The Case for CASE—Saving Time and Effort with Computer-Assisted Structure Elucidation in Routine Analysis

Mikhail Elyashberg, Patrick Wheeler, Arvin Moser, and Steve Hayward

Advanced Chemistry Development, Inc.
Toronto, ON, Canada
www.acdlabs.com

Introduction

Too often, the elucidation of unknown structures, especially those with novel moieties found in natural products, resulted in incorrect initially published structures, which then require either exhaustive spectroscopic analysis, full chemical synthesis, or both to prove the correct structure. In many cases, both the initial incorrect structure and subsequent analytical work could be avoided by employing Computer-Assisted Structure Elucidation (CASE) methods alongside the original investigations; many examples of which we have previously reported on [1].

In this work, we apply standard CASE methods to one recent example, and also show how ACD/Structure Elucidator Suite can determine the best structure, based on one proposed structure with assigned $^{13}$C and $^1$H chemical shifts.

Background

Julianti et al. [2] isolated a novel modified base, acremonin (1), from the culture broth of the marine fungus acremonium stric tum.

Based on the original combined spectroscopic analyses, the structure of this compound was determined to be a methyl guanine base containing an isoprene unit. The presence of a $^1$H-azirine moiety is unprecedented among natural products, as noted by the authors. [2]
Januar and Molinski [3] were intrigued by this report, as the structure 1 assigned to acremol seemed highly surprising to them. They suggested that such an unstable compound would be unlikely to exist in nature, and hypothesized that the acremol molecule could have one of two possible structures—2 or 3:

Independently, Banert [4] proposed the same hypothesis and suggested that structure 2 better fits the NMR data for acremol, based on analysis of predicted $^{13}$C NMR chemical shift increments for 2 and 3, but without experimental evidence.

**Figure 1.** Synthesis performed by Januar and Molinski.
To resolve this issue, Januar and Molinski performed a synthesis of compound 2 and tried (without success) to synthesize compound 3 based on the N2,3-ethenoguanine skeleton. The scheme for a five-step synthesis suggested by researchers is shown in Figure 1. (Structure 2 is denoted as 5a and structure 3 as 5b.)

It turned out that the $^1$H NMR (DMSO-d$_6$, 500 MHz) and $^{13}$C NMR spectra of 2 (5a) and other data (UV-VIS, FTIR, HMBC, and HRMS) were identical in every respect with those reported by Julianti and co-workers for acremolin. [2] However compound 5a raised questions of regioisomerism. The location of the isopropyl group in 5a followed from the mechanism of formation of N2,3-ethenoguanine bases. Conceivably, each of the isomers 5a and 5b could be formed.

Doubts regarding the structures of 5a and 5b were removed upon analysis of the experimental 5a $^1$H-$^{15}$N HMBC spectrum (Figure 2)—where critical correlations colored in red confirm regioisomer 5a and exclude 5b.

**Figure 2.** $^1$H-$^{15}$N HMBC correlations for structure 5a.

This total synthesis of a hypothetical alternative structure 2 (5a) and its spectroscopic confirmation represented a huge amount of work done by the authors [3] to disprove the original structure 1. As such, we attempted to answer the following question: what solution would be obtained if the original authors [2] used ACD/Structure Elucidator Suite for processing the spectroscopic data of the unknown?
Applying CASE

With the above question in mind, we used \(^1\)H and \(^{13}\)C NMR and \(^1\)H-\(^{13}\)C HSQC and HMBC data of acremolin into ACD/Structure Elucidator Suite (see Table 1).

**Table 1.** Spectroscopic NMR data used for structure elucidation of acremolin.

<table>
<thead>
<tr>
<th>Label</th>
<th>(\delta^C)</th>
<th>XHn</th>
<th>(\delta^H)</th>
<th>M(J)</th>
<th>C HMBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>C 1</td>
<td>142.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 2</td>
<td>141.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 3</td>
<td>108.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 4</td>
<td>152.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 5</td>
<td>140.5</td>
<td>CH</td>
<td>8.16</td>
<td>u</td>
<td>C 3, C 4, C 2</td>
</tr>
<tr>
<td>C 6</td>
<td>28.9</td>
<td>(\text{CH}_3)</td>
<td>3.57</td>
<td>u</td>
<td>C 1, C 2</td>
</tr>
<tr>
<td>C 7</td>
<td>103.2</td>
<td>(\text{CH, NH})</td>
<td>7.38</td>
<td>u</td>
<td>C 8, C 1</td>
</tr>
<tr>
<td>C 8</td>
<td>148</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C 9</td>
<td>27.7</td>
<td>CH</td>
<td>2.88</td>
<td>u</td>
<td>C 8, C 10, C 7</td>
</tr>
<tr>
<td>C 10</td>
<td>22.1</td>
<td>(\text{CH}_3)</td>
<td>1.25</td>
<td>u</td>
<td>C 9, C 8</td>
</tr>
</tbody>
</table>

From the input data, a Molecular Connectivity Diagram (MCD, Figure 3) was created:

![Molecular Connectivity Diagram](image)

