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

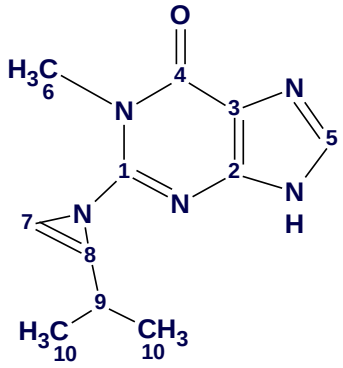
Introduction

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

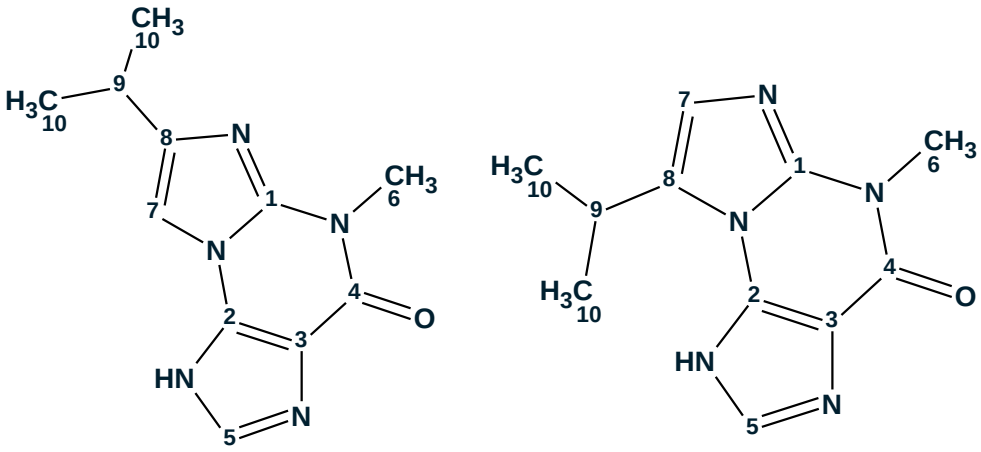
Julianti *et al.* [2] isolated a novel modified base, acremolin (1), from the culture broth of the marine fungus acremonium strictum.



1

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 acremolin seemed highly surprising to them. They suggested that such an unstable compound would be unlikely to exist in nature, and hypothesized that the acremolin molecule could have one of two possible structures—2 or 3:



2 3

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

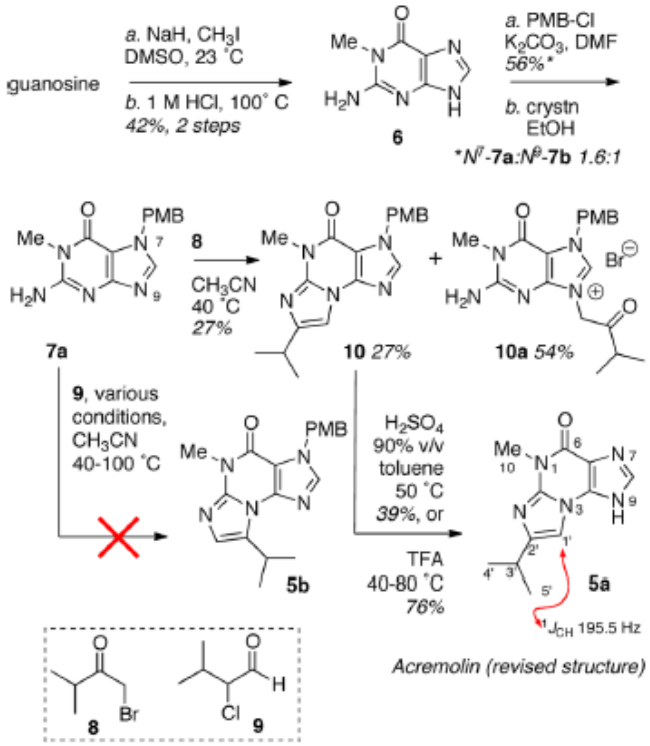


Figure 1. Synthesis performed by Januar and Molinski.

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**.

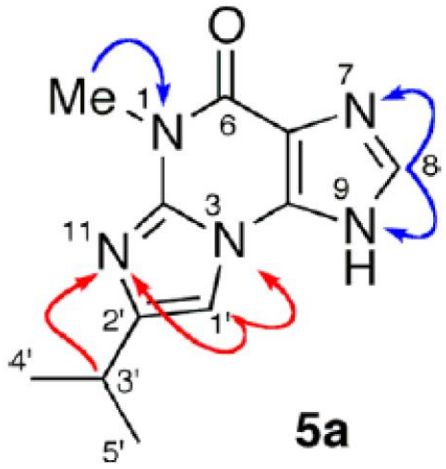


Figure 2. ^{1}H-^{15}N HMBC correlations for structure **5a**.

