Applying New Technologies to old ^{13}C-NMR Reference Data – the Advantage of Automatic Peer-Reviewing using the CSEARCH-Protocol

Wolfgang Robien

Department of Organic Chemistry

University of Vienna

Währingerstrasse 38

A-1090 Vienna, Austria

Wolfgang.Robien@univie.ac.at

ORCID: 0000-0001-8565-7316

Abstract

A systematic investigation of the experimental ^{13}C-NMR spectra published in the period from 1977 to 2006 in Planta Medica with respect to their quality using CSEARCH-technology is described. In a surprisingly large number of cases reasonable alternative structure proposals can be elucidated by reinterpretation of the published chemical shift values. A final decision in case of doubt can only be done by reinspection of the raw data showing the urgent need for a repository holding the experimental data used for structure elucidation. It is shown that the systematic application of the CSEARCH-Robot-Referee during the peer-reviewing process is able to prohibit at least the most trivial assignment errors and wrong structure proposals. The comparison of the most significant key numbers derived for Planta Medica against those of other important journals in the field of natural product chemistry show a quite similar level of quality for all publishers responsible for the five journals under investigation.

Introduction

NMR-Spectroscopy is an important technique providing a massive amount of information during the structure elucidation process at the level of the constitution, the configuration as well as the conformation of an unknown compound. The tremendous development of pulse techniques, high-field NMR-equipment and automatic sample changers during the last three decades has dramatically shifted the earlier bottleneck of the amount of time necessary for acquiring the experimental data to the new bottleneck of spectrum interpretation. This leads to the effect of quite frequent misinterpretations of the experimental data in terms of wrong structure proposals. A systematic analysis of structure revisions was published by Nicolaou [1] pointing out the importance of structure proofs by organic synthesis. Pauli and coauthors [2] demonstrated the necessity of a searchable public domain repository holding the raw spectral data used during the structure elucidation process. A comprehensive review [3] on computer-assisted peer reviewing and subsequent fully automatic structure revisions verified the tremendous effect of automatic quality control of spectral data.

The flood of experimental NMR-data has stimulated the development of computer-software helping the spectroscopist with data interpretation. Carbon-NMR spectroscopy is well-suited for this purpose because of its large range of chemical shift values and its simplicity based on usually missing coupling patterns. A powerful computer-assisted technology for spectrum prediction named HOSE-code was already introduced in 1978 by Bremser [4]; later on 'Neural Networks' [5] were used; this technology has been reprogrammed recently in a similar way and is now called 'Machine Learning' [6] using specialized hardware architecture. The HOSE-code in its basic version represents

only constitutional properties of the molecule (except the easy case of cis/trans-isomerism) and was later on expanded to handle also stereochemical features [7,8]. Meanwhile stereochemistry can also be used for spectrum prediction in the open-access system nmrshiftdb2 [9]. Since the late 1970's a large number of databases collecting the reference material published in the public domain chemical literature was built (ACD/NMR-Predictor [10], CAST/CNMR [11], CSEARCH [12], NMRPREDICT [13], NMRSHIFTDB [14], SPECINFO [15], Spectrabase [16]). Spectrum prediction engines were also implemented into drawing programs for chemical structures [17], as well as into programs for processing measured NMR-data [18,19].

NMR-Spectroscopy is an 'atom-centered' type of spectroscopy allowing to assign a measured chemical shift value to a specific atom. Understanding this correlation is essential because the value of the chemical shift is influenced by the structural environment of the atom under investigation. This relationship between a structural property and the value of the chemical shift is the basis for all database-oriented prediction tools. Independent from the mathematical model behind the prediction we need to start from correct data to describe this correlation. From this simple strategy behind spectrum prediction it can be immediately derived that publishing simply the list of peaks is not sufficient, because the central information – the 1:1-relationship between structural environment and chemical shift value – is not furthermore available then. Such an unassigned peak list is only a fingerprint for a given structure sufficient for spectral similarity searches, but completely useless for spectrum prediction.

Despite the tremendous development in the field of pulse-techniques available in NMR-spectroscopy many structure elucidation problems are solved by comparison of the new measurement with already published reference material taken from the public domain chemical literature. The main problem with

existing reference material is, that there is usually no information available upon the quality of these data [20]. During the last two decades many journals published so-called “Supplementary Material” to their articles – mainly as PDFs - showing in principle only pictures of the spectra efficiently prohibiting reprocessing of the underlying experimental data. With respect to NMR-measurements there is an ongoing initiative [21,22] to agree on a common vendor-independent format for the raw data in order to add them to the supplementary material of the publication. This strategy allows to start a necessary reprocessing and reinterpretation from the original data sets in case of doubt. A well-defined exchange format is furthermore a prerequisite to create community-driven databases [23,24].

It is interesting to observe that many errors in the literature are introduced by misinterpretation of the spectra obtained by state-of-the-art 2D-techniques. A leading role with respect to this effect, seem to have HMBC-spectra showing frequent misinterpretation of ^{2}J (or ^{4}J)–couplings as ^{3}J-couplings necessarily leading to wrong structure proposals [24]. This kind of wrong interpretation of 2D-information is frequently preferred against application of simple spectrum prediction efficiently showing the inconsistency between structure proposal and spectral data.

The automatic structure verification based on ^{13}C-NMR chemical shift data is the central procedure within the “CSEARCH-Robot-Referee” [3,26,27,28]. The knowledge base behind consists of some 340,000 curated ^{13}C-NMR spectra taken from the public domain literature. A database holding 430 million of predicted spectra is additionally used to allow efficient structure dereplication. The workflow applied consists of the following steps:

- Formal check of the structure (valency, charge, stereocenter)

- Create table to link structure to other databases (e.g. PUBCHEM [29,30], Chemspider [31], eMolecules [32])
- Check formal correctness of supplied peak list (symmetry, exchangeable assigned signals)
- Perform spectrum prediction by HOSE-code [4] and NN [5]
- Perform statistical analysis based on underlying data used during the spectrum prediction process to allow evaluation of the quality of the result
- Assign signals if unassigned signals are given
- Calculate and visualize coincidence between structure proposal and given experimental data
- Perform search for identical structures contained in the underlying knowledge base
- Perform search for identical spectral pattern associated with different structures
- Detect positions with large deviation between experimental and predicted value
- Start structure generator which exclusively modifies the topology of the given structure at the positions having a large deviation
- Perform dereplication based on the given peak list using the database of 430 million predicted spectra

This workflow described before was applied to 3,621 spectra taken from "Planta Medica" between 1988 and 2006. The publications having assigned ^{13}C-NMR data within this range of years were extracted and these data sets used to build a CSEARCH-database. The reason for selecting this period of time is simple, that we were not able to extract more than approximately 25K spectra per year from the literature and we had to switch to other journals in order to cover as many

journals as possible. When comparing “Planta Medica” with other prominent journals in the field of natural product chemistry like “Chemical and Pharmaceutical Bulletin”, “Fitoterapia”, “Journal of Natural Products”, “Phytochemistry” a nearly identical number of compounds can be found, which seem to be in error. It should be mentioned that the manually performed extraction of the data from the journal is already quite selective, because data sets which seem to be in error at a first glance are completely ignored anyway leading to an improved impression for the respective journal. This effect influences all journals in a quite similar and positive way according to our experience. The ^{13}C-NMR data published in these five before mentioned journals (**Table 1**) are coming from compounds of similar size having an average molecular weight between 479 and 520 amu, the average deviation between experimental and predicted chemical shift values is also very similar (1.61 – 1.85 ppm), the number of compounds having at least one signal more than 20ppm away from the predicted chemical shift value is between 1.34 and 2.68% of the entries available in the CSEARCH-collection. A nearly identical finding is given, when the prediction is restricted by a partial structure search; the examples selected here contain either a chromone-fragment or a steroid skeleton. In both cases the average deviation is again in a very narrow range starting at 1.40ppm and 1.27ppm, respectively, with a maximum at 1.74 and 1.54ppm, respectively. From the data compiled (**Table 1**) it can be concluded that all five journals show a very similar level of quality with respect to the ^{13}C-NMR data contained therein.

Table 1 Comparison of the quality of ^{13}C-NMR data published in "Planta Medica" (PM) versus "Chemical and Pharmaceutical Bulletin" (CPB), "Fitoterapia", "Journal of Natural Products" (JNP) and "Phytochemistry "(PC)".

