A few Remarks on wrong Structures in the Literature

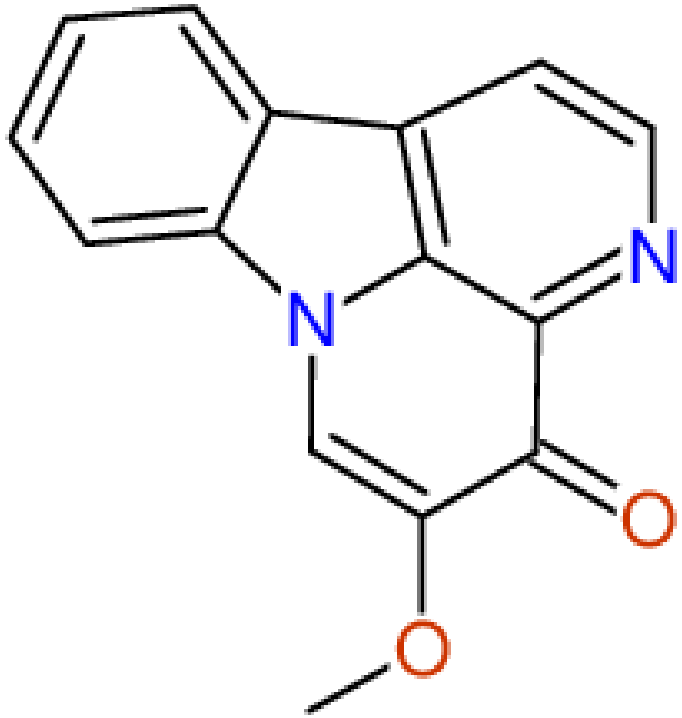
Wolfgang Robien

Wolfgang.Robien@univie.ac.at

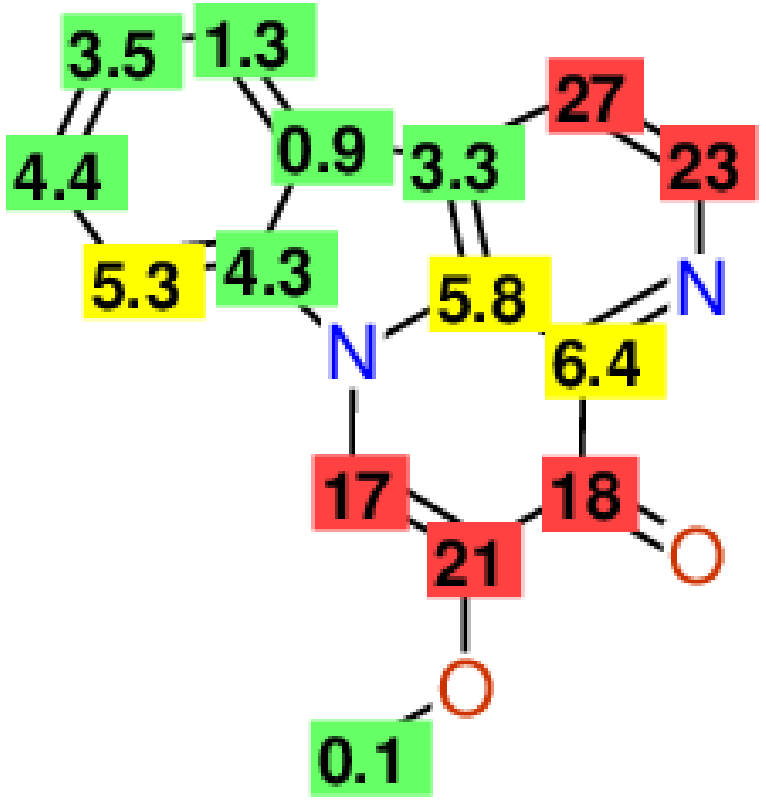
University of Vienna, Department of Organic Chemistry
A-1090 Vienna / Austria

