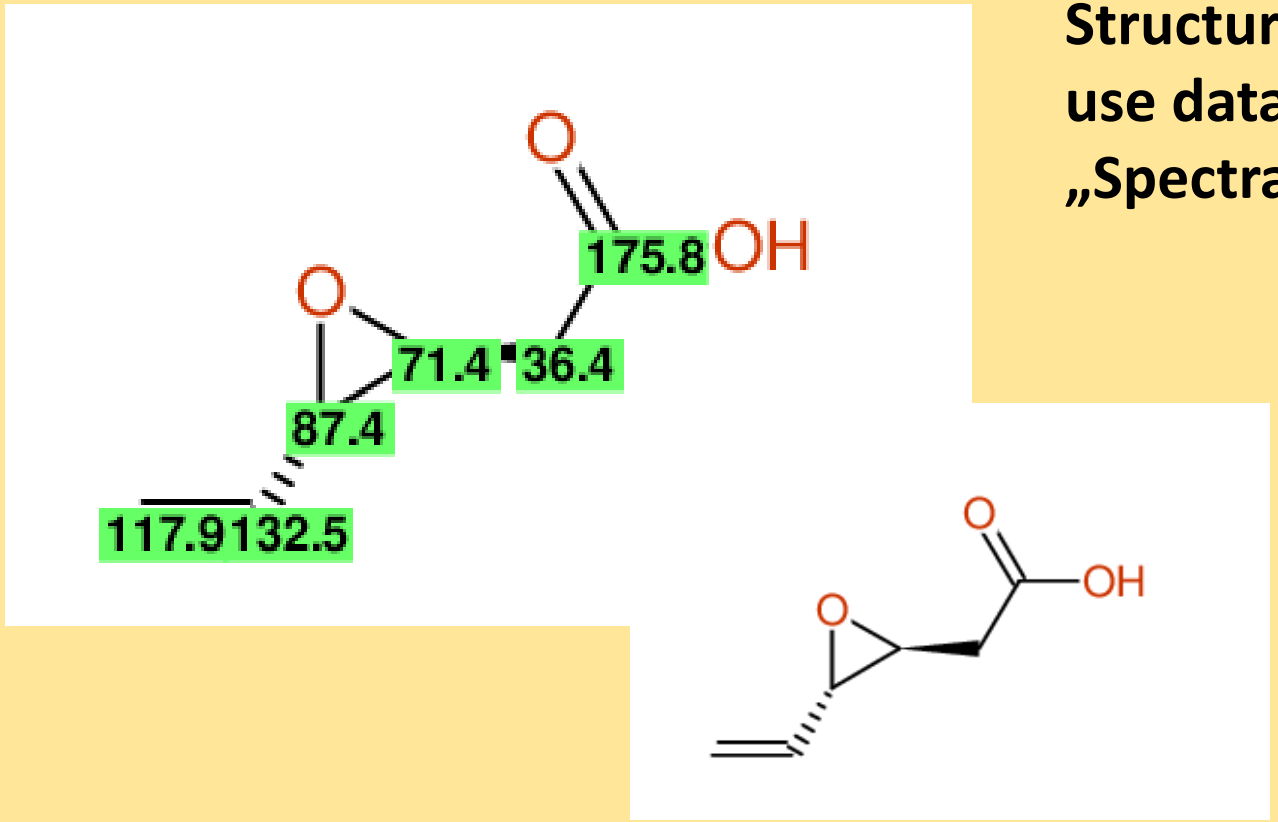
CSEARCH-Workshop: Berlin 2021 (via Zoom): A structure revision

Fully automatic Structure Revision – use the following form:
https://nmrpredict.orc.univie.ac.at/c13robot/robot.php



Structure proposal and assignment as published – use data for the „CSEARCH-Robot-Referee“ and a „Spectral Similarity Search“

- **Draw structure, enter projectname, compoundname and your email**
- **Assign the 6 shift values as given**
- **Skip all other forms by clicking „Continue“**
- **„Submit“ request**
- **Wait for a few minutes and check your mailbox**
- **Access data are given at the very end of the email received**

CSEARCH Workshop: Berlin 2021 – via Zoom

EXAMPLE #1: Spectrum Prediction

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

Obligatory for all people having NOT registered with their email!

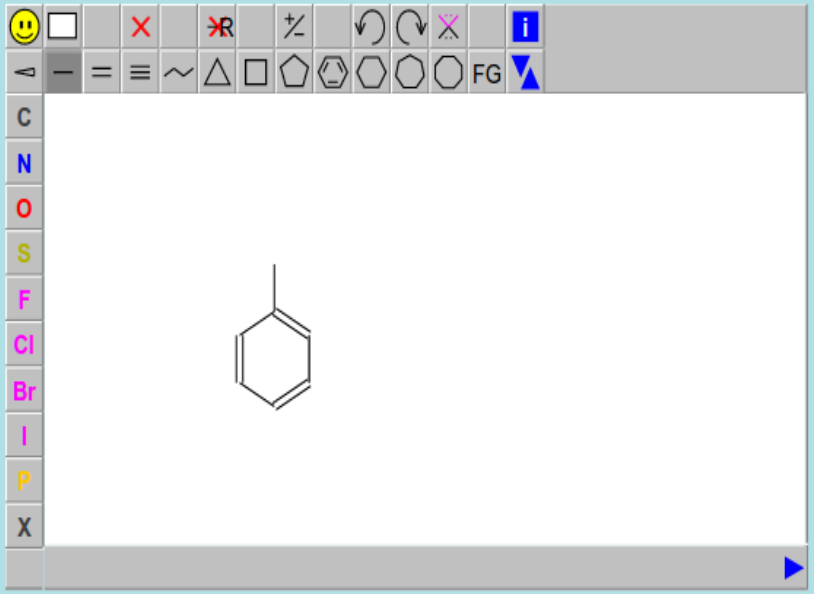
Open webpage, fill in basic data, draw a structure of your choice and skip the following forms by clicking the ‚CONTINUE'-button – at the end click ‚SUBMIT'