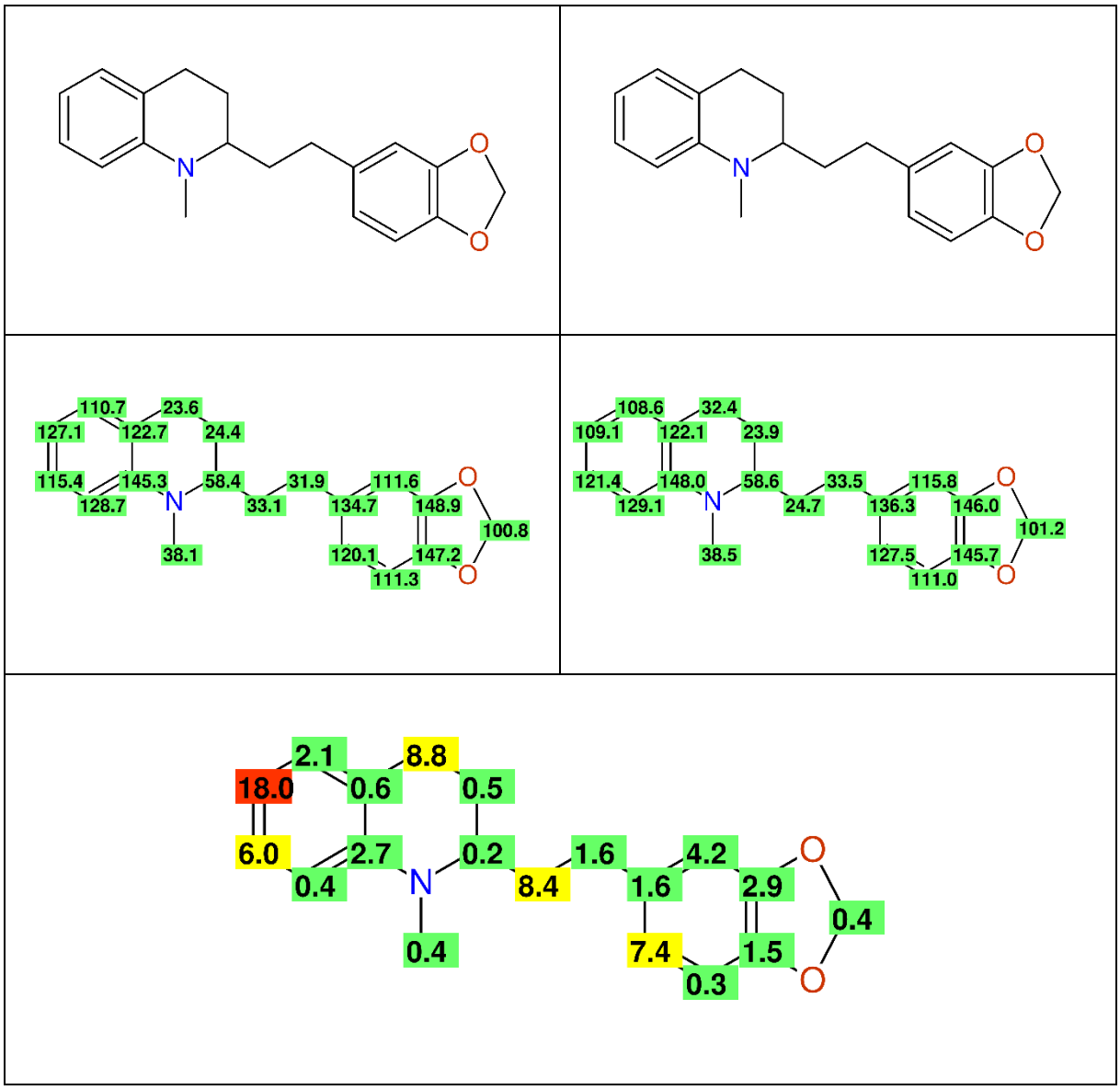
Journal	**CPB**	**Fitoterapia**	**JNP**	**PC**	**PM**
Entries	17,863	1,568	34,933	38,379	3,621
Period	1977-2016	1998-2012	1979-2013	1976-2015	1977-2006
Average MWT (amu)	520	497	481	483	479
$\Delta\delta_C$ (ppm)	1.61	1.85	1.79	1.66	1.72
$\Delta\delta_C$>20ppm	239/1.34%	42/2.68%	542/1.55%	533/1.39%	64/1.77%
$\Delta\delta_C$ (ppm) Chromones	1.40	1.74	1.68	1.45	1.57
$\Delta\delta_C$ (ppm) Steroids	1.27	1.54	1.47	1.37	1.36

Alkaloids from *Galipea officinalis*

In 1998 the new alkaloid named galipinine was isolated from the bark of *Galipea officinalis* and its structure was determined by various 2D-NMR techniques at 400MHz including HMBC-spectra [33]. In 1999 the same compound was published [34] and named allocuspareine. The structure elucidation was performed by application of HMQC-spectra and long range ^{1}H-^{13}C-NMR coupling experiments. In both papers the authors derived the identical structure proposal from nearly identical overall spectral patterns using state-of-the-art NMR techniques. The measurements were performed in $CDCl_3$ as solvent, the chirality of the structure remains undefined in both cases. The main difference between the 2 sets of experimental data is the signal assignment – despite 2D-NMR - with differences of up to 18.0ppm (**Fig. 1**), which is quite surprising. Both publications have no supplementary information attached, therefore performing a reinterpretation of the experimental data is impossible.

Another structure pair consisting of compound 4 from [33] and compound D1c from [34] - showing the well-known alkaloid cuspareine - has a similar inconsistency in the signal assignment. The finding that the structure itself is correct and the inconsistency in the data is purely based on an assignment problem is strongly supported when applying the structure generator to these experimental data. In all cases the published structures are ranked first with no reasonable alternative proposals among thousands of structures generated.

Fig. 1 Experimental ^{13}C-NMR data for galipinine (compound 5 in [33], CAS-RN: 221054-46-2) and allocuspareine (compound D1b in [34], CAS-RN: 223610-06-8) having identical constitution, but undefined chirality. The signal assignment is extremely different as shown in the last row.



Alkaloids from *Quassia amara*

2-Methoxycanthin-6-one was isolated from the stem wood of *Quassia amara* and its structure was elucidated by ^{1}H-, ^{13}C-, ^{1}H-COSY and ^{1}H-^{13}C-chemical shift correlation experiments [35]. In this article it is claimed that "A new alkaloid, 2-methoxycanthin-6-one (1) has been isolated from the methanol extract of the stem wood of *Quassia amara*." This compound was also obtained a few years earlier [36] during the structure elucidation of 2-hydroxycanthin-6-one, when a methylation was performed to allow NOE-measurements. Analyzing the ^{13}C-NMR data published for the structure elucidation of this compound a significant difference is observed for some carbons in the pyridine ring system (**Fig. 2**). Comparing the ^{13}C-NMR data given for compound 1 and compound 5 results in excellent agreement with respect to the pyridine-moiety, but with the methoxy-group in position 1 (compound 5 in [35]) instead of position 2 (compound 1 in [35]).

Using the structure proposal of 2-methoxycanthin-6-one together with the given ^{13}C-NMR peak list as starting point for the structure generator the 1-methoxycanthin-6-one is selected as best existing alternative among 3,049 structures generated (**Table 2**). The 1-methoxycanthin-6-one derivative has been already published in 1976 [37]; the carbon chemical shift values can be found in [38]. it should be noted that the published values in [38] differ from the values published in [35].

Fig. 2 2-Methoxycanthin-6-one from [35] (CAS-RN: 116353-93-6, PUBCHEM-CID: 10,106,139) and assigned ^{13}C-NMR data showing the large difference between experimental and predict chemical shift values.

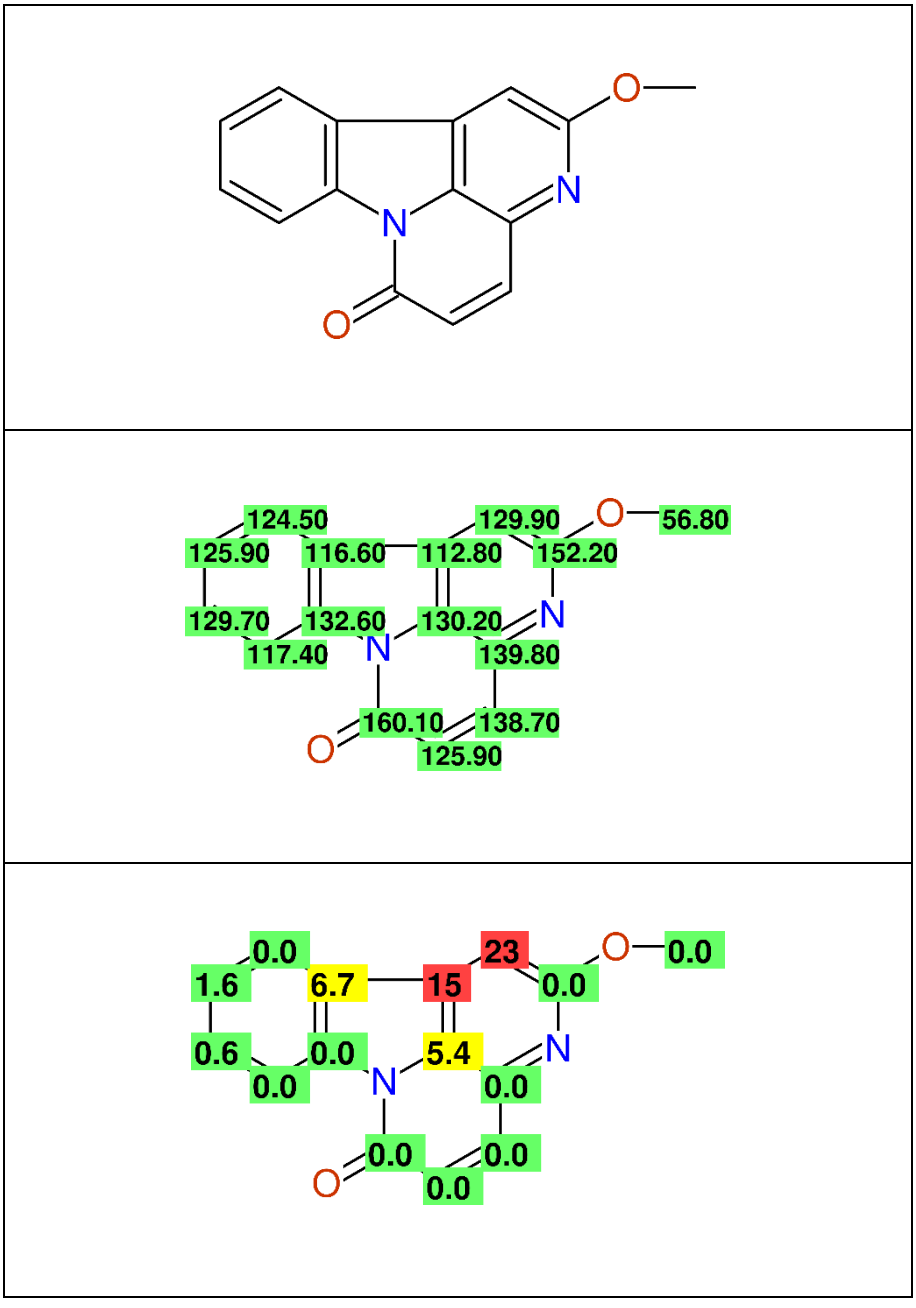
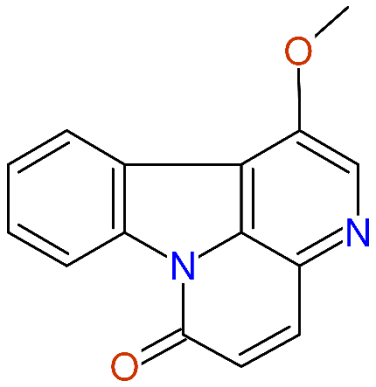
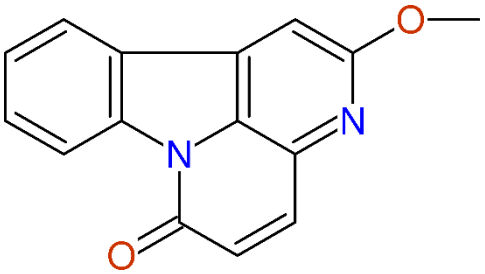


Table 2 The ranked hitlist shows the given structure and the assigned ^{13}C-NMR data at position 120, whereas the 1-methoxycanthin-6-one is located at position 13 with a significantly smaller average deviation.