Drymaritin from JNP-2004: 1D/2D at 600MHz incl. NH-HMBC

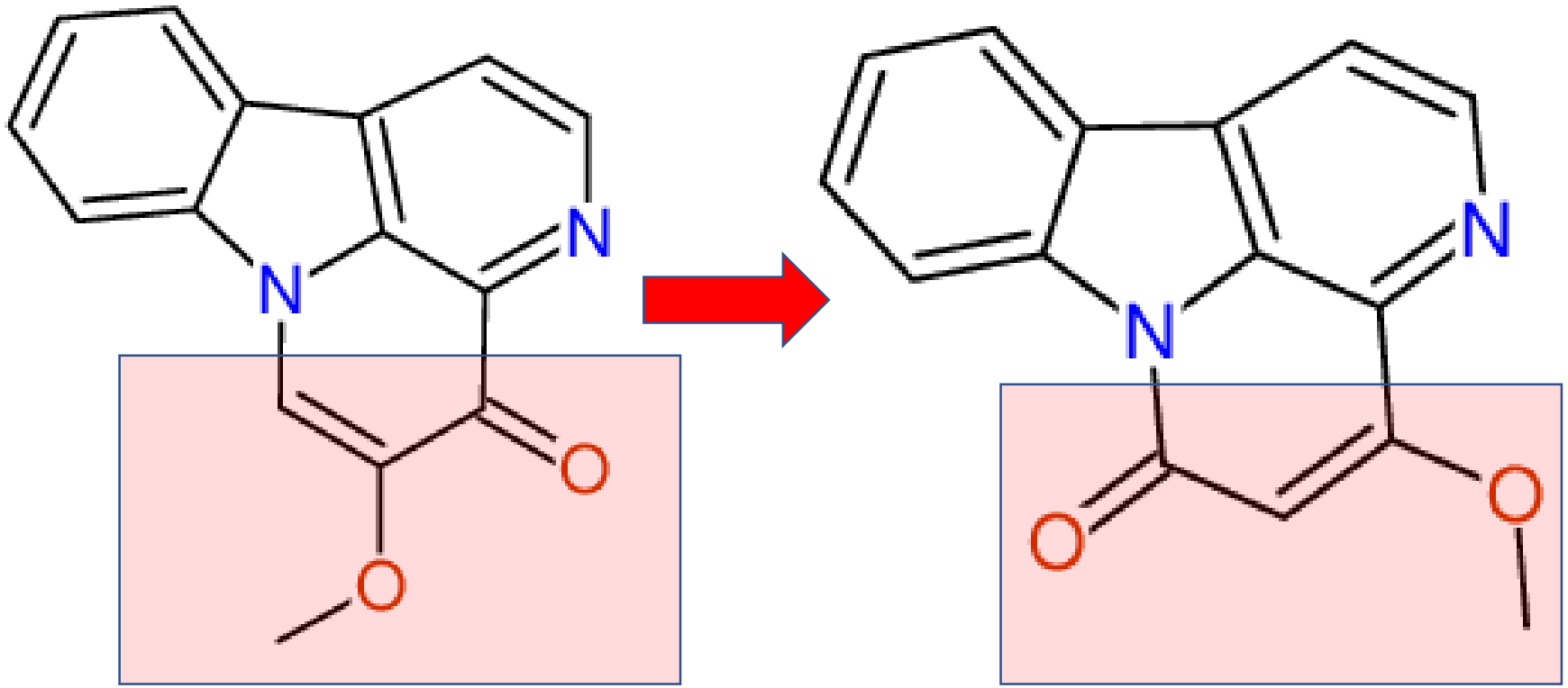
Proposed Structure



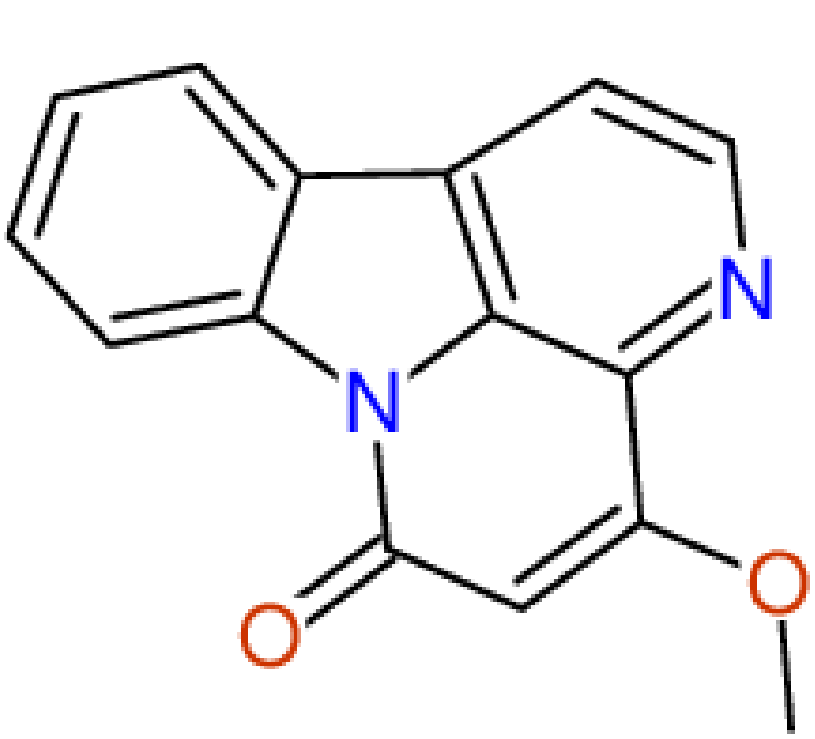
Deviation: δ_{exp}-δ_{calc}

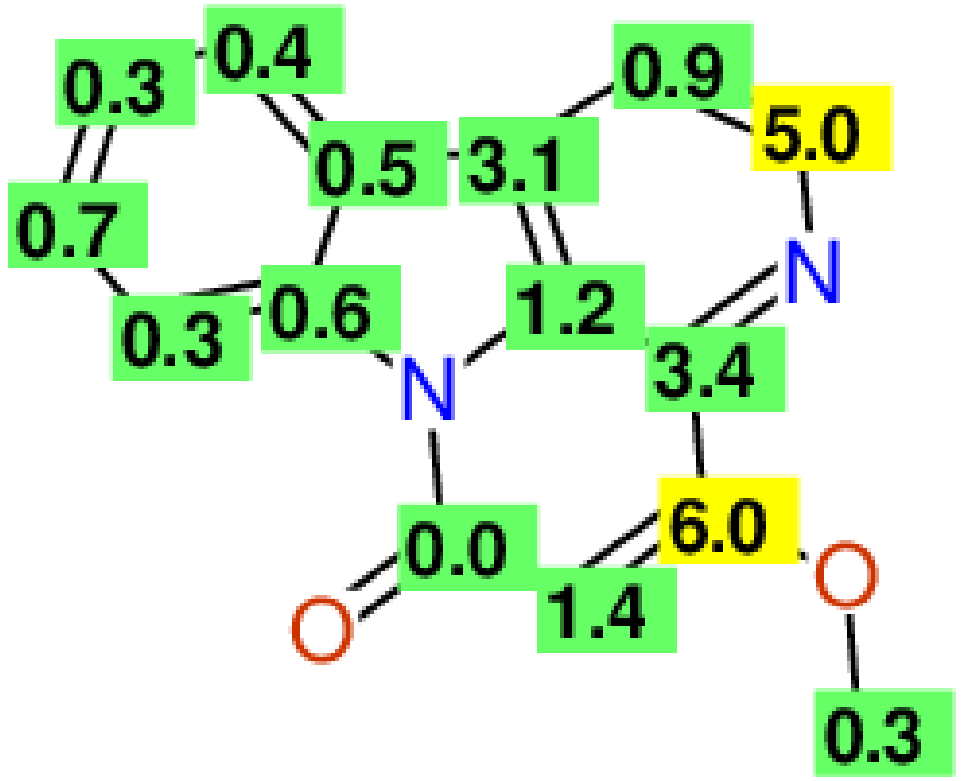


Structure generation starting from the WRONG PROPOSAL and the C-13 NMR peaklist using CSEARCH-technology gives the correct structure within 1 minute of CPU-time

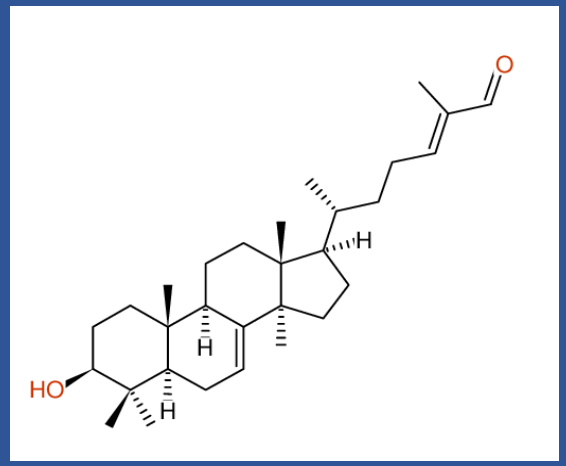


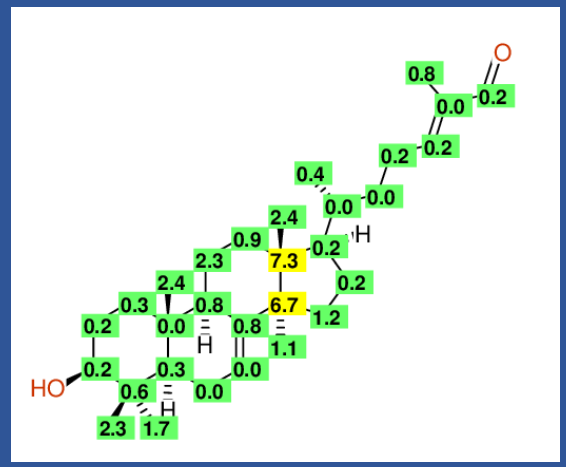
Correct structure – proven by synthesis (Bracher, JNP-2009)

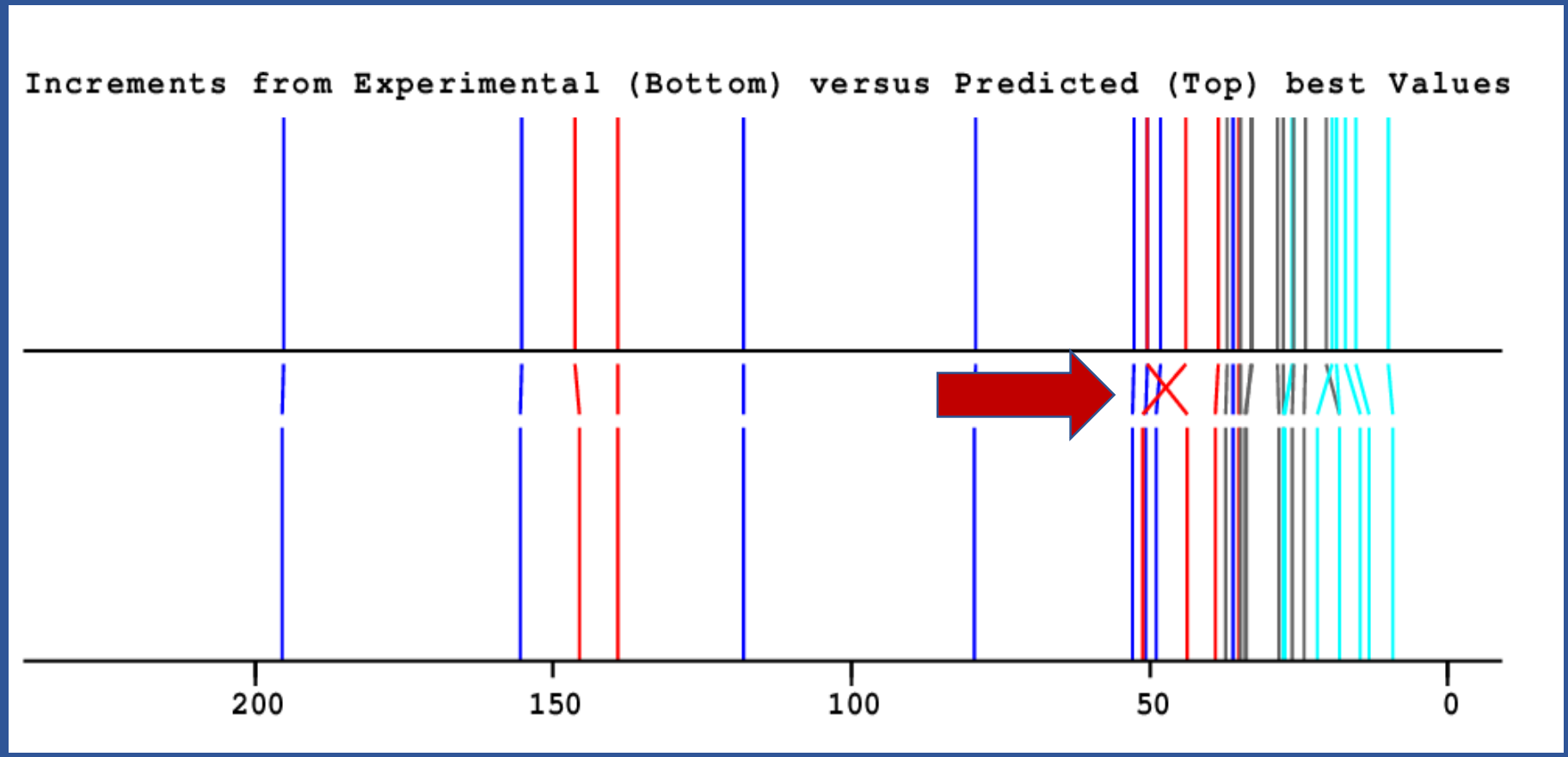




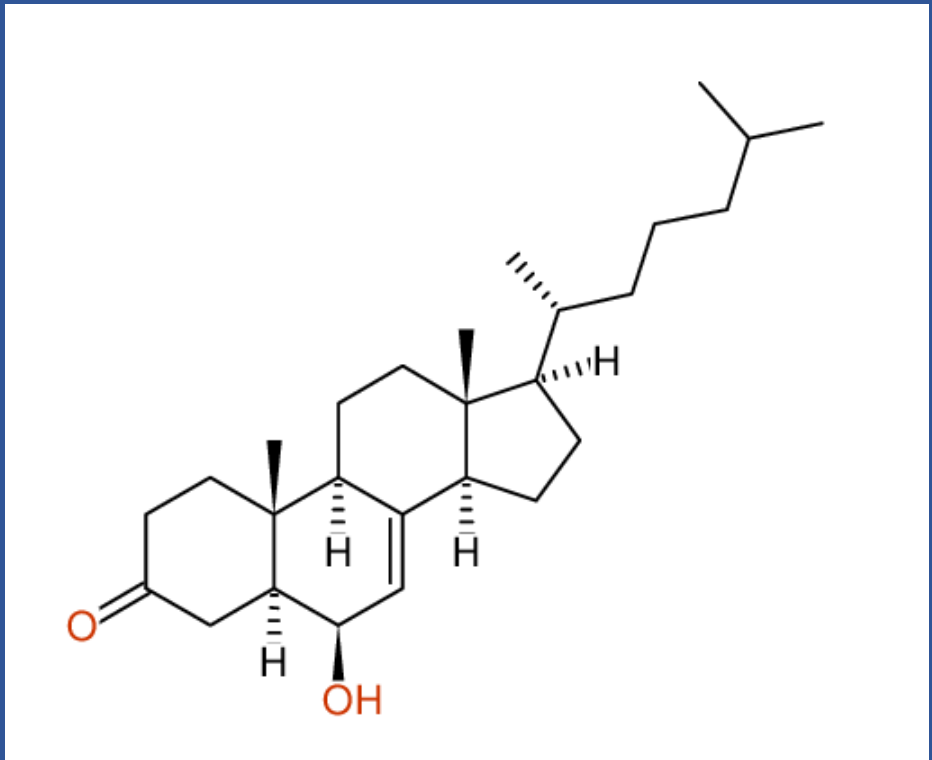
Kiusianin A – Assignment error / CPB, 62, 937 (2014)

