

# **Chasing Molecules that Were Never There: Misassigned Natural Products and the Role of CASE Systems in Modern Structure Elucidation**

**Mikhail Elyashberg**

***Advanced Chemistry Development,  
Moscow Department.***

***Institute of Geochemistry and Analytical  
Chemistry, Russian Academy of Sciences***

# Chasing Molecules That Were Never There: Misassigned Natural Products and the Role of Chemical Synthesis in Modern Structure Elucidation

*K. C. Nicolaou\* and Scott A. Snyder*



*K. C. Nicolaou was born in Cyprus and educated in England and the USA. He is currently Chairman of the Department of Chemistry at The Scripps Research Institute, and is also Professor of Chemistry at the University of California, San Diego. His impact on chemistry, biology, and medicine is reflected in nearly 600 publications and 57 patents, and he has trained hundreds of graduate students and postdoctoral fellows. His Classics in Total Synthesis series, co-authored with Erik J. Sorensen and Scott A. Snyder, is a source of inspiration for students and organic chemists around the world.*



*Scott A. Snyder, born in Palo Alto, California, received his BA in chemistry from Williams College in 1999. He completed his PhD in May 2004 at The Scripps Research Institute with K. C. Nicolaou on the total synthesis of diazonamide A and is currently an NIH postdoctoral fellow with E. J. Corey at Harvard University. He is co-author of Classics in Total Synthesis II and has contributed to over 30 publications, review articles, book chapters, and patents. He received predoctoral fellowships from the National Science Foundation, Pfizer, and Bristol-Myers Squibb.*

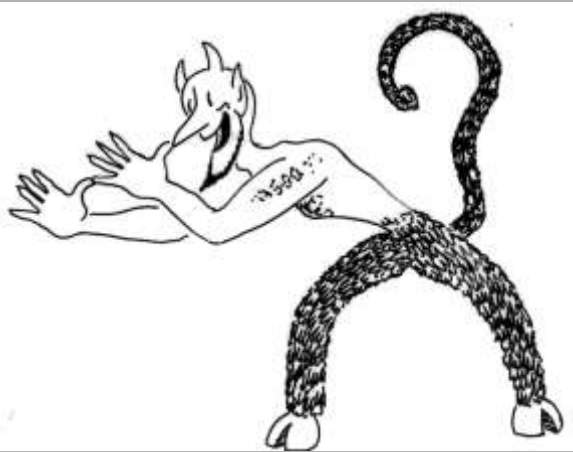
**K.C. NICOLAOU**

**S.A. SNYDER**

**Angewandte Chemie, Int. Ed., 2005, 44, 1012-1044**

**About 1000 articles were published last 15 years,  
where originally determined structures were revised**

## **Journal Of Erroneous Structure**



## **Journal Of Revised Structure**



# The cost of erroneous structure assignment



**40 - 45 issues  
of J. Erron. Struct.**



**> 40-50 issues  
of J. Revis. Struct.**

**There exists a scientific problem:  
Reduce the stream of articles containing  
mistaken structures.**

**«...There is a long way to go before natural product  
characterization can be considered a process  
devoid of ... unavoidable pitfalls”.**

**K.C. NICOLAOU, S.A. SNYDER**

# Main Postulates of the Structure Elucidation Philosophy

- Statements used for Structure Elucidation can be considered as either “axioms” or hypotheses.
- Setting axioms and hypotheses associated with a given problem is equivalent to the creation of some particular axiomatic theory
- To obtain a **valid** solution to the problem, the set of axioms must be **true, consistent, and complete.**

# I. Axioms and hypotheses reflecting characteristic spectral features

***A typical axiom confirmed by scientific experience:***

***If a molecule contains fragment  $A_j$  then its signals are observed in spectral ranges  $X_1, X_2, \dots, X_j, \dots, X_m$  which are characteristic for this fragment.***

**CH<sub>2</sub> → [1450 cm<sup>-1</sup>], CH<sub>3</sub> → [1380 cm<sup>-1</sup>]  $\wedge$  [1450cm<sup>-1</sup>]**

**“no band at [1450 cm<sup>-1</sup>] → fragment CH<sub>2</sub> is absent”**

***A typical hypothesis related to a given problem only:***

***If a signal is observed in a spectrum range  $X_j$  then the molecule contains at least one fragment of the set***

***$A_i(X_j), A_k(X_j), \dots, A_l(X_j).$***

**1450 cm<sup>-1</sup> → CH<sub>2</sub>  $\vee$  CH<sub>3</sub>**

## II. Axioms (hypotheses) of 2D NMR Spectroscopy.

### *“Standard Correlations”*

#### COSY

- If a peak (H-1, H-2) is observed in COSY, then a molecule contains the chemical bond



#### HMBC

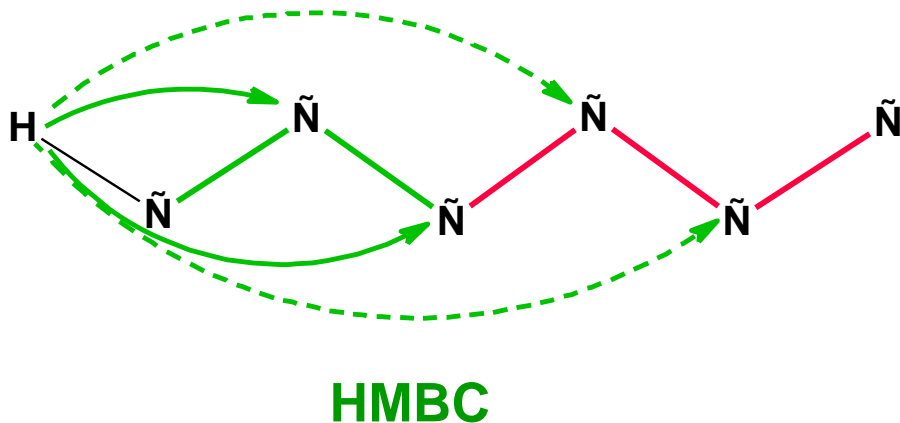
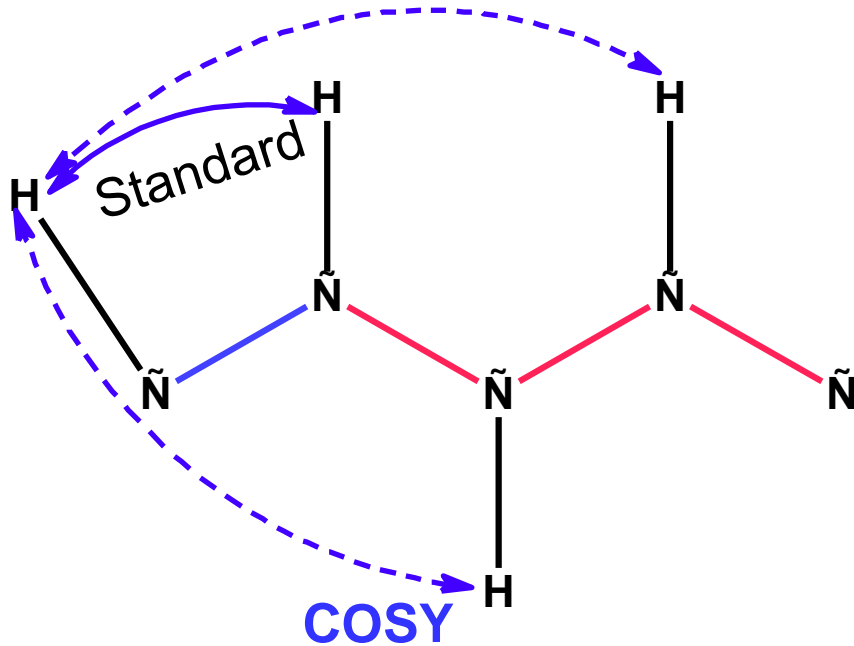
- If a peak (H-1, C-2) is observed in HMBC, then atoms C-1 and C-2 are separated in the structure by ONE or TWO chemical bonds:



#### NOESY

- If a peak (H-1, H-2) is observed in NOESY, then the distance between H-1 и H-2 in space is less than 5Å

## II. Axioms (hypotheses) of 2D NMR Spectroscopy. “Nonstandard Correlations”, NSCs



- Signals of “*Standard*” and “*Nonstandard*” correlations are experimentally not distinguishable in 2D NMR.
- If the 2D NMR data contain *Standard* and *Nonstandard* correlations the set of axioms becomes **contradictory**.

### III. Structural axioms

*necessary for structure assembling*

- Allowable chemical composition: **CHO, CHNO, CHNOS...**
- Possible molecular formula (formulae)
- Valences of atoms: **N(3,4,5), S(2, 4, 6), P(3, 5)**
- Hybridization of **each** carbon atom :  
***sp<sup>3</sup>, sp<sup>2</sup>, sp, sp<sup>3</sup> or sp<sup>2</sup>, not defined***
- Carbon neighborhood with heteroatoms (C-X):  
***fb* (X is forbidden), *ob* (X is obligatory), *not defined***
- List of obligatory fragments
- List of forbidden fragments

# Properties of information used for the structure elucidation

- Information is **fuzzy by the nature** (2 or 3 bonds between H and C in HMBC)
- Not all expected correlations are observed in spectra, i.e. information is **incomplete**
- Presence of nonstandard correlations (NSCs) frequently makes information **contradictory**
- Number of NSCs and their lengths are unknown. Signal overlapping leads to the appearance of ambiguous correlations. Information is otherwise **uncertain**
- Information can be **false** if a mistaken hypothesis is suggested
- “Structural axioms” reflects subjective opinion of the researcher. Information is in addition **subjective**.

# Consequences:

- The human expert is frequently unable to explicitly take into account all “axioms” and derive all plausible structural hypotheses.
- As a result different researchers may come to different structures from the same experimental data!
- Algorithms and software are necessary that would be capable of inferring a **valid solution** from an **equivocal information**.

## **Expert system Structure Elucidator - a Tool for Deriving and Verifying Structural Hypotheses from Equivocal Information.**

- The system was specially developed for processing the **fuzzy, contradictory, incomplete, uncertain, subjective** and **false** information.
- In frames of the system, each problem is formulated in terms of a **partial axiomatic theory**.
- The system includes an **inference machine** capable of generating all (without any exception) **structural hypotheses** in a reasonable time.
- The **Most Probable Structure** is selected using fast algorithms of  $^{13}\text{C}$  NMR spectrum prediction

**See Structure Elucidator overview in:**

**“Computer-assisted methods for  
molecular structure elucidation:  
Realizing a spectroscopist’s dream”.**

M. Elyashberg, K. Blinov, S. Molodtsov,  
Y. Smurnyy, A. Williams, T. Churanova

***Journal of Cheminformatics* 2009, 1:3**

**doi:10.1186/1758-2946-1-3**

**(free assessed)**

# **Structure Elucidator as a “Polygraph Detector.”**

## **EXAMPLES**

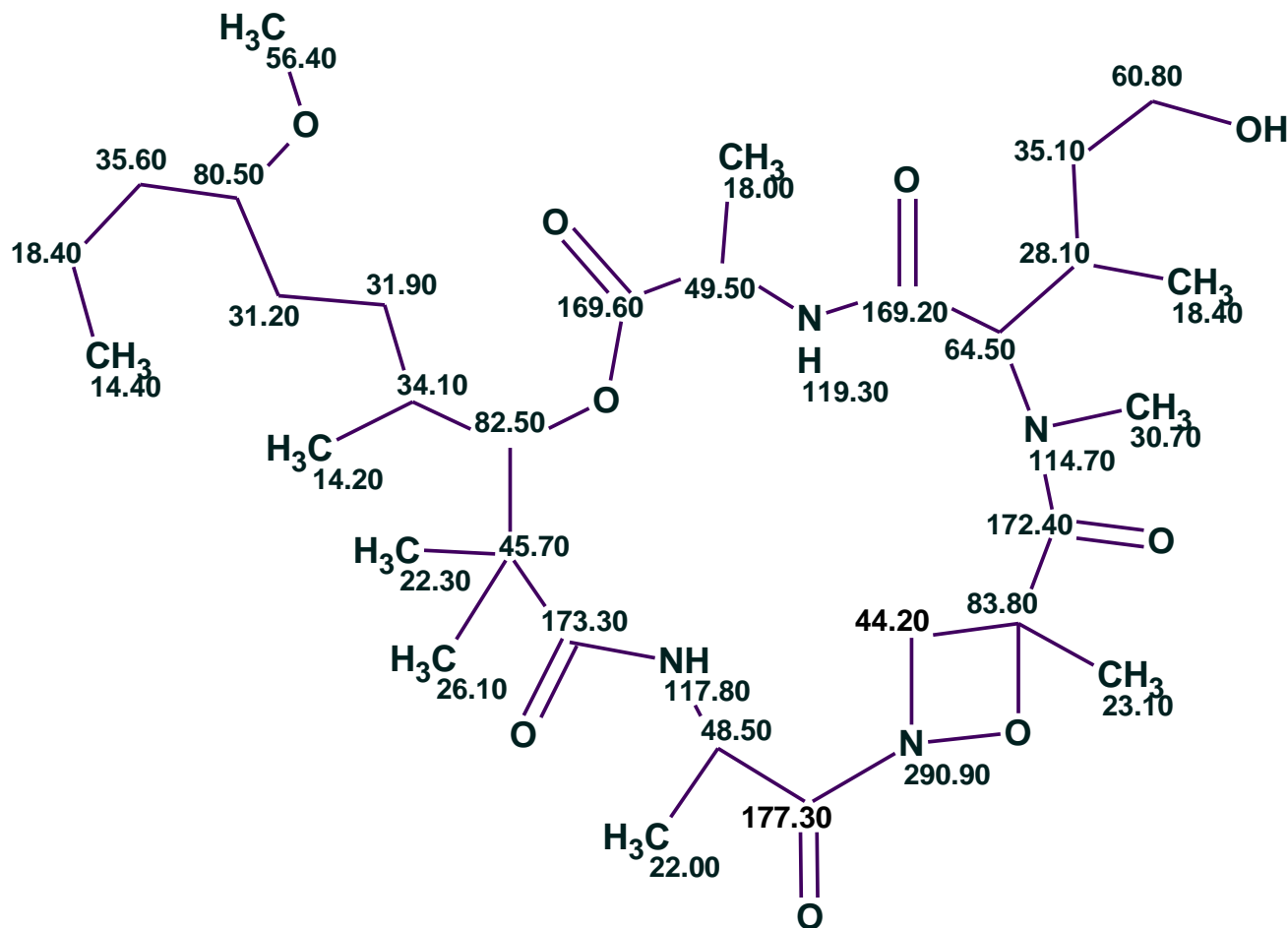
**“Halipeptins A and B: Two **Novel** Potent  
Anti-inflammatory Cyclic Depsipeptides from the  
Vanuatu Marine Sponge *Haliclona* species”.**

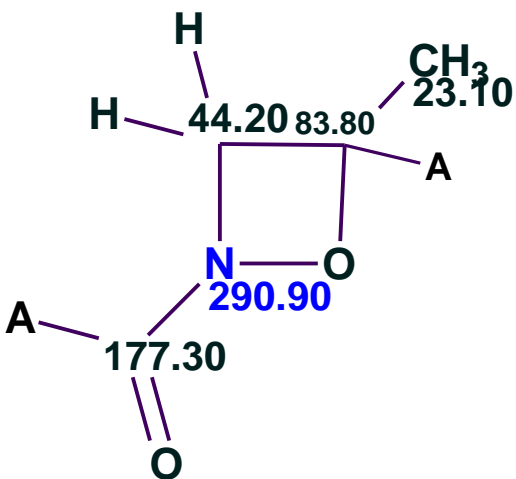
**A. Randazzo, G. Bifulco, C. Giannini, M. Bucci, C. Debitus,  
G. Cirino, L. Gomez-Paloma.**

**[J. Am. Chem. Soc.](#) 2001, 123, 10870**

Hypotheses utilized by authors:  
Elemental composition **CHNO**, **C<sub>31</sub>H<sub>54</sub>N<sub>4</sub>O<sub>9</sub>**, **N(3)**  
1D and 2D NMR data

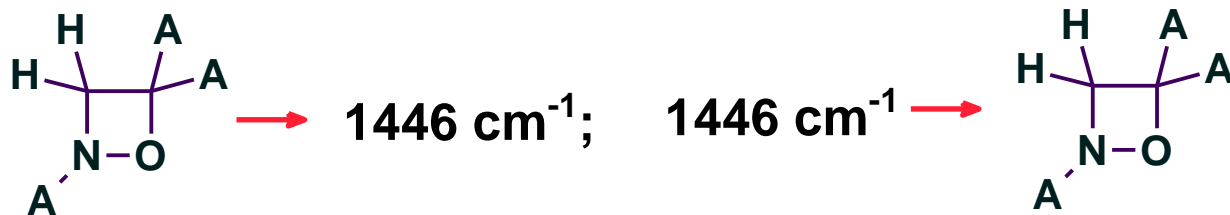
The derived structure:





“...a rather **intriguing** moiety, being an **unprecedented four-membered N-O** cyclic form ...”

1. “The presence of **N-O** was inferred from a **characteristic IR** band at **1446 cm<sup>-1</sup>** (stretchings in this range have already been observed in these systems <**axiom**>”).