Software is adjusted to perform ‚AUTO-REGISTRATION' – all basic functionalities are available, structure revision is NOT ACTIVE ! Password is given at the very end of the email received for a specific request - password not necessary for accessing the result of a „Spectral Similarity Search" ! See email !

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

Structure and basic information – use **your** email !!

Enter Basic Data	Enter / Correct Assigned Lines	Enter / Correct Unassigned Lines	Submit / Restart



Your e-mail address: workshop@robien.at

Name of project: Project: Workshop

Name of compound: Demo assignment

- (•) Keep entry as confidential
- ( ) Donate entry to CSEARCH after 6 months

Continue

Either draw / edit your molecule in the applet or use this form for text input in MDL molfile format

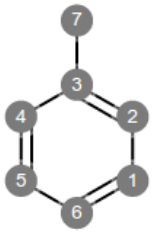
https://nmrpredict.orc.univie.ac.at/c13robot/robot.php

Enter assigned lines if available – use multiplicity, if known

Enter Basic Data | Enter / Correct Assigned Lines | Enter / Correct Unassigned Lines | Submit / Restart

Your input

E-mail address: workshop@robien.at
Project: Project Workshop
Compound: Demo assignment
Confidentiality: Dataset is private



Definitely Assigned Carbon Shift Values:
Please try to assign 7 lines to the 7 carbons given in the structure

Enter ^{13}C chemical shift value (ppm) and multiplicity of the signals:
s = singulet, d = doublet, t = triplet, q = quartet
e = even (CH, CH_3), o = odd (C, CH_2)
use the 'Exchflag' (a-z) for exchangeable assignments

Atom	Shift (ppm)	Multiplicity	Exchflag
C-1:		unknown	
C-2:		unknown	
C-3:	135	unknown	
C-4:		unknown	
C-5:		unknown	
C-6:		unknown	
C-7:	21	q = CH3	

Continue

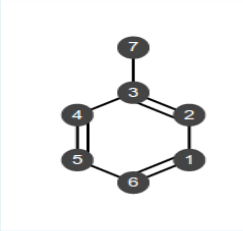
https://nmrpredict.orc.univie.ac.at/c13robot/robot.php

Enter unassigned lines (if necessary) – any combination assigned/unassigned is allowed

Enter Basic Data | Enter / Correct Assigned Lines | Enter / Correct Unassigned Lines | Submit / Restart

Your input

E-mail address: workshop@robien.at
Project: Project Workshop
Compound: Demo assignment
Confidentiality: Dataset is private



Assigned ^{13}C-NMR lines

```
C-3  = 135.00
C-7  = 21.00  q
```

Continue

Further unassigned lines:

Please enter additional lines belonging to this structure, which have not been assigned so far.

Shift (ppm)	Multiplicity
132	d = CH
	unknown
	unknown
	unknown
	unknown

Clear input

Copy/Paste peaklist

Be aware you need one value per carbon
When entering more than 5 unassigned values, you disable the automatic assignment feature of CSEARCH

Shift [mult] Shift [mult] .....

```
131
128
```

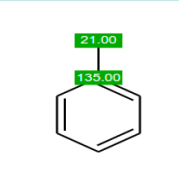
https://nmrpredict.orc.univie.ac.at/c13robot/robot.php

Additional requests & Summary of data – Submit when ready

Enter Basic Data | Enter / Correct Assigned Lines | Enter / Correct Unassigned Lines | Submit / Restart

Your input

E-mail address: workshop@robien.at
Project: Project Workshop
Compound: Demo assignment
Confidentiality: Dataset is private



Forget all

Assigned ^{13}C-NMR lines

```
C-3   = 135.00
C-7   = 21.00   q
```

Unassigned ^{13}C-NMR lines

```
C-999   = 132.00   d
C-999   = 131.00
C-999   = 128.00
```

Select the databases to be used for this request - optionally you can use a set of your own data

- CSEARCH only
- CSEARCH + All your previously performed evaluations classified as "ACCEPT" (show entries)
- CSEARCH + All your previously performed evaluations classified as "ACCEPT" or "MINOR REVISION" (show entries)
- CSEARCH + A manually selected set of your previously performed evaluations (select entries)

Be aware of the fact, that your private entries might severely influence the result of this request. Activate only entries with superior quality ! Manual selection recommended - your selection will be automatically stored for further requests.