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.

Label	δC	XHn	δH	M(J)	C HMBC
C 1	142.3	C			
C 2	141.6	C			
C 3	108.9	C			
C 4	152.8	C			
C 5	140.5	CH	8.16	u	C 3, C 4, C 2
C 6	28.9	CH_3	3.57	u	C 1, C 4
C 7	103.2	CH, NH	7.38	u	C 8, C 1
C 8	148	C			
C 9	27.7	CH	2.88	u	C 8, C 10, C 7
C 10	22.1	CH_3	1.25	u	C 9, C 8

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

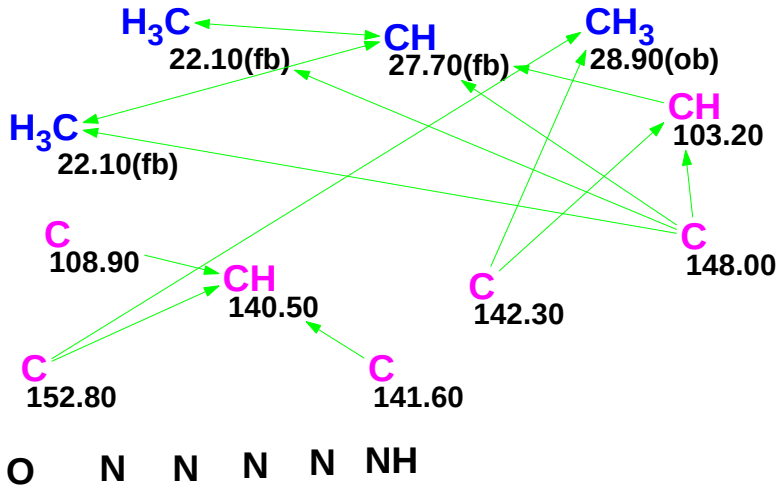


Figure 3. Molecular Connectivity Diagram (MCD) for acremolin. ^{13}C chemical shifts are shown.

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 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 $4J_{CH}$ type. The result: k = 17004 → 12468 → 498, t_g= 1 m 43 s.

Next we performed ^{13}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.

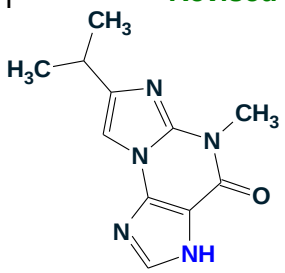
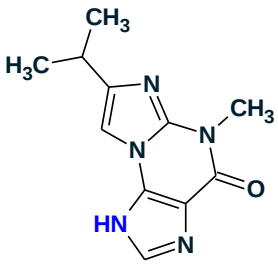
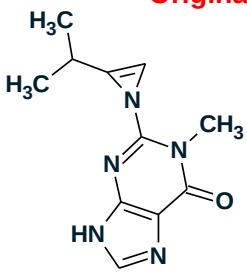
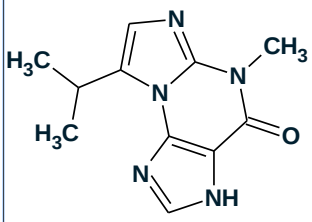
1 Revised	2 Revised	10 Original	15 Competing
			
$d_A(^{13}C)$: 1.430 $d_N(^{13}C)$: 1.827 $d_I(^{13}C)$: 1.919 max_$d_A(^{13}C)$: 3.160	$d_A(^{13}C)$: 1.482 $d_N(^{13}C)$: 2.941 $d_I(^{13}C)$: 3.498 max_$d_A(^{13}C)$: 6.490	$d_A(^{13}C)$: 4.213 $d_N(^{13}C)$: 5.417 $d_I(^{13}C)$: 4.650 max_$d_A(^{13}C)$: 13.860	$d_A(^{13}C)$: 4.768 $d_N(^{13}C)$: 4.611 $d_I(^{13}C)$: 5.117 max_$d_A(^{13}C)$: 21.900

Figure 4. Selected structures of the ranked output file.[1]

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 10th position with a difference in calculated chemical shift deviations between the two structures Δ=d(10)-d(1) ~ 3 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 15th 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}H-^{15}N HMBC

[1] Abbreviations for structural rankings:
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

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).