	
1-Methoxycanthin-6-one Deviation: $\Delta\delta_C$=1.55ppm Position: 13 CAS-RN: 60755-86-4 PUBCHEM-CID: 483,522	2-Methoxycanthin-6-one Deviation: $\Delta\delta_C$=2.27ppm Position: 120 CAS-RN: 116353-93-6 PUBCHEM-CID: 10,106,139

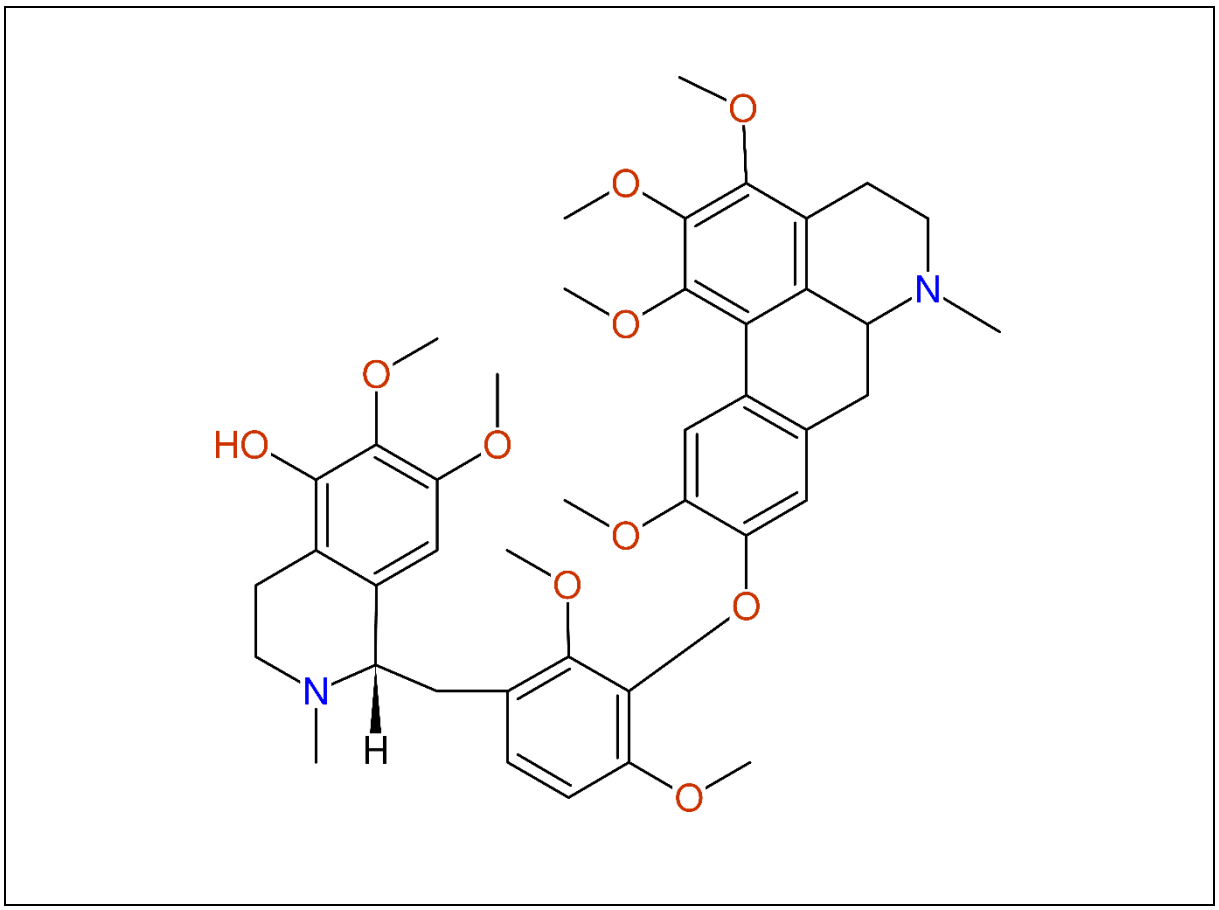
Alkaloids from *Thalictrum przewalskii*

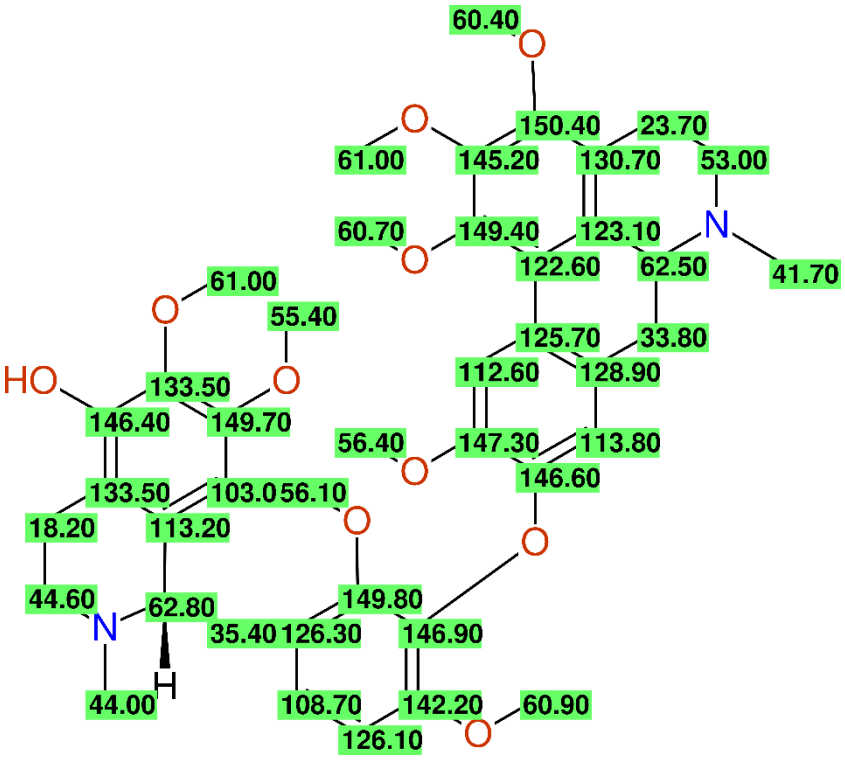
Nine new alkaloids were isolated from *Thalictrum przewalskii* Maxim. together with a few already known alkaloids. Their structures were determined by extensive use of 1D- and 2D-NMR techniques including ^{1}H- and ^{13}C-NMR as well as CH-COSY, CH-COLOC, HMQC and HMBC [39]. Compounds 1-7 from [39] have either a 1,2,3-trihydroxy-4-methyl-benzene moiety or a 1,2,4-trihydroxy-5-methylbenzene-motif in common within the central non-fused ring system. Compounds 1, 2, 4 and 6 have the first substructure in common, whereas substances 3, 5 and 7 belong to the second class of compounds. Przewalstidine has been chosen to represent the problems with the signal assignment of the 4 compounds having the 1,2,3-trihydroxy-4-methylbenzene motif. Two reasonable alternative structures having nearly identical similarity measures (1.25ppm and 1.46ppm, respectively instead of 1.26ppm) are found during the structure generation process. This example clearly shows the necessity to inspect the original experimental data in order to decide between an assignment error or a different structure proposal; from the ^{13}C-NMR data alone it can be only derived that there is some severe inconsistency in the publication (**Fig. 3**). The given data for Przewalstidine (compound 6 in [39]) are consistent with the spectral patterns of the other very similar compounds in this series, therefore a further detailed discussion will give neither more insight nor allow a final decision without reinspection of the raw data and is therefore omitted.

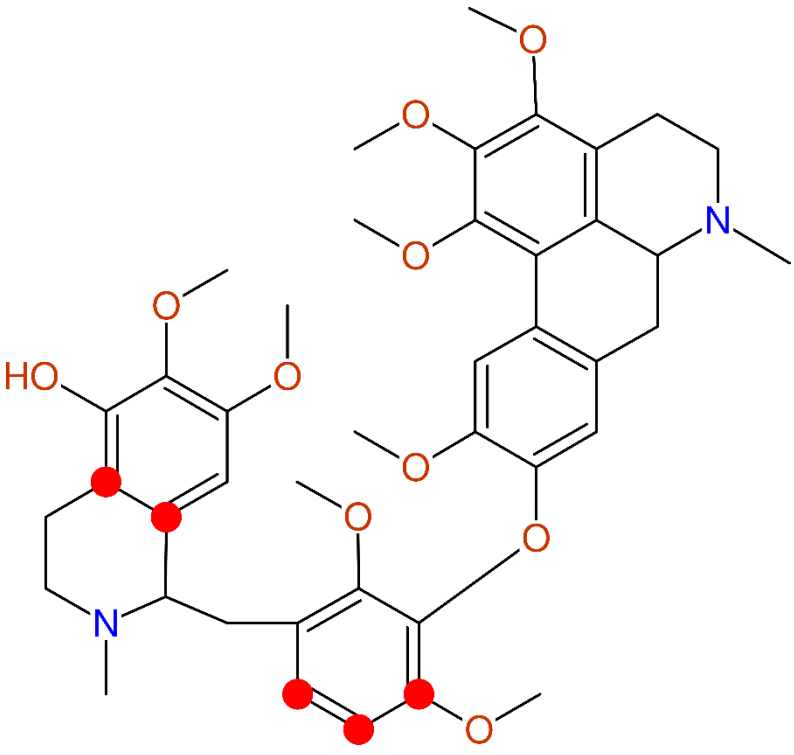
The second series of compounds having the 1,2,4-trihydroxy-5-methylbenzene fragment is analyzed in detail only for one representative (compound 7 in [39]) out of this series of three similar compounds because of consistent signal assignments therein. The conclusion drawn from this example is again a significant difference between experimental and predicted chemical shift values

(**Fig. 4**); a reasonable alternative structure proposal could not be derived from ^{13}C-NMR data alone showing again the need of inspection of the original experimental data. It is important to note that the signal assignment within the two series of compounds is consistent making them highly reliable showing the need for a second completely independent technology for spectrum prediction, because the well-established HOSE-code method severely tends to reproduce also obviously wrong assignments whereas neural networks are significantly more error-tolerant.