An overview of all your previously performed requests since March 1st, 2015 is given here

Launch a *'Spectral Similarity Search'* using the given peaklist over 74 millions of predicted spectra

- No, thanks
- Yes, In case the evaluation recommends "Minor Revision" or worse
- Yes, in case the evaluation recommends "Major Revision" or worse
- Yes, in case the evaluation recommends "Reject"

Submit

CSEARCH Workshop: Berlin 2021 – via Zoom

EXAMPLE #2: Spectral Similarity Search

Use http://c13nmr.at/similar/eval.php

Open webpage, fill in data, click ‚SEARCH‘

Use the following data:

39.4 137.1 115.5 115.3 143.5 120.9 141.7 115.7 113.1
Molweight: 120 to 190
Number of signals: 7 to 14
Include O
Exclude P,F,Cl,Br, other

Spectral Similarity Search with Ranking
Accessing 285,861,408 predicted ^{13}C-NMR spectra

Enter CARBON chemical shift value (ppm) and MULTIPLICITY of your signals:

Allowed chemical shift range for ^{13}C-signals is between -399.0 and +399.0 ppm

s = singulet (C_{quat}), d = doublet (CH), t = triplet (CH_2), q = quartet (CH_3),
e = even (CH or CH_3), o = odd (C_{quat} or CH_2), p = protonated (CH or CH_2 or CH_3) , ? = unknown (any type)

Couplings to other nuclei than protons (e.g. ^{31}P, ^{19}F) are ignored - enter only the shift values !
e.g. a CF_3-group should be given as "122.8 s 2.5"

Deviation: 1.0 to 5.0 ppm allowed, useful values are somewhere around 3 ppm

Linenumber	Chemical Shift (ppm)	Multiplicity	Deviation (ppm)
1	9	q = CH3	
2	14	q = CH3	
3		unknown	
4		unknown	
5		unknown	
6		unknown	
7		unknown	
8		unknown	
9		unknown	
10		unknown	

As an alternative: Copy/Paste peaklist
Shiftvalue [Multiplicity [Deviation]]; one per line

```
27 t
60 t
174
```

Global pattern of reference-spectrum: No other significant lines ○ Other significant lines allowed ◉

Molecular weight from 90 to 110 ○ Use constraint for display only ◉ Use constraint already during search

Number of signals from 4 to 6

	N	O	P	S	F	Cl	Br	I	other
Element must be absent:	○	○	◉	◉	◉	◉	◉	◉	◉
Element may be present:	◉	○	○	○	○	○	○	○	○
Element must be present:	○	◉	○	○	○	○	○	○	○

Email for receiving result: workshop@robien.at
Name of Project: Workshop Uppsala
Name of Peaklist: Example