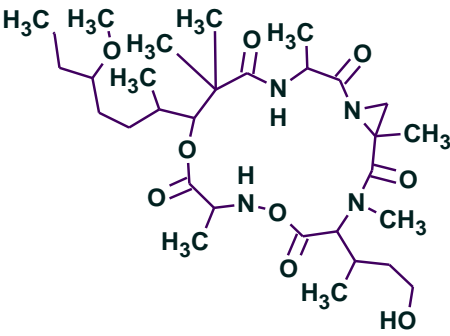
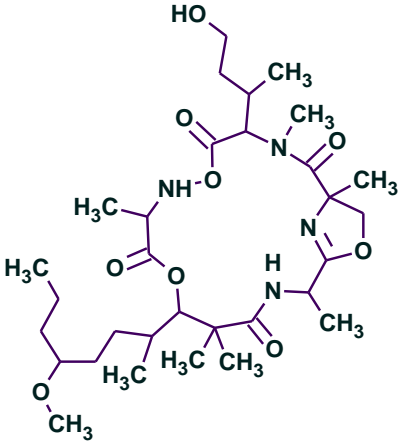
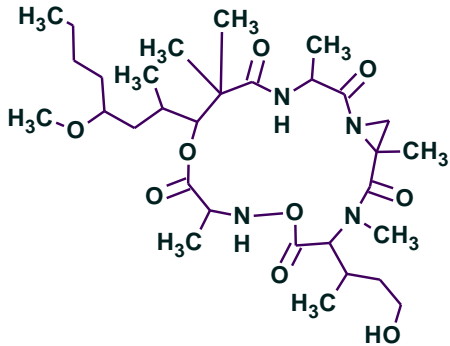
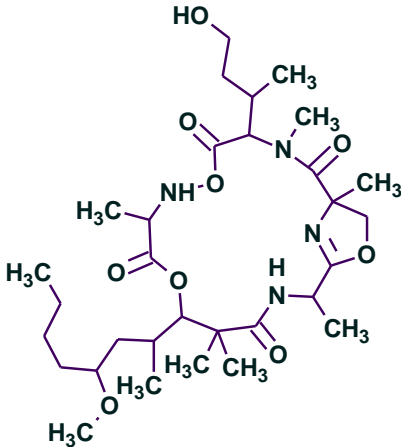
1446 cm<sup>-1</sup> → CH<sub>2</sub> V CH<sub>3</sub>

2. “The **unusual** chemical shift of nitrogen (**291**, **typical 110-120 ppm**) could be explained in terms of an effect by **ring strain**”



# Strict Structure Generation using hypotheses: $C_{31}H_{54}N_4O_9$ , no NSCs, N(3), No $X\equiv Y$ , N-O bond allowed.

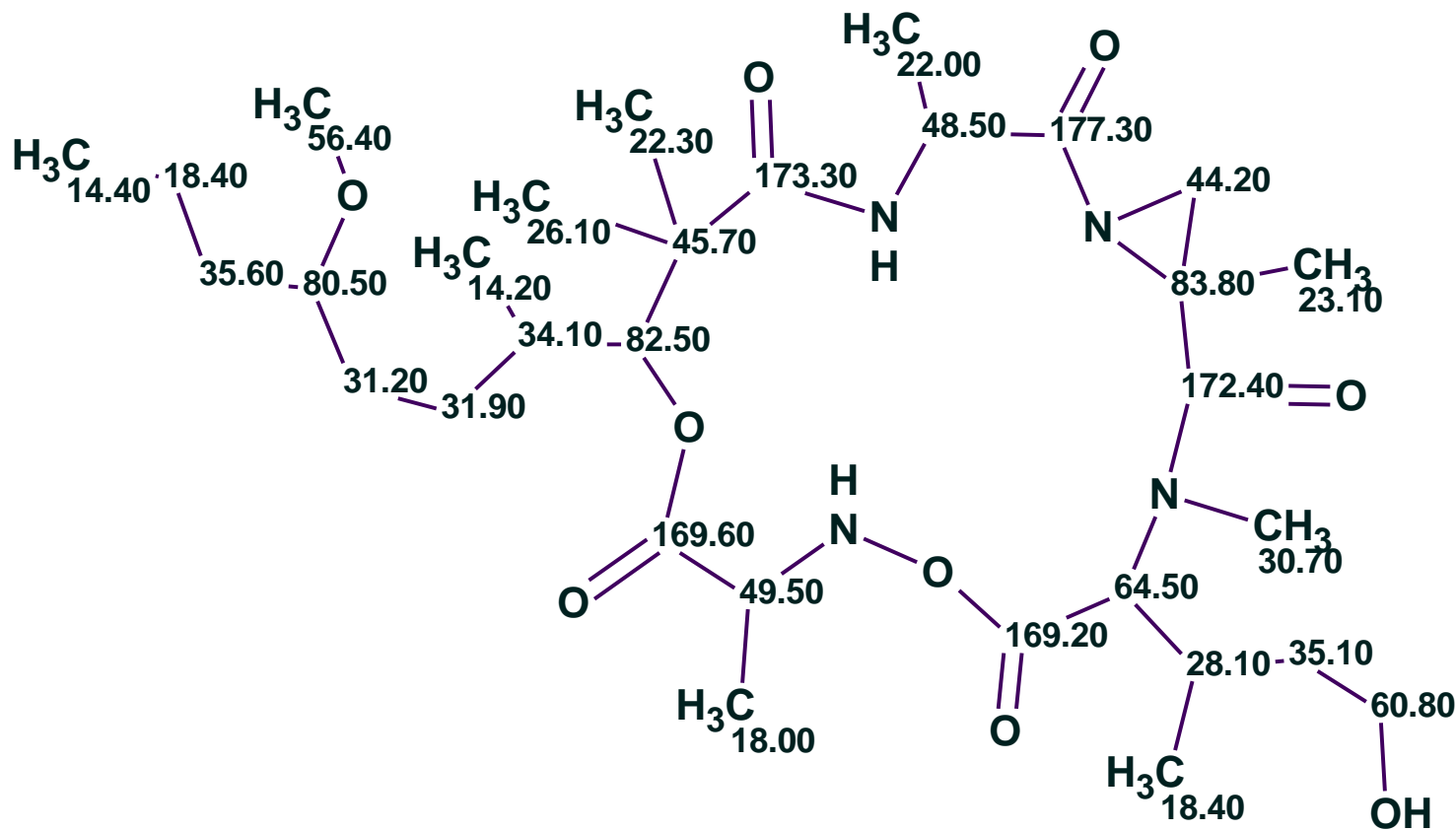
Results:  $k = 6 \rightarrow 4$ ,  $t_g = 0.1$  s

|  |  |
|--|--|
| <p>1 <b>Best</b></p>  | <p>2</p>   |
| <p><math>d_A(^{13}C) : 3.606</math></p>  | <p><math>d_A(^{13}C) : 3.725</math></p>  |
| <p>3</p>             | <p>4</p>  |
| <p><math>d_A(^{13}C) : 4.323</math></p>  | <p><math>d_A(^{13}C) : 4.452</math></p>  |

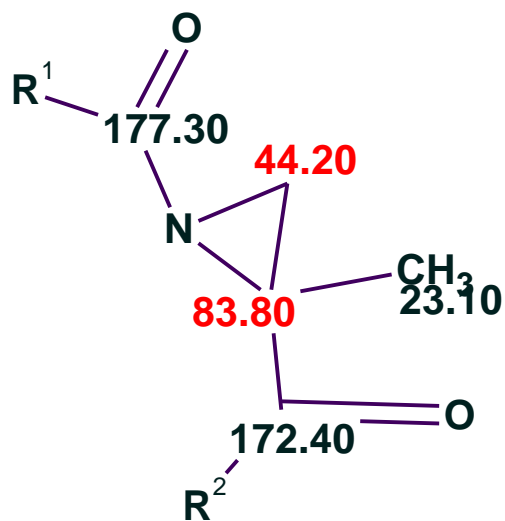
# Average deviations of the Best Structure in the ranked file:

$d_A = d(\text{HOSE}) = 3.60 \text{ ppm}$   
 $d_N = d(\text{NN}) = 3.62 \text{ ppm}$   
 $d_I = d(\text{Inc}) = 3.43 \text{ ppm}$

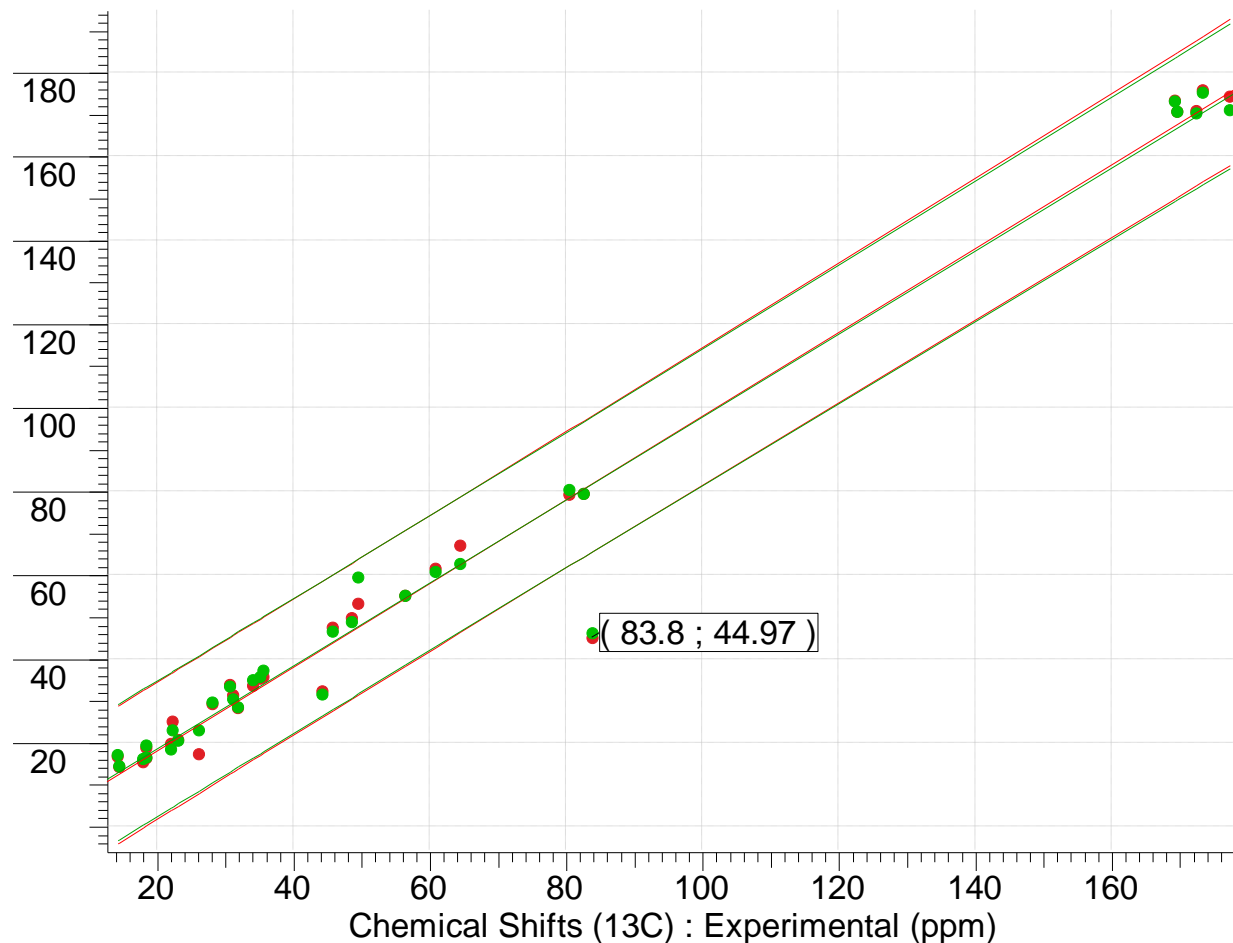
Prediction  
Accuracy  
1.6-1.8 ppm



# Calculated $^{13}\text{C}$ chemical shifts (Y) vs. experimental ones (X).



$d_A^{\max} = 40 \text{ ppm} !$



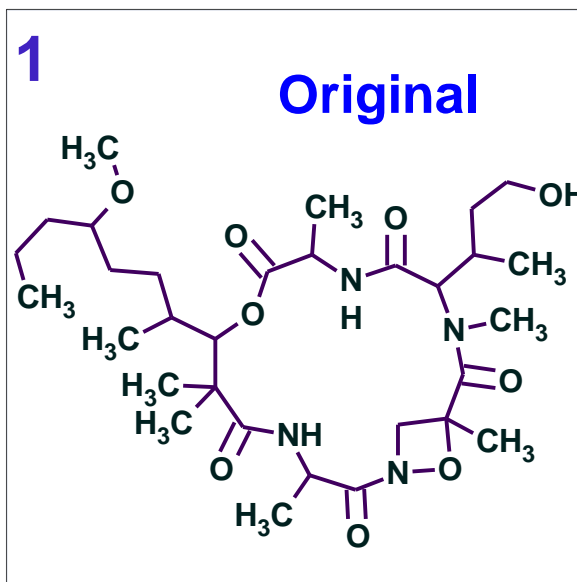
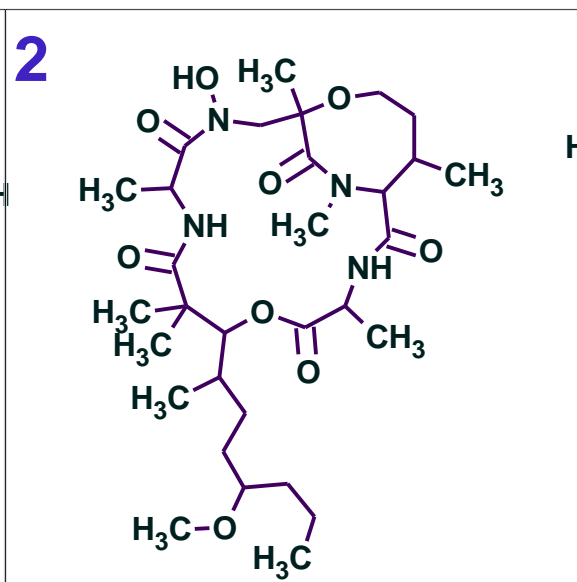
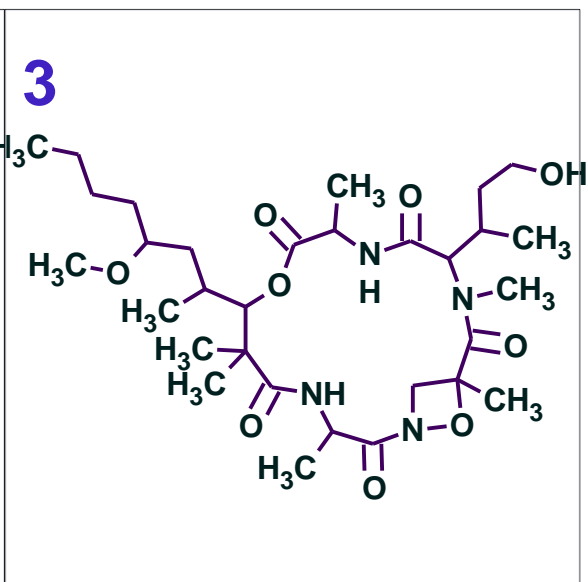
Database: Generated Molecules

- Chemical Shifts ( $^{13}\text{C}$ ) : HOSE Calc. (ppm) (Current Record) (31 pts)
- Chemical Shifts ( $^{13}\text{C}$ ) : NN Calc. (ppm) (Current Record) (31 pts)

A hint to the presence of **NSCs** in 2D NMR!

# Fuzzy Structure Generation: $m=1-15$ , $a=x$

## Results: $k=304 \rightarrow 284 \rightarrow 183$ , $t_g = 35$ s

|   |   |   |
|---|---|---|
| <p><b>1</b></p> <p><b>Original</b></p>              | <p><b>2</b></p>                                   | <p><b>3</b></p>                                  |
| <p><math>d_A(^{13}\text{C})</math>: 1.526<br/> <math>q(^{13}\text{C})</math>: 1.714<br/> <math>d_N(^{13}\text{C})</math>: 1.879</p> | <p><math>d_A(^{13}\text{C})</math>: 2.140<br/> <math>q(^{13}\text{C})</math>: 2.217<br/> <math>d_N(^{13}\text{C})</math>: 2.207</p> | <p><math>d_A(^{13}\text{C})</math>: 2.218<br/> <math>q(^{13}\text{C})</math>: 2.456<br/> <math>d_N(^{13}\text{C})</math>: 2.480</p> |

$d_A(^{13}\text{C})$ : 1.526  
 $q(^{13}\text{C})$ : 1.714  
 $d_N(^{13}\text{C})$ : 1.879

**Prediction  
 Accuracy  
 1.6-1.8 ppm**

**“Structural Revision of Halipeptins: Synthesis of the Thiazoline Unit and Isolation of Halipeptin”.**

**C. D. Monico, A. Randazzo, G. Bifulco, P. Cimino,  
M. Aquino, I. Izzo, F. De Riccardis, L. Gomez-Paloma.**

**Tetrahedron Lett. 2002, 43, 5707.**



# What would happen if StrucEluc was used from the very beginning?

Hypotheses (constraints) on elemental composition:  
C(31), H(52-56), O(0-10), N(0-10), S(0-2)

Modest MS resolution,  $m/z$  (M+H)<sup>+</sup> = 627.4073

**Molecular Formula Generator**

Allowed Elements

| Name | Mass      | Ran... | RDBE | Val... |
|------|-----------|--------|------|--------|
| C    | 12.000000 | 31     | 1.0  | 4      |
| H    | 1.007825  | 52-56  | -0.5 | 1      |
| O    | 15.994915 | 0-10   | 0.0  | 2      |
| N    | 14.003074 | 0-10   | 0.5  | 3      |
| S    | 31.972070 | 0-2    | 0.0  | 2      |

Add Element... Predict Limits Restore Data

Remove Element Clear Limits

Monoisotopic Mass (Da)

m/z Value (Da) 627.4073 ± 0.1

M+H Ion  M-H Ion

Rings and Double Bonds Equivalent

RDBE 0-100

Allow Integer Values  Allow Fractional Values

MF (Selected/Displayed/Total - 0/3/3)

| # | "Classic" MF                        | MF   | Mass       | Mass Differ... | RDBE |
|---|-------------------------------------|--|------------|----------------|------|
| 1 | <input type="checkbox"/>            | C <sub>31</sub> H <sub>54</sub> O <sub>5</sub> N <sub>4</sub> S <sub>2</sub> | 626.353561 | -0.045914      | 7.0  |
| 2 | <input type="checkbox"/>            | C <sub>31</sub> H <sub>54</sub> O <sub>7</sub> N <sub>4</sub> S <sub>1</sub> | 626.371320 | -0.028155      | 7.0  |
| 3 | <input checked="" type="checkbox"/> | C <sub>31</sub> H <sub>54</sub> O <sub>9</sub> N <sub>4</sub>                | 626.389079 | -0.010396      | 7.0  |

Select All Unselect All

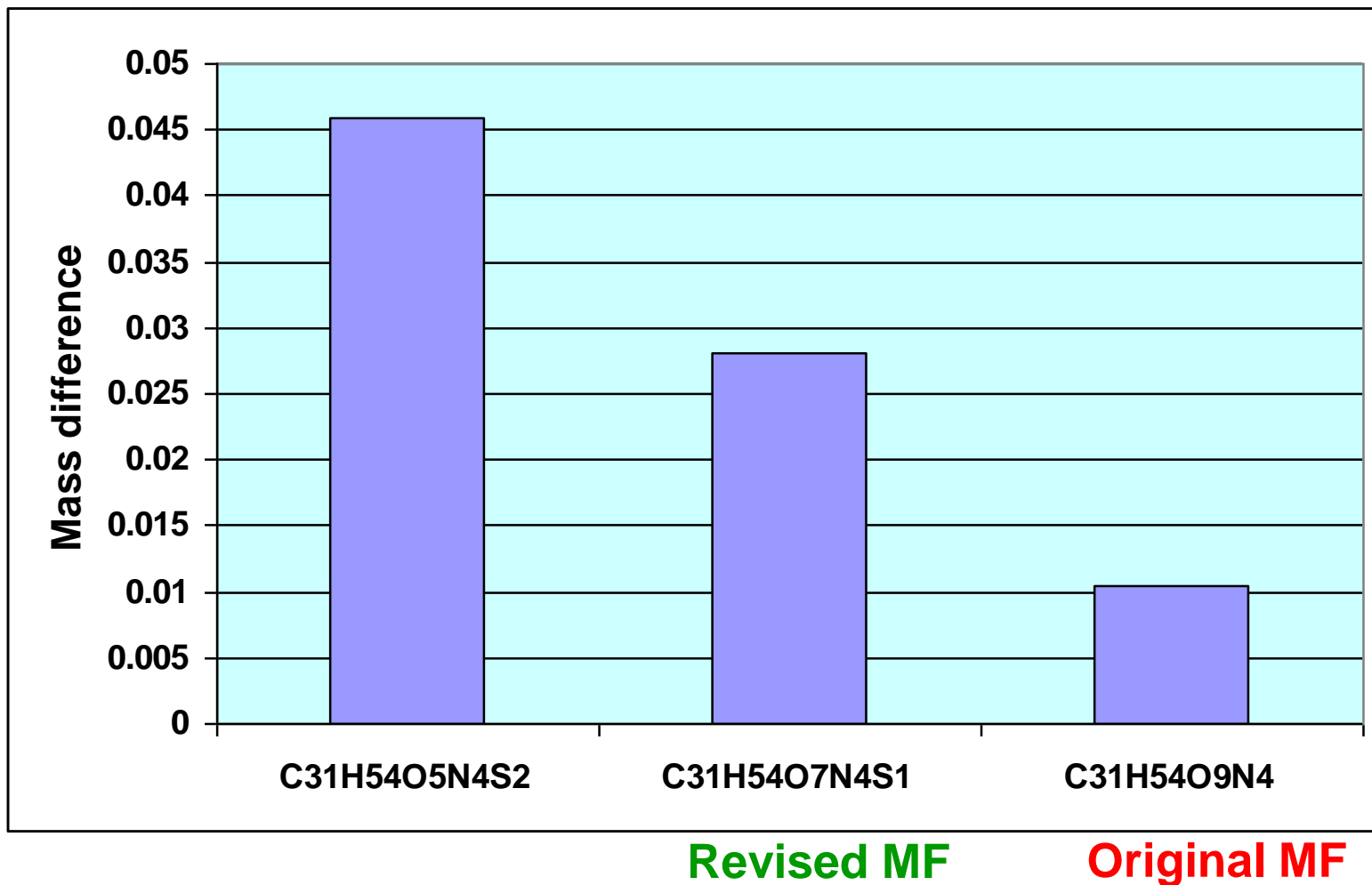
Generator Options

Apply Nitrogen Rule

Generate MF

OK Cancel Help

# Mass differences $|m/z_{(\text{exp})} - M_{(\text{calc})}|$ for three possible molecular formulae



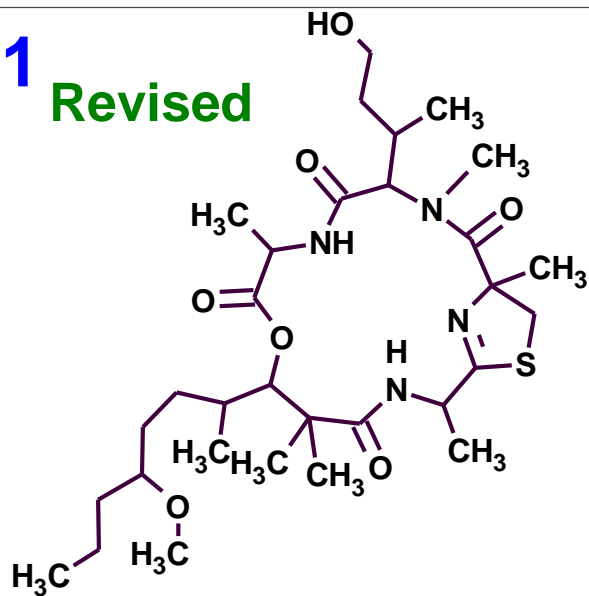
# Structure generation from $C_{31}H_{54}N_4O_7S$ and $C_{31}H_{54}N_4O_9$

Hypotheses: N(3),S(2). No  $X\equiv Y$ . N-O possible.