Fig. 3 Przewalstidine (compound 6 in [39], CAS-RN: 205520-62-3, PUBCHEM-CID: 101,937,318) and the assigned ^{13}C-NMR data resulting in five questionable positions. Two possible alternative structure proposals, a yet unknown compound having $\Delta\delta_C$=1.25ppm (bottom/left) and a compound known with PUBCHEM (CID: 101,937,319) having $\Delta\delta_C$=1.46ppm (bottom/right).




60.40
O
O
150.40
23.70
61.00
145.20
130.70
53.00
N
60.70
149.40
123.10
O
122.60
62.50
41.70
61.00
O
55.40
125.70
33.80
HO
133.50
O
112.60
128.90
146.40
149.70
56.40
147.30
113.80
O
146.60
133.50
103.0
56.10
O
18.20
113.20
O
44.60
62.80
149.80
N
35.40
126.30
146.90
H
44.00
108.70
142.20
60.90
126.10
O


O
O
O
N
HO
O
O
O
N
O
O

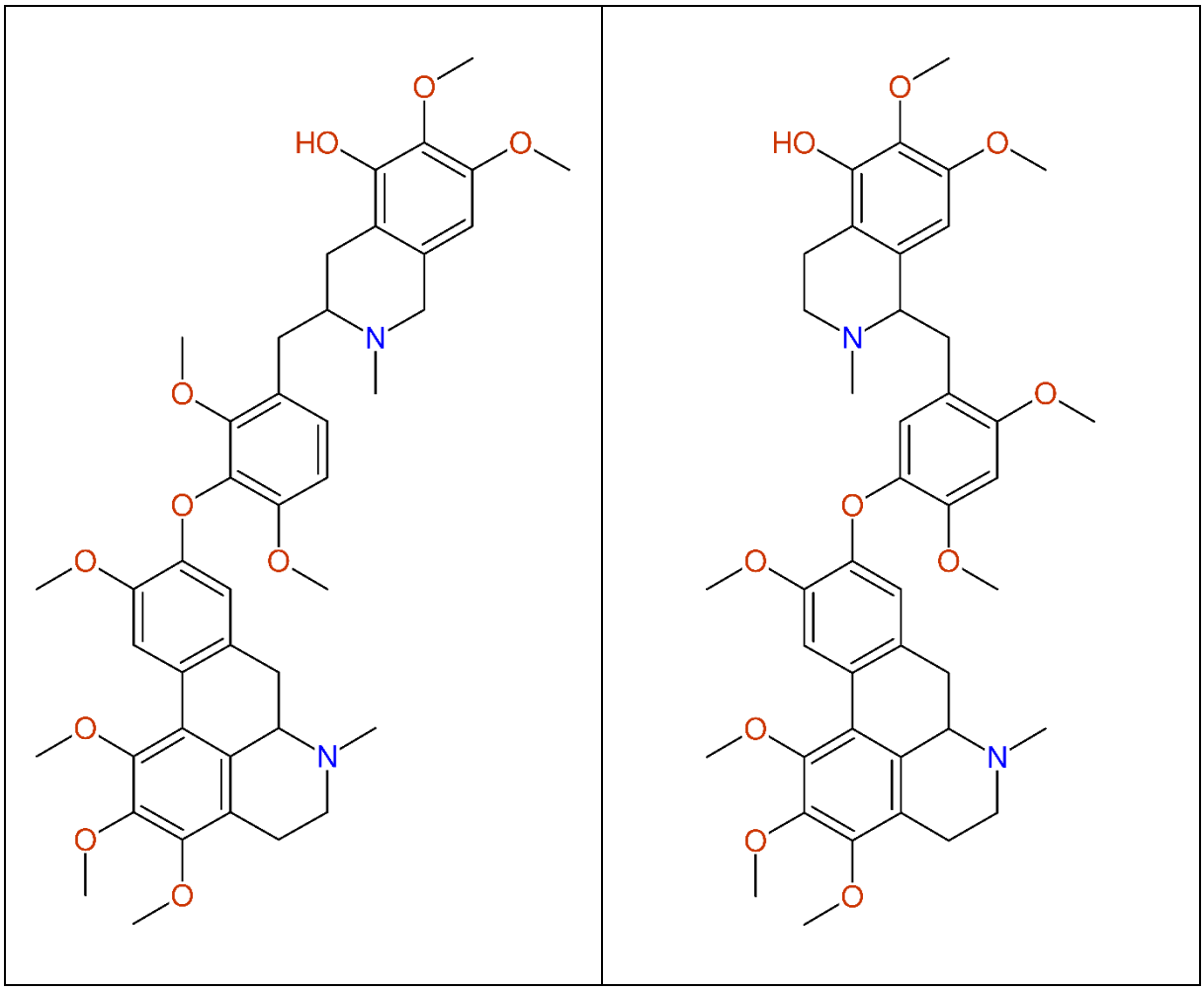
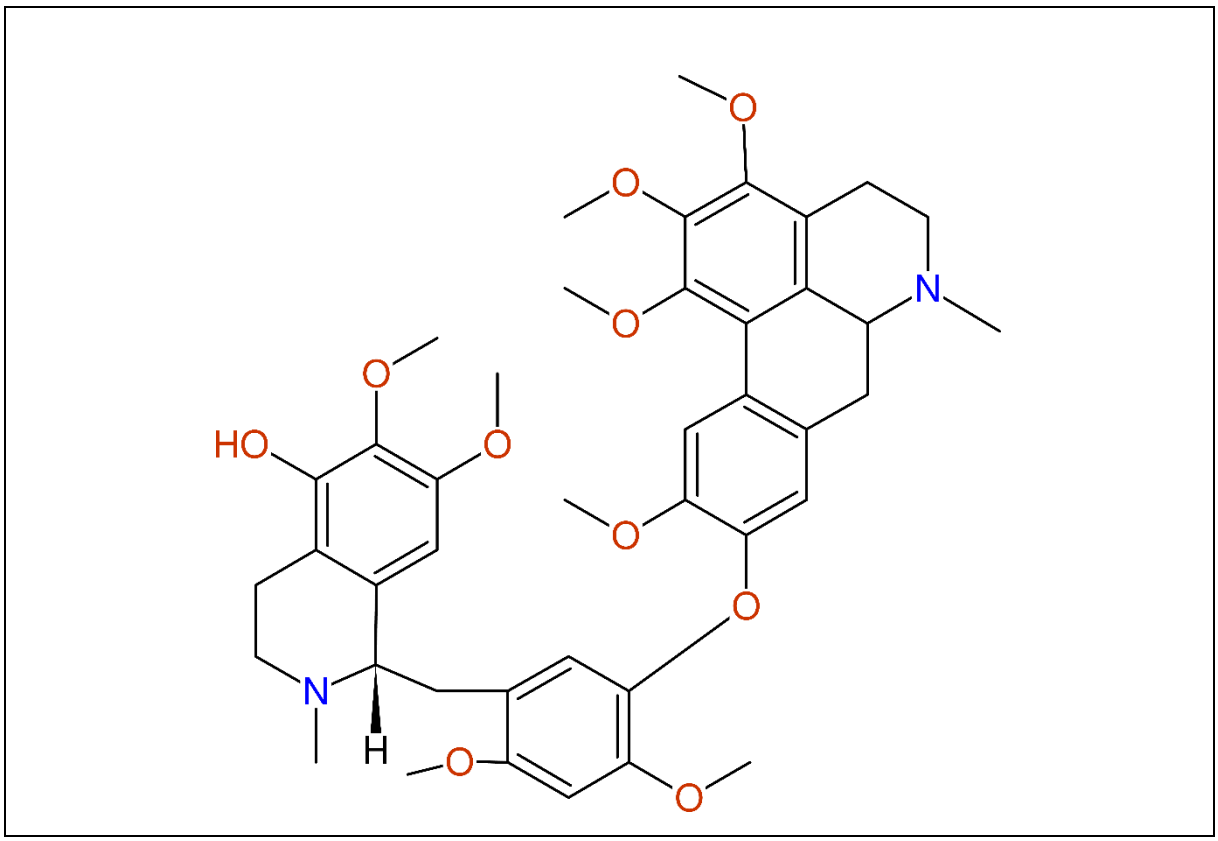
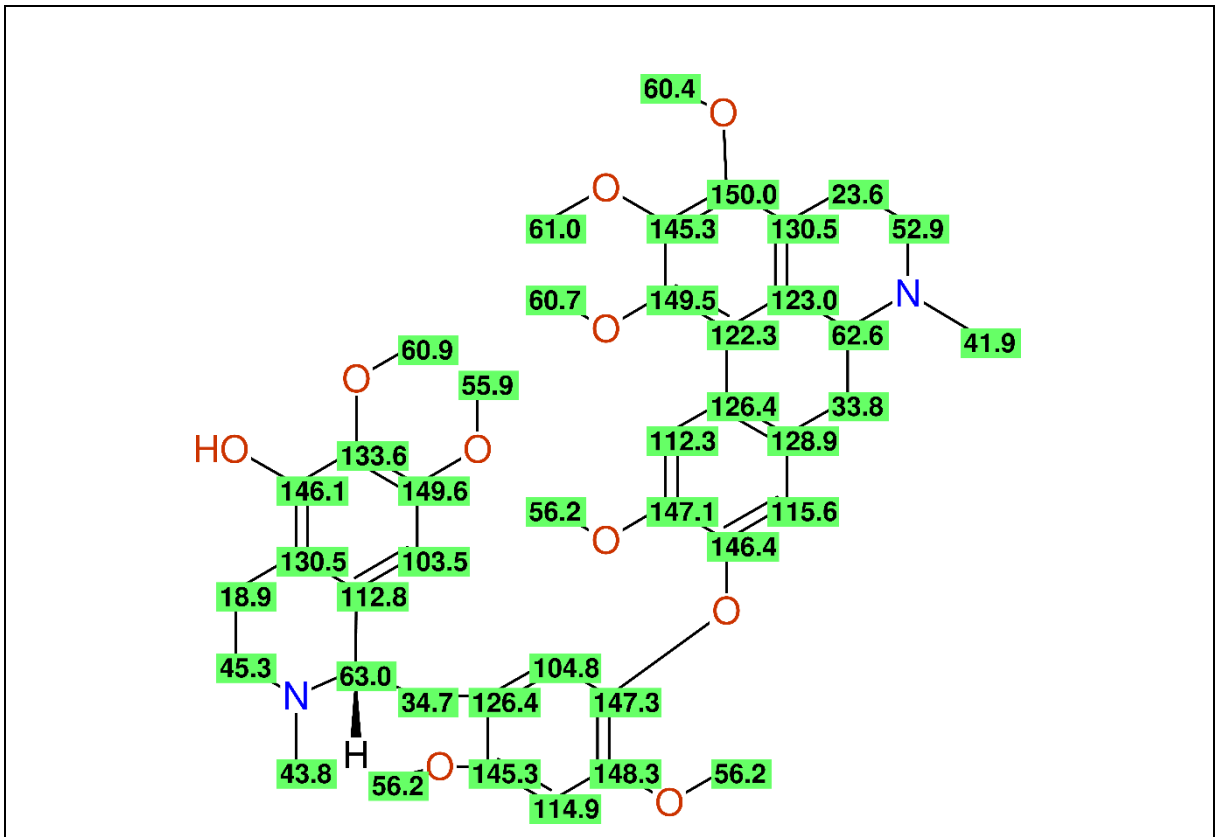
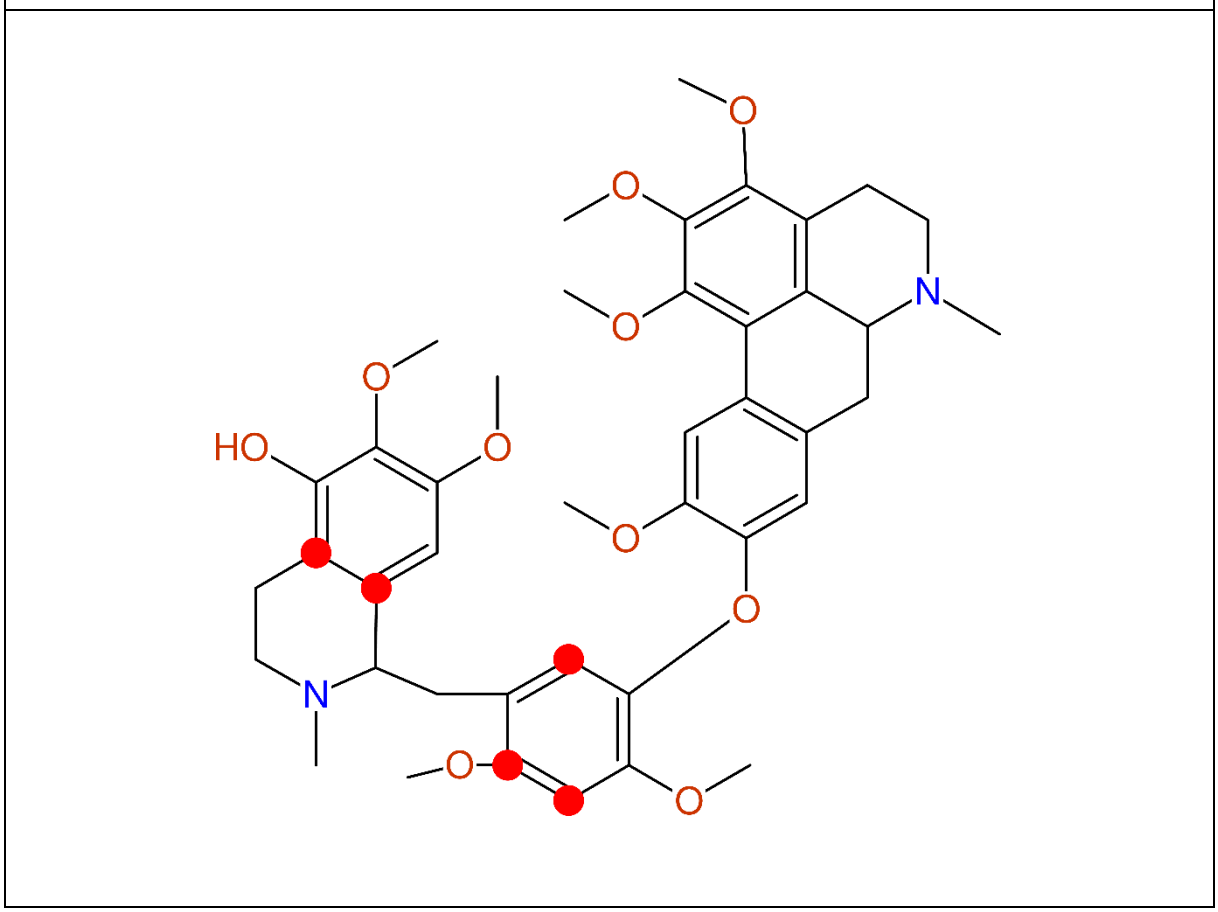