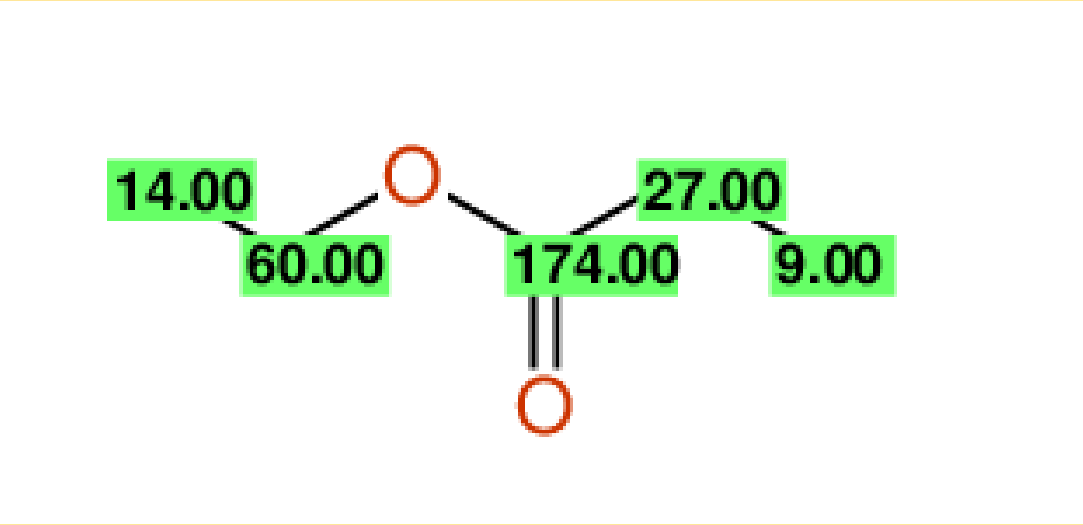
Add more lines | Clear input | Search

CSEARCH Workshop: Berlin 2021 – via Zoom

EXAMPLE #3: Propionic acid, ethyl ester – correct assignment

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

Start from the given structure and the assigned values

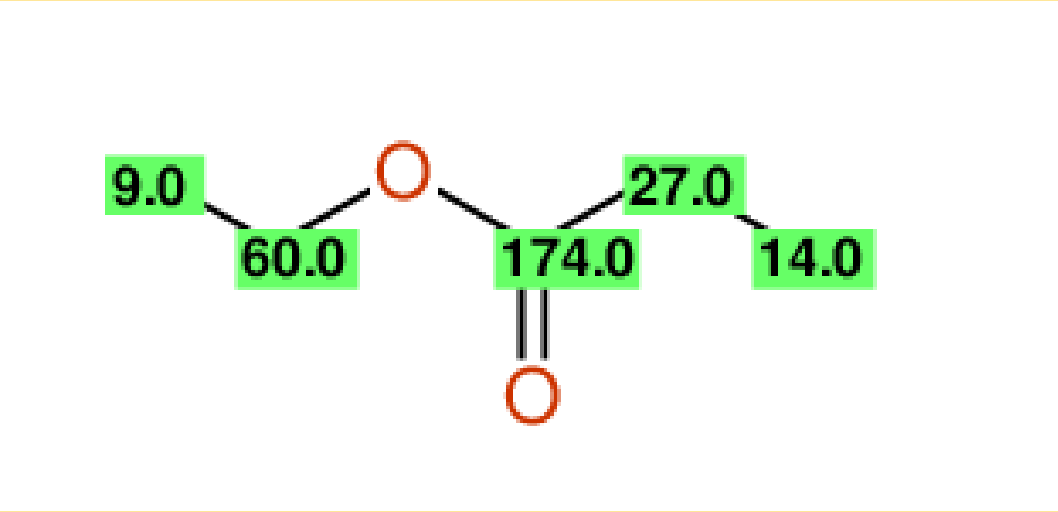


CSEARCH Workshop: Berlin 2021 – via Zoom

EXAMPLE #4: Propionic acid, ethyl ester – Me-assignment wrong

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

Start from the given structure and the assigned values

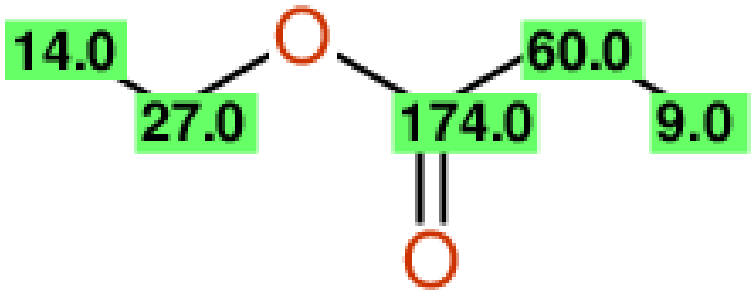


CSEARCH Workshop: Berlin 2021 – via Zoom

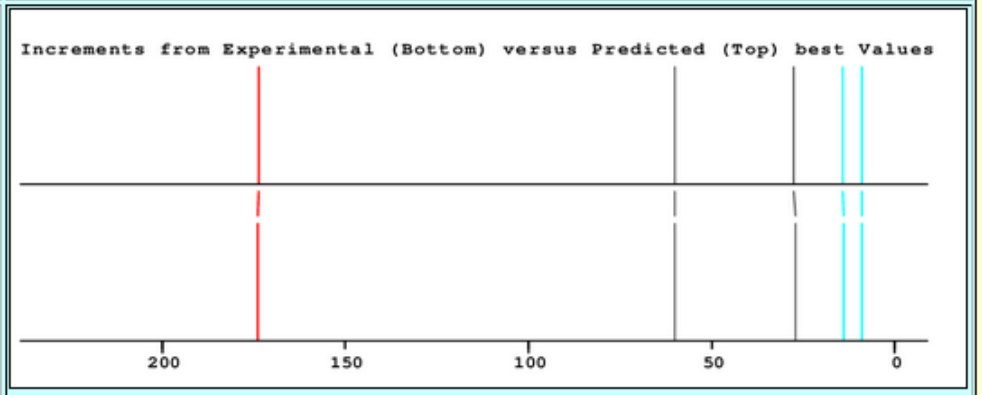
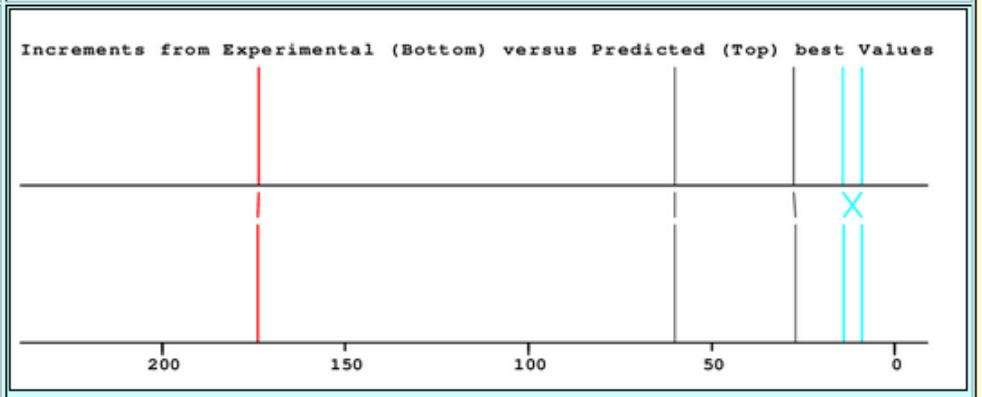
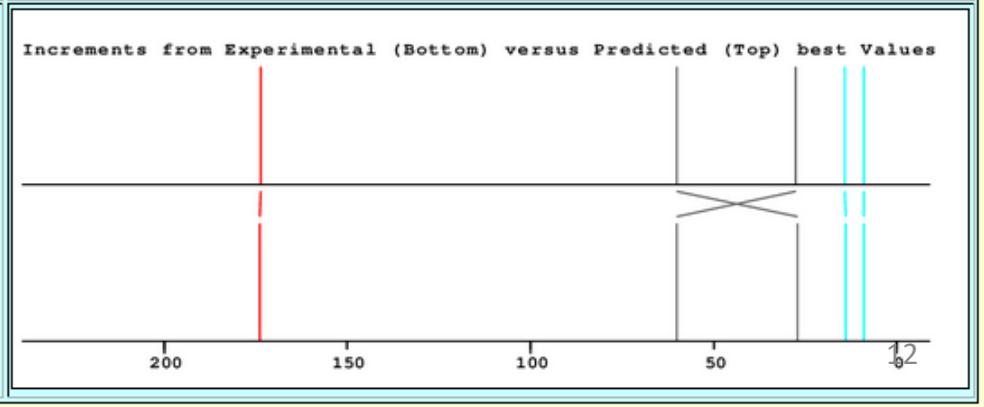
EXAMPLE #5: Propionic acid, ethyl ester – CH_2-assignment wrong