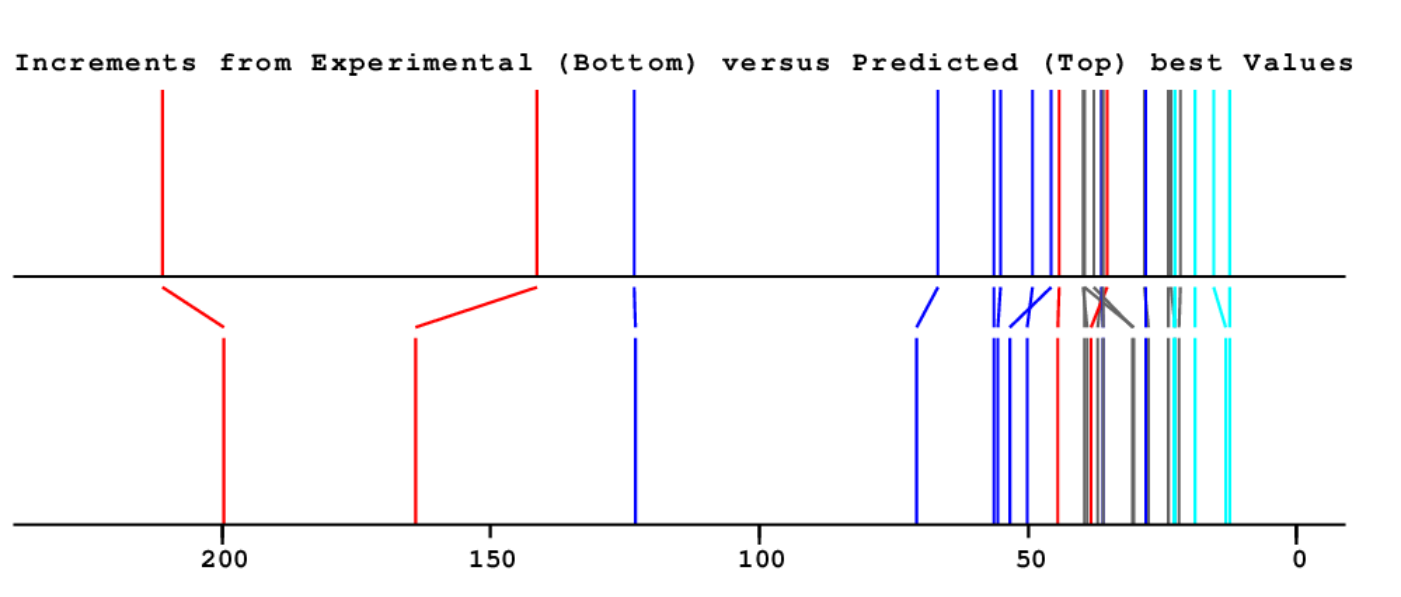




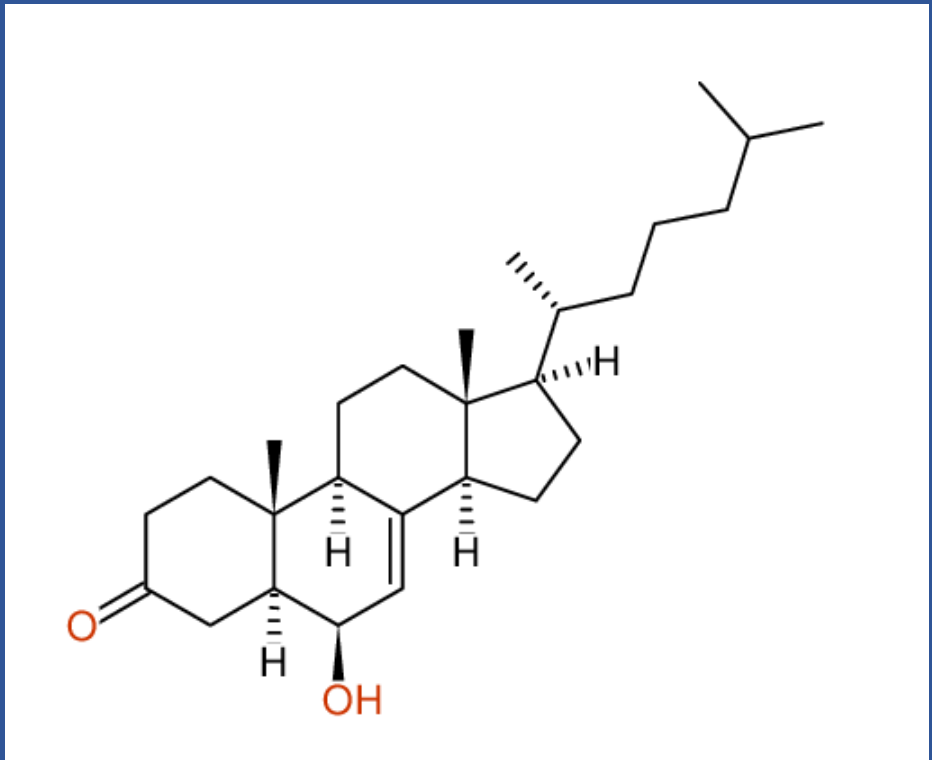


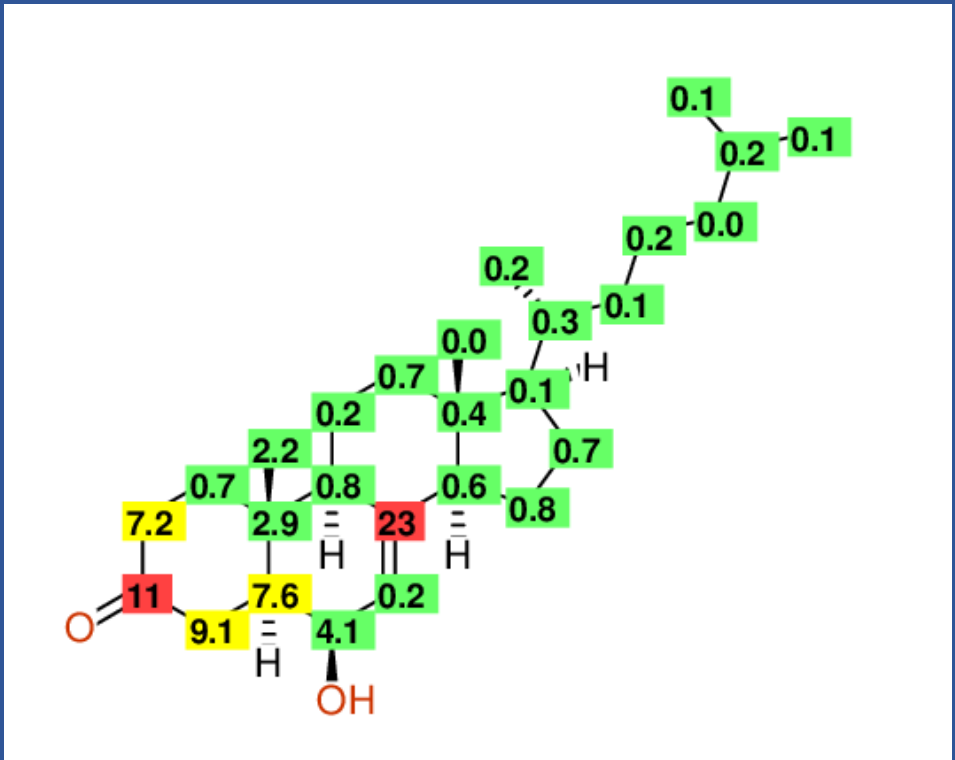
Kiusianin C – Wrong structure proposal / CPB, 62, 937 (2014)