HO
O
O
N
O
O
O
O
O
O
O
O
HO
O
O
N
O
O
O
O
O
O
O
O

Fig. 4 Przewalstidinine; compound 7 from [39] (CAS-RN: 205520-63-4, PUBCHEM-CID: 101,937,319) and the assigned ^{13}C-NMR data resulting in five questionable positions. No reasonable alternative structure could be generated.



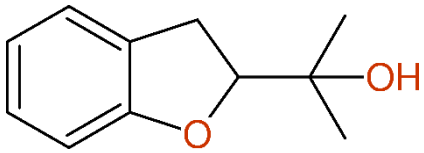
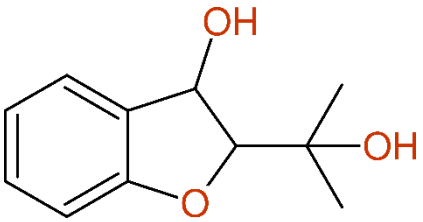
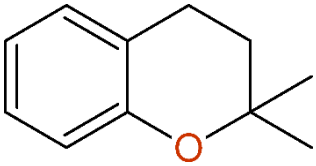
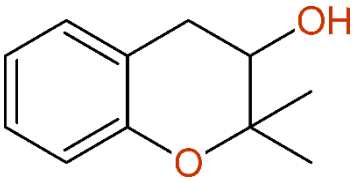
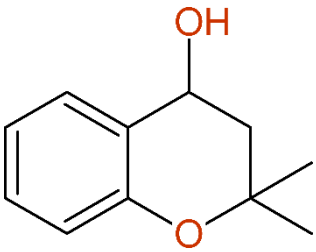
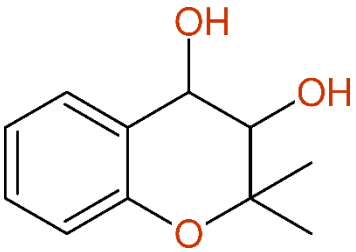

60.4
O
O
150.0
23.6
61.0
145.3
130.5
52.9
N
60.7
O
149.5
123.0
122.3
62.6
41.9
60.9
O
55.9
126.4
33.8
HO
133.6
O
112.3
128.9
146.1
149.6
56.2
O
147.1
115.6
130.5
103.5
146.4
18.9
112.8
O
45.3
N
63.0
104.8
34.7
126.4
147.3
H
43.8
56.2
O
145.3
148.3
O
56.2
114.9


O
O
O
N
HO
O
O
O
N
O
O

Benzofurans from *Eupatorium aschenbornianum* - Chromanes versus Dihydrobenzofuranes