No User intervention.

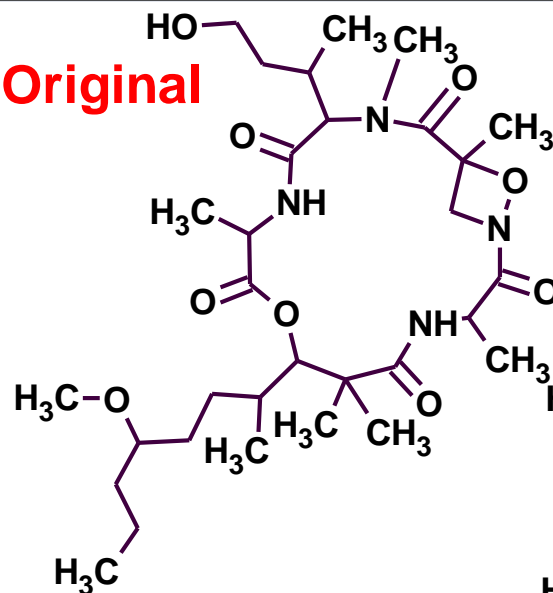
Results:  $k=303$ ,  $t_g= 36$  s

**1** Revised



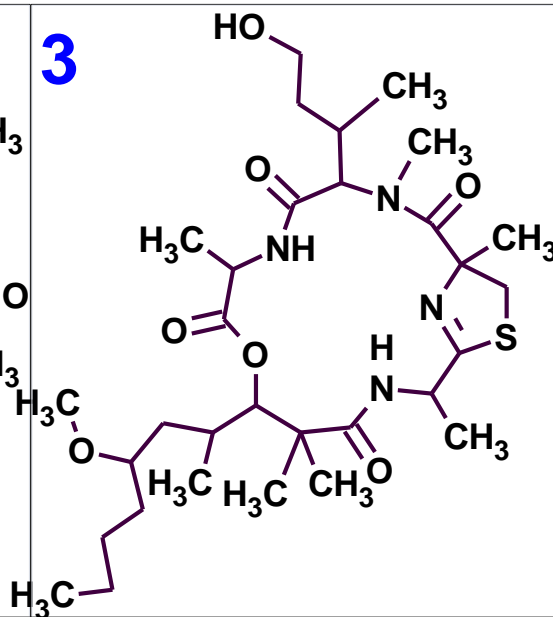
$d_A(^{13}C)$ : 1.112  
 $d(^{13}C)$ : 1.642  
 $d_N(^{13}C)$ : 1.539

**2** Original



$d_A(^{13}C)$ : 1.526  
 $d(^{13}C)$ : 1.714  
 $d_N(^{13}C)$ : 1.879

**3**



$d_A(^{13}C)$ : 1.803  
 $d(^{13}C)$ : 2.342  
 $d_N(^{13}C)$ : 2.142

$C_{31}H_{54}N_4O_7S$

$C_{31}H_{54}N_4O_9$

$C_{31}H_{54}N_4O_7S$



**“A [New Inhibitor](#) of 5'-Hydroxyaverantin  
Dehydrogenase, an Enzyme Involved in  
Aflatoxin Biosynthesis, from  
*Trichoderma hamatum*”**

**E. Sakuno, K. Yabe, T. Hamasaki,  
H. Nakajima**

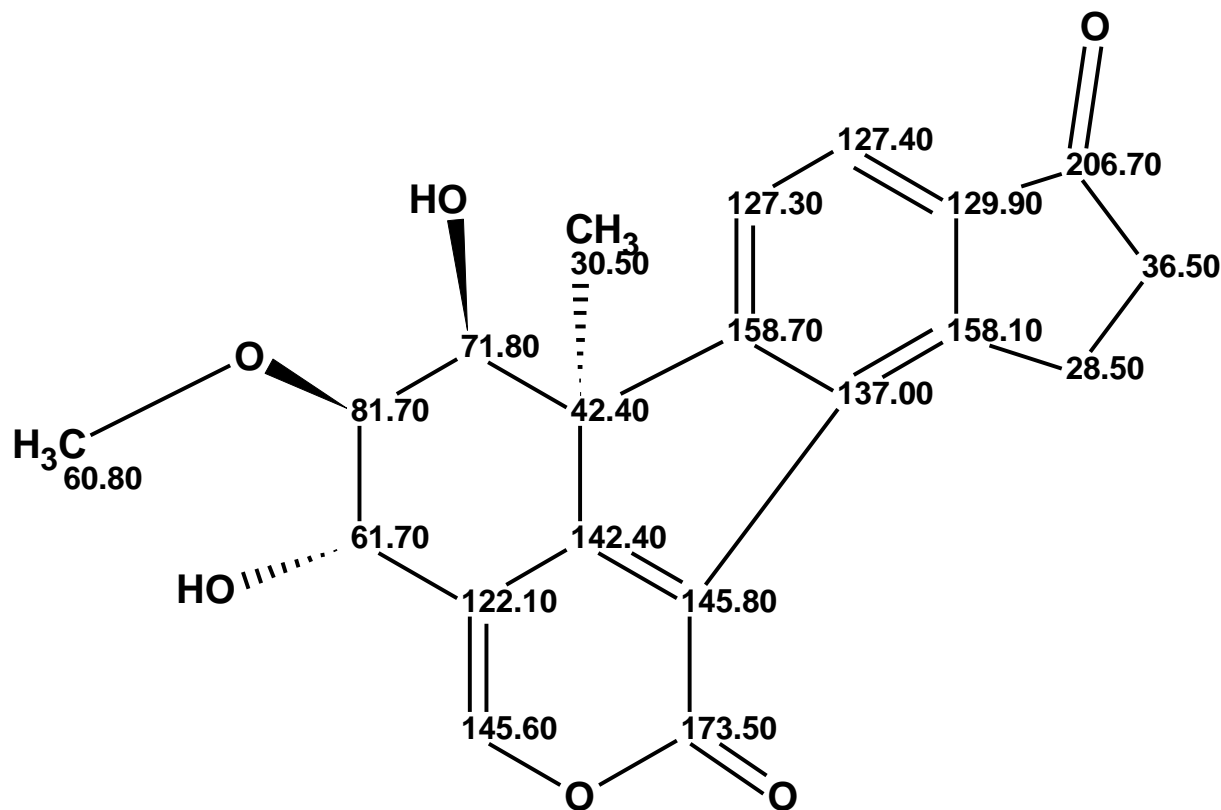
**[J. Nat. Prod.](#) 2000, 63, 1677-1678**

“One ketone and one ester group were indicated by carbon resonances at  $\delta$  206.7 and 173.4 and IR bands at 1671 and 1707  $\text{cm}^{-1}$ .”

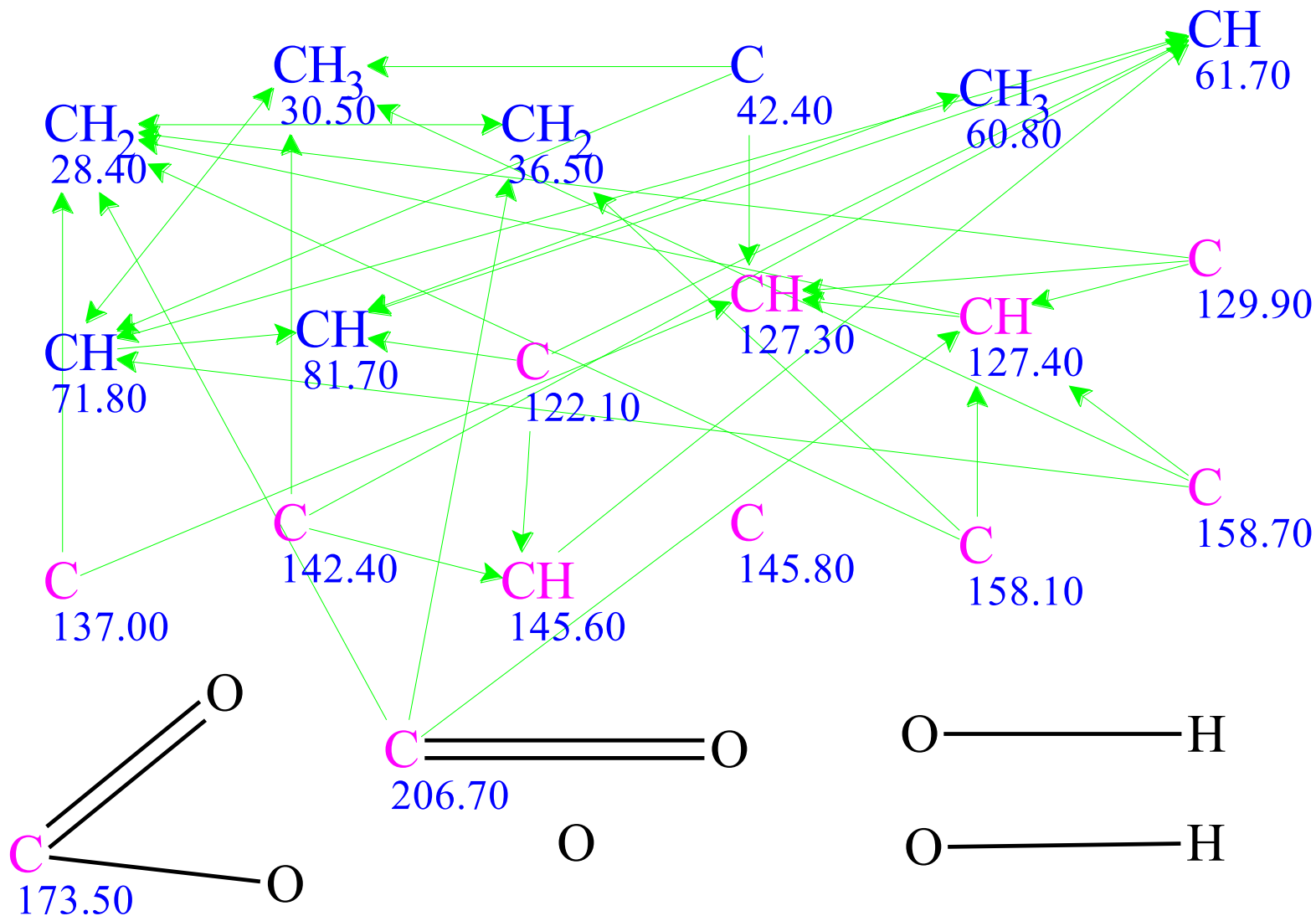
Hypotheses: **O-C=O**, (C)-C=O(C), 2 (O-H) are present



The structure derived by authors:




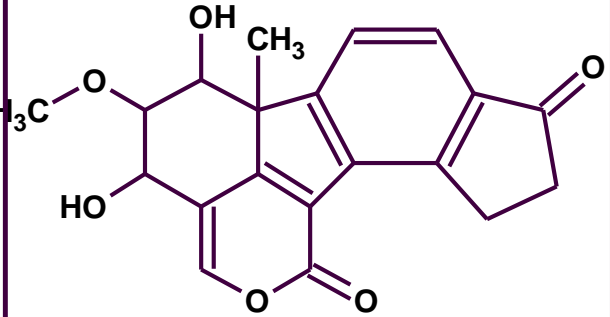
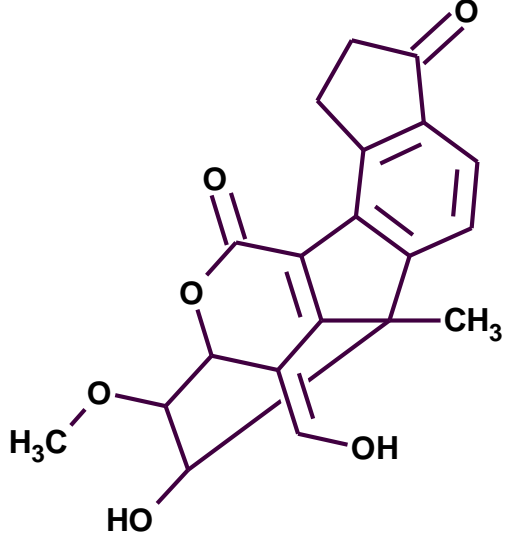
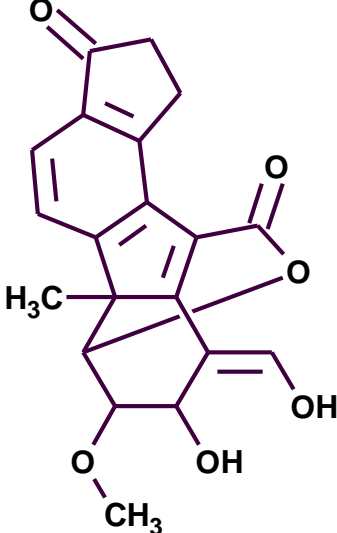
# MCD corresponding to the hypotheses postulated by Sakuno et al.



# Fuzzy Structure Generation:

$k=174 \rightarrow 80 \rightarrow 60$ ,  $t_g = 30$  s

Three best structures:

|   |  |   |
|---|--|---|
| <p>1 (ID:14)</p> <p><b>Original</b> </p>  | <p>2 (ID:13)</p>  | <p>3 (ID:15)</p>  |
| <p><math>d_{(^{13}\text{C})}</math>: 4.486<br/><math>d_{\text{N}}(^{13}\text{C})</math>: 4.137</p>  | <p><math>d_{(^{13}\text{C})}</math>: 6.176<br/><math>d_{\text{N}}(^{13}\text{C})</math>: 5.453</p>   | <p><math>d_{(^{13}\text{C})}</math>: 6.410<br/><math>d_{\text{N}}(^{13}\text{C})</math>: 5.565</p>    |

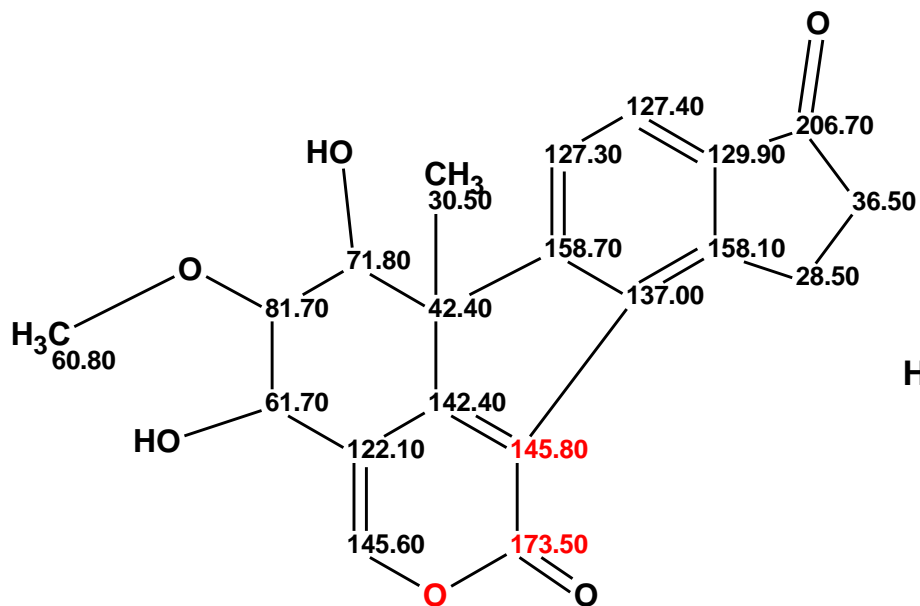
**“Structure Reassignment of the Fungal Metabolite TAEMC161 as the Phytotoxin Viridiol”**

**P. Wipf, A.D. Kerekes**

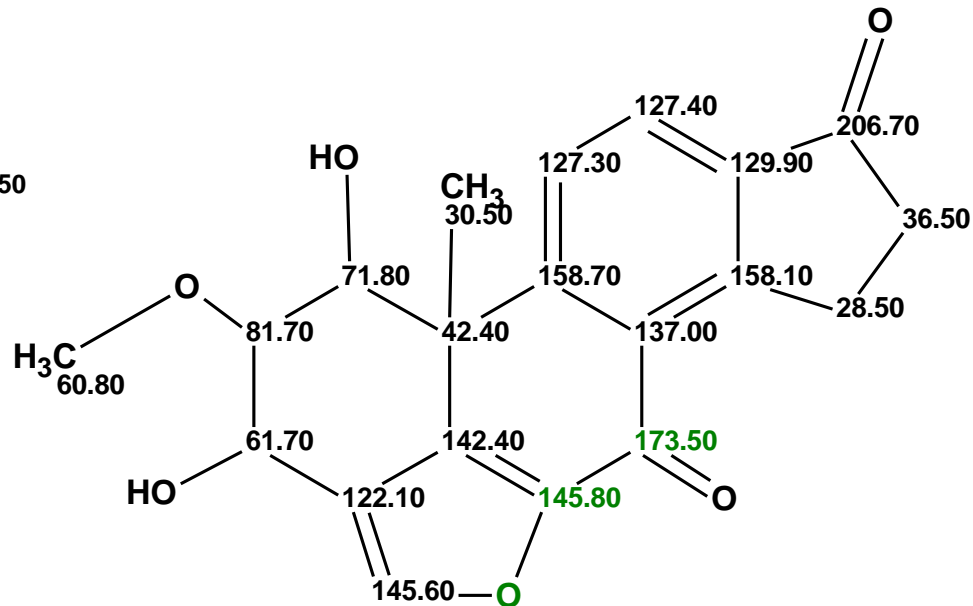
**[J. Nat. Prod.](#) 2003, 66, 716-718**

