Characterization of Leptazolines A–D, Polar Oxazolines from the Cyanobacterium *Leptolyngbya* sp., Reveals a Glitch with the “Willoughby–Hoye” Scripts for Calculating NMR Chemical Shifts

Jayanti Bhandari Neupane, Ram P. Neupane, Yuheng Luo, Wesley Y. Yoshida, Rui Sun, and Philip G. Williams*

Department of Chemistry, University of Hawai‘i at Mānoa, 2545 McCarthy Mall, Honolulu, Hawaii 96822, United States

Supporting Information

ABSTRACT: The bioactivity-guided examination of a *Leptolyngbya* sp. led to the isolation of leptazolines A–D (1–4), from the culture media, along with two degradation products (5 and 6). Density functional theory nuclear magnetic resonance calculations established the relative configurations of 1 and 2 and revealed that the calculated shifts depended on the operating system when using the “Willoughby–Hoye” Python scripts to streamline the processing of the output files, a previously unrecognized flaw that could lead to incorrect conclusions.

As part of our long-standing interest in cyanobacterial natural products, we recently began screening strains within our culture collection against pancreatic adenocarcinoma (PANCA). This screen flagged several media extracts. Whereas metabolites from cells are well studied, there are comparatively few reports of cyanobacterial secondary metabolites excreted in the culture medium to any appreciable degree. Excreted metabolites are likely to be more polar and structurally different from those isolated from the cell mass. We report here the results from the examination of strain O-2-5, a *Leptolyngbya* sp., that is, the isolation and characterization of leptazolines A–D (1–4) and two hydrolysis products (5 and 6).

The assignment of the relative configurations of 1 and 2 involved calculating nuclear magnetic resonance (NMR) chemical shifts using a widely cited protocol outlined by Willoughby et al., which revealed a surprising operating system dependence on the calculated values due to issues with one of the Python scripts.

The molecular formula of leptazoline A (1) was established as C_{13}H_{14}ClNO_{5} from its high-resolution electrospray ionization mass spectrometry (HRESIMS) signal at *m/z* 300.0637 [M + H]^+ and isotopic patterns indicative of a chlorine atom, that is, a 3:1 ratio of *m/z* 300 and 302. The ^1^H and ^13^C NMR spectra (Table 1 and Table S1; MeOH-d$_4$) revealed the presence of three aromatic proton signals consistent with a 1,2,4-trisubstituted benzene ring, that is, δ$_{\text{H}}$ 7.02 (d, *J* = 8.8 Hz), 7.47 (dd, *J* = 8.8, 2.7 Hz), and 7.53 (d, *J* = 2.7 Hz). The analysis of the NMR data suggested that a chlorine atom, an oxygen atom, and a sp$^2$-hybridized carbon were attached to this ring at C-12, -9, and -8, respectively. These assignments were consistent with the carbon chemical shifts reported for 5-chlorosalicylic acid, with the same substitution pattern, and yielded the lowest mean average error (MAE) of the calculated ^13^C NMR shifts given the other possible isomers (Table S2).

A correlated spectroscopy (COSY) correlation connected the lone methyl doublet (H-6) to H-5 (4.77 ppm), whereas a heteronuclear multiple bond correlation (HMBC) from H-6 to C-4 (δ$_{\text{C}}$ 75.7) as well as a COSY correlation between H-5 and H-4 (δ$_{\text{H}}$ 3.93) extended the chain. A further network of COSY correlations from H-4 to H-3 and then to H-2 completed the
proton-coupled spin system. The large geminal coupling constant (>15 Hz) for the methylene protons H-2 indicated an n-acceptor substituent; presumably, a carbonyl moiety, given the molecular formula, was at C-1, \textsuperscript{173.8} which was supported by an HMBC correlation from H-3 to C-1 (\deltaC 173.8), giving a hexanoyl fragment.