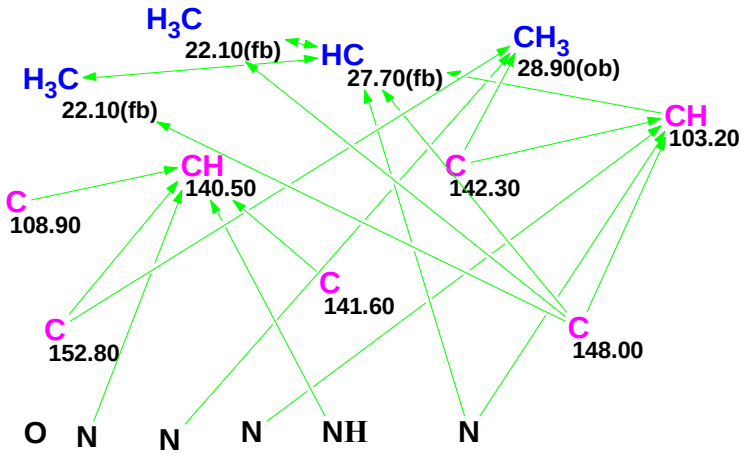
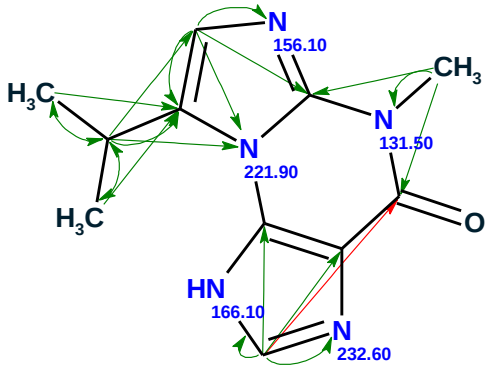
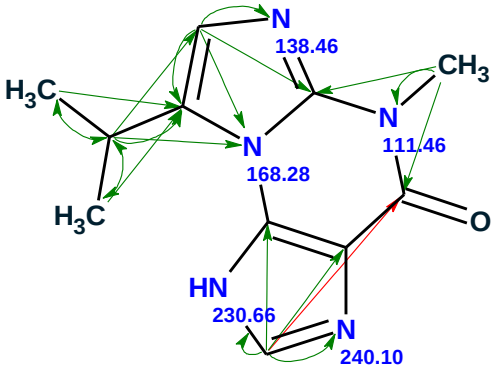


Figure 5. New Molecular Connectivity Diagram. Connectivities corresponding to ^{1}H-^{13}C and ^{1}H-^{15}N HMBC are shown.

Fuzzy Structure Generation from this new MCD gave the following results: k=67 → 65 → 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:



3



3a

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):

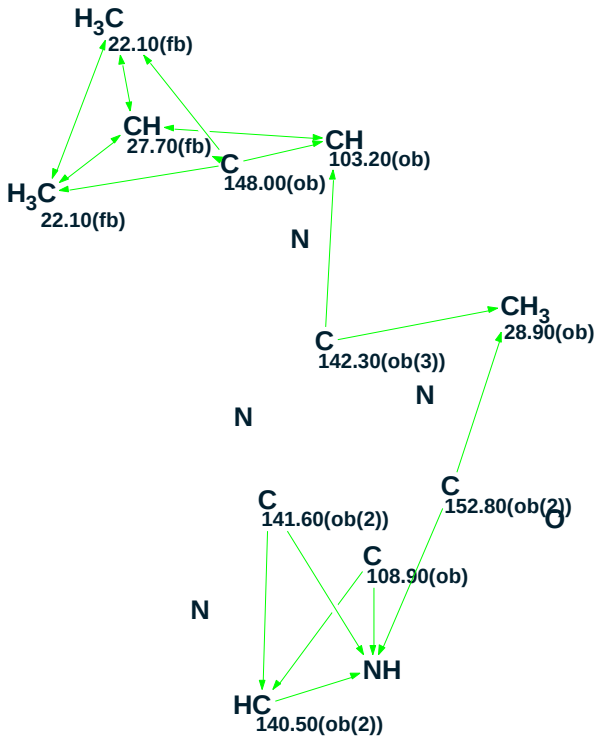


Figure 6. Artificial MCD created from structure **1**.

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 → 4, t_g= 0.06 s, and the ranked output file is presented in Figure 7.