# Structure revision

Original

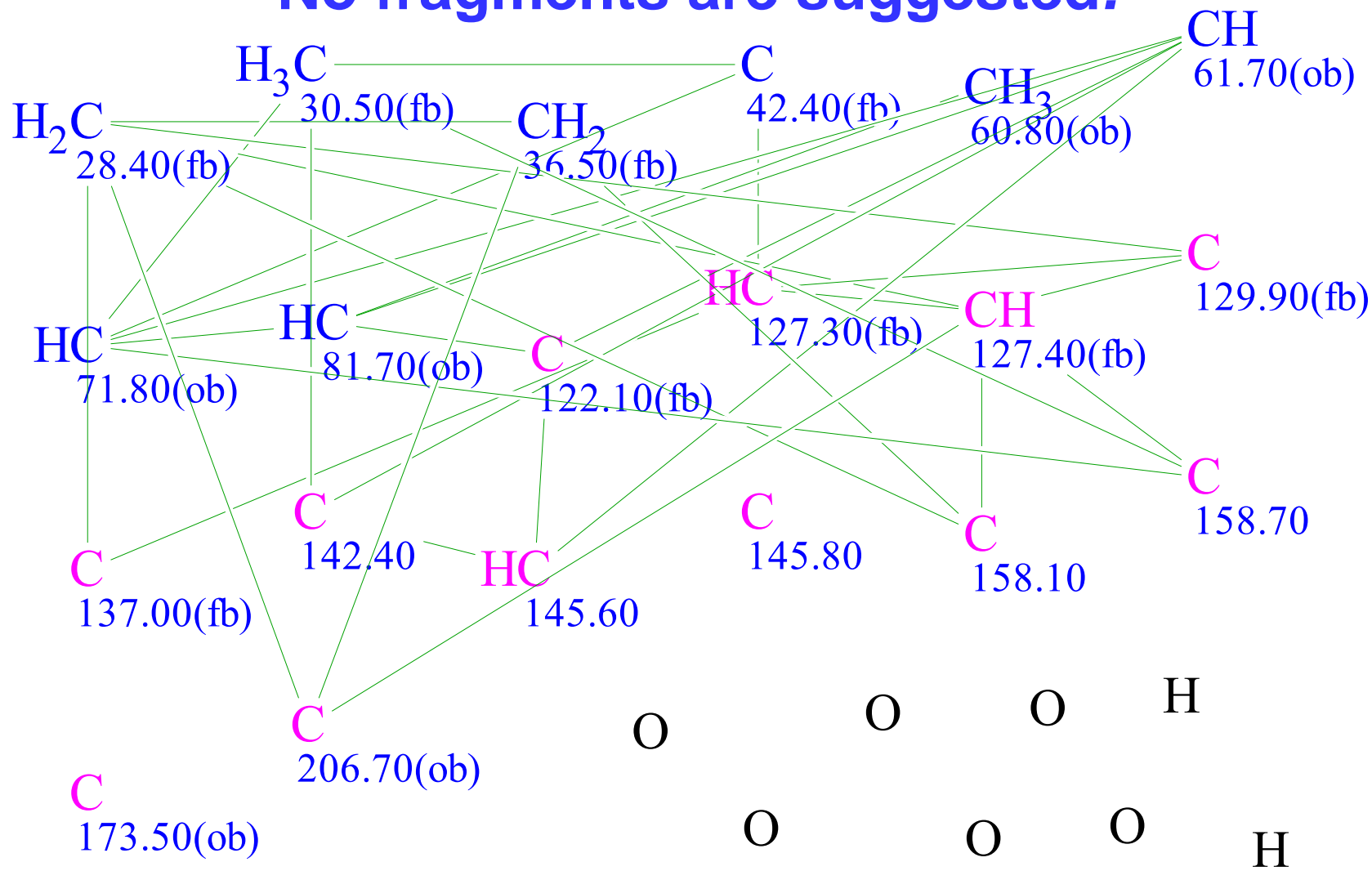


Revised using QM  
<sup>13</sup>C NMR prediction



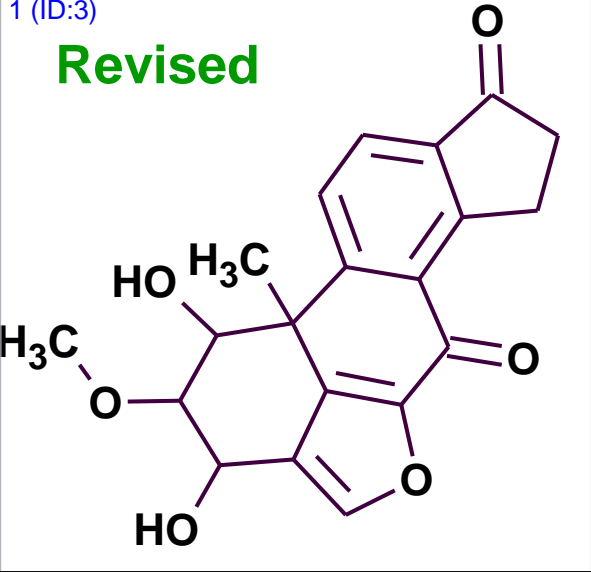
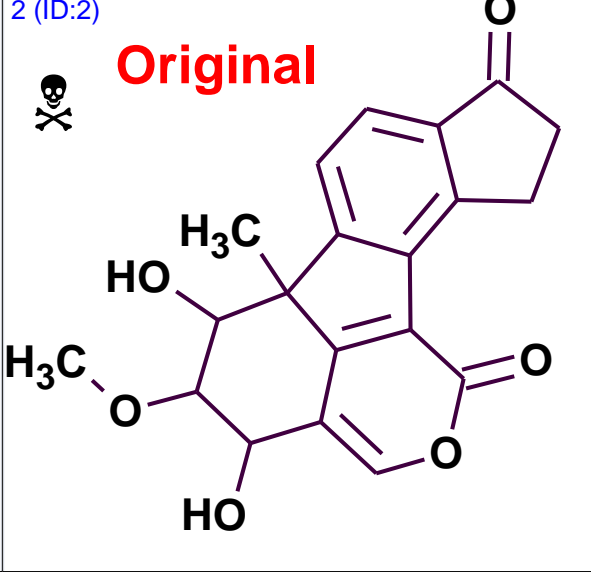
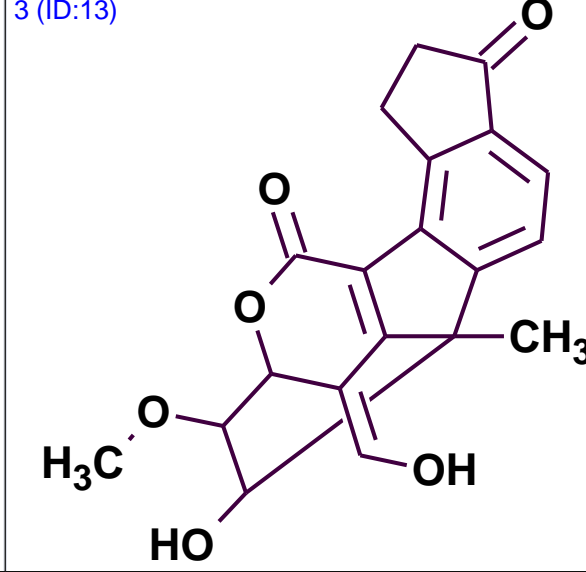
# What would happen if StrucEluc was used from the very beginning?

No fragments are suggested.



# Fuzzy Structure Generation:

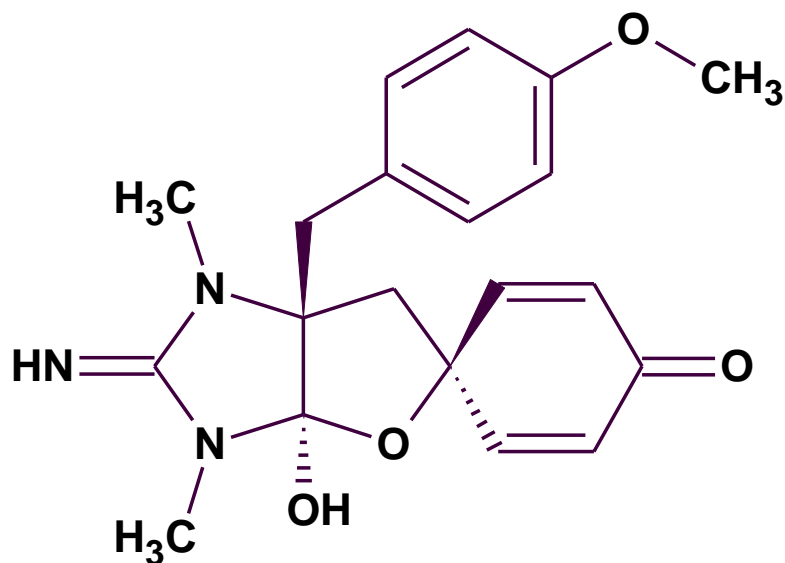
$m=0-20$ ,  $a=x$ , stop at  $m=m_g$   
 $k=494 \rightarrow 398 \rightarrow 272$ ,  $t_g = 1 \text{ m } 40 \text{ s}$

|   |  |   |
|---|--|---|
| <p>1 (ID:3)</p> <p><b>Revised</b></p>  | <p>2 (ID:2)</p> <p><b>Original</b></p>  | <p>3 (ID:13)</p>  |
| <p><math>d(^{13}\text{C})</math>: 2.137</p> <p><math>d_{\text{N}}(^{13}\text{C})</math>: 2.316</p>                      | <p><math>d(^{13}\text{C})</math>: 4.486</p> <p><math>d_{\text{N}}(^{13}\text{C})</math>: 4.137</p>                         | <p><math>d(^{13}\text{C})</math>: 6.383</p> <p><math>d_{\text{N}}(^{13}\text{C})</math>: 5.453</p>    |

**“A New Structural Theme in the Imidazole-Containing Alkaloids from a Calcareous Leucetta Sponge”**

**P. Ralifo, P. Crews, University of California, Santa Cruz**

**[J. Org. Chem. 2004, 69, 9025-9029](#)**

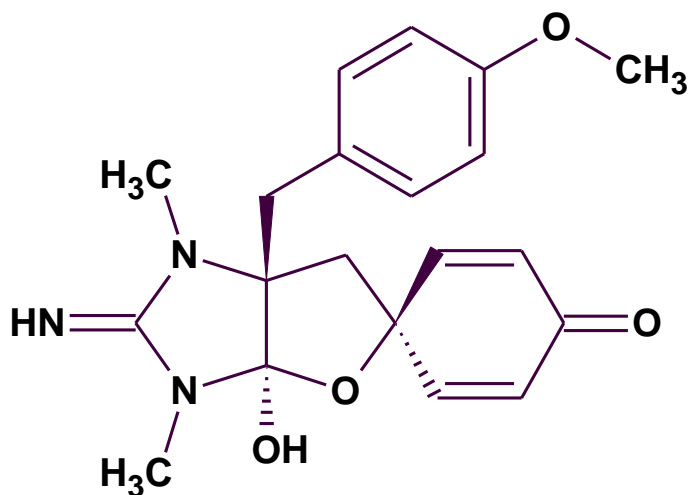


**$C_{20}H_{23}N_3O_4$ , DBE = 11**

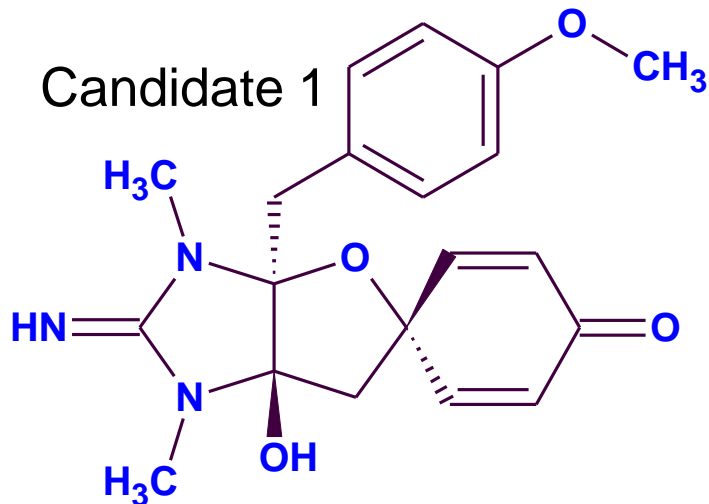
**“NMR analysis allowed the unambiguous 2D structural assignment of the molecule, its relative stereochemistry was determined by ROESY data.”**

1. Li, C. M.; Danishefsky, S. J. *Tetrahedron Lett.* 2006, 47, 385.
2. Chang, J. J.; Chan, B.; Ciufolini, M. A. *Tetrahedron Lett.* 2006, 47, 3599.
3. Aberle, N.; Ovenden, S. P. B.; Lessene, G.; Watson, K. G.; Smith, B. J. *Tetrahedron Lett.* 2007, 48, 2199.

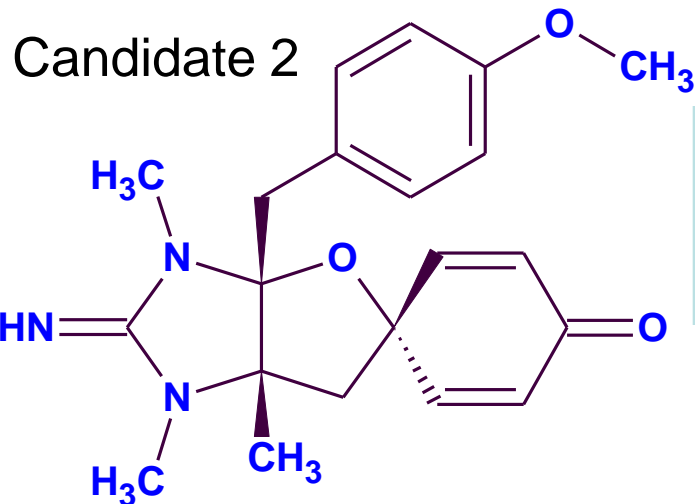
**Original**



[1-2]



[3]



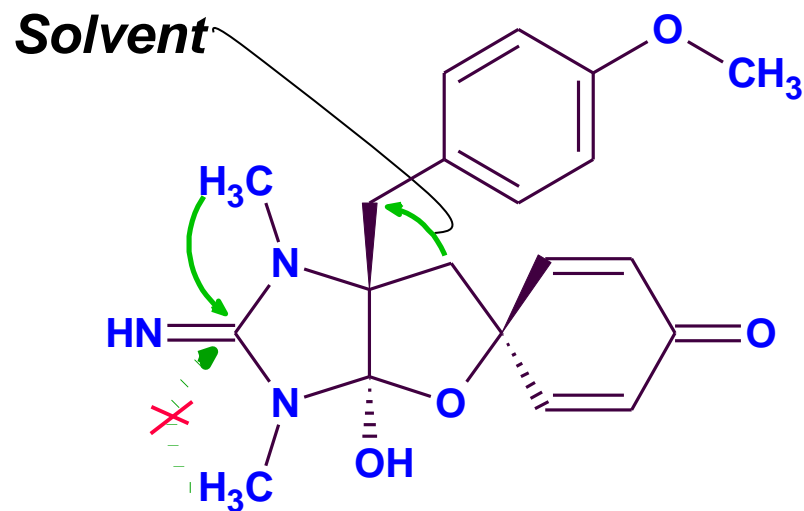
[3]

**“Structure Revision of Spiroleucettadine, a Sponge Alkaloid with a Bicyclic Core Meager in H-Atoms”**

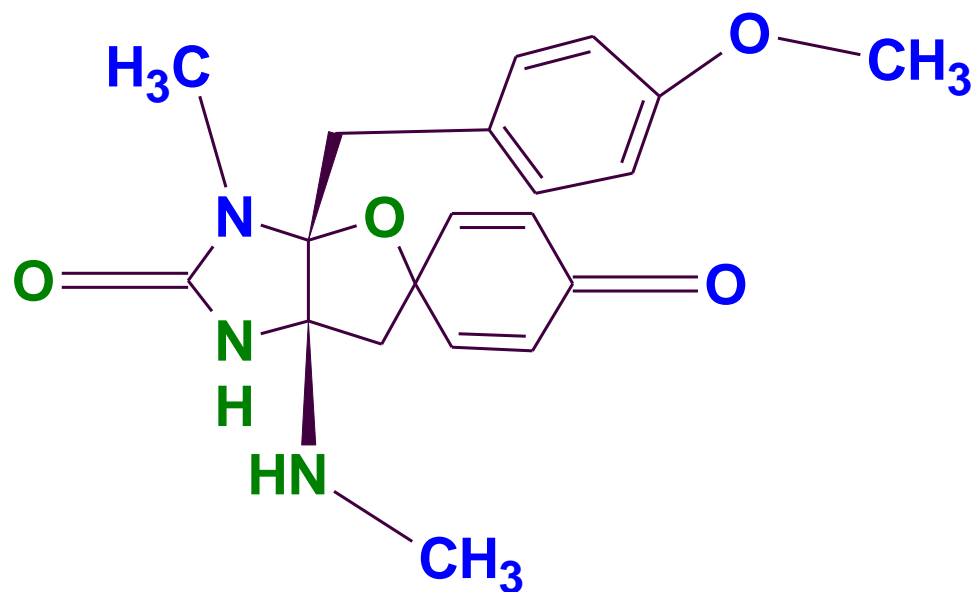
K. White, T. Amagata, A. Oliver, K. Tenney, P. Wenzel, P. Crews  
[J. Org. Chem. 2008, 73, 8719–8722](#)

Structure was reisolated and determined by X-ray.  
Confirmed by 2D NMR data and QM DFT  $^{13}\text{C}$  calculations

**Original**

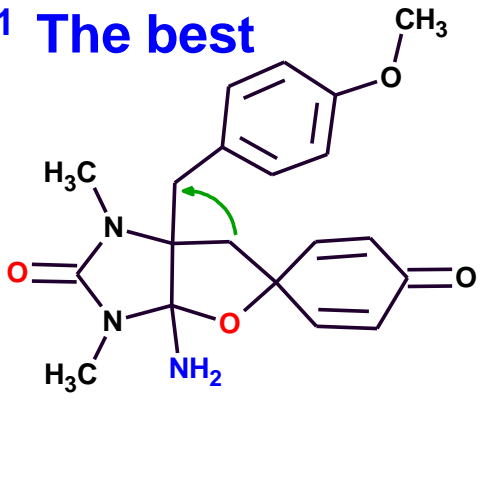
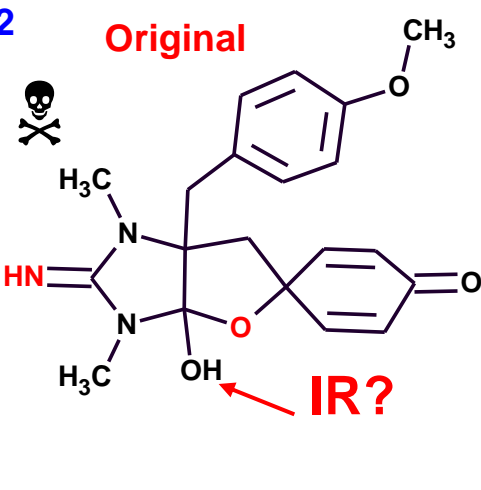
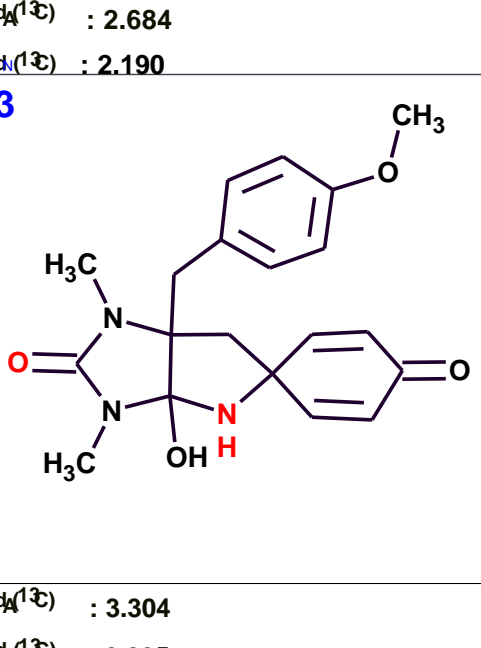
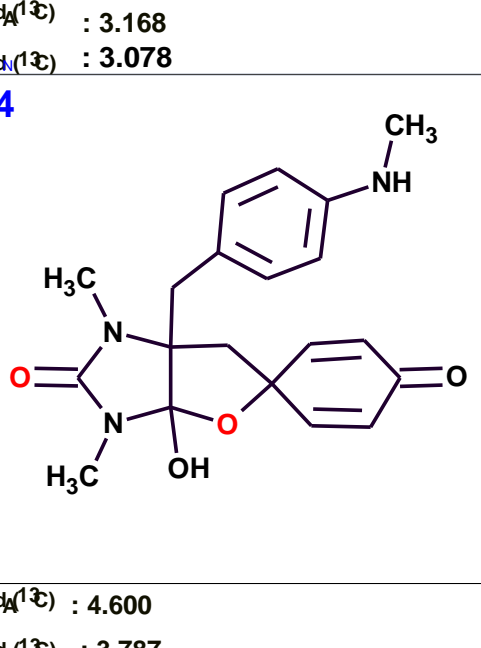