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

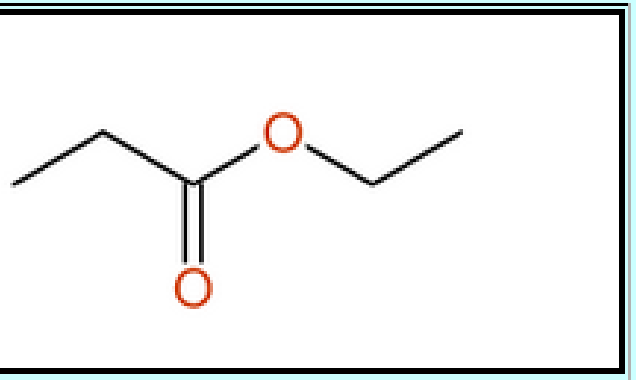
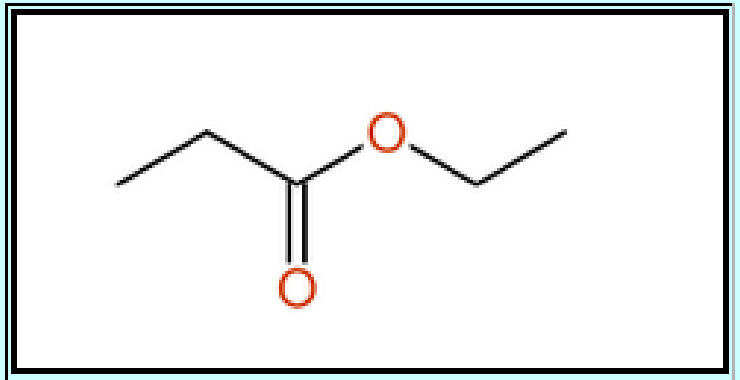
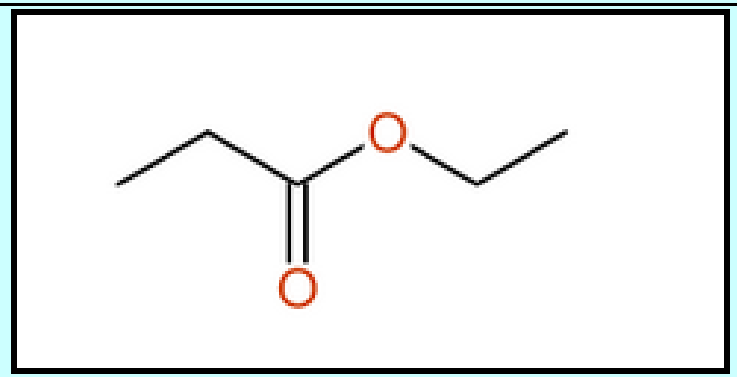
Start from the given structure and the assigned values



History of your requests for this compound

Date/Time	Result	Method	Assigned Lines	Unassigned Lines	Stereoisomer	Permanent URL	Remark	Comparison of experimental and predicted data (Evaluation only)
2019-11-24 10:27:09	Accept	Evaluation	5	0	NO			
2019-11-24 10:27:36	Minor	Evaluation	5	0	NO			
This request	Major	Evaluation	5	0				

Alternative structures created – No better alternative found
„Only“ assignment error !

Your proposal	Best proposal	Best proposal (out of 29) which exists either in CSEARCH or PUBCHEM
Deviation = 1.04 ppm Position = 1	Deviation = 1.04 ppm	Deviation = 1.04 ppm Position = 1
		
FKRCODPIKNYEAC	FKRCODPIKNYEAC	FKRCODPIKNYEAC
$C_5H_{10}O_2$	$C_5H_{10}O_2$	$C_5H_{10}O_2$

CSEARCH-Robot-Referee: Compatibility-Check of a proposed Structure versus the ^{13}C-NMR data

What is checked ?

The assignment (if not given: assignment is done by software)

Logic behind exchange-flags

Symmetry

Definition of stereocenters

Query structure compared against database

Query structure compared against PUBCHEM-collection (138 million compounds)

Multiplicity check (experimentally determined versus from structure derived)

Experimental shift values versus predicted shift values

2 independent techniques: HOSE-code prediction and Neural Network prediction

Quality of underlying data is also analyzed and influences the result

Spectral data are used as query to detect different structures
associated with the same spectrum

CSEARCH-Robot-Referee: Compatibility-Check of a proposed Structure versus the ^{13}C-NMR data

What is the result ?