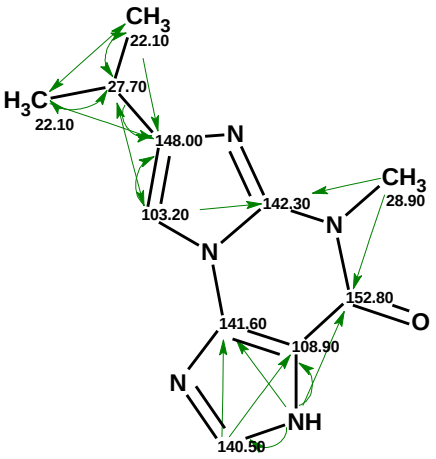
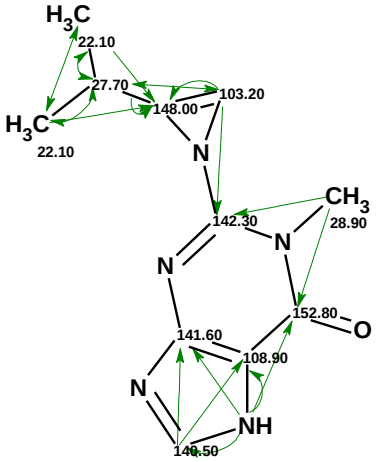
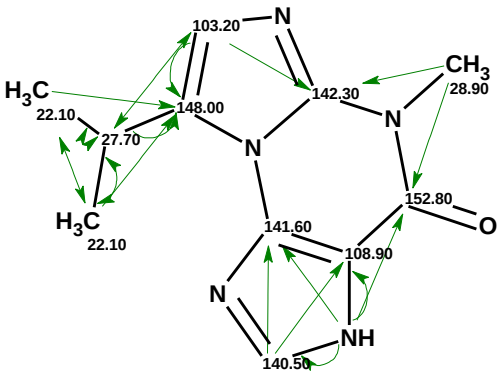
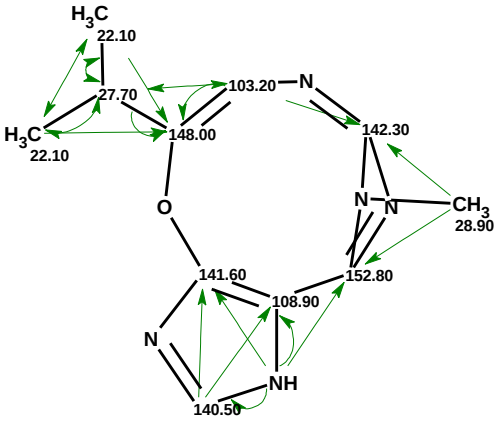
1  	2 
$d_A(^{13}C)$: 1.430 $d_N(^{13}C)$: 1.827 $d_I(^{13}C)$: 1.919 max_$d_A(^{13}C)$: 3.160	$d_A(^{13}C)$: 4.769 $d_N(^{13}C)$: 5.779 $d_I(^{13}C)$: 4.994 max_$d_A(^{13}C)$: 13.860
3 (ID:2)  	4 (ID:1) 
$d_A(^{13}C)$: 4.768 $d_N(^{13}C)$: 4.611 $d_I(^{13}C)$: 5.117 max_$d_A(^{13}C)$: 21.900	$d_A(^{13}C)$: 7.555 $d_N(^{13}C)$: 7.951 $d_I(^{13}C)$: 6.522 max_$d_A(^{13}C)$: 21.640

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.**

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

1. Elyashberg, M.; Williams, A. J.; Blinov, K. **(2010)** Structural revisions of natural products by Computer-Assisted Structure Elucidation (CASE) systems. *Nat. Prod. Reports, 27*:1296–1328.
2. Julianti, E.; Oh, H.; Lee, H.-S.; Oh, D.-C.; Oh, K.-B.; Shin, J. **(2012)** Acremolin, a new 1H-azirine metabolite from the marine-derived fungus Acremonium strictum. *Tetrahedron Lett., 53*:2885–2886.
3. Januar, L. A.; Molinski T. F. **(2013)** Acremolin from Acremonium strictum is N^2,3-Etheno-20-isopropyl-1-methylguanine, not a 1H-Azirine. Synthesis and Structural Revision. *Org. Lett., 15*(10):2370–2373.
4. Banert, K. **(2012)** Acremolin, a stable natural product with an anti-aromatic 1H-azirine moiety? A structural reorientation. *Tetrahedron Letters, 53*:6443–6445.