### Table 1. NMR Spectroscopic Data of Leptazoline A (1) in DMSO-\textit{d}_6

<table>
<thead>
<tr>
<th>Compound</th>
<th>(\deltaC) type</th>
<th>(\deltaH) (J)</th>
<th>COSY</th>
<th>HMBC or CIGAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>173.8, C</td>
<td>2.46, dd (15.4, 3.8)</td>
<td>1, 3, 4</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>39.0, CH\textsubscript{2}</td>
<td>2.37, dd (15.3, 8.6)</td>
<td>3, 4, 5</td>
<td>1, 2, 5</td>
</tr>
<tr>
<td>3</td>
<td>68.0, CH</td>
<td>3.99, dt (8.6, 3.8)</td>
<td>2, 4</td>
<td>1, 2, 3, 5</td>
</tr>
<tr>
<td>4</td>
<td>75.7, CH</td>
<td>3.93, dd (6.3, 3.8)</td>
<td>2, 3, 6, 7, 8\textsuperscript{c}</td>
<td>4, 5</td>
</tr>
<tr>
<td>5</td>
<td>77.7, CH</td>
<td>4.77, p (6.3)</td>
<td>4, 6</td>
<td>3, 4, 7</td>
</tr>
<tr>
<td>6</td>
<td>20.5, CH\textsubscript{2}</td>
<td>1.38, d (6.3)</td>
<td>4, 5</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>163.4, C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>111.7, C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>158.2, C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>118.5, CH</td>
<td>7.02, d (8.8)</td>
<td>11</td>
<td>7, 8, 9, 12, 13, 12\textsuperscript{d}</td>
</tr>
<tr>
<td>11</td>
<td>133.2, CH</td>
<td>7.47, dd (8.8, 2.7)</td>
<td>8, 9, 12, 13</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>122.2, C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>126.6, CH</td>
<td>7.53, d (2.7)</td>
<td>11\textsuperscript{d}</td>
<td>7, 9, 10, 11, 12</td>
</tr>
<tr>
<td>NH</td>
<td>–170.4\textsuperscript{a,b}</td>
<td>12.28, brs</td>
<td>11\textsuperscript{d}</td>
<td>7, 9, 10, 11, 12</td>
</tr>
</tbody>
</table>

\(\textsuperscript{a}\)Nitrogen chemical shift from \(\text{H}^\text{\textit{1}}\text{N} \text{HMBC} \delta^{15}\text{N} \text{ referenced to MeNO}_2 \text{ (IUPAC std). Referenced to NH}_3, \text{ it resonates at 210.1 ppm} \text{.} \textsuperscript{b}\)CIGAR only. \textsuperscript{c}Weak correlations.

The hexanoyl fragment and the 5-chlorosalicylic acid derivative were connected by an oxygen atom between C-5 and C-7, based on their chemical shifts, and an HMBC correlation from H-5 to C-5. The molecular formula of the molecule was C\textsubscript{16}H\textsubscript{21}ClN\textsubscript{2}O\textsubscript{7} and the molecular ion was observed at m/z 404.1 in the high-resolution electrospray ionization mass spectrometry (HRESIMS) spectrum. The oxygenated carbons C-2 and C-7 were connected via an oxygen atom and C-3 and C-9 had acetylated, as shown by the HMBC correlation from H-3 to C-1 (\deltaC 173.8), giving a hexanoyl fragment.

The unassigned ring must therefore cyclize via some combination of C-1, C-4, and C-7. Thus the main structural framework of the molecule was established, but the additional heteroatoms, exchangeable protons, and the remaining degree of unsaturation—a ring—could not be placed with certainty based on the available spectroscopic data. The carbon chemical shifts of C-3, C-4, and C-7, based on their chemical shifts, and an HMBC correlation from H-3 to a nitrogen atom resonating at 210.1 ppm suggested the attachment to an oxygen atom, but it was also in reasonable agreement for a carbon atom in an oxazoline ring (Figure 1).

Whereas the cyanobacteria of the genus \textit{Leptolyngbya} are known to produce various bioactive metabolites, such as the palmyrolides\textsuperscript{11} and the coibamides,\textsuperscript{12} it does not appear that oxazoline-containing compounds have been previously reported from this genus.

Because of the degradation of I, the relative configuration was established through density functional theory (DFT) calculations of the NMR chemical shifts.\textsuperscript{2} The lowest energy conformers for each diastereomer of I were identified using MacroModel (Maestro/Schrodinger), optimized in Gaussian 09,\textsuperscript{13} and their zero-point energy calculated before the NMR shielding tensors of the \(\text{H}^\text{1}\) and the \(\text{C}^\text{13}\) nuclei were computed. The chemical shift values for each diastereomer were obtained using published intercepts and slopes in consideration of the Boltzmann distribution of the conformers.

A comparison of the \(\text{H}^\text{1}\) and \(\text{C}^\text{13}\) chemical shifts (Table S3) shows that the experimental values are most closely aligned with the computed values for the \(3\text{R}^\text{*},4\text{S}^\text{*},5\text{S}^\text{*}\)-diastereomer (1d). On the basis of the DP4+ probabilities\textsuperscript{14} computed from these data sets (Table S3), the predicted NMR shifts of the diastereomer 1d deviate the least from the experimental data (96.7% probability considering \(\text{H}^\text{1}\) and \(\text{C}^\text{13}\)). These calculations suggest that the \(3\text{R}^\text{*},4\text{S}^\text{*},5\text{S}^\text{*}\)-diastereomer (1d) reflects the correct relative configuration of compound I.