A webpage holding all the details of this evaluation

A summary holding the final result and the most decisive information

Excellent visualization of each analysis

The result is given as:

Accept as it is – very rare case
Minor Revision
Major revision
Reject

The response of the evaluation is adjusted (intentionally) as extremely sensitive to errors (or even warnings)

A few technical specifications

All pages are electronically signed in order to prohibit editing/manipulation
This signature can be recalled by everybody knowing the URL of the page

The sequence „Predict spectrum by CSEARCH“- „Do an automatic assignment“ – „Redo evaluation“ → will give „Accept as it is“, but this is again visible to everybody knowing the URL – the complete history of a compound processed by an email-address is given at the end of each page

Doing this using 2 different email-addresses is in principle possible – the same tests that are running for one single user are invisibly running in the background over all email-addresses

Transfering the page to a mobile device is easily possible, simpy take a snapshot of the provided QR-code.

CSEARCH Workshop: Berlin 2021 – via Zoom

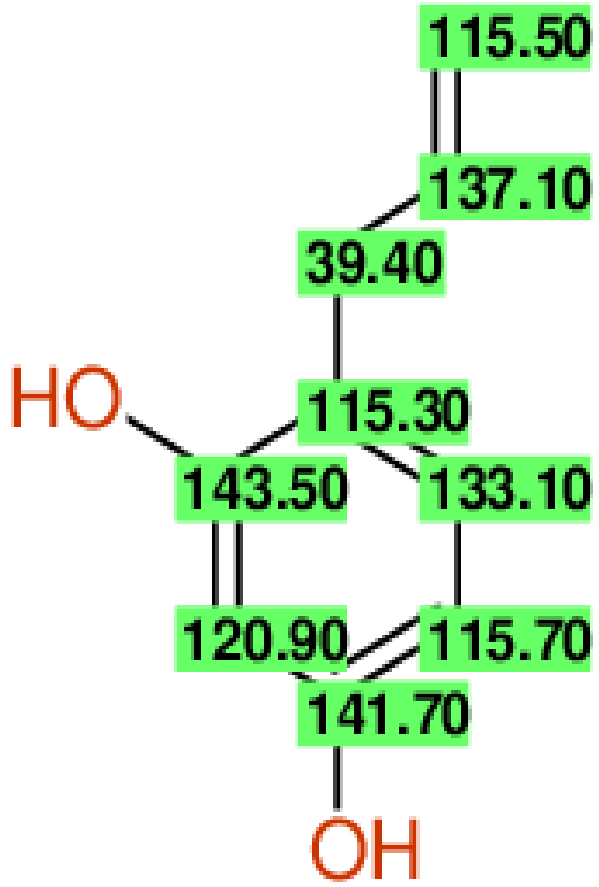
EXAMPLE #6: Fully automatic Structure Revision

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

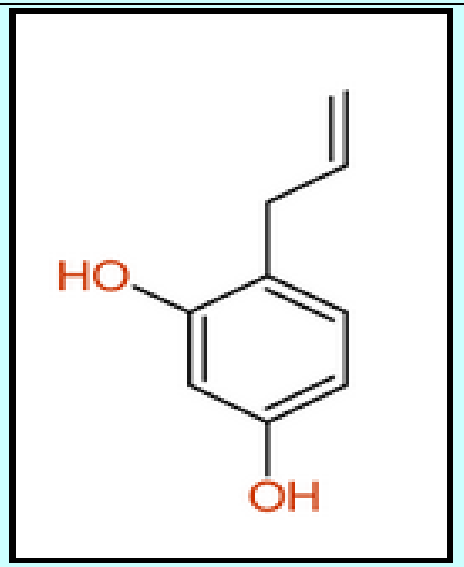
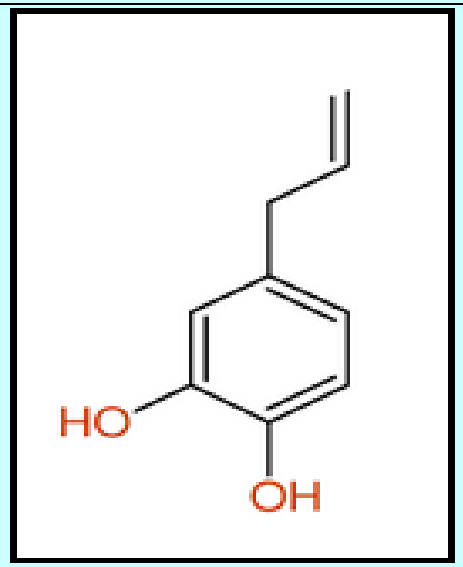
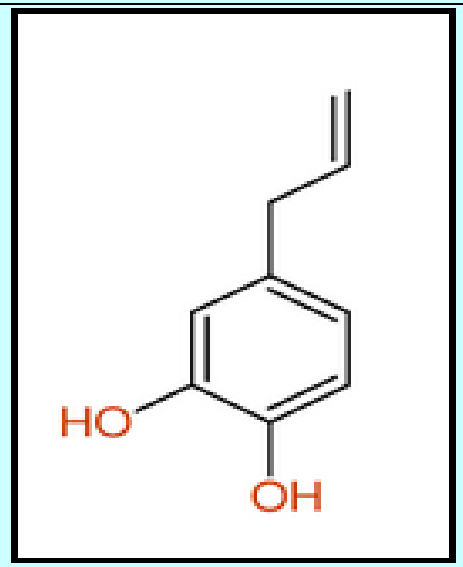
Start from the given structure and the assigned values (as published in the literature!)