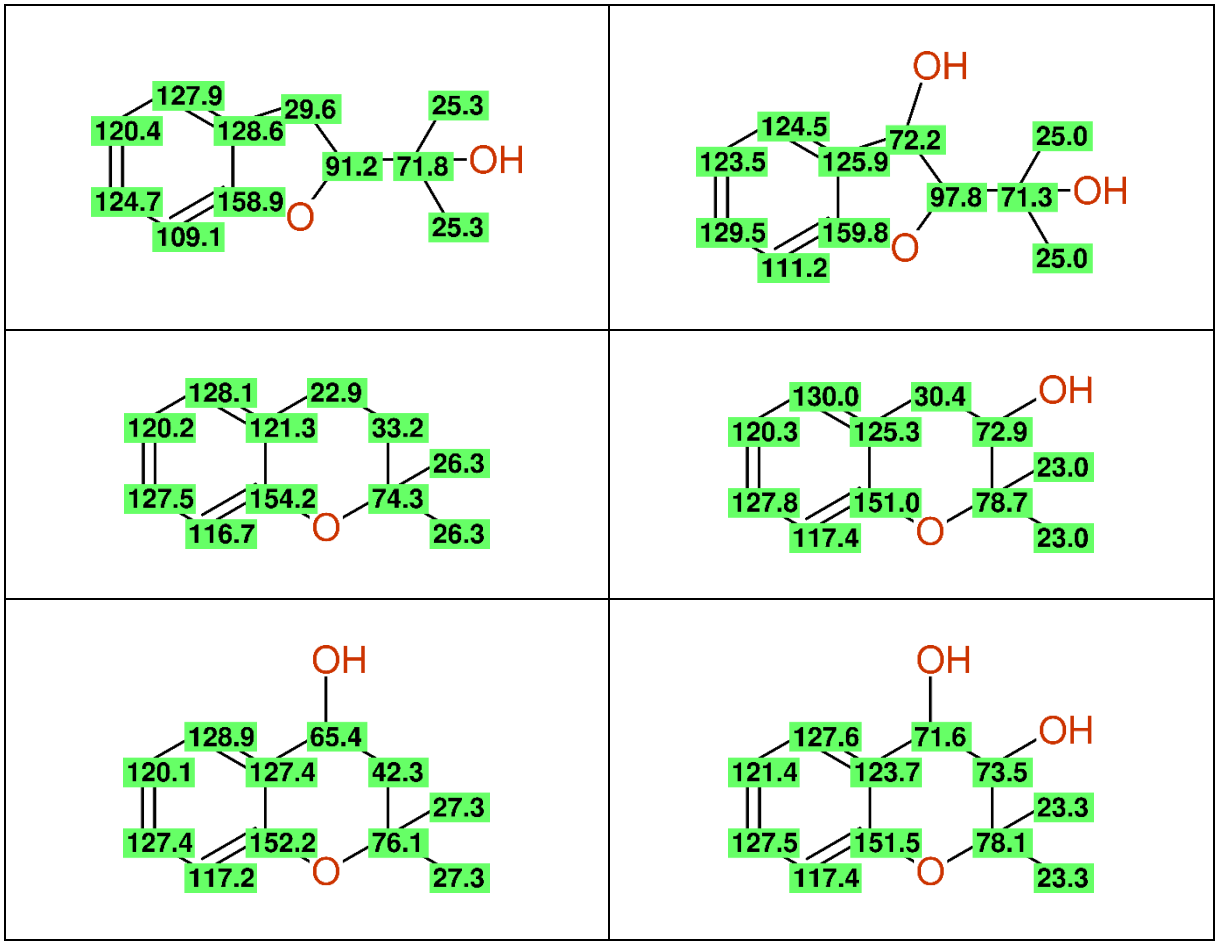
Chromanes and 2,3-dihydrobenzofuranes are frequently occurring motifs in natural products. These two classes of compounds are frequently mixed up during the structure elucidation process because of high similarity in their proton NMR data and their spin-systems. Even with 2D-NMR – when neglecting simple chemical shift arguments – it is difficult to distinguish between them. In order to get an impression of the scope of the problem a partial structure search using the CAS-registry file for six frequently occurring structural fragments having only additional substitution at the aromatic ring system allowed, was performed and reveals approximately 16,000 known compounds (**Table 3**).

Table 3 The six partial structures used for a SCIFINDER-Search and the number of their occurrence in the CAS-registry file (Search performed on April, 23rd,2020). As an additional constraint only further substitution at the aromatic ring system is allowed.

	
935	130
	
13,589	602
	
857	442

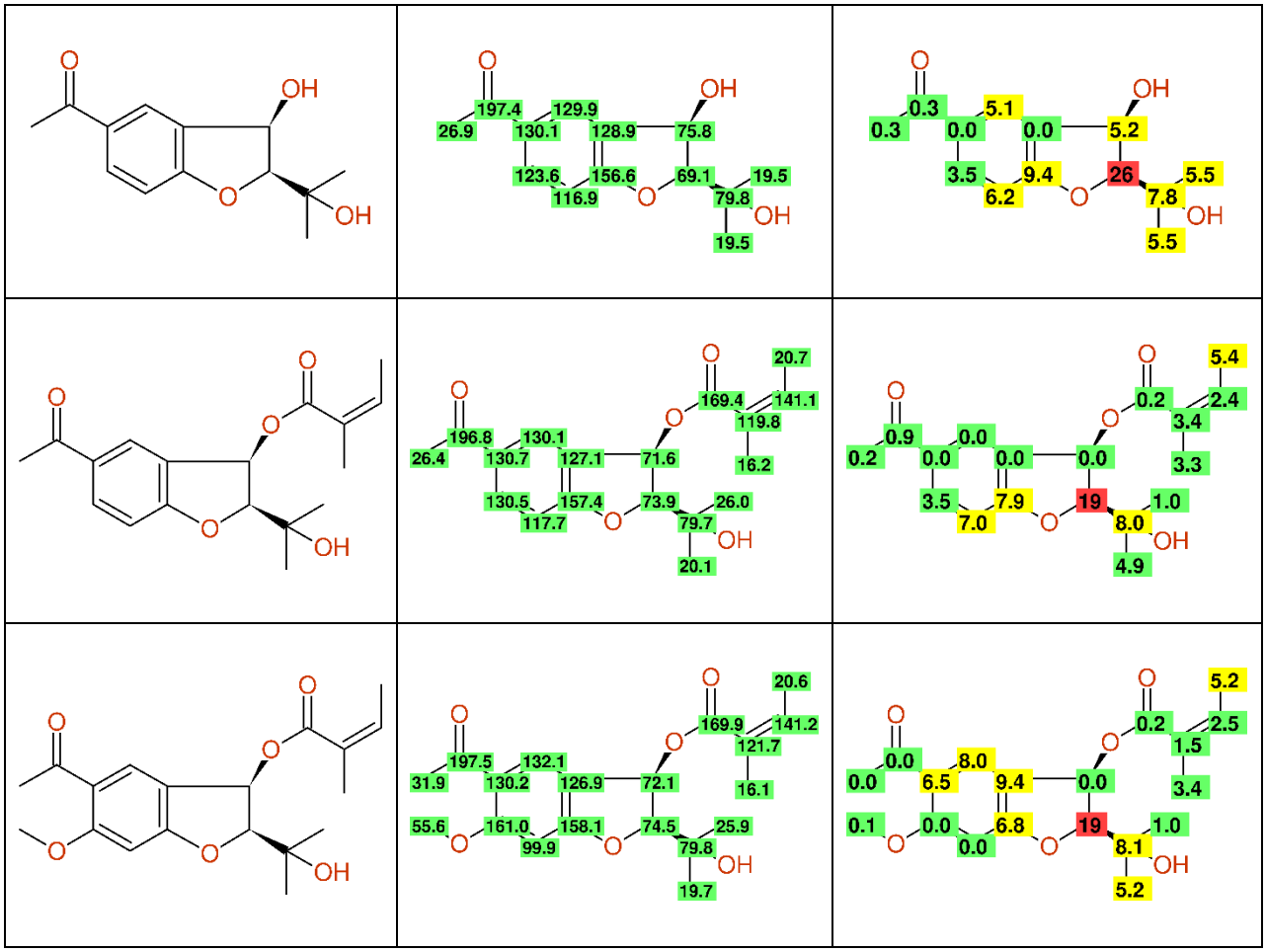
The most significant indicator of a dihydrofuran-derivative is the carbon chemical shift value at position 2 in the dihydrofuran-ring system, which resonates within the range between 89 and 99ppm, whereas all chemical shift values of the sp^3-carbons in the pyranoid-type derivatives resonate below 80ppm. Three independent spectrum predictions using the CSEARCH-Robot-Referee [26] based on 340,554 reference spectra, NMRSHIFTDB [14] based on 52,452 spectra and the predicted spectra contained in the CAS-Registry file, which were calculated with the ACD-software [40], strongly support this finding (**Fig. 5**).

Fig. 5 Predicted chemical shift values using the CSEARCH-Robot-Referee [26]. The chemical shift values of the sp^3-carbons are below 80ppm for the pyranoid-derivatives, whereas the furanoid-derivatives resonate above 88ppm allowing easy distinction between these 2 classes of compounds.



When searching the CSEARCH-NMR database for chemical shift values more than 15ppm away from the values to be expected, a publication [41] was found describing two new benzofuranes isolated from *Eupatorium aschenbornianum*. The structures described therein were elucidated by means of ^{1}H-, ^{13}C-NMR, DEPT, HMQC and HMBC measurements. The interpretation of the NMR-data together with IR- and MS-measurements leads to the published benzofuran-derivatives (**Fig. 6**).

Fig. 6 Three benzofuran-derivatives (compounds 5(top), 1(middle) and 2(bottom)) and their ^{13}C-NMR signal assignment as published in [41]. Left: Published structure. Middle: Experimental ^{13}C-NMR chemical shift values. Right: Difference between experimental and predicted chemical shift values.