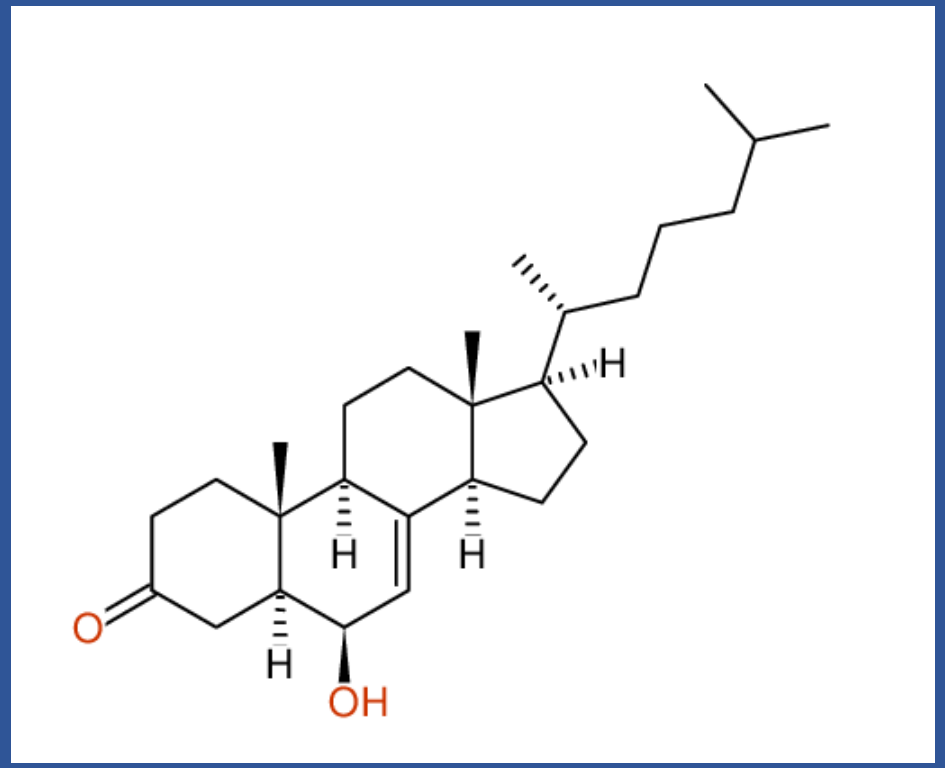


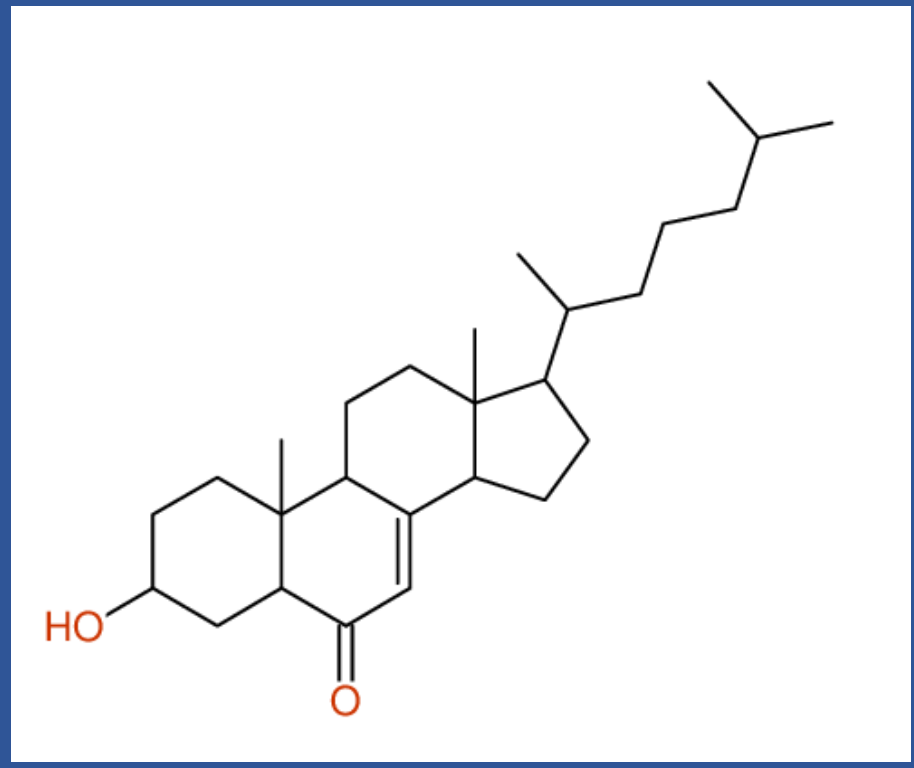
Kiusianin C – Wrong structure proposal / CPB, 62, 937 (2014)





Kiusianin C – Wrong structure proposal / CPB, 62, 937 (2014)
Automatic Revision only from 1D-13C-Peaklist & Wrong Proposal



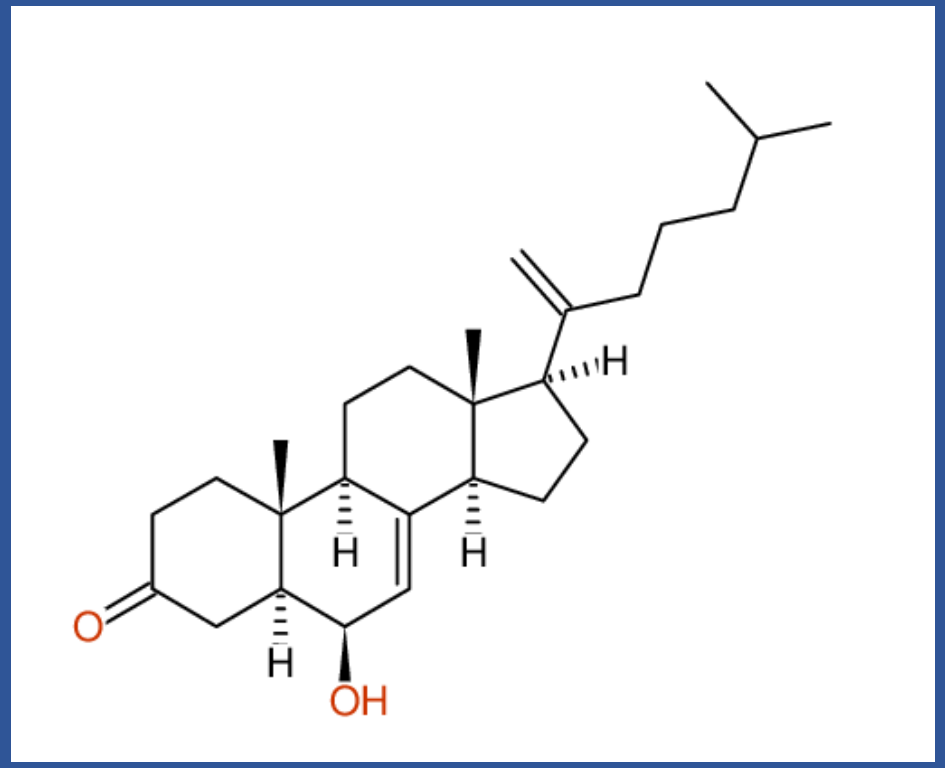


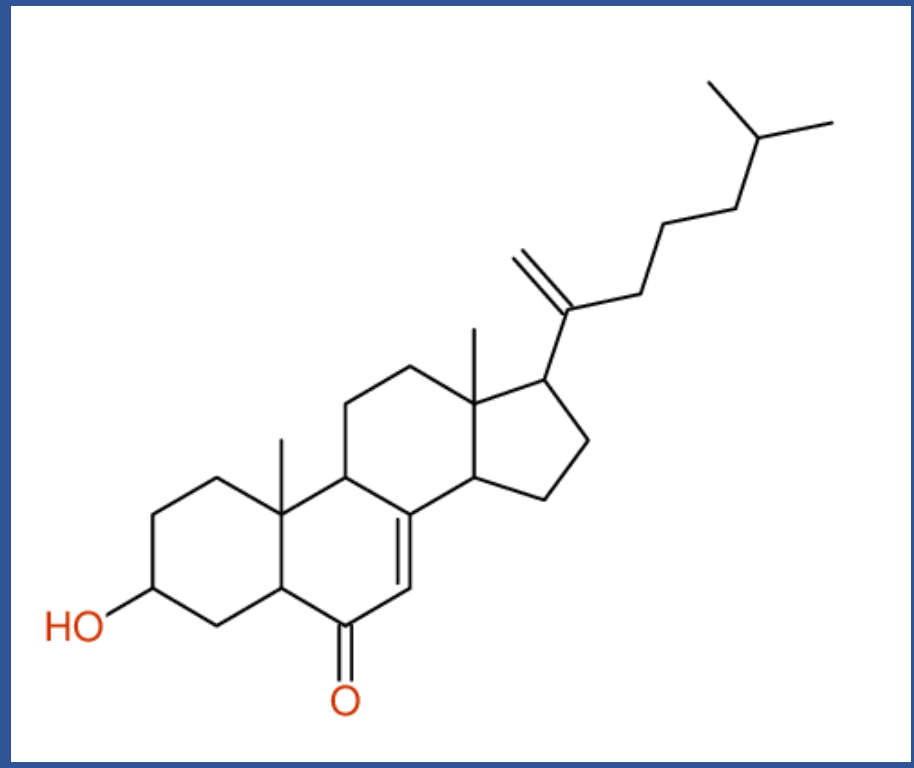
Hitlist

Position: #857 / 4.27 ppm

Position: #1 / 0.79 ppm

Kiusianin D – Wrong structure proposal / CPB, 62, 937 (2014)
Automatic Revision only from 1D-13C-Peaklist & Wrong Proposal



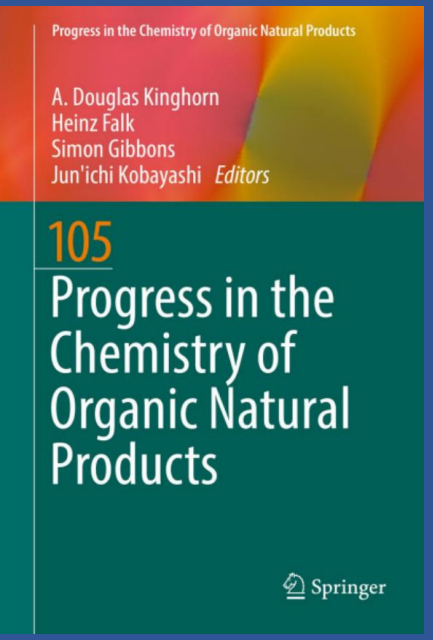


Hitlist

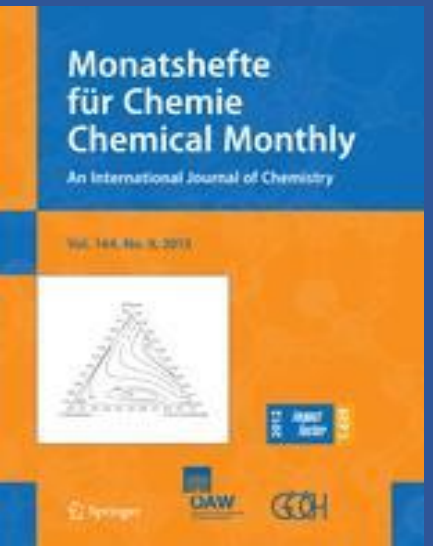
Position: #307 / 2.92 ppm

Position: #7 / 1.46 ppm

More Examples for Automatic Structure Revisions



Robien W. A Critical Evaluation of the Quality of Published 13C NMR Data in Natural Product Chemistry Progress in the Chemistry of Organic Natural Products, eds. Kinghorn AD, Falk H, Gibbons S, Kobayashi JI 2017; 105:137-215