Example #7: Use the given shiftvalues for a Spectral Similarity Search

Use multiplicity information – see what happens
Maybe restrict by Molweight (eg. 100 to 200)
Maybe restrict by number of signals (eg. 6 to 14)
Exclude Cl,Br,F,P,other

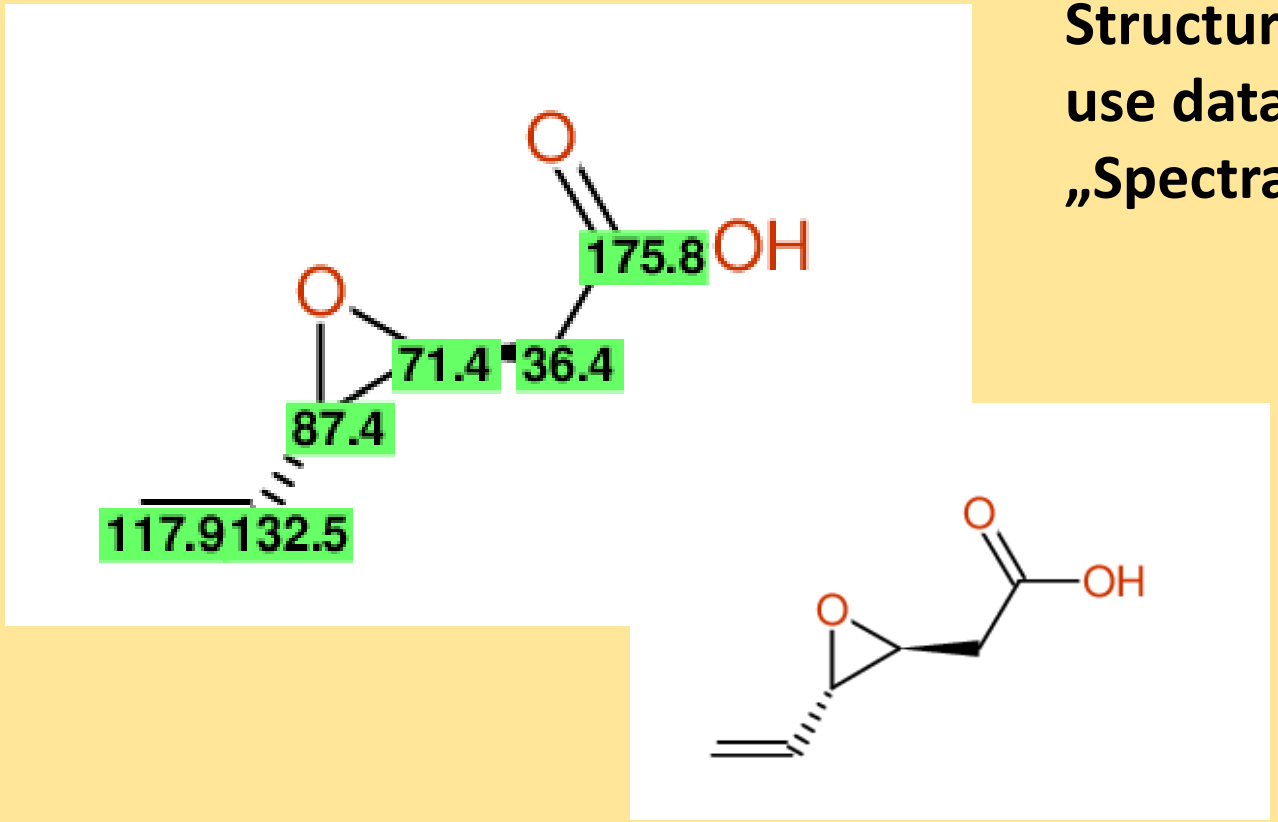


Revised structure proposal

Your proposal	Best proposal	Best proposal (out of 25) which exists either in CSEARCH or PUBCHEM
Deviation = 6.02 ppm **Position = 60**	**Deviation = 1.39 ppm**	**Deviation = 1.39 ppm** **Position = 1**
Number of reasonable alternative structures: 54 **Project:** Structure_revision **Compound:** Natural_product_meta-OH		
		