**Revised**

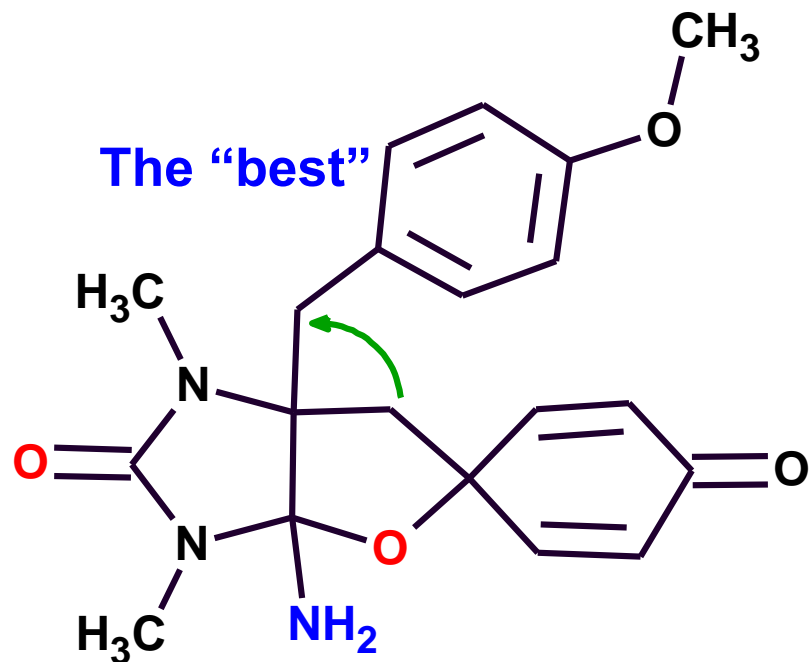


# Structure Generation from **OLD (erroneous)** 2D NMR data

$k=117 \rightarrow 83 \rightarrow 79$ ,  $t_g = 10$  s

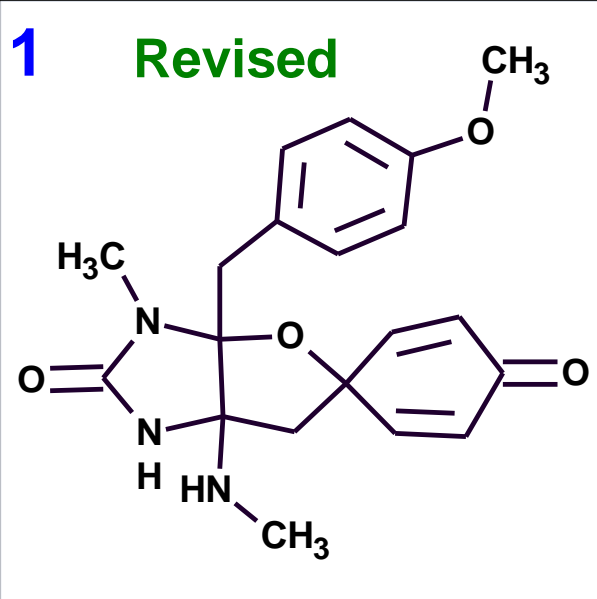
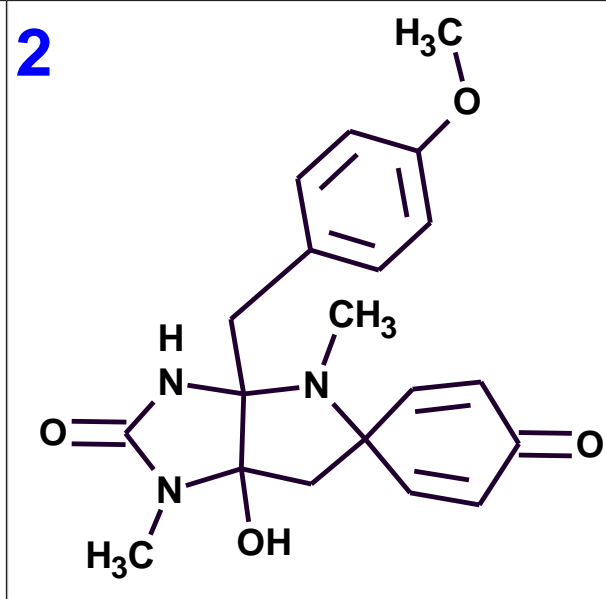
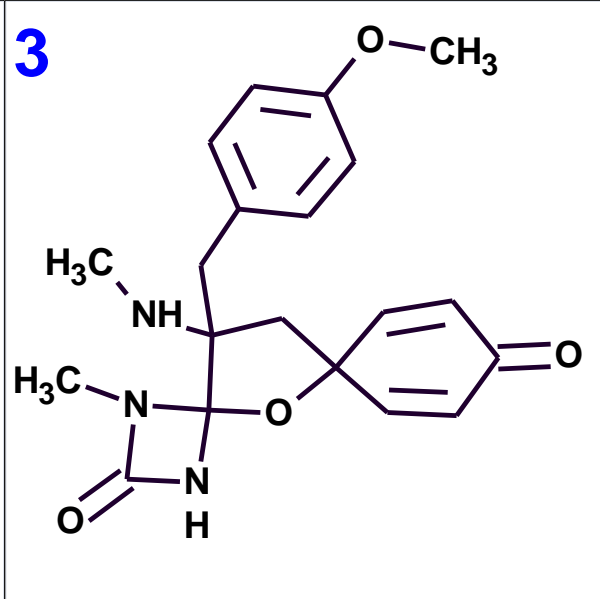
|  |  |
|--|--|
| <p><b>1 The best</b></p>  <p><math>d_A(^{13}\text{C})</math> : 2.684<br/><math>d_N(^{13}\text{C})</math> : 2.190</p> | <p><b>2 Original</b></p>  <p><math>d_A(^{13}\text{C})</math> : 3.168<br/><math>d_N(^{13}\text{C})</math> : 3.078</p> |
| <p><b>3</b></p>  <p><math>d_A(^{13}\text{C})</math> : 3.304<br/><math>d_N(^{13}\text{C})</math> : 3.395</p>         | <p><b>4</b></p>  <p><math>d_A(^{13}\text{C})</math> : 4.600<br/><math>d_N(^{13}\text{C})</math> : 3.787</p>         |

Program detected that the **Original Structure was wrong.** Hypotheses should be revised



$d(1) = 2.69, 2.19$  ppm  
 $d(2) = 3.17, 3.08$  ppm

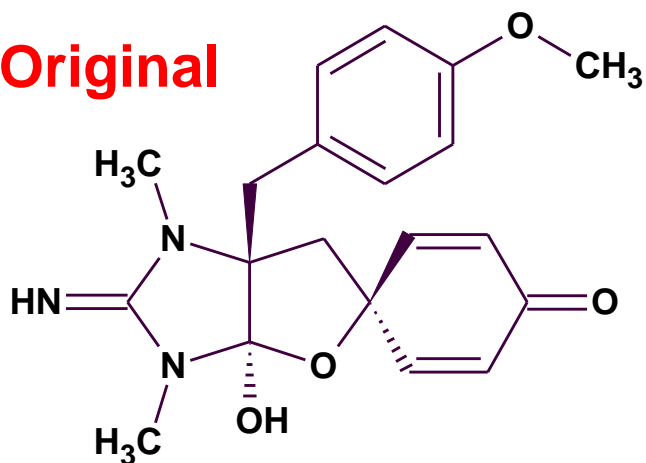
Structure Generation from **NEW (correct)** 2D NMR data with “liberal” atom properties at DBE = 11.  
k = 342 → 256, t<sub>g</sub> = 8 h 2 min (“overnight”)

|   |   |  |
|---|---|--|
| <p><b>1</b> Revised</p>  | <p><b>2</b></p>  | <p><b>3</b></p>  |
| <p>d<sub>A</sub>(<sup>13</sup>C) 1.762<br/>d<sub>N</sub>(<sup>13</sup>C) 1.513</p>                        | <p>d<sub>A</sub>(<sup>13</sup>C) 2.745<br/>d<sub>N</sub>(<sup>13</sup>C) 2.290</p>                  | <p>d<sub>A</sub>(<sup>13</sup>C) 2.681<br/>d<sub>N</sub>(<sup>13</sup>C) 2.395</p>                   |

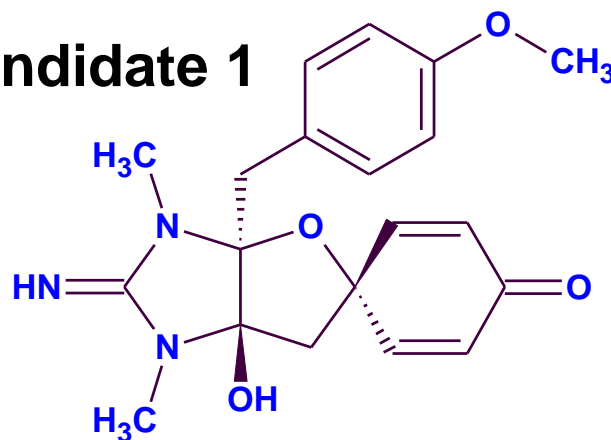
# The Revised Structure and its Stereochemistry.

QM chemical shift calculations were performed by Crews group for 16 stereoisomers of all involved proposed structures.

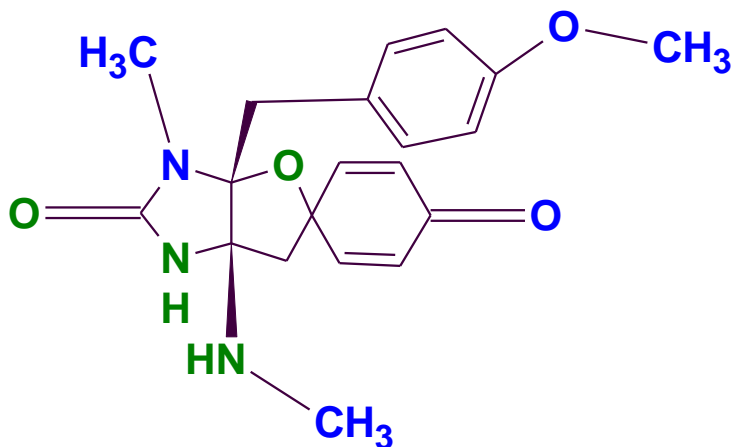
**Original**



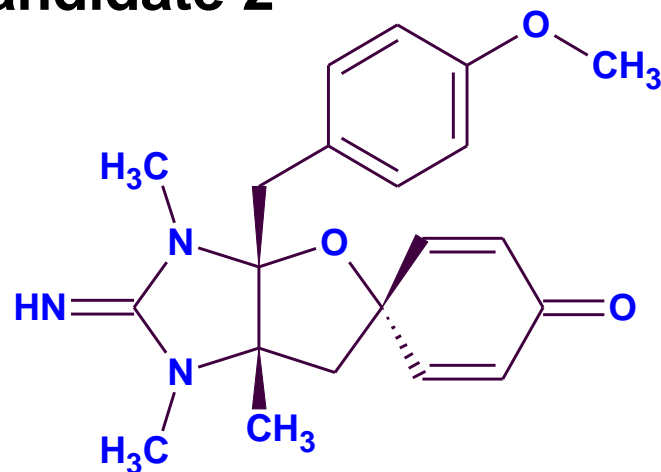
**Candidate 1**



**Revised**



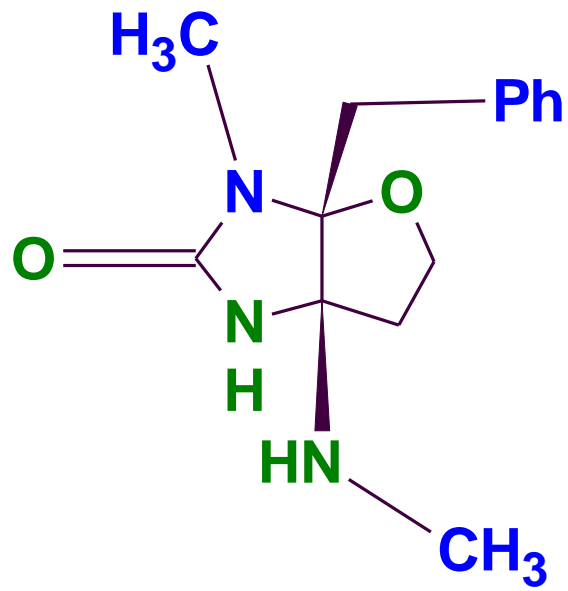
**Candidate 2**



# A HARD WORK...



# Results of $^{13}\text{C}$ NMR chemical shift prediction by the aid of StrucEluc

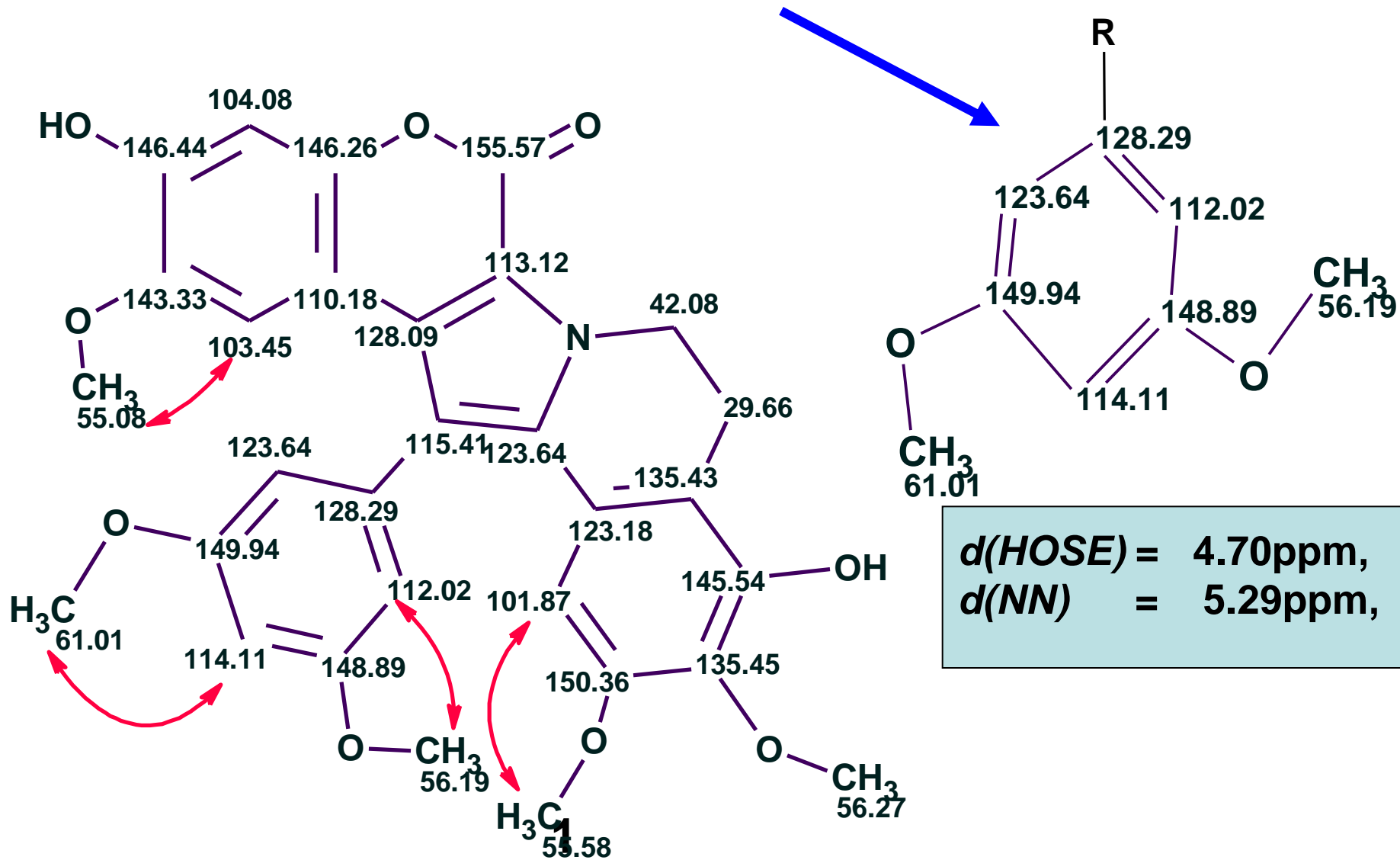


|                                   |                            |                            |                              |
|-----------------------------------|----------------------------|----------------------------|------------------------------|
| 1 (ID:9)<br><b>I, Revised</b><br> | 2 (ID:12)<br><b>L</b><br>  | 3 (ID:11)<br><b>K</b><br>  | 4 (ID:10)<br><b>J</b><br>    |
| $d_A(^{13}\text{C})$ 1.799        | $d_A(^{13}\text{C})$ 1.811 | $d_A(^{13}\text{C})$ 1.853 | $d_A(^{13}\text{C})$ 1.952   |
| 5 (ID:14)<br><b>N</b><br>         | 6 (ID:13)<br><b>M</b><br>  | 7 (ID:15)<br><b>O</b><br>  | 8 (ID:16)<br><b>P</b><br>    |
| $d_A(^{13}\text{C})$ 2.152        | $d_A(^{13}\text{C})$ 2.178 | $d_A(^{13}\text{C})$ 2.412 | $d_A(^{13}\text{C})$ 2.506   |
| 9 (ID:1)<br><b>A</b><br>          | 10 (ID:7)<br><b>G</b><br>  | 11 (ID:5)<br><b>E</b><br>  | 12 (ID:3)<br><b>C</b><br>    |
| $d_A(^{13}\text{C})$ 2.666        | $d_A(^{13}\text{C})$ 2.783 | $d_A(^{13}\text{C})$ 2.783 | $d_A(^{13}\text{C})$ 2.877   |
| 13 (ID:8)<br><b>H</b><br>         | 14 (ID:6)<br><b>F</b><br>  | 15 (ID:4)<br><b>D</b><br>  | 16<br><b>B, Original</b><br> |
| $d_A(^{13}\text{C})$ 2.937        | $d_A(^{13}\text{C})$ 2.937 | $d_A(^{13}\text{C})$ 2.970 | $d_A(^{13}\text{C})$ 3.039   |