Robien W. Computer-assisted peer reviewing of spectral data: The CSEARCH protocol. Monatsh Chem 2019; 150, 927-932 https://doi.org/10.1007/s00706-019-02407-5

Request

Structure
Peaklist
Structure & Peaklist

CSEARCH
Robot-Referee

Knowledge base

340,000 experimental (curated collection) C13-NMR Spectra

- **Predicted C13-NMR Spectrum**
- **Structure Proposals**
- **Structure Verification & Alternative Structures & Similarity Search**

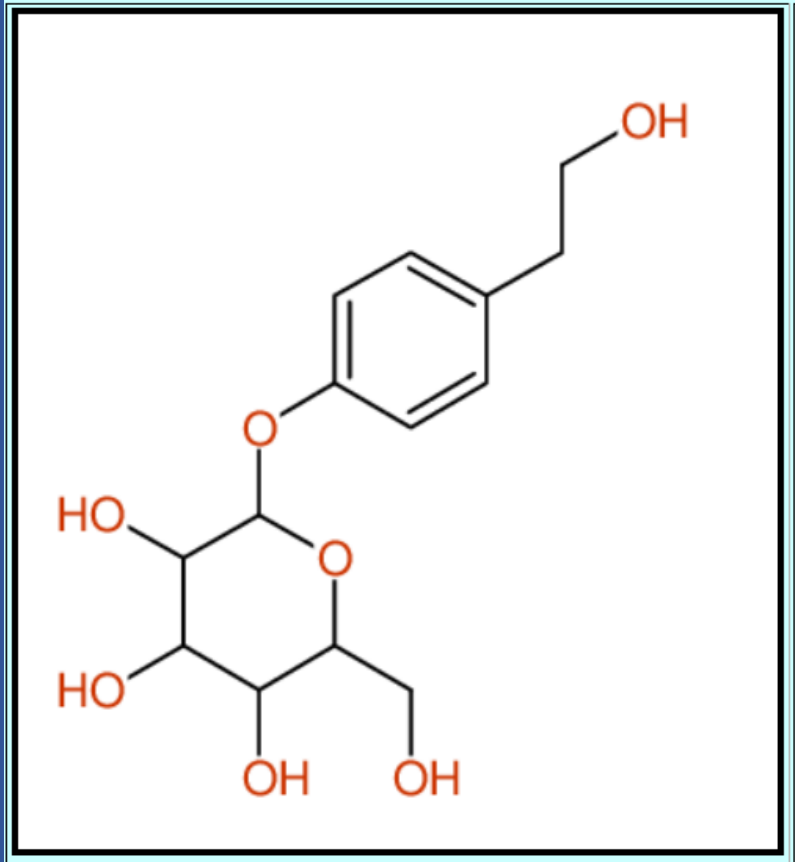
286,000,000 predicted C13-NMR Spectra 65,000,000 in preparation

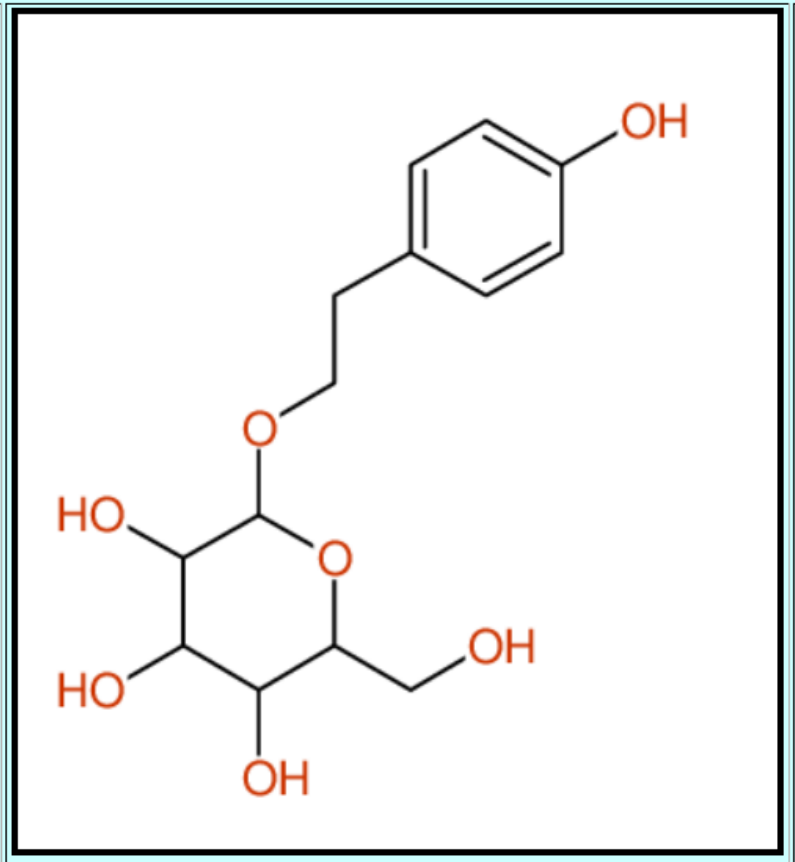
Isomer generator	Structure generator
Create all possible isomers exactly once Complete scan of structural space Combinatorial explosion – use constraints Based on graph-theory	Create a certain set of similar structures Incomplete scan of structural space Maximal number of alternatives can be given Based on predefined rules
Perfect solution – if molecular formula is correct Must be restarted with another molecular formula if necessary	Pragmatic solution – creates also non-isomeric structures, depending on rule set
More than 1 solution – rank hitlist (isomer generation=exact mathematics, ranking based on spectrum prediction=probability) OR do additional experiment(s)	Structure generation & Ranking = Probability -N versus N-oxide; -S- versus –S(=O)- / -S(=O)(=O)- / -O- versus –O-O- → depends on rule set

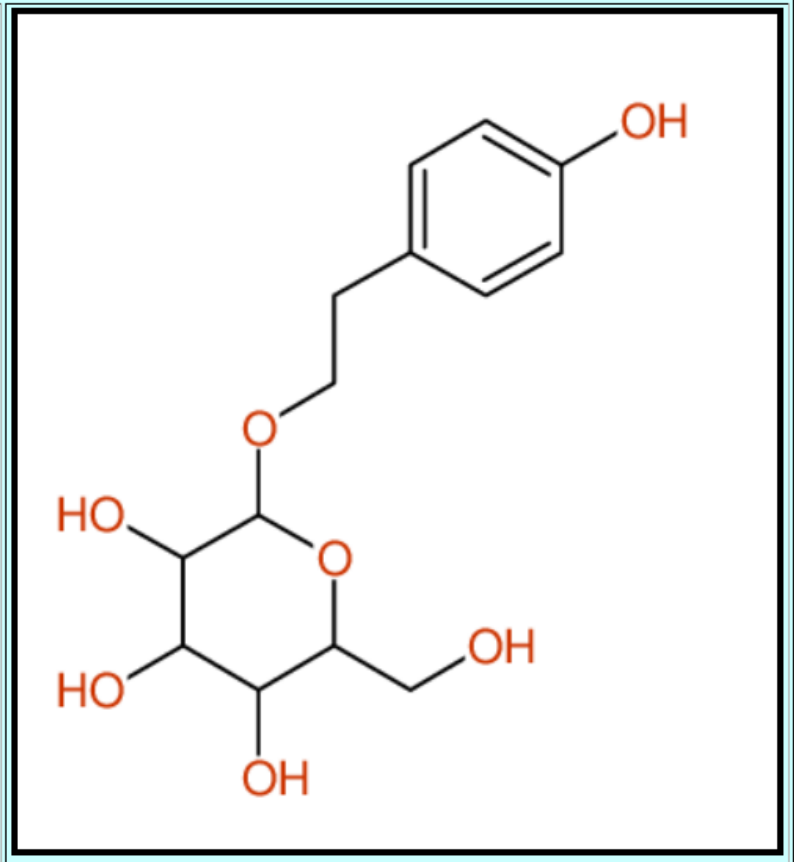
Your proposal	Best proposal	Best proposal (out of 3) which exists either in CSEARCH or PUBCHEM
Deviation = 1.50 ppm Position = 13	Deviation = 0.99 ppm	Deviation = 0.99 ppm Position = 1

Number of reasonable alternative structures: 11

Project: AS.J.TRAD.MED.,2,70[2007]CSEARCH-NUMBER:UWLU081228-20150627
Compound: ICARIDIDE-D2_COMPOUND-3