The molecular formula of 2 was established as C\textsubscript{16}H\textsubscript{14}ClN\textsubscript{2}O\textsubscript{5} from its HRESIMS signal at m/z 389.1110 [M + H]\textsuperscript{+}. The spectral data for 2 (Table S4) had many similarities to those of I, but there were a few noticeable differences, including an additional C\textsubscript{12}H\textsubscript{7}NO\textsubscript{3} in the molecular formula. An analysis of the 2D NMR data established that C-2 (7.35 ppm) was oxygenated and C-1 and C-1' were linked through an amide bond. Evidence to support the amide bond includes the absence of the broad O–H bond observed in 1 in the infrared (IR) spectrum (see Figure S22) and a second nitrogen atom observed at ~263.0 ppm in the \(\text{H}^\text{1}–15\text{N}\) HMBC NMR spectrum. Oxygenated carbons C-2' and...
C-3’ were magnetically equivalent and attached to C-1’ based on COSY correlations from the respective protons, which formed a 2-aminopropan-1,3-diol moiety that has been found in other natural products.15,16 These changes accounted for the difference in the molecular formula.

The assignment of the relative configuration of 2 relied on calculated NMR shifts. Whereas we hypothesized that the three chiral centers shared with 1 would remain unchanged in 2, we were able to compute the 1H and 13C NMR chemical shifts of all eight diastereomers using minimal computing resources by truncating to the primary amide (2a–h), which reduced the number of conformers of each diastereomer from >250 to <10. The calculated data for the diastereomers 2a–h and their DP4+ probabilities suggest conflicting results when 1H and 13C chemical shifts (see Table S5) are considered separately, but 2d, which corresponds to (2S*,3S*,4S*,5S*)-2, appears as the clear choice when both data sets are taken into account.

Analogs 3–6 were also isolated. Although there was insufficient material to record 13C NMR data, the planar structures and relative configurations could be determined from the comparison of their high-resolution mass spectrometry (HRMS) and 1H NMR data with 1 to 2. For example, 3 and 4 had 1H NMR spectra indicative of a 1,2-disubstituted benzene ring without a chlorine atom, which was consistent with the MS data. The hydrolyzed products 5 and 6 were determined from the characteristic increase of 18 mass units in the mass spectra and the examination of the 1D and 2D NMR data and are likely the result of the inclusion of formic acid in the high-performance liquid chromatography (HPLC) mobile phase.

Compounds 1 and 2 were isolated in sufficient quantities to be evaluated in our PANC-1 assay system. Whereas 1 did not significantly inhibit the proliferation of Panc-1 cells, 2 did show the inhibition of growth (GI50 10 μM). Further biological evaluation was, however, hindered due to the degradation of the molecules.

While preparing a manuscript, to our surprise, attempts by team members to replicate these results produced different calculated NMR chemical shifts despite using the same Gaussian files and the same procedure outlined by Willoughby et al. For example, all attempts concluded that 1d was the correct diastereomer, but these conclusions were based on chemical shifts that appeared to depend on the computer system on which step 15 of that protocol was performed (Table 2).

Published in 2014, this Nature Protocols manuscript provides detailed instructions aimed at enabling those with minimal theoretical knowledge of the subject area to calculate gauge-including atomic orbital (GIAO) NMR chemical shifts and includes Python scripts to streamline the process. It has been cited over 130 times in the last 5 years. Detailed investigations traced the source of the discrepancies to those Python scripts that summarize the conformers, free energies, Boltzmann distributions, and isotopic tensors from the various Gaussian output files and then use this information to calculate Boltzmann averaged chemical shifts (step 15). Specifically, the script “nmr-data compilation” extracts the free-energy information of each conformer and the isotopic tensors for each atom from two files, for example, 1b-opt_freq-conf-16.out and 1b-opt_NMR-conf-16.out, respectively, and performs the necessary calculations.

In theory, the chemical shift of each atom from each isomer should be properly weighted according to the free energy of the conformer, but this was not consistently the case. In the end, the inconsistency was traced to differences in the default file-sorting algorithm in Python across platforms, as shown in Figure 2, and the fact that, as written, the script “nmr-data Compilation” assumes that the frequency and NMR files are sorted in the same order. The Boltzmann contribution calculated from the first frequency file in the directory is simply applied to the chemical shifts calculated in the first NMR file. For example, in Windows 10, the script works as designed: Two groups of output files, *opt_freq-*.ID.out containing free energies and *nmr-*.ID.out containing chemical shifts, are automatically sorted by the system such that the conformers are paired correctly (Figure 2B).

However, in LINUX there is no default sorting of file names because this depends on the local settings, and the scripts, as designed, do not check that the frequency and the NMR files are properly matched. As a result, the free energies could be paired with chemical shifts from different conformers (Figure 2A), which leads to incorrectly calculated chemical shifts that could surely lead to a wrong final conclusion.