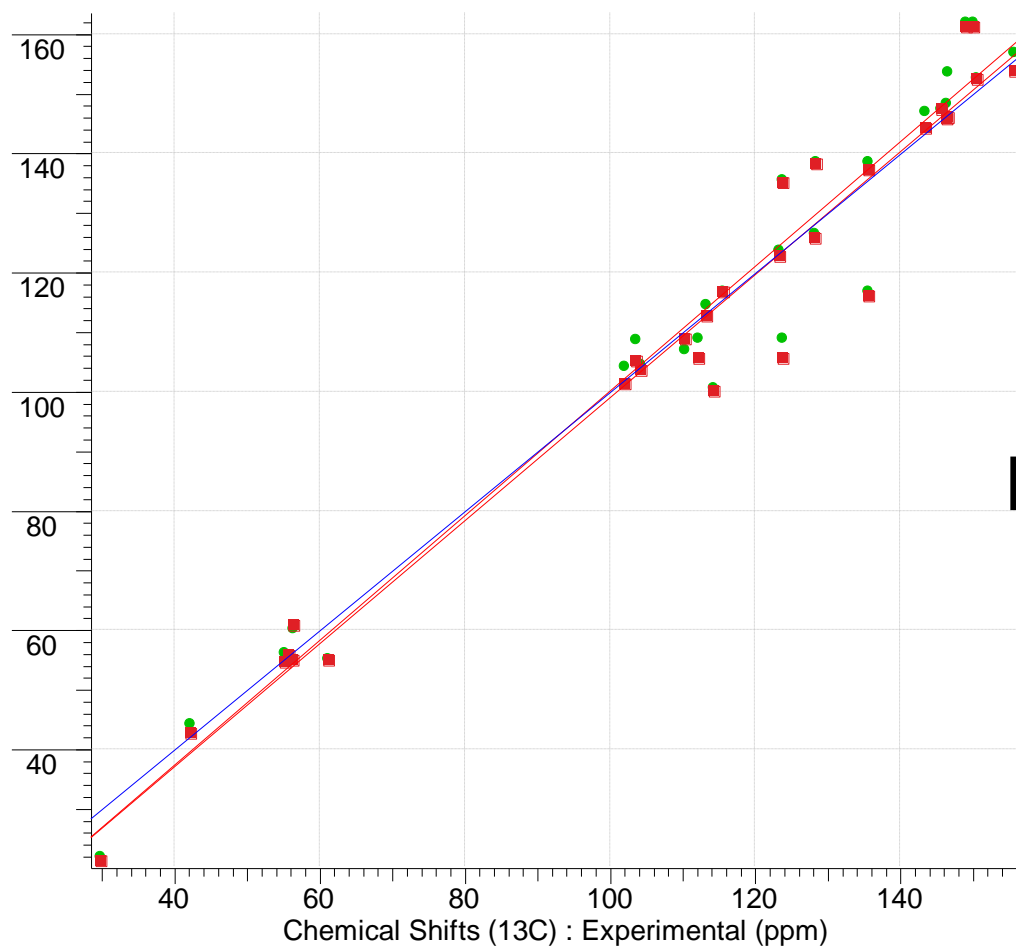
**“New lamellarin alkaloids from the indian *ascidian didemnum obscurum* and their antioxidant Properties”**

**Krishnaiah, P.; Reddy, V. L. N.; Venkataramana, G.; Ravinder, K.; Srinivasulu, M.; Raju, T. V.; Ravikumar, K.; Chandrasekar, D.; Ramakrishna, S.; Venkateswarlu, Y.**  
**J. Nat. Prod. 2004, 67, 1168-1171.**

# Assumptions: All 2D NMR correlations are standard Symmetry violation is allowed (?)



# Correlation between calculated $^{13}\text{C}$ chemical shifts (Y) and experimental ones (X).



$$R^2 = 0.965$$

Database: Proposed Structures

- Chemical Shifts ( $^{13}\text{C}$ ) : NN Calc. (ppm) (Current Record) (30 pts)
- Chemical Shifts ( $^{13}\text{C}$ ) : HOSE Calc. (ppm) (Current Record) (30 pts)

# Applications of Computer-Aided Methods of Structure Elucidation and Revision.

## I. Structure Revision of Lamellarin $\gamma$ .

M. E. Elyashberg, K. A. Blinov, S.G. Molodtsov, T.S. Churanova, A.W. Williams.

***ChemSpider Journal of Chemistry***. 2009,

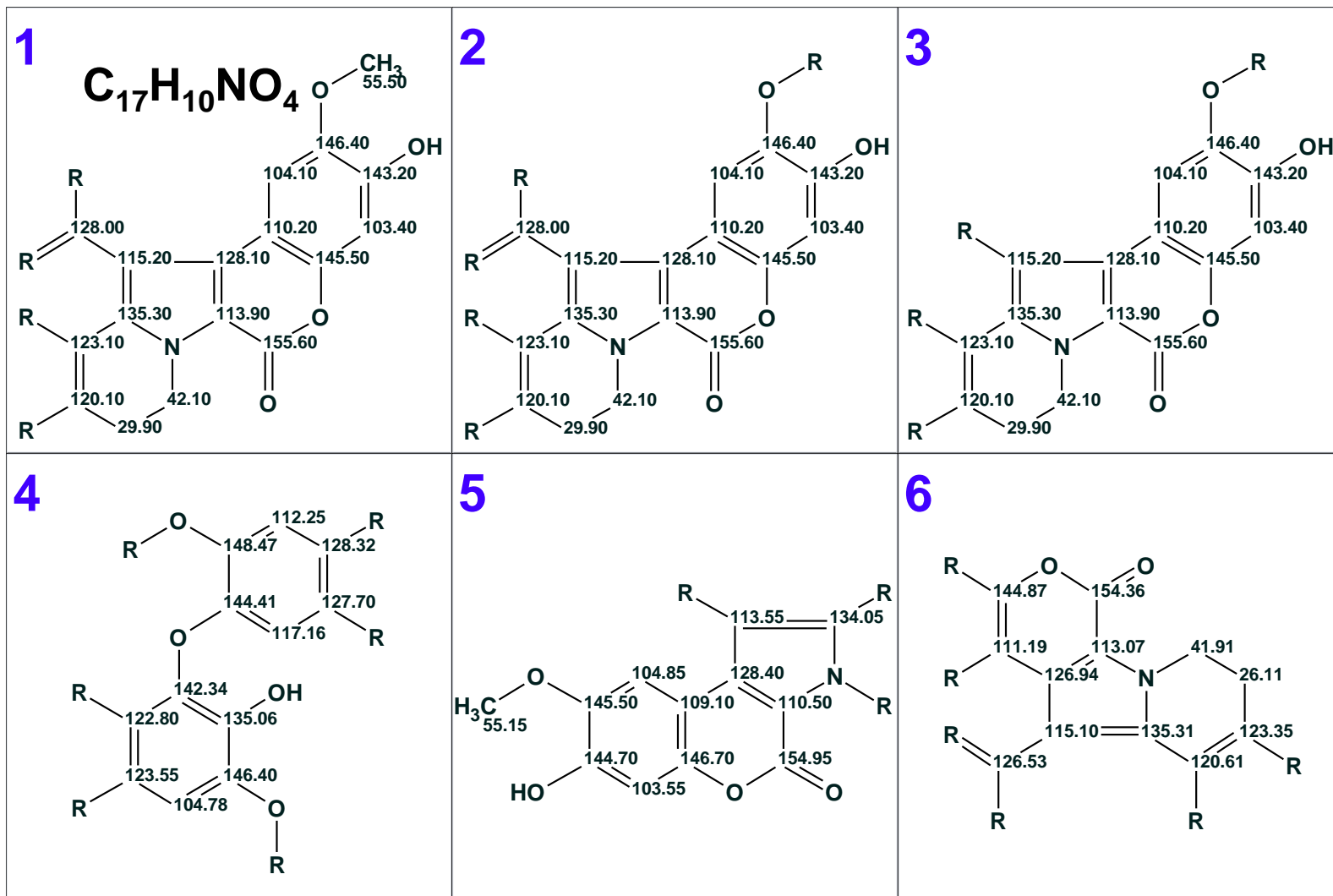
**Published:**3/18/2009 11:20:33 AM

<http://www.chemmantis.com/Article.aspx?id=889>

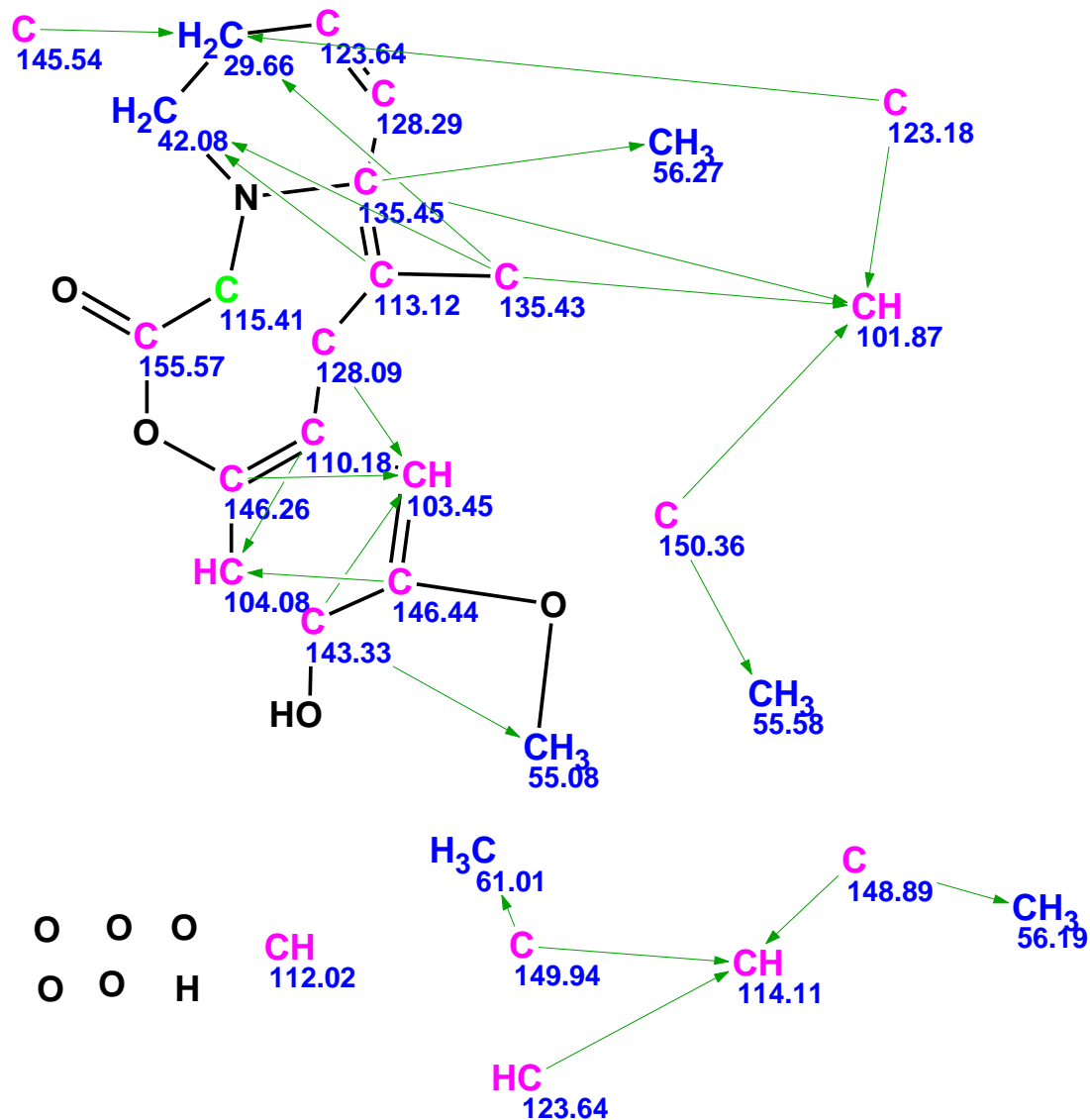
**Molecular formula: C<sub>30</sub>H<sub>27</sub>O<sub>9</sub>N, DBE = 18**

**Assumption: Fragment DB contains an appropriate fragment.**

**Number of Fragments Found by <sup>13</sup>C: L = 2 318; p = 10; n(MCD) = 192**



# One of 192 MCDs having different shift assignments of fragment atoms.

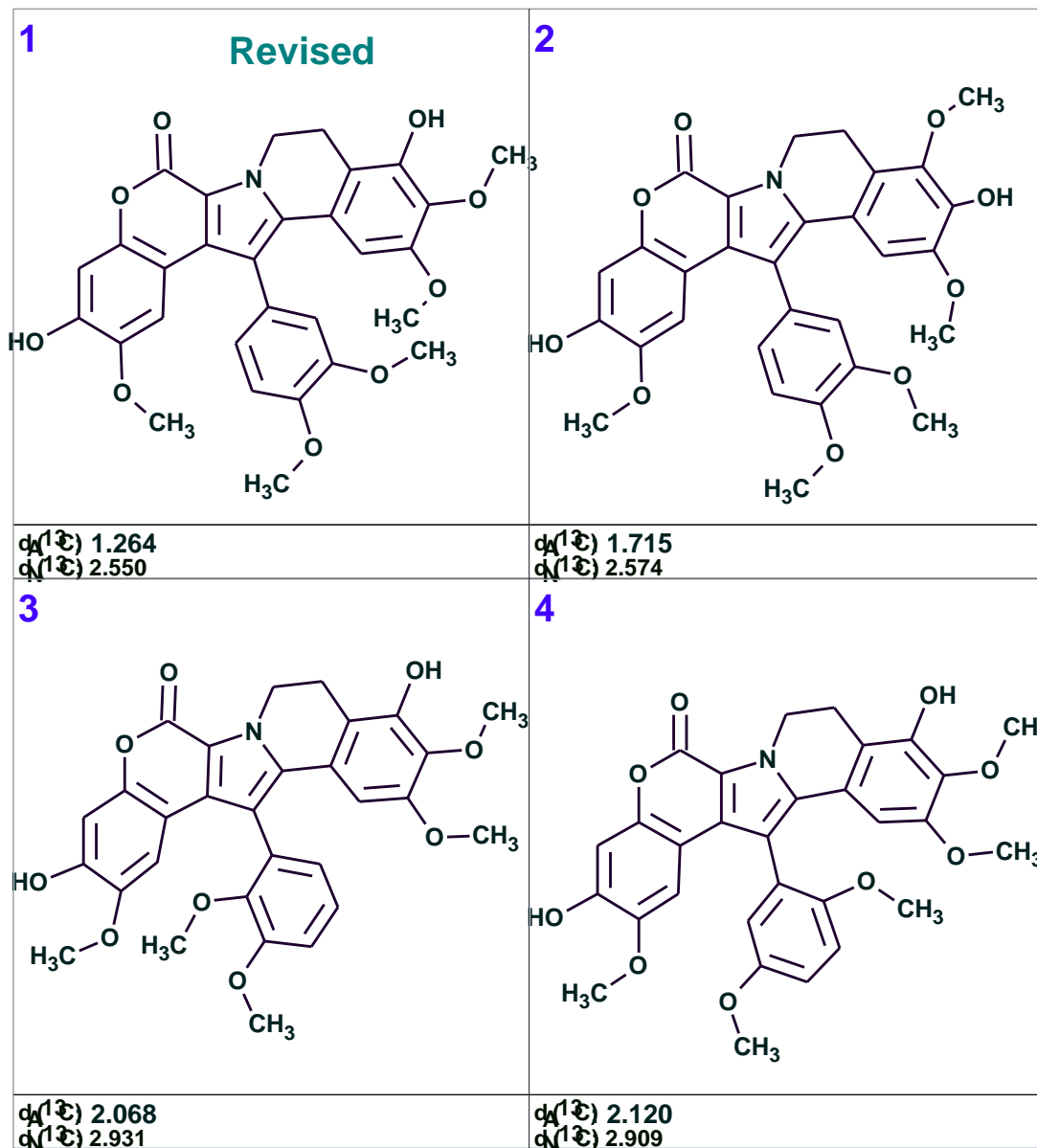


# Fuzzy Structure Generation from 192 MCDs

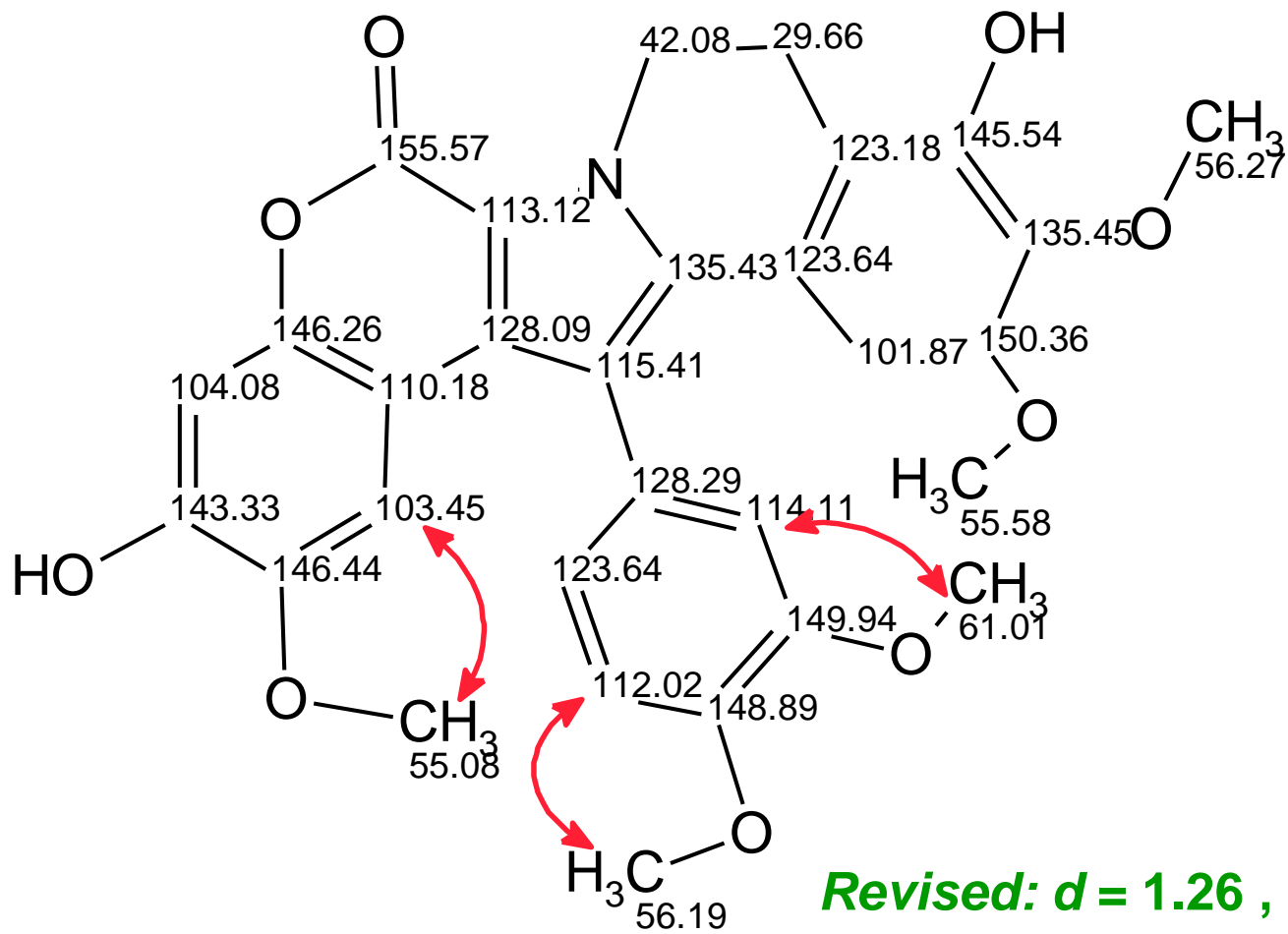
**Result:  $k = 133, 504 \rightarrow 120, 816$ ;  $t_g = 11$  min**