**Figure 3.** Molecular Connectivity Diagram (MCD) for acremolin. \(^{13}\)C chemical shifts are shown.

No manual edits of the initial, automatically created MCD were made except one: a label “ob” (obligatory bond to a heteroatom) was removed from C 103.2 to avoid unnecessary constraints on structure generation. The software checked the MCD for any inconsistencies, and passed. Structure
Generation was initiated. The program produced 81 exotic structures with large average chemical shift deviations\(^1\) (the best one with \(d=3-4 \text{ ppm}\)) in 2 seconds. This result hinted at the presence of latent nonstandard correlations (NSC). Therefore the next run was performed using Fuzzy Generation Mode with parameters \(m=1, a=1\), i.e., we checked the simplest hypothesis that HMBC data contained at least one NSC of \(4J_{\text{CH}}\) type. The result: \(k = 17004 \rightarrow 12468 \rightarrow 498, t_p = 1 \text{m 43 s.}\)

Next we performed \(^{13}\text{C}\) chemical shift prediction and ranked the output file in ascending order of average deviation values. Both the original structure 1 and the conceivable regioisomer of the revised structure 2 (5a) were generated and saved. Their positions in the ranked structural file are shown in Figure 4.

The revised structure 2 (5a) was ranked second, while its tautomer had minimal calculated deviations and was placed in the first position. The original structure 1 was placed in 10\(^{th}\) position with a difference in calculated chemical shift deviations between the two structures \(\Delta = d(10) - d(1) \sim 3 \text{ ppm}\). Therefore the wrong original structure 1 was reliably rejected by Structure Elucidator Suite. The competing regioisomer 5b of the correct structure was placed in 15\(^{th}\) position and also was rejected by the software.

Our investigation showed that application of Structure Elucidator Suite could determine the correct acremolin structure with a high reliability, almost instantly and without the application of \(^1\text{H}-^{15}\text{N}\) HMBC

---

\(^1\) Abbreviations for structural rankings:
- \(d_A\) = average deviation of the calculated \(^{13}\text{C}\) spectra using the fragment method
- \(d_I\) = average deviation of the calculated \(^{13}\text{C}\) spectra using the incremental method
- \(d_N\) = average deviation of the calculated \(^{13}\text{C}\) spectra using the neural net method
- \(\text{max}_A\) = maximum \(^{13}\text{C}\) chemical shift deviation using the fragment method
data. If the program were used from the very beginning, structures 1 and 5b would rank poorly [2] due to the very large average deviations calculated for them. Moreover, arduous labor—the five-step synthesis of the revised structure—would be unnecessary for the authors [3], saving the time of highly qualified chemists.

For completeness, we added the $^1$H-$^{15}$N HMBC correlations shown in Figure 2 to the set of experimental data and created a new MCD (Figure 5).

![New Molecular Connectivity Diagram](image)

**Figure 5.** New Molecular Connectivity Diagram. Connectivities corresponding to $^1$H-$^{13}$C and $^1$H-$^{15}$N HMBC are shown.

Fuzzy Structure Generation from this new MCD gave the following results: $k=67 \rightarrow 65 \rightarrow 55$, $t_g = 0.5$ s. As expected, the additional structural information carried by H-N HMBC led to a dramatic reduction in both the number of generated structures and processing time (from 100 sec to 0.5 sec). The first two structures shown in Figure 4 and competing regioisomer 3 were again ranked in the same order, but the original structure 1 was not generated because H-N HMBC correlations contradict it. The structures of regioisomer 3 with $^{15}$N experimental and calculated (3a) chemical shifts and connectivities are displayed below:

![Structures 3 and 3a](image)
We see that structure 3 can be rejected not only by its large average $^{13}$C deviations (~5 ppm), but also by the large deviations calculated for nitrogen atoms N221.9 and N166.1.

**An Easier Solution?**

We have demonstrated how the acremolin structure was determined using standard CASE methodology. However we can also examine the simplest way to disprove the original structure 1 and find the correct one, using the command Structure/Create Project for Structure. The command was applied after drawing structure 1, where $^{13}$C and $^1$H chemical shifts were assigned. As a result an artificial MCD was created (Figure 6):

In this MCD, all theoretically possible HMBC correlations were drawn by the program. The question was posed: which structures except structure 1 will be generated from the artificial MCD? Structure generation gave the following results: $k=96 \rightarrow 4$, $t_g=0.06$ s, and the ranked output file is presented in Figure 7.
Figure 7. The ranked output file generated from the artificial MCD. All theoretically possible connectivities are shown.
Figure 7 gives answers to all of the questions which were posed by Januar and Molinski. [3] The complete solution to the problem was found in a second while requiring neither synthesis nor additional NMR experiments, and without the application of standard CASE methodology.

In conclusion, we believe that although the classic dictum “synthesis is the ultimate proof of structure” continues to hold true, it should be complemented by the following statement: before starting total synthesis for structure revision it is very desirable to take into account results delivered by a CASE system.

To learn more about ACD/Structure Elucidator Suite, please visit www.acdlabs.com/se.

A complete elucidation package that speeds up the elucidation process and ensures that no candidate is overlooked.

Abbreviations

CASE: Computer-Assisted Structure Elucidation
MCD: Molecular Connectivity Diagram
d_A: average deviation of the calculated $^{13}$C spectra using the fragment method
d_I: average deviation of the calculated $^{13}$C spectra using the incremental method
d_N: average deviation of the calculated $^{13}$C spectra using the neural net method
max_d_A: maximum $^{13}$C chemical shift deviation using the fragment method

References