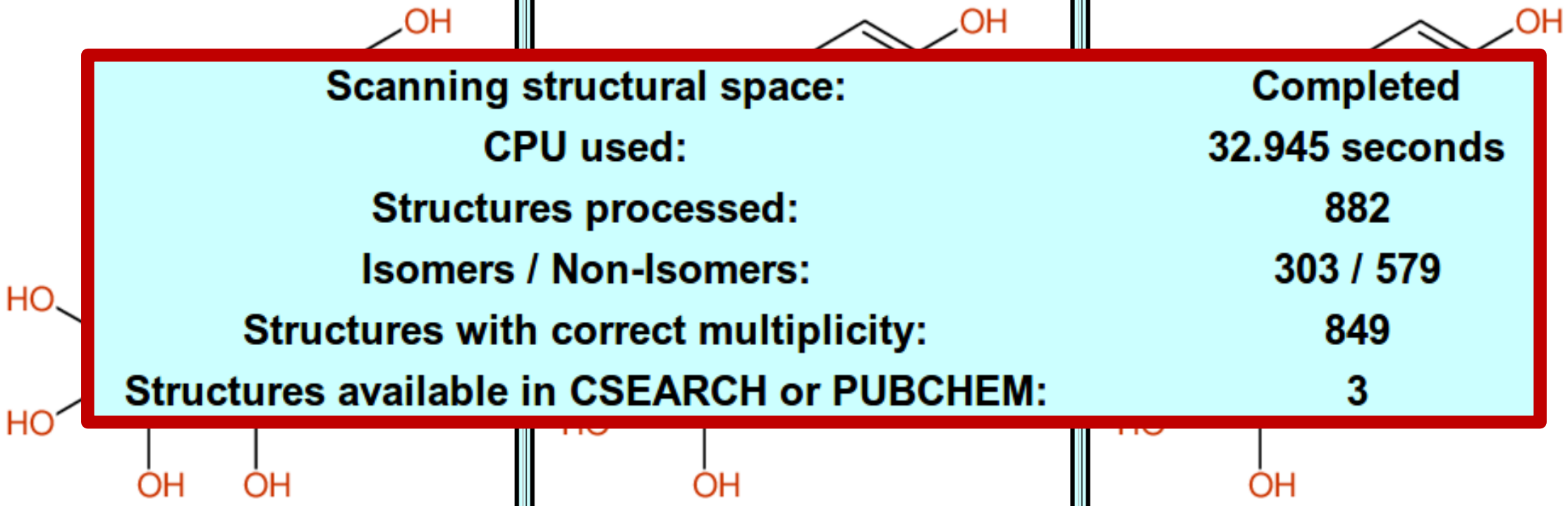




Your proposal	Best proposal	Best proposal (out of 3) which exists either in CSEARCH or PUBCHEM
Deviation = 1.50 ppm Position = 13	Deviation = 0.99 ppm	Deviation = 0.99 ppm Position = 1

Number of reasonable alternative structures: 11

Project: AS.J.TRAD.MED.,2,70[2007]CSEARCH-NUMBER:UWLU081228-20150627
Compound: ICARIDIDE-D2_COMPOUND-3



Scanning structural space:	Completed
CPU used:	32.945 seconds
Structures processed:	882
Isomers / Non-Isomers:	303 / 579
Structures with correct multiplicity:	849
Structures available in CSEARCH or PUBCHEM:	3

How to efficiently hide Structure Revisions ?
(But not efficient enough for CSEARCH !)

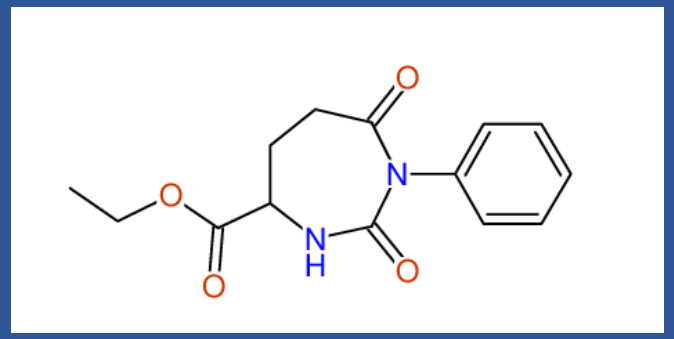
Strategy:

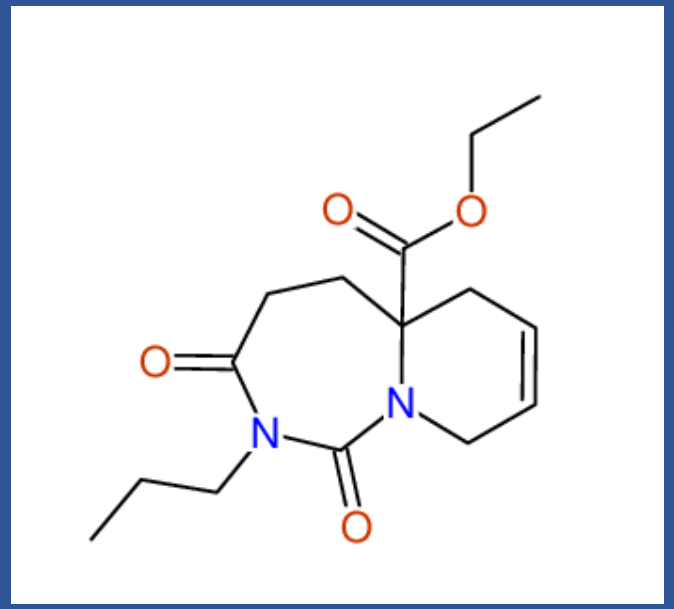
Search for identical spectra within a given tolerance having different structures associated coming from different literature citations but having at least one common author

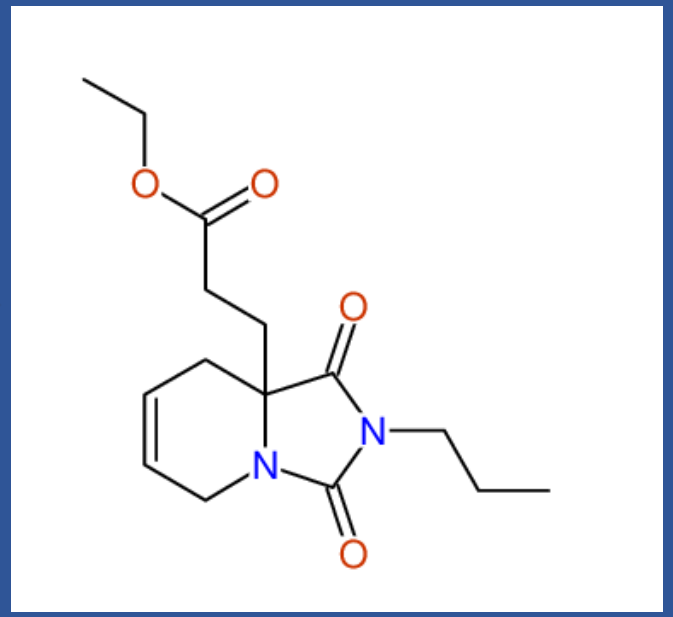
Result:

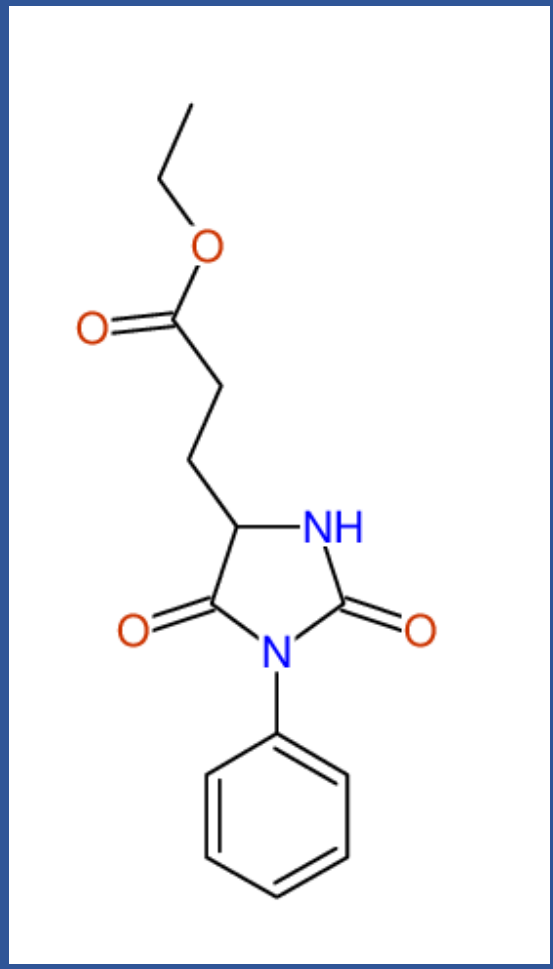
a) **The later paper clearly states a structure revision or an erratum → all is fine / errors might happen**
b) **Seems to be interesting for a detailed inspection**

29 Wrong Structures published in 2 articles (2 examples shown)









Paper #1: 8 wrong structures (Org.Lett.)
Paper #2: 21 wrong structures (Chem.Comm.)
Paper #3: 29 revisions + ca. 20 other structures (Eur.J.Org.Chem.)

a) Published in another journal among some 20 other compounds
b) 29 structure revisions
c) Experimental data identical (including a typo), obviously copy/paste
d) Neither paper #1 nor paper #2 was cited by paper #3
e) No hint that there are 29 structure revisions
f) Identical arguments are used to elucidate now 5-membered rings instead of the previously elucidated 7-membered rings – even the wording to interprete the spectral data is identical
g) All publications passed the peer-reviewing – no chance to find this problem using e.g. SciFinder (Erratum: EurJOrgChem/2018)

A few final remarks:

1) The literature contains many assignment errors
2) The literature holds a lot of wrong structures
3) The peer-reviewing as it is now does not really work
4) We need validated and curated repositories
5) The repository and all tools must be available 24/7
6) A prerequisite is a data format supported by (many/all) software vendors
7) NMReDATA is a FORMAT-definition; must be vendor-independent AND accepted by the community – includes acceptance of depositing the experimental data
8) Any vendor-specific definition/tag must be avoided
9) We need a retrospective refurbishment of existing data – original experimental data (FID) are usually NOT available, we have to rely on published tables and „pictures“ (PDF)
10) We need high-quality verification tools to collect reliable reference material