- Time of  $^{13}\text{C}$  spectrum calculation for ~121 thousand structures - 11 min**
- Removal of duplicates:  $120\ 816 \rightarrow 1\ 530$**

# Top structures of the ranked output file



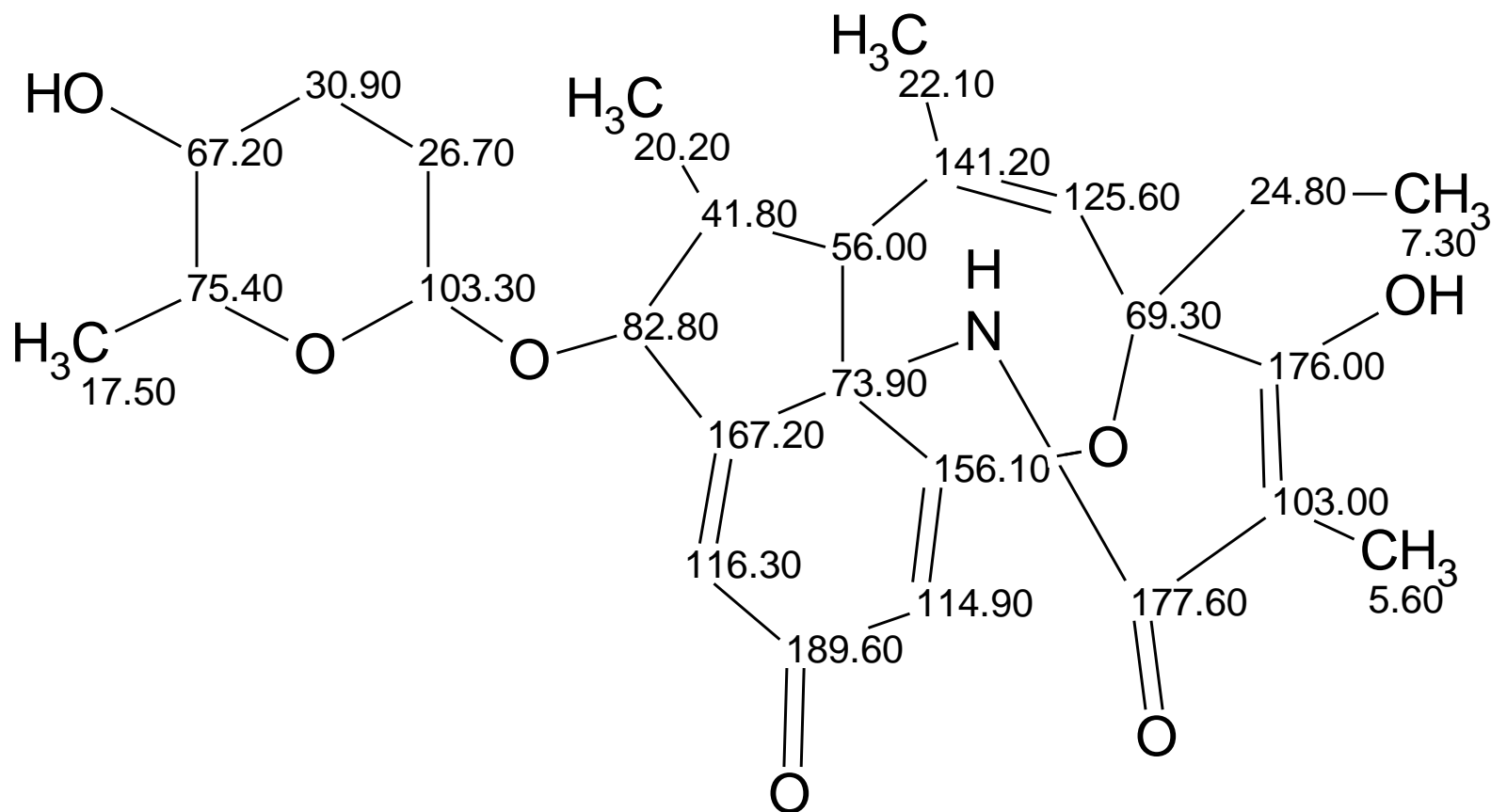
# Structure meets HMBC, NOE and Symmetry.



**Revised:  $d = 1.26, 2.55$  ppm;**  
**Original:  $d = 4.70, 5.29$  ppm**

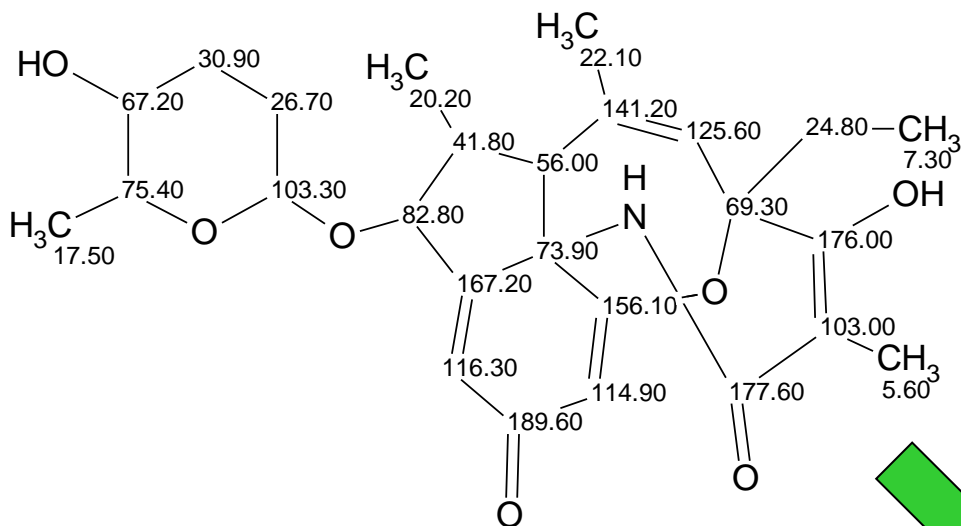
Komoda, T.; Sugiyama, Y.; Abe, N.; Imachi, et al.  
Tetrapetalone A, a novel lipoxygenase inhibitor from  
*Streptomyces sp.*

*Tetrahedron Letters* 44 (2003) 1659-1661

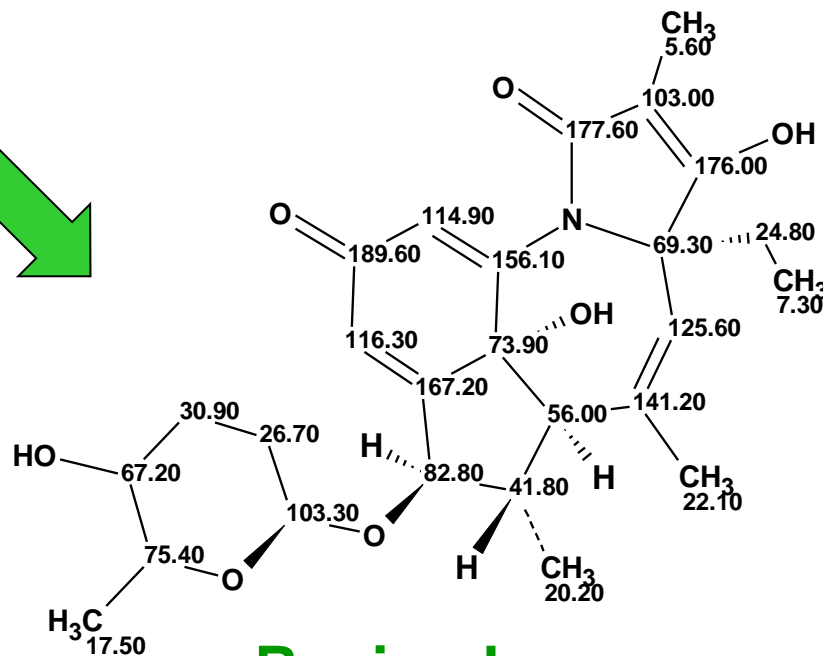


**Komoda, T.; Sugiyama, Y.; Abe, N.; Imachi, M. et al**  
**Revised structure of tetrapetalone A and its absolute stereochemistry.**

**Tetrahedron Letters 44 (2003) 7417-7419**

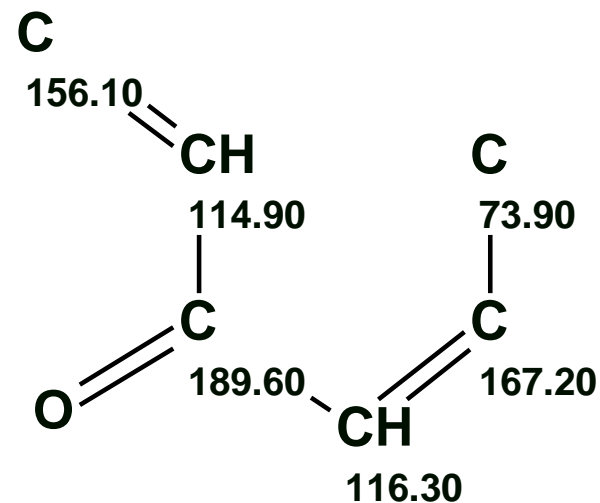
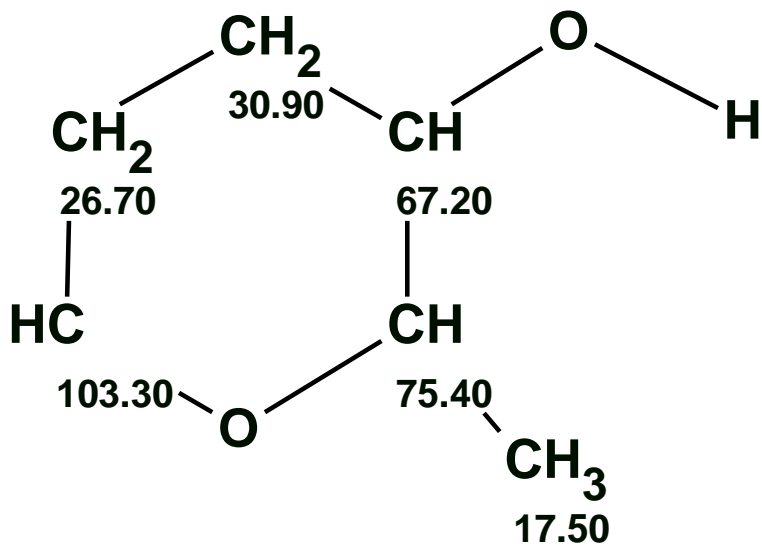
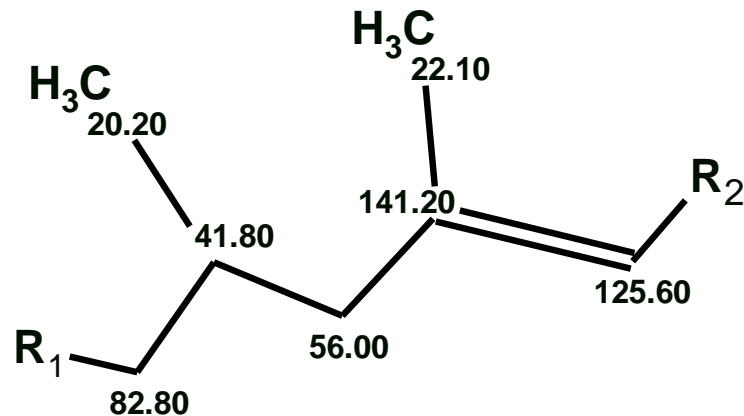
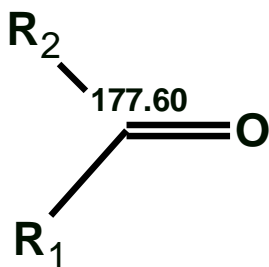
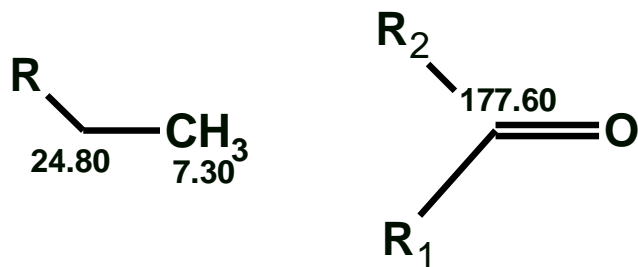


**Original**



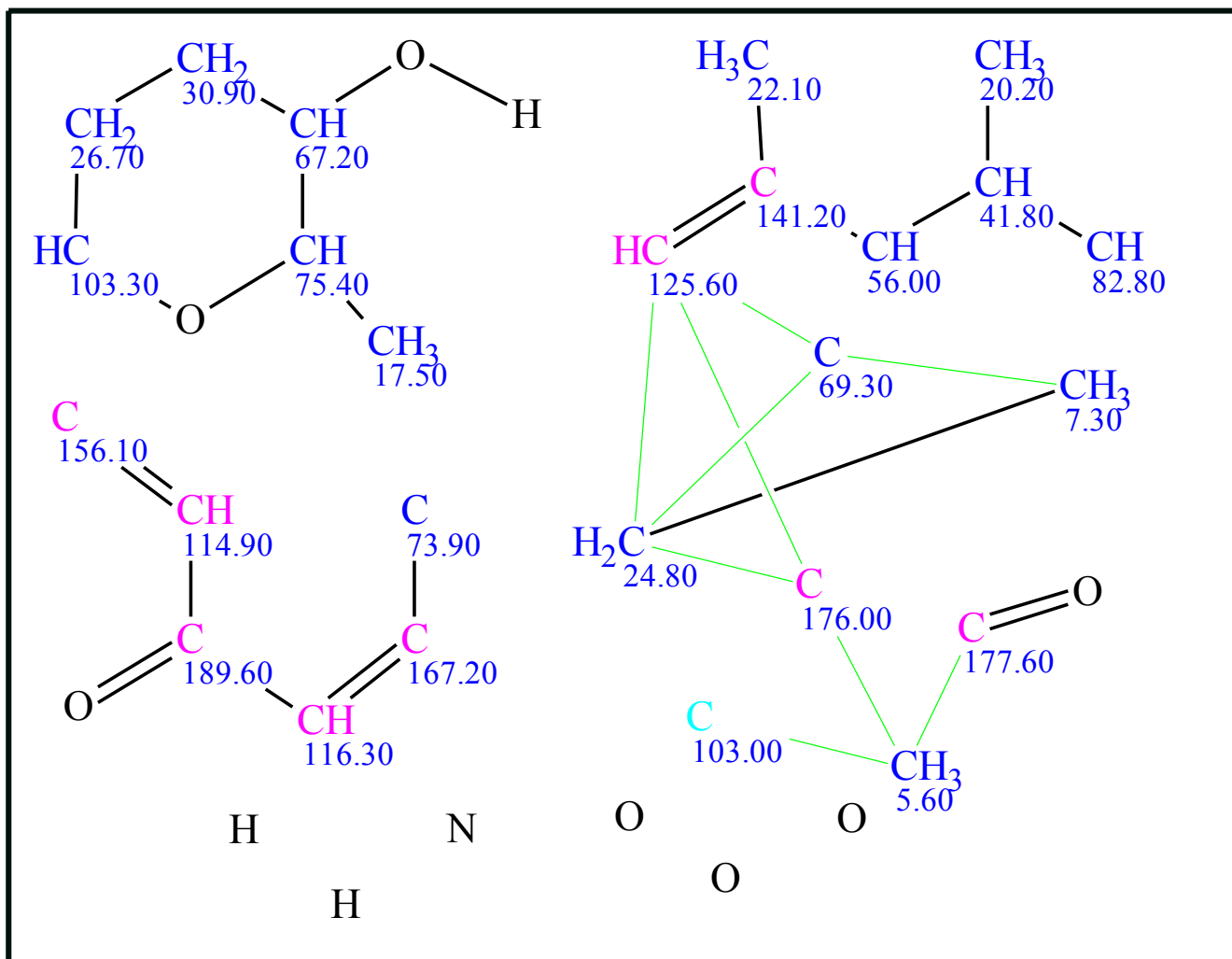
**Revised**

# Fragments deduced by authors of ORIGINAL structure



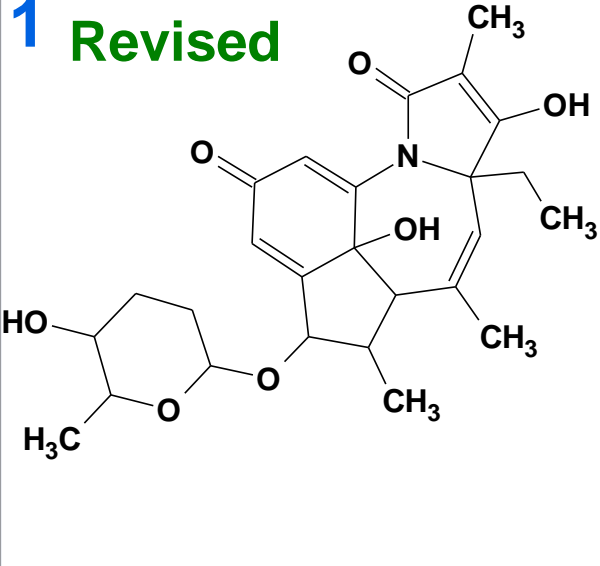
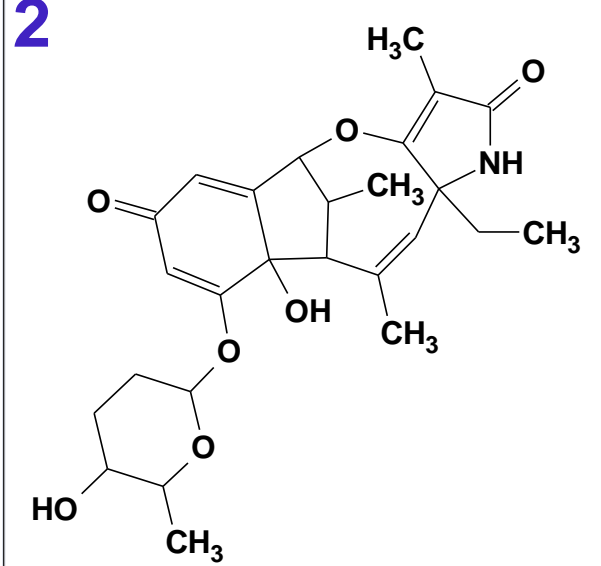
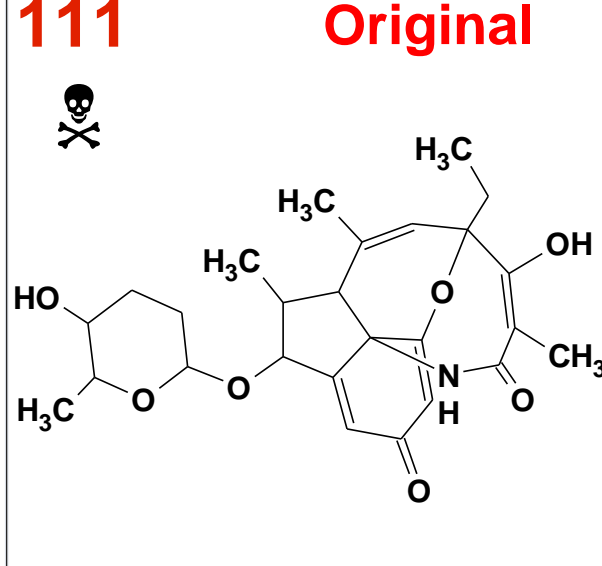
# Hypothesis: 5 fragments inferred by authors from 2D NMR data are present.