VXUCMXVCGIAHHR	FHEHIXJLCWUPCZ	FHEHIXJLCWUPCZ
$C_9H_{10}O_2$	$C_9H_{10}O_2$	$C_9H_{10}O_2$

Another structure revision

EXAMPLE #8: Fully automatic Structure Revision



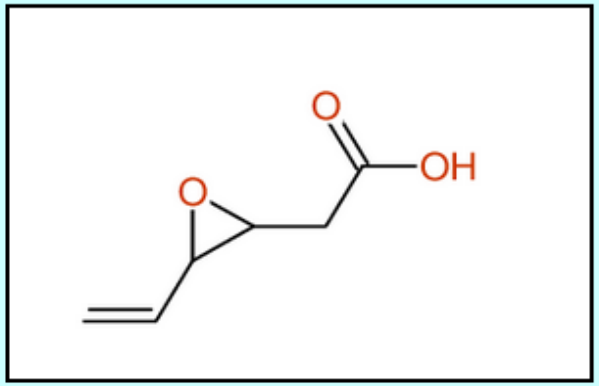
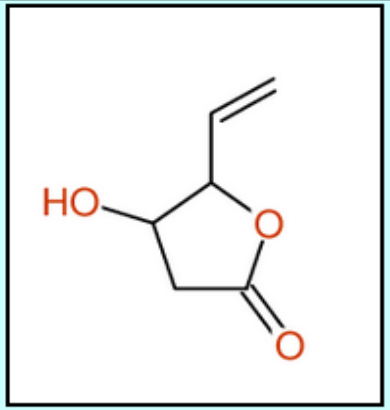
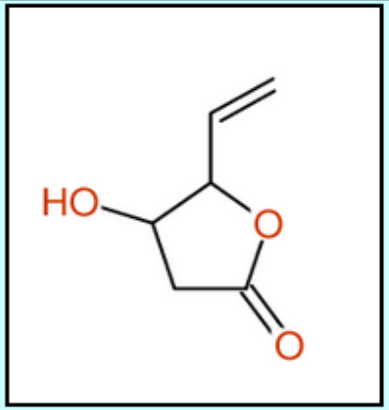
Structure proposal and assignment as published – use data for the „CSEARCH-Robot-Referee“ and a „Spectral Similarity Search“

Best alternative structure is:

Your proposal	Best proposal	Best proposal (out of 7) which exists either in CSEARCH or PUBCHEM
Deviation = 11.29 ppm Position = 46	Deviation = 3.43 ppm	Deviation = 3.43 ppm Position = 1

Number of reasonable alternative structures: 27

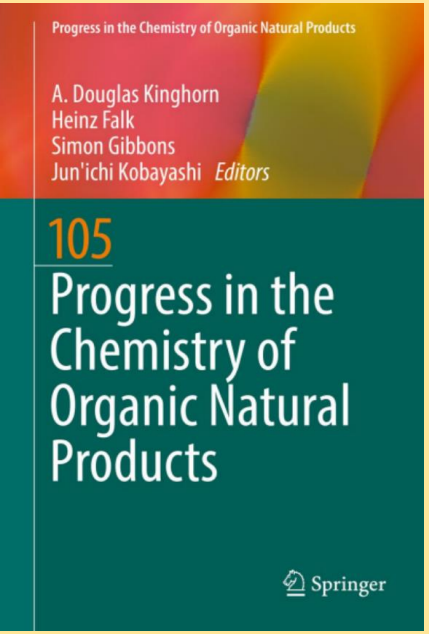
Project: J.AM.CHEM.SOC.,118,10971[1996]CSEARCH-NUMBER:UWLU053538-20110301
Compound: TRANS-3,4-EPOXY-5-HEXENOIC-ACID_COMPOUND-14

		
ZQOISUAETPFTAA	UHZDTSAFZSJJTG	UHZDTSAFZSJJTG
$C_6H_8O_3$	$C_6H_8O_3$	$C_6H_8O_3$

Scanning structural space:	Completed
CPU used:	9.577 seconds
Structures processed:	253
Isomers / Non-Isomers:	171 / 82
Structures with correct multiplicity:	100
Structures available in CSEARCH or PUBCHEM:	7

CSEARCH-Database: Version 9.4.0 CSEARCH-Robot-Referee from 2017:06:10

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

[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

Spectral Similarity Search – The Background

Request consists of:

- Peaktable – might hold solvent signals
- Multiplicity – optional (S,D,T,Q – O,E – P – unknown / any combination)
- Deviation for every line
- Range for number signals
- Range for molecular weight
- Elemental composition (Present, Absent or Unknown)

Resulting hitlist may consist of:

- No structure proposal found
- Between 1 and 10,000 proposals found
- More than 10,000 proposals found

Results are summarized as ZIP-files ready for upload to the „Supplementary Material“ of your publication

Spectral Similarity Search – The Background

Run request „as it is“

Multiplicity given	**→ rerun without multiplicity**
Solvent detected	**→ rerun with solvent lines eliminated**
Too many answers	**→ rerun with smaller deviations** **→ rerun with smaller range for number of signals (and combinations thereof)**
No answer found	**→ rerun with larger deviations** **→ rerun with larger range for number of signals (and combinations thereof)**
Worst case scenario:	**14 Requests searching 510 million spectra**
Total:	**Search 7 Billion of spectra** **Search 14x9 Billion of shift values**

Hardware and Software

Computer:	64GB Memory EightCore CPU (8 x 2 x 4.7GHz)
Storage:	58TB NAS (7 x 8TB HDD + 2GB SSD)
Software:	OS: Linux CSEARCH-Package / Mol2PS (N.Haider)
Robot:	Usually 6-8 sec CPU-consumption
Similarity:	20 – 200 sec depending on query-peaklist Massive parallel processing (up to 30 processes) Requests coming from TOPSPIN or CMC-se are processed in a separate privileged queue
Structure Generator:	A few seconds up to a few minutes of CPU-usage
Processes:	More than 100 jobs started by the „crontab“-utility in regular intervals – stable over months
URL:	http://c13nmr.at https://nmrpredict.orc.univie.ac.at