 H. Gao, K. Hong, G.-D. Chen, Ch.-X. Wang, J.-Sh. Tang, Y. Yu, M.-M. Jiang, M.-M. Li, N.-L. Wang, X.-Sh. Yao. *Magn. Reson. Chem.* 2010, **48**, 38.

21 E. M. Terefe, F. A. Okalebo, S. Derese, J. Muriuki, W. Rotich, E. Mas-Claret, N. Sadgrove, G. F. Padilla-González, T. A. K. Prescott, H. Siddique, M. K. Langat. *J. Nat. Prod.* 2022, https://doi.org/10.1021/acs.jnatprod.1c01013

22 Y. Hou, Q. Li, M. Chen, H. Wu, J. Yang, Z. Sun, X. Xu, G. Ma. *Fitoterapia* 2022, 105251, https://doi.org/10.1016/j.fitote.2022.105251

23 H. K. Le, T. T. Phan, T. m. N. Nguyen, T. C. Nguyen, C. L. Tran, P. E. Hansen, Q. T. That *Nat. Prod. Res.* 2022, https://doi.org/10.1080/14786419.2022.2094374

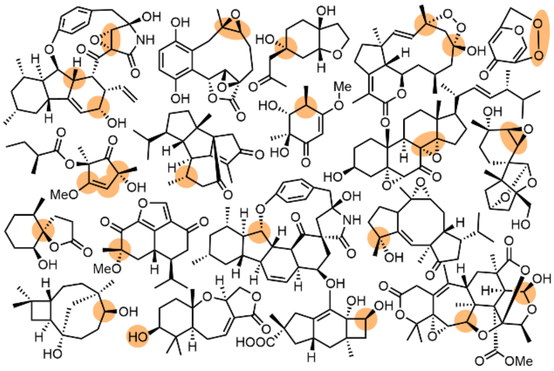
24 G. R. Pettit, C. L. Herald, M. S. Allen, R. B. Von Dreele, L. D. Vanell, J. P. Y. Kao, W. Blake *J. Am. Chem. Soc.* 1977, **99**, 262.

25 R.-F. Li, C.-G. Zhu, C.-B. Xu, Q.-L. Guo, J.-G. Shi *J. Asian Nat. Prod. Res.* 2021, **23**, 513.

26 W.-D. Xuan , H.-S. Chen , J.-L. Du , S. Liang , T.-Z. Li, D.-G. Cai. *J. Asian Nat. Prod. Res.* 2006, **8**, 719.

27 Z. Liu, L. Liu, A. Wang, L. Li, S. Zhao, Y. Wang, Y. Sun. *Mar. Drugs* 2022, **20**, 375.

TOC Graphics



Biographies

Ivan M. Novitskiy is a postdoctoral research associate in the Department of Chemistry and Biochemistry of University of Denver. His current research focuses on the development of photoassisted methods of synthesis and solution structure elucidation of complex organic molecules. He received both his undergraduate and doctorate degrees from Lomonosov Moscow State University in 2014 and 2019.

Andrei G. Kutateladze is John Evans Professor in the Department of Chemistry and Biochemistry and Dean of the College of Natural Sciences and Mathematics of University of Denver. His research interests are in experimental and computational photochemistry, complexity building photoassisted synthesis of polyheterocyclic molecular architectures and solution structure elucidation of complex molecules.