The real-world situation as it is now ……

Any vendor-specific definition/tag must be avoided – No name-dropping

<NMREDATA_CERTIFICATION_NMRSHIFTDB2>
Software=nmrshiftdb2
Version=1.0 \hl{DJ:I propose to add a verion number}

Should be:

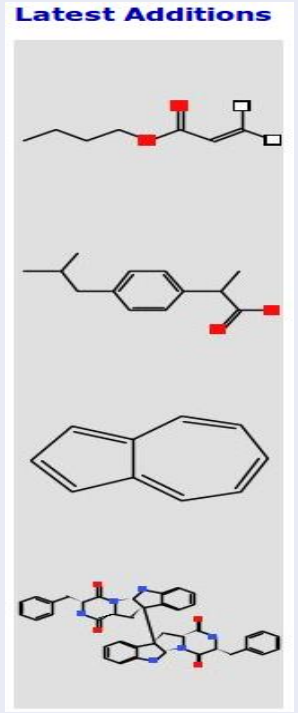
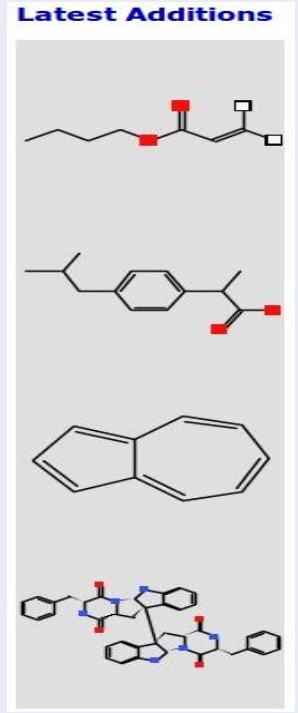
<NMREDATA_CERTIFICATION>
Software=nmrshiftdb2
Version=1.0 \hl{DJ:I propose to add a verion number}

Same valid for keywords holding “ACD“/“BIORAD“/“BRUKER“/“CSEARCH“/
“JEOL“/“MESTRELAB“/“MODGRAPH“/“NMRPREDICT“/“……whatever…..“

We need validated and curated repositories
accepted by the community

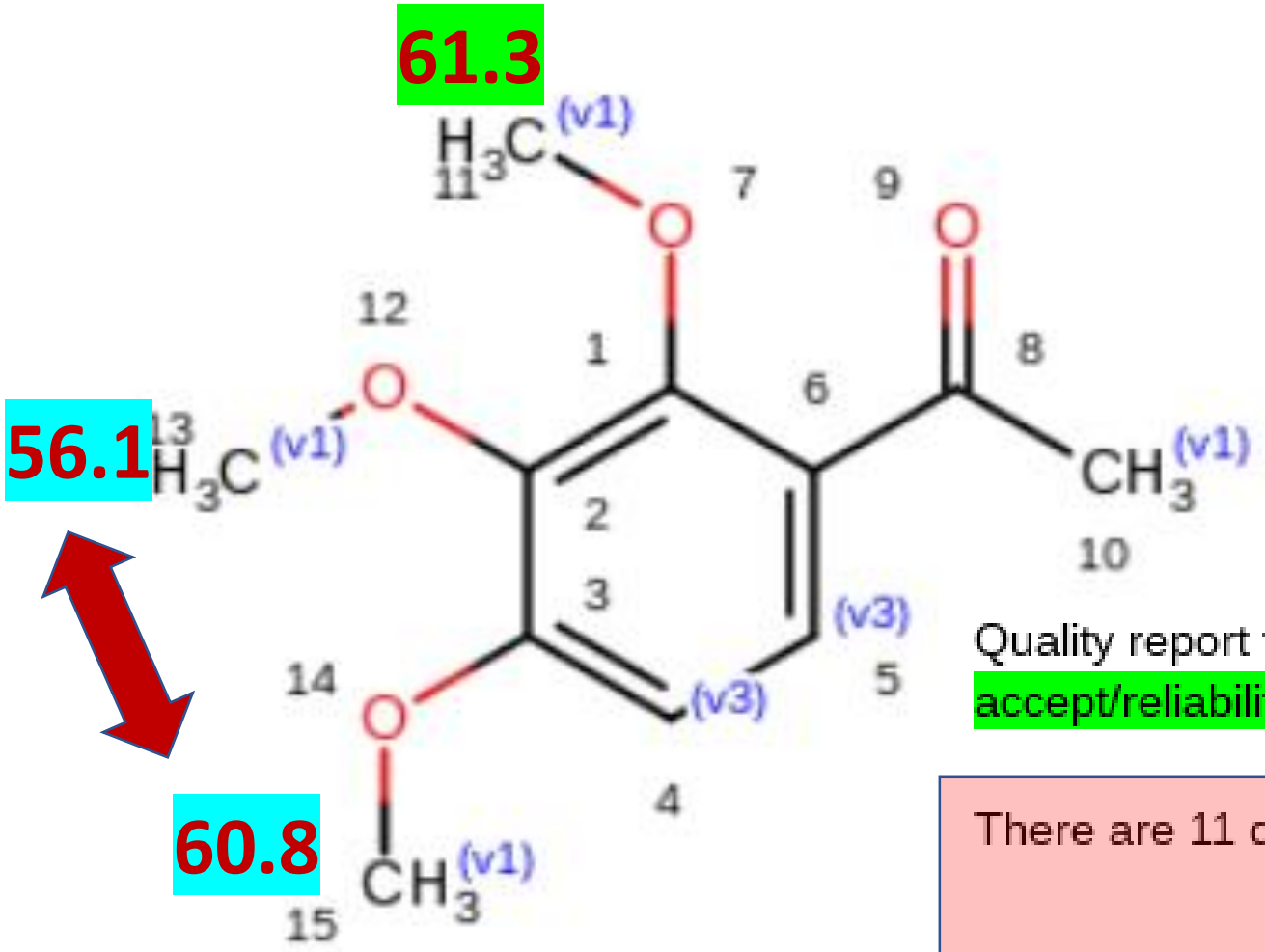
Deposition of experimental Data at NMRShiftDB

CAS: 15,000 per day / ca. 2 million NMRShiftDB: ZERO

Jun 5th, 2019	Sep 20th, 2019
52,223	**52,223**
	

We need validated and curated repositories accepted by the community

We need high-quality verification tools to collect reliable reference material



Response is:

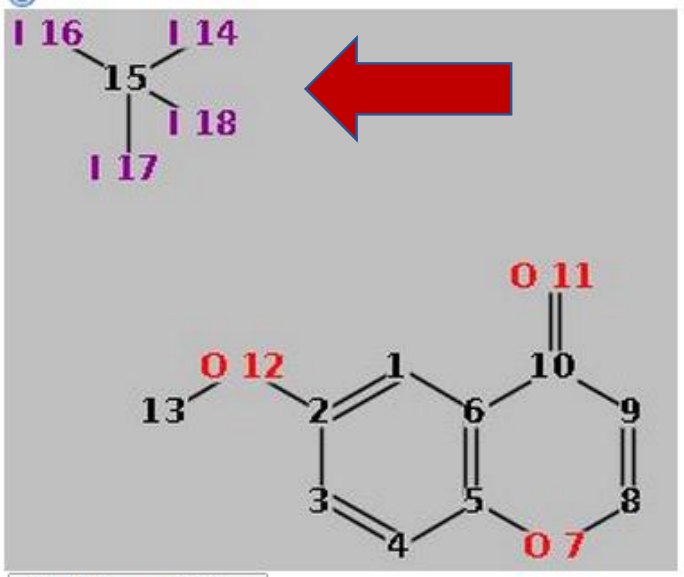
ACCEPT / EXCELLENT

Sorry - See textbook !

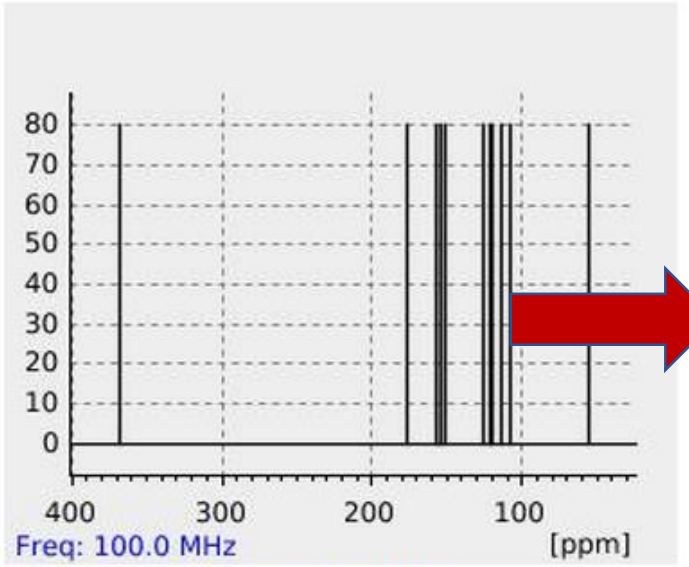
Quality report for carbon spectrum: 9: nmrshiftdb2 quality check: accept/reliability: excellent (Show full report)

There are 11 different carbons, but 11 shifts are given!

Dont understand ?!