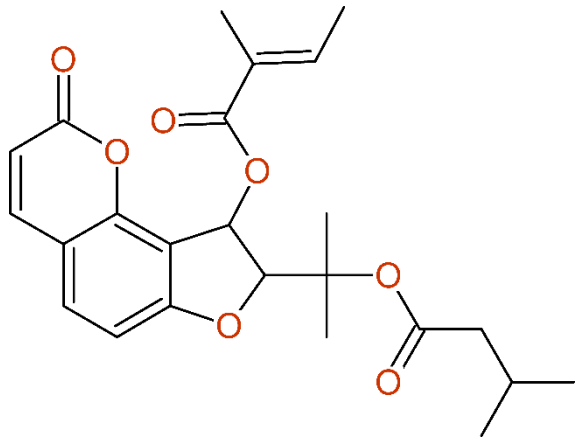
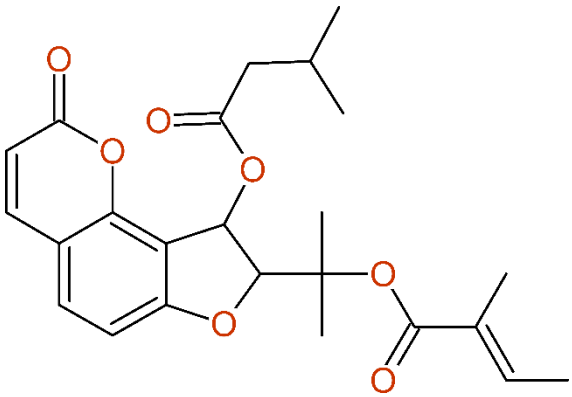
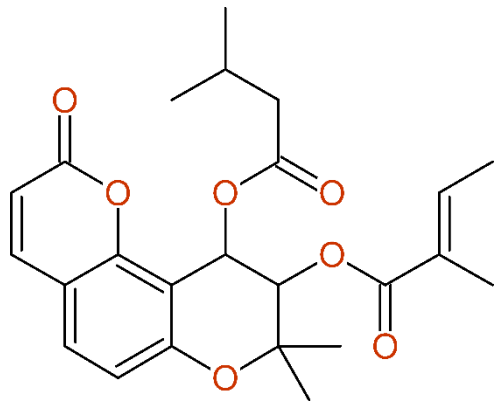
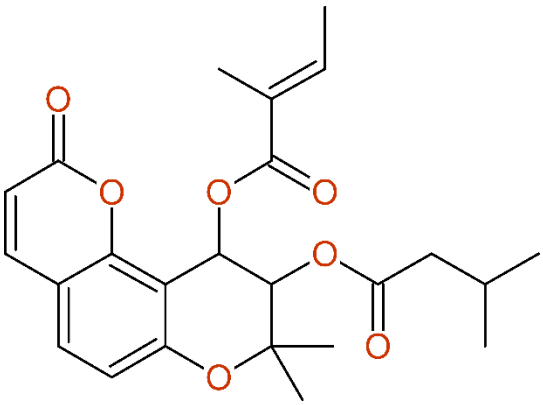


According to the authors the presence of the hydroxyisopropyl sidechain was verified by (mis)interpretation of the singlet at 79.7ppm. The typical chemical shift value for this carbon in benzofurane-derivatives is found around 71ppm, whereas 79ppm is characteristic for the carbon at position 2 in 2,2-dimethylated benzopyran-derivatives. Based on the very detailed discussion of the structure elucidation process of compound 1 in [41], the structure of compound 2 is elucidated by similarity consideration and HMQC and HMBC-data as well as by comparison with literature data. For this reason 3 literature citations [42,43,44] are given. In [42] and [43] only ^{1}H-data are used to characterize the isolated compounds; in [44] two furanocoumarines are described by their ^{1}H- and ^{13}C-data together with DEPT-measurements and HETCOR-correlations. These two compounds exactly show the characteristic line at 89ppm for C_2 as to be expected for a dihydrofuran-moiety, the typical line around 71ppm for the hydroxyisopropyl-fragment is shifted downfield here because of further substitution with isovaleric acid and angelic acid, respectively. The ^{13}C-NMR data in the cited literature [44] are correctly interpreted and benzofurans are elucidated thereof, whereas in [41] a significantly different spectral pattern is used to derive similar structure proposals having also a benzofuran moiety each. In the first step of the computer-assisted analysis, the data in the cited article [44] were processed using the CSEARCH-Robot-Referee in order to verify them as valid reference material. When starting from the ^{13}C-NMR data together with the given structure proposal of 2'-angeloyl-3'-isovaleryl vaginate (compound 1a in [44]) the automatic evaluation recommends "Minor Revision" mainly based on the fact that the underlying reference data taken from the literature used for spectrum prediction show some inconsistency. The structure generation process creates 1912 alternative structures, three out them are contained in the PUBCHEM-collection [30]. The subsequent ranking of the 1912 proposals puts the published structure at position 21 in the final hitlist with an average deviation

of 1.98ppm, whereas the other known structures can be found at positions 77 and 129 having a deviation of 2.24ppm and 2.40ppm, respectively. There is only 1 benzofurane derivative with a smaller deviation of only 1.79ppm, which has the 2 acid-residues exchanged. It is interesting to note that starting from the obviously correct benzofuran-derivative as published in [44] both possible alternative benzopyrane-derivatives were generated, both are ranked behind the benzofurane-analogue in the final hitlist (**Table 4**). From this evaluation the structure proposal and the given ^{13}C-NMR data can be assumed to be consistent with a very high degree of probability. The positions of the two acid fragments were analyzed in the literature by comparison mainly of the ^{1}H-NMR data with reference compounds – the mentioned differences in the ^{13}C-NMR chemical shift values are far below the usual preciseness of spectrum prediction technologies.

Application of the identical workflow using the data of compound 1b (=Archangelicin) from [44] creates 410 alternative structures, 2 out of them are known compounds in the PUBCHEM-collection [30]. The given structure of Archangelicin (PUBCHEM-CID: 5,281,371) is ranked at position 4 in the final hitlist, with no other reasonable alternative having a smaller average deviation. At position 22 the pyranoid-analogue can be found having a larger average deviation of 2.32ppm. These findings show again the consistency between the published structure proposal and the measured ^{13}C-NMR-data in the Planta Medica paper from 1992 [44] which is used as reference material in [41].

Table 4 Automatic structure verification for compound 1a from [44], which was used as reference material in [41] – during the structure generation process the configuration is ignored when creating display coordinates (tiglic acid shown instead of angelic acid)

	
Compound 1a from [44] Deviation: $\Delta\delta_C$=1.98ppm Position in hitlist: 21 PUBCHEM-CID: 56,776,263	Alternative Structure from CSEARCH Deviation: $\Delta\delta_C$=1.79ppm Position in hitlist: 4 PUBCHEM-CID: unknown
	
Alternative Structure from CSEARCH Deviation: $\Delta\delta_C$=2.24ppm Position: 77 PUBCHEM-CID: 13,592,770	Alternative Structure from CSEARCH Deviation: $\Delta\delta_C$=2.40ppm Position: 129 PUBCHEM-CID: 132,606,937

Application of the CSEARCH-protocol to the structures contained in the cited literature [44] verified the quality of this reference material and also supported the conclusions withdrawn from chemical shift arguments. Now the same workflow will be applied to the two new benzofurans from *Eupatorium aschenbornianum*. Their structure elucidation is based on state-of-the-art 2D-NMR techniques but ignores the arguments derived from chemical shift analysis [41].

The evaluation for compound 5 starts from the published structural proposal together with the ^{13}C-NMR signal assignment showing massive deviations between experimental and predicted chemical shift values (**Fig. 6**) mainly located at the dihydrofuran-system. This evaluation recommends “Major Revision”. It should be mentioned that these obviously wrong data (structure and signal assignment) are contained in the underlying CSEARCH-database used for this analysis; based on this fact the prediction using the HOSE-code technology [4] exactly reproduces the wrong spectral data, whereas the NN-technology predicts more reliable chemical shift values. This example dramatically shows the importance of at least two independent prediction technologies in order to obtain reliable results. Starting the structure generator program integrated in CSEARCH from the wrong structure proposal with the information of questionable carbon positions derived from the previously mentioned spectrum prediction creates 2,169 proposals. 1,975 out of them are isomers and 1,314 structures have the same signal multiplicity as the published proposal. Subsequent spectrum prediction for each proposal using NN-technology allows ranking of 2,169 structures created. The published structure can be found at position 277 having an average deviation of 4.85ppm, the pyranoid analogue has a deviation of only 2.20ppm and can be found at position 2 (**Table 5**). The best alternative at position 1 in the ranked hitlist is a highly strained structure

incompatible with the proton NMR-data. From a spectral similarity search ignoring the multiplicity information, the identical benzopyran-analogue as proposed by the structure generator is found within the CSEARCH-database associated with two other publications [45,46]. This finding strongly supports the result obtained by the CSEARCH-Robot-Referee.

Applying the same workflow to compound 1 from [41] decreases the average deviation from 3.51ppm for benzofuran-derivative to 1.63ppm for benzopyran-analogue (**Table 5**). The analysis of the similar benzofurane/benzopyrane structure pair created when starting from the data of compound 2 in [41] shows the opposite effect. The benzofurane-derivative is slightly preferred giving an average deviation of 2.54ppm whereas the benzopyrane-derivative shows 2.68ppm. Visual comparison of the experimental spectrum from [41] versus the predicted spectra of both compounds (**Table 6**) showing that the increased similarity in the sp^3-region (70 – 100ppm) for the benzopyrane-derivative is compensated by a decreased similarity in the sp^2-region (105 – 135ppm) leading to the overall-effect, that the published benzofurane is ranked better (position 224 versus 379 (**Table 5**)) than the benzopyrane-analogue. The difference in the sp^2-region in the measured spectrum of compound 2 from [41] needs a more detailed investigation, because a similar natural product named 6-methoxytrementone (compound 1 in [47]) having the identical substitution pattern in the benzene ring shows considerably deviating values, even when reassigning the published data.

From this analysis it can be concluded with a high degree of confidence, that obviously the 3 published benzofurans are in error and they seem to be benzopyrans [41]. The cited paper [44] holds 2 furocoumarins which seem to be correct. The reference compound 1 from [47] has obviously an assignment error and furthermore the ^{13}C-NMR spectral data substantially differ despite the

substitution pattern compared with compound 2 in [41] of the benzene ring is quite similar. For both papers there is no “Supplementary Information” available prohibiting reinspection of the measured spectra. It should be mentioned that [41] has already been cited in 21 other publications holding 109 different structures without detecting this inconsistency in the ^{13}C-NMR spectral data. Among these 109 structures not even one 2,2-dimethyl-3,4-chromandiol can be found, only 4 structures out of these 109 have a benzopyran-4-one moiety which is not suitable as reference material for ^{13}C-NMR spectroscopy.

Table 5 Automatic structure revision for compounds 5 and 1 from [41] – in both cases the benzopyran-derivative is significantly preferred, whereas when starting from compound 2 the situation is more complicated (**Table 6**).