To overcome this issue, we amended the script to include a line in the read_gaussian_outputfiles subroutine that forces

| Table 2. Variability in Calculated Carbon Chemical Shifts of 1b |
|-----------------|-----------------|-----------------|-----------------|
| no.             | LINUX (Ubuntu16) | Windows (ver. 10) | Mac (Mavericks) | Mac (Mojave) |
| 1               | 172.4            | 173.2            | 173.2           | 172.7        |
| 2               | 36.0             | 37.7             | 37.7            | 39.3         |
| 3               | 68.3             | 68.4             | 68.4            | 69.0         |
| 4               | 70.6             | 70.5             | 70.5            | 71.2         |
| 5               | 79.0             | 78.4             | 78.4            | 79.0         |
| 6               | 15.5             | 13.3             | 13.3            | 13.3         |
| 7               | 162.4            | 162.5            | 162.5           | 161.8        |
| 8               | 110.4            | 109.8            | 109.8           | 110.3        |
| 9               | 155.0            | 156.5            | 156.5           | 155.5        |
| 10              | 116.2            | 115.5            | 115.5           | 116.0        |
| 11              | 131.6            | 131.6            | 131.6           | 131.7        |
| 12              | 127.1            | 126.6            | 126.6           | 127.0        |
| 13              | 126.7            | 125.6            | 125.6           | 126.3        |

Figure 2. Inconsistent sorting of files.
sorting before pairing and a longer file-matching check function that alerts the user when there is a potential file-matching issue (see the Supporting Information). We have tested the revised script across platforms (i.e., MacOS, LINUX, and Windows) running different versions of Python, and it yields consistent results.

This simple glitch in the original script calls into question the conclusions of a significant number of papers on a wide range of topics in a way that cannot be easily resolved from published information because the operating system is rarely mentioned. In the first half of 2019 alone, the protocol was referenced/used during the elucidation of several natural products, and it yields consistent results.

Authors who used these scripts should certainly double-check their results and any relevant conclusions using the modified scripts in the SI. Ultimately, this example serves as a reminder of the principle Caveat emptor and that users should validate noncommercial software on their system prior to use on new applications.

**ASSOCIATED CONTENT**

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.orglett.9b03216.

General experimental details, NMR spectra for 1−6, NMR calculation for 2, and picture of the producing organism (PDF)

Raw NMR data for all new compounds (ZIP)

Revised NMR-data Compilation Python scripts and instructions (ZIP)

**AUTHOR INFORMATION**

Corresponding Author

*E-mail: philipwi@hawaii.edu.

ORCID

Ram P. Neupane: 0000-0002-9310-7426

Rui Sun: 0000-0003-0638-1353

Philip G. Williams: 0000-0001-8987-0683

**Author Contributions**

J.B.N. and P.G.W. designed the experiments and analyzed the NMR and biological testing data. W.Y.Y. acquired the NMR data and assisted with its interpretation. R.P.N., Y.L., and R.S. performed and analyzed the NMR calculations. R.S. and Y.L. revised the original Python scripts.

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

We gratefully acknowledge the advanced computing resources provided by the University of Hawaii Information Technology Service Cyberinfrastructure. We thank Dr. Charles O’Kelly (Cyanotech, HI) for the taxonomic analysis. This work was supported by NIH grant S001AG039468 to P.G.W. Funds for the upgrades of the NMR instrumentation were provided by the CRIF program of the National Science Foundation (CH E9974921), the Elsa Pardee Foundation, and the University of Hawaii at Manoa. The purchase of the Agilent TOF LC-MS was funded by grant W911NF-04-1-0344 from the Department of Defense, and the purchase of the Agilent QTOF LC-MS was funded by MRI grant 1532310 from the National Science Foundation. We thank T. Hoye (University of Minnesota) and P. Willoughby (Ripon College) for the helpful discussions about the Python scripts.

**REFERENCES**


3. Compounds 1−6 were never detected by LC−MS in the cell mass, which was also inactive in the original screen against PANC1.


9. Typical ¹⁵N NMR chemical shifts for amides are between −210 and −300 ppm. For an amine, typical shifts are between −280 and −380 ppm.

10. Initial data were collected in CD₃OD (see Table S1), and hence this signal was not detected until later.


(17) The corresponding author of ref 3 has written to us the following, “My coauthors and I are pleased to see that this flaw in our original script has been uncovered. We appreciate that the researchers here shared this information with us prior to publication and fully endorse their publishing the revised script here. Once this publication has appeared, we plan to update and correct the original protocol (ref 3) to further minimize the chance of compromise of future use of the method by other investigators. -T.R.H.”

(18) Glitch, a term of art in computer science, describes the unexpected result of a minor malfunction.