Modify prediction



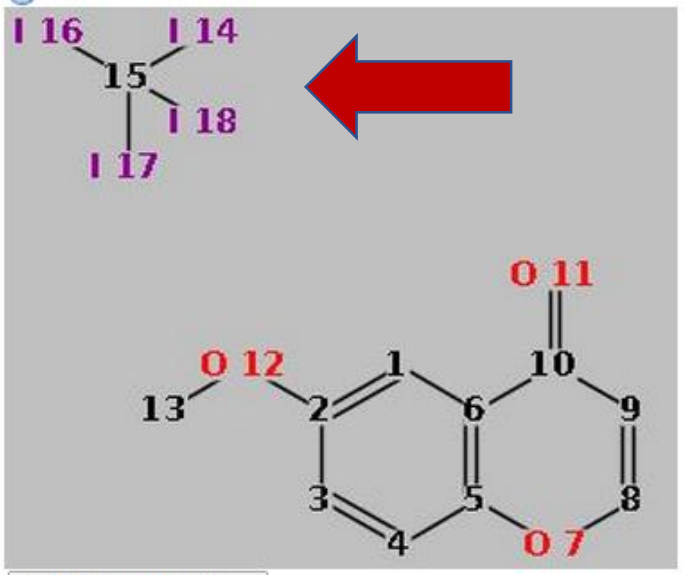
Show structure with shifts

No.	Shift	HOSE code used
1	107.44	
2	157.52	
3	121.93	
4	119.20	
5	151.05	
6	125.14	
8	154.86	
9	113.31	
10	176.42	
13	56.13	
15	368.57	

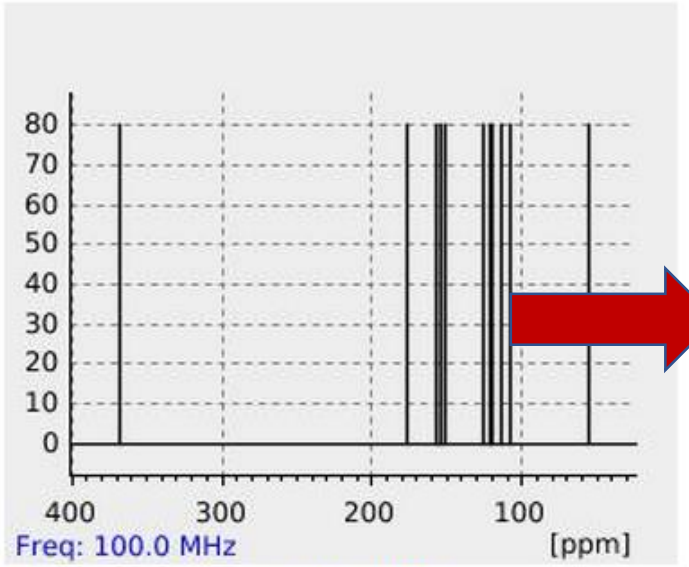
New prediction

Known literature value is: ca. -290ppm / Prediction: 369ppm, ca. 660ppm error

Basic knowledge at textbook-level !



Modify prediction



Show structure with shifts

No.	Shift	HOSE code used
1	107.44	
2	157.52	
3	121.93	
4	119.20	
5	151.05	
6	125.14	
8	154.86	
9	113.31	
10	176.42	
13	56.13	
15	368.57	

New prediction

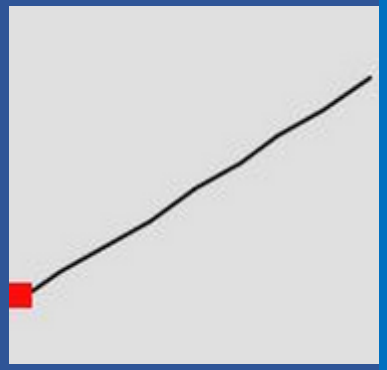
Explanation:

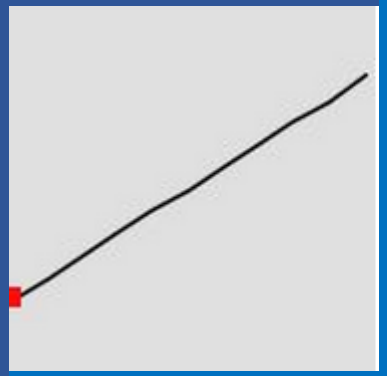
Method not trained to predict iodine-containing compounds

Even for a mediocre-talented programmer it should be possible to check the occurence of elements leading to an appropriate warning

Every user reads a few hundred pages in the handbook before typing one single character into „Word“ – NO ! A program must be self-explanatory !

We need validated and curated repositories accepted by the community





Which compounds are shown?

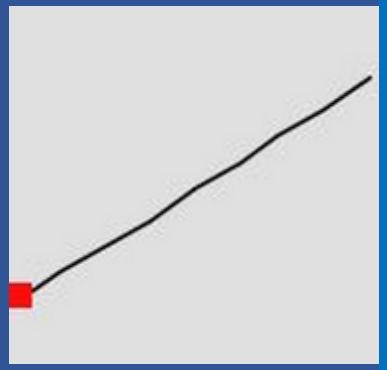
1-Hexanol, 1-Heptanol, 1-Oktanol, 1-Nonanol, 1-Decanol, 1-Undecanol, ….. ?

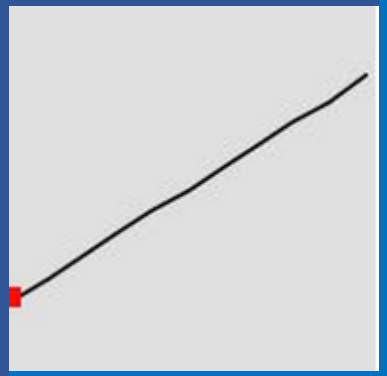
Strange coordinates brought to attention during summer 2014 !

These Screenshots are from September 2019 ! – Nothing done

Is this really what the NMR-community deserves ?

Which compounds are shown?





1-Hexanol, 1-Heptanol, 1-Oktanol, 1-Nonanol, 1-Decanol, 1-Undecanol, ….. ?

Strange coordinates brought to attention during summer 2014 !

These Screenshots are from September 2019 ! – Nothing done

Finding errors can be done automatically – curation must be done (mostly) manually afterwards – its simply ***WORK***

The bad news:

- We suffer from missing quality of our reference data
- We criticize missing professionalism during structure elucidation and signal assignment in the literature
- It seems that we repeat this missing professionalism by computer programs making the conclusions virtually more realiable

We need a general and flexible format-definition

The bad news:

- We suffer from missing quality of our reference data
- We criticize missing professionalism during structure elucidation and signal assignment in the literature
- It seems that we repeat this missing professionalism by computer programs making the conclusions virtually more realiable

We need a general and flexible format-definition

We do not need: Name-dropping
We do not need: Vanity Fair

The bad news:

- We suffer from missing quality of our reference data
- We criticize missing professionalism during structure elucidation and signal assignment in the literature
- It seems that we repeat this missing professionalism by computer programs making the conclusions virtually more reliable

We suffer from a „mediocracy“ in the literature
We know this, we criticize this
We continue with a computer-assisted „mediocracy“

The bad news:

- We suffer from missing quality of our reference data
- We criticize missing professionalism during structure elucidation and signal assignment in the literature
- It seems that we repeat this missing professionalism by computer programs making the conclusions virtually more realiable

The good news:

I will retire soon – it is probably my last participation in a SMASH-conference

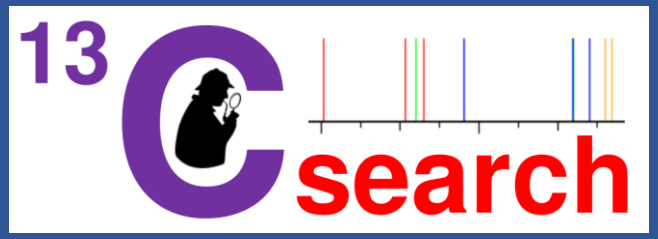
Good luck community – thank you for your attention !

CSEARCH engines ? (Free of charge, no registration)

Structure verification: https://nmrpredict.orc.univie.ac.at/c13robot/robot.php

QR-Code for mobile devices

Spectral Similarity Search: https://nmrpredict.orc.univie.ac.at.at/similar/eval.php



QR-Code for mobile devices

Directly accessible from Bruker‘s TOPSPIN & CMC-se programs

[1] Nicolaou KC, Snyder SA. Chasing molecules that were never there: misassigned natural products and the role of chemical synthesis in modern structure elucidation. Angew Chem Int Ed 2005; 44:1012-1044

[2] McAlpine JB, Chen SN, Kutateladze A, MacMillan JB, Appendino G, Barison A, Beniddir MA, Biavatti MW, Bluml S, Boufridi A, Butler MS, Capon RJ, Choi YH, Coppage D, Crews P, Crimmins MT, Csete M, Dewapriya P, Egan JM, Garson MJ, Genta-Jouve G, Gerwick WH, Gross H, Harper MK, Hermanto P, Hook JM, Hunter L, Jeannerat D, Ji NY, Johnson TA, Kingston DGI, Koshino H, Lee HW, Lewin G, Li J, Linington RG, Liu M, McPhail KL, Molinski TF, Moore BS, Nam JW, Neupane RP, Niemitz M, Nuzillard JM, Oberlies NH, Ocampos FMM, Pan G, Quinn RJ, Reddy DS, Renault JH, Rivera-Chávez J, Robien W, Saunders CM, Schmidt TJ, Seger C, Shen B, Steinbeck C, Stuppner H, Sturm S, Taglialatela-Scafati O, Tantillo DJ, Verpoorte R, Wang BG, Williams CM, Williams PG, Wist J, Yue JM, Zhang C, Xu Z, Simmler C, Lankin DC, Bisson J, Pauli GF. The value of universally available raw NMR data for transparency, reproducibility, and integrity in natural product research. Nat Prod Rep 2019; 36:35-107

[3] Robien W. A Critical Evaluation of the Quality of Published 13C NMR Data in Natural Product Chemistry. Progress in the Chemistry of Organic Natural Products, eds. Kinghorn AD, Falk H, Gibbons S, Kobayashi JI 2017; 105:137-215