## The task: to complete the structure



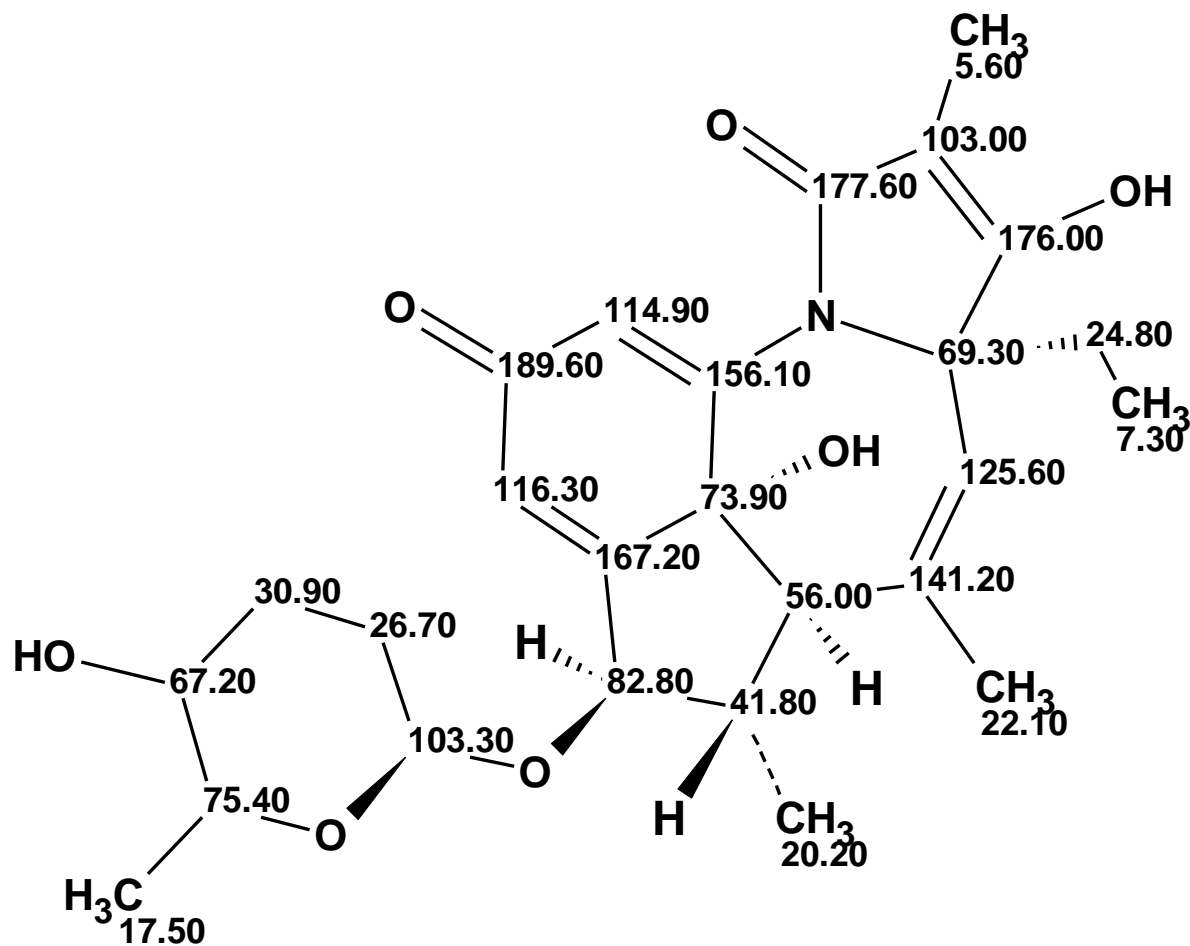
# Strict structure generation

$k = 16,465 \rightarrow 13,672 \rightarrow 9,203, t_g = 61 \text{ s}$

| <b>1 Revised</b><br> | <b>2</b><br> | <b>111 Original</b><br> |
|---|---|---|
| $d_A(^{13}\text{C})$ 2.160<br>$d_N(^{13}\text{C})$ 2.836  | $d_A(^{13}\text{C})$ 2.922<br>$d_N(^{13}\text{C})$ 3.022  | $d_A(^{13}\text{C})$ 3.370<br>$d_N(^{13}\text{C})$ 4.424  |

# Relative stereochemistry determination

All 256 stereoisomers have been generated by StucEluc and right stereochemistry was determined by  $^{13}\text{C}$  chemical shift prediction.

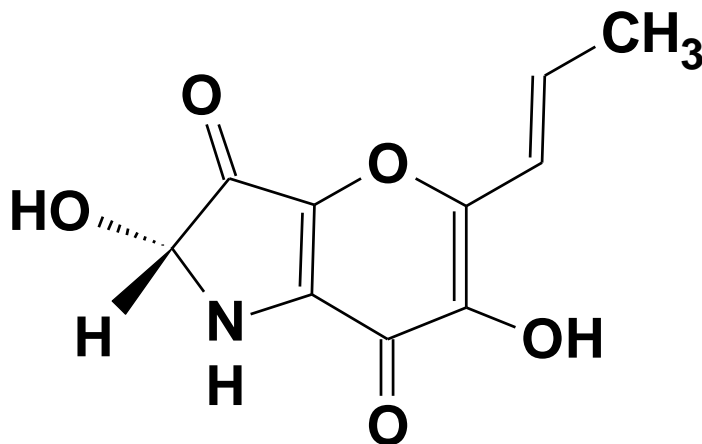


Hiort, J.; Maksimenka, K.; Reichert, M.; Perović-Ottstadt, S.; Lin, W. H.;  
Wray, V.; Steube, K.; Schaumann, K.; Weber, H.; Proksch, P.; Ebel, R.;  
Müller, W. E. G.; Bringmann, G.

**New Natural Products** from the Sponge-Derived Fungus  
**Aspergillus niger**

*J. Nat. Prod.* 2004, 67, 1532-1543

Structure was inferred from  $^1\text{H}$ ,  $^{13}\text{C}$ , COSY, HMQC, HMBC, NOESY data.



***Pyranonigrin A*,  $\text{C}_9\text{H}_{10}\text{NO}_5$ , EBD = 7**

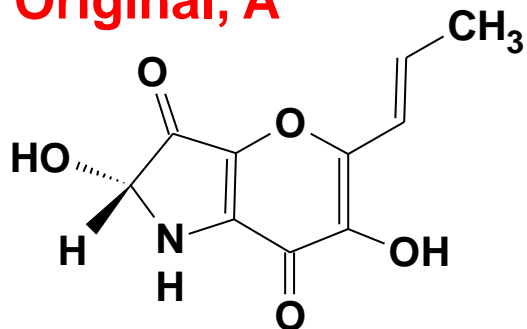
The absolute configuration was established by **Quantum Mechanical** calculations of circular dichroism (CD) spectra.

Schlingmann, G.; Taniguchi, T.; He, H.; Bigelis, R.; Yang, H. Y.;  
Koehn, F.E.; Carter, G.T.; Berova, N.

**Reassessing** the Structure of *Pyranonigrin*.

*J. Nat. Prod.*, 2007, 70, 1180 -1187

**Original, A**

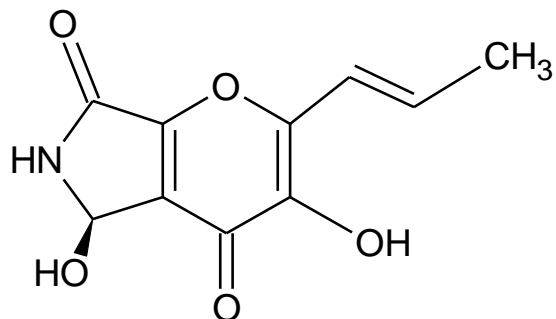


**The substance was reisolated.**

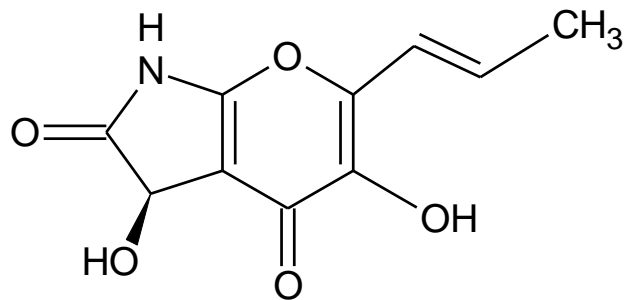
**New 2D NMR data:**

**COSY, ROESY,  $^1\text{H}$ - $^{13}\text{C}$  HMBC +  $^1\text{H}$ - $^{15}\text{N}$  HMBC.**

**3 Proposed Structures**

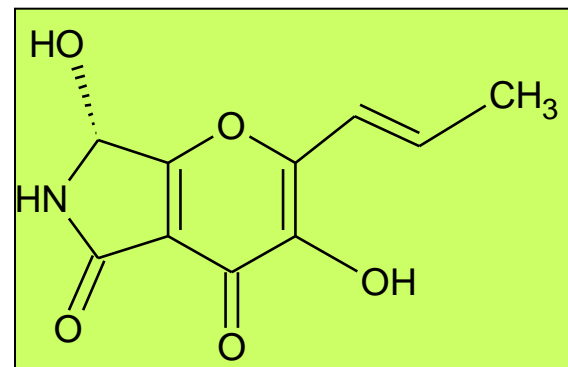


**B**



**C**

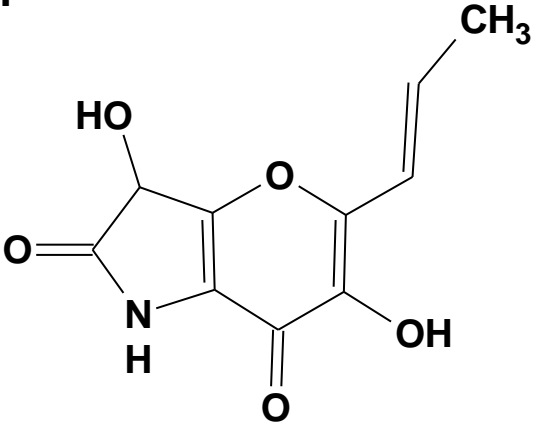

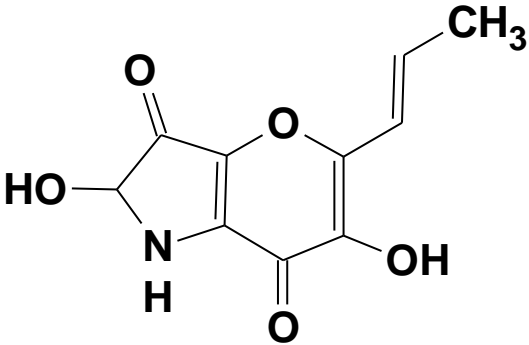
**UV/CD QM calculations**



**D**

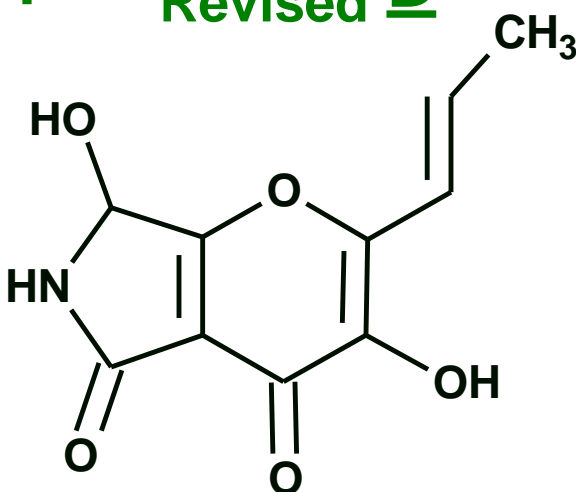
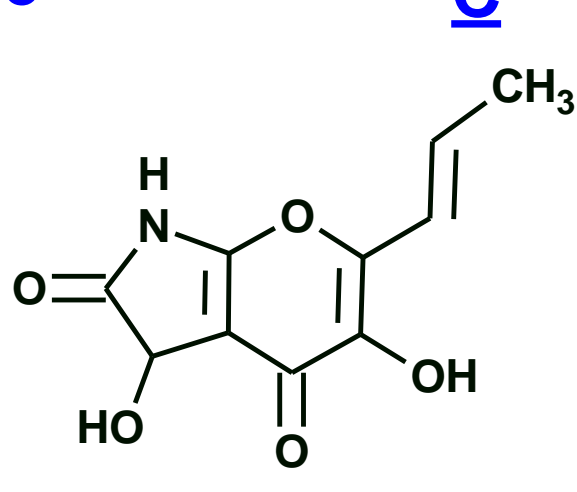
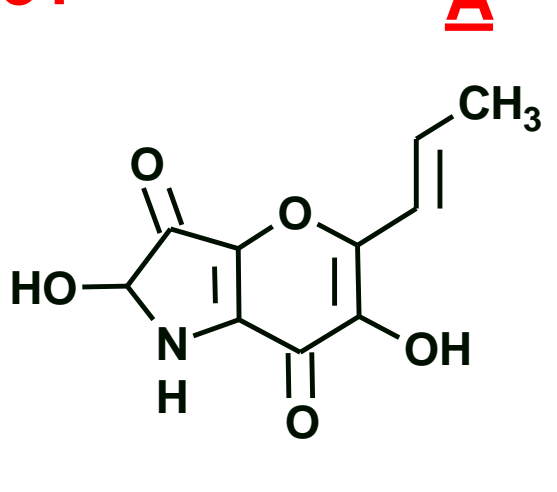
# StrucEluc application to the initial NMR data used for inferring the original structure

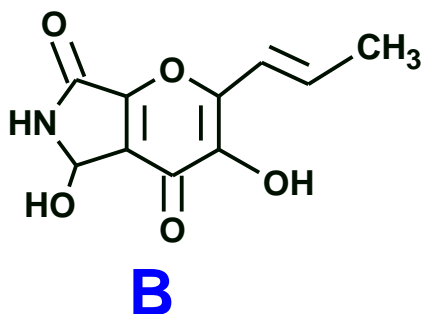
Step 1. Hypothesis: there is no NSC in 2D NMR data.  
Strict Structure Generation:  $k=109 \rightarrow 81 \rightarrow 72$ ,  $t_g = 0.3$  s

|  |   |
|--|---|
| <p>1</p>                                     | <p>6  <b>Original, A</b></p>  |
| <p><math>d_A(^{13}\text{C})</math> 5.603<br/><math>q_N(^{13}\text{C})</math> 5.311<br/><math>q(^{13}\text{C})</math> 7.695</p> | <p><math>d_A(^{13}\text{C})</math> 10.664<br/><math>q_N(^{13}\text{C})</math> 7.689<br/><math>q(^{13}\text{C})</math> 10.029</p>  |

# Step 2. Fuzzy structure Generation, $m=1$ , $a=x$

Result:  $k = 3024 \rightarrow 2130 \rightarrow 1144$ ,  $t_g = 14$  s.

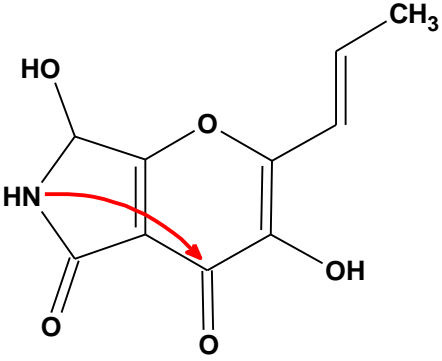
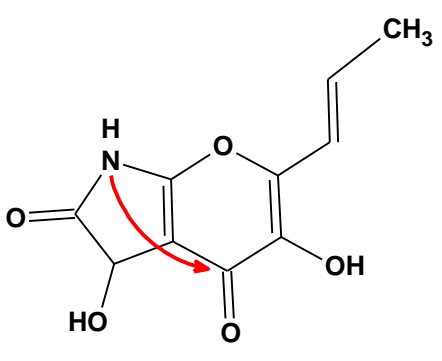
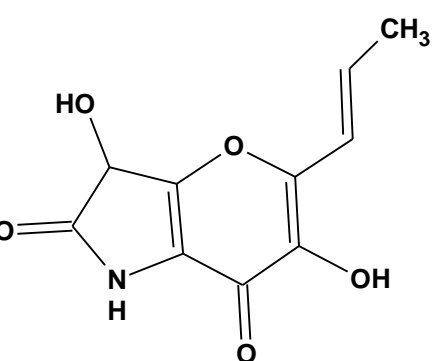
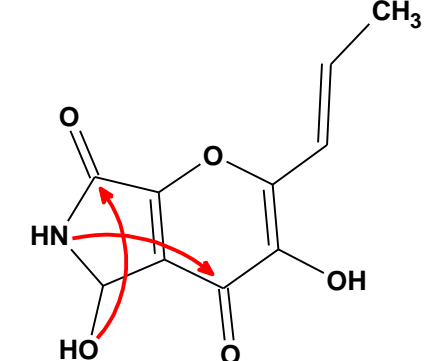
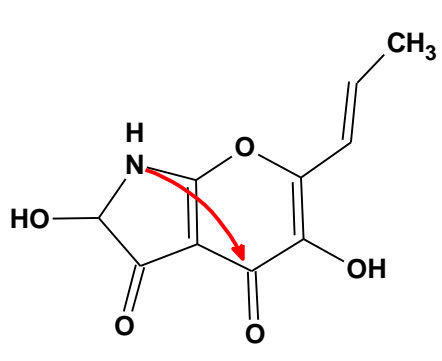
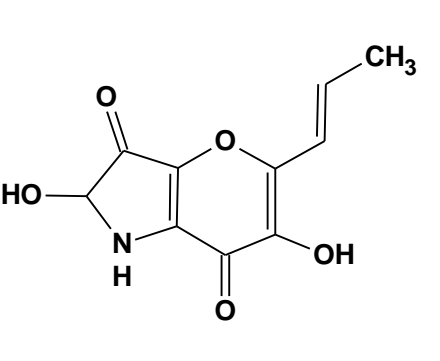
|  |  |  |
|--|--|--|
| <p><b>1</b> Revised <b><u>D</u></b></p>  <p><math>d_A^{(13C)} : 2.027</math></p> | <p><b>5</b> <b><u>C</u></b></p>  <p><math>d_A^{(13C)} : 5.264</math></p> | <p><b>31</b> <b><u>A</u></b></p>  <p><math>d_A^{(13C)} : 9.003</math></p> |
|--|--|--|



Structure **B** was not generated

# Step 3. Fuzzy structure Generation, $m=2$ , $a=x$

Result:  $k = 18275 \rightarrow 10725 \rightarrow 3506$ ,  $t_g=2m$  23 s

|   |   |  |
|---|---|--|
| <p><b>1</b> <b>D</b></p>   | <p><b>6</b> <b>C</b></p>  | <p><b>7</b></p>             |
| <p><math>\alpha(1\mathcal{C}) : 2.027</math></p>  | <p><math>\alpha(1\mathcal{C}) : 5.264</math></p>  | <p><math>\alpha(1\mathcal{C}) : 5.603</math></p>   |
| <p><b>8</b> <b>B</b></p>  | <p><b>57</b></p>         | <p><b>95</b> <b>A</b></p>  |
| <p><math>\alpha(1\mathcal{C}) : 5.633</math></p>  | <p><math>\alpha(1\mathcal{C}) : 8.094</math></p>  | <p><math>\alpha(1\mathcal{C}) : 10.664</math></p>  |

# Conclusions

Application of StrucEluc allows one:

- Automatically generate all structural hypotheses from the set of axioms adopted by the chemist.
- Automatically choose the most probable structure.
- Automatically establish influence of any axiom on the final result.
- **Avoid “unavoidable” pitfalls!**

# Structure Elucidator Team



**Sergey Molodtsov, Mikhail Elyashberg, Tatiyana Churanova, Kirill Blinov**