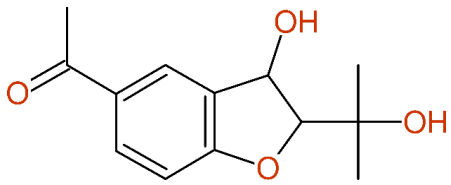
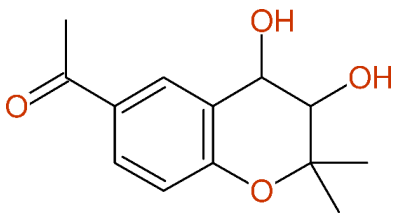
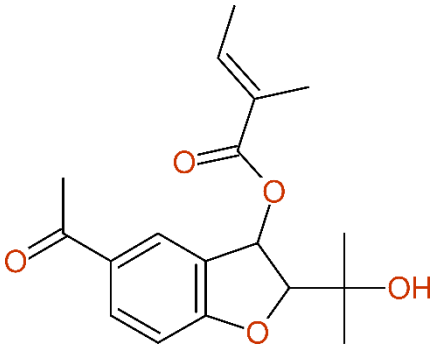
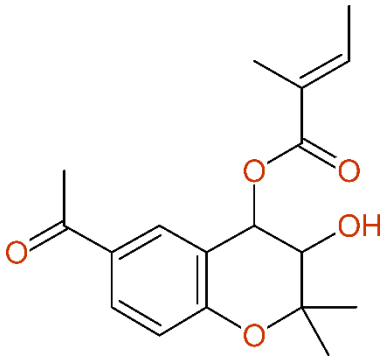
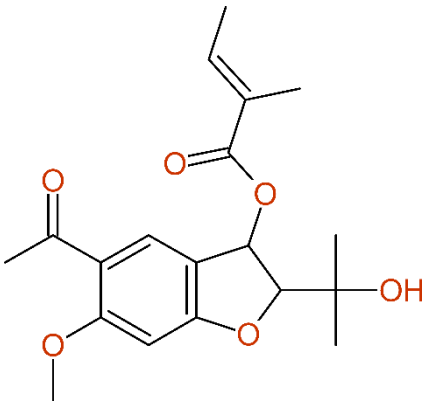
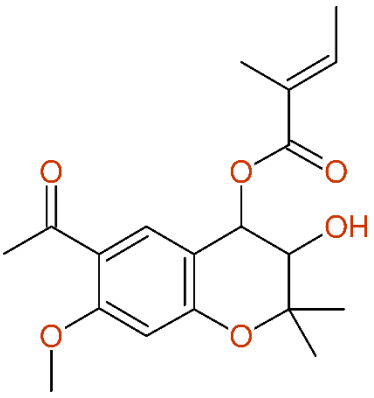
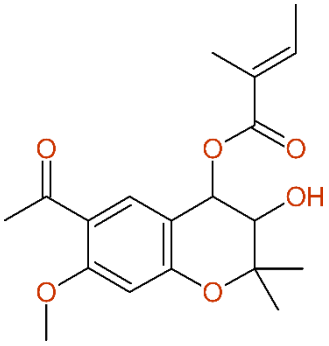
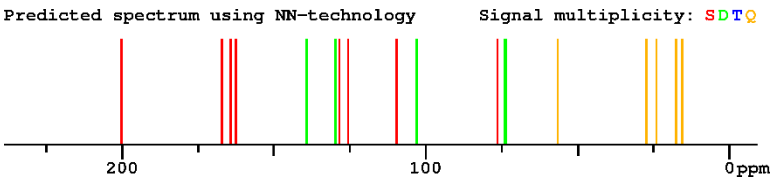
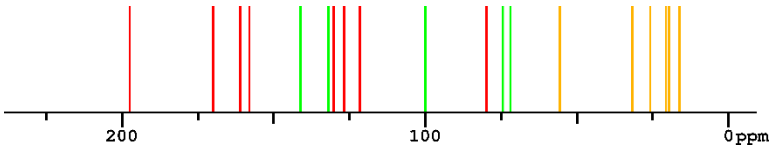
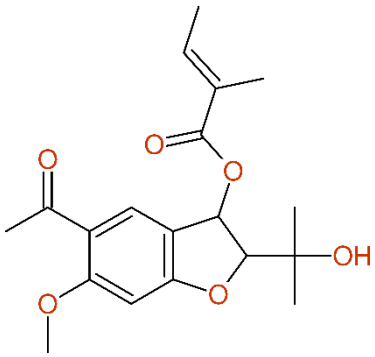
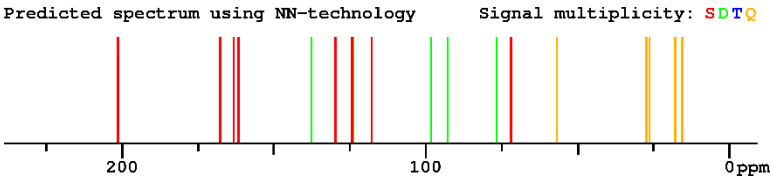
	
Compound 5 from [41] $\Delta\delta_C$=4.85ppm / Position 277 PUBCHEM-CID: 25,750,696	Best alternative structure [45,46] $\Delta\delta_C$=2.20ppm / Position 2 PUBCHEM-CID: 636,909
	
Compound 1 from [41] $\Delta\delta_C$=3.51ppm / Position 1077 PUBCHEM-CID: 78,412,635	Alternative structure $\Delta\delta_C$=1.63ppm / Position 12 PUBCHEM-CID: unknown
	
Compound 2 from [41] $\Delta\delta_C$=2.54ppm / Position 224 PUBCHEM-CID: 3,009,227	Alternative structure $\Delta\delta_C$=2.68ppm / Position 379 PUBCHEM-CID: unknown

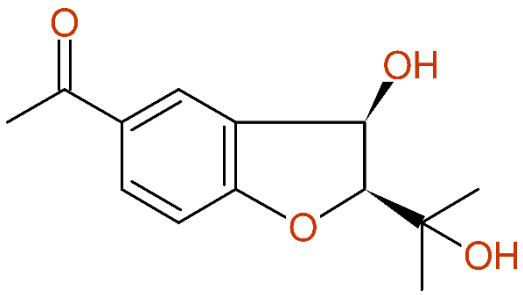
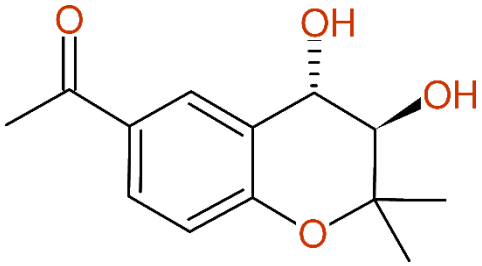
Table 6 Visual comparison of the predicted ^{13}C-NMR chemical shift values versus the experimental data of compound 2 from [41].

 Alternative structure proposed by CSEARCH	 Deviation: $\Delta\delta_C$=2.68ppm
Experimental spectrum of compound 2 as given in [41]	
 Published structure of compound 2 in [41]	 Deviation: $\Delta\delta_C$=2.54ppm

Using the experimental chemical shift values of compound 5 from [41] without multiplicity information allowing a deviation of +/-1ppm for a spectral similarity search the benzopyran-analogue is found again supporting the result obtained from the structure generation process. The peak lists of both compounds show nearly identical chemical shift values with inconsistent multiplicities, but in case of compound 13 from [45] no multiplicity determination was done (**Table 7**).

The final conclusion deduced from this detailed analysis is that there is in general a massive structure revision necessary in these two classes of compounds – the selected paper [41] is only one example among many others [48,49,50,51,52,53,54].

Table 7 Comparison of the published ^{13}C-NMR peaklists for the benzofuran-derivative and its benzopyran-analogue

Compound 5 from [41] Multiplicity determination: DEPT	Compound 13 from [45] Multiplicity determination: No details given
	
19.51 Q **19.51 Q** 26.93 Q 69.11 D 75.75 D 79.82 S 116.92 D **123.60 D** 128.85 S 129.95 D **130.13 S** 156.64 S 197.38 S	19.50 Q **26.90 Q** **27.00 Q** 69.20 D 75.80 D 80.20 S 117.20 D **124.10 S** 129.30 S 130.20 D **130.20 D** 157.20 S 198.30 S

Flavonoids from *Bolusanthus speciosus*

A series of four similar pterocarpans, named bolucarpan A to D were isolated from the root bark of *Bolusanthus speciosus* and their structures elucidated by application of ^{1}H-NMR, ^{13}C-NMR, COSY, DEPT, HMQC and HMBC-spectra [55]. The ^{13}C-NMR chemical shift values of the trihydroxylated benzene ring enumerated "A" in these compounds with signals around 136, 147 and 150ppm in all four derivatives are incompatible with the given 1,3,5-trihydroxy pattern. The CSEARCH-Robot-Referee is applied to the data of Bolucarpan B (**Fig. 7**), for the three other Bolucarpans analogous modifications must be done accordingly. It is interesting to note, that the deviations between experimental and predicted ^{13}C-NMR chemical shift values are extremely small; this is based on the fact, that the data from the HOSE-code prediction are predominantly taken for statistical reasons leading to a possible error propagation based on the mathematical model behind this method. When adding the evaluation obtained from the neural network technology which is based on a model with better strategies for generalizing substituent effects, the questionable carbons are automatically detected and included into the algorithm for structure generation as positions where structure changes are necessary in order to get better coincidence between experimental and predicted chemical shift values.

The structure generator creates 4540 alternative structure proposals, three alternatives are known compounds (**Table 8**). Eight proposals have a smaller average deviation between experimental and predicted ^{13}C-NMR chemical shift values compared to the published structure. The same chemical shift values as used to elucidate bolucarpan B [55] - proposal 167 (**Table 8**) - were published again a few years later in [56] (**Table 9**), the later paper does not cite the earlier

one and a different structure is elucidated - proposal 69 (**Table 8**) - from nearly identical data.

This example shows in an impressive way the urgent need for a public-domain and easily searchable repository.

Fig. 7 Bolucarpan B, published ^{13}C-NMR chemical shift values, deviation between experimental and predicted chemical shift values and highlighted questionable positions in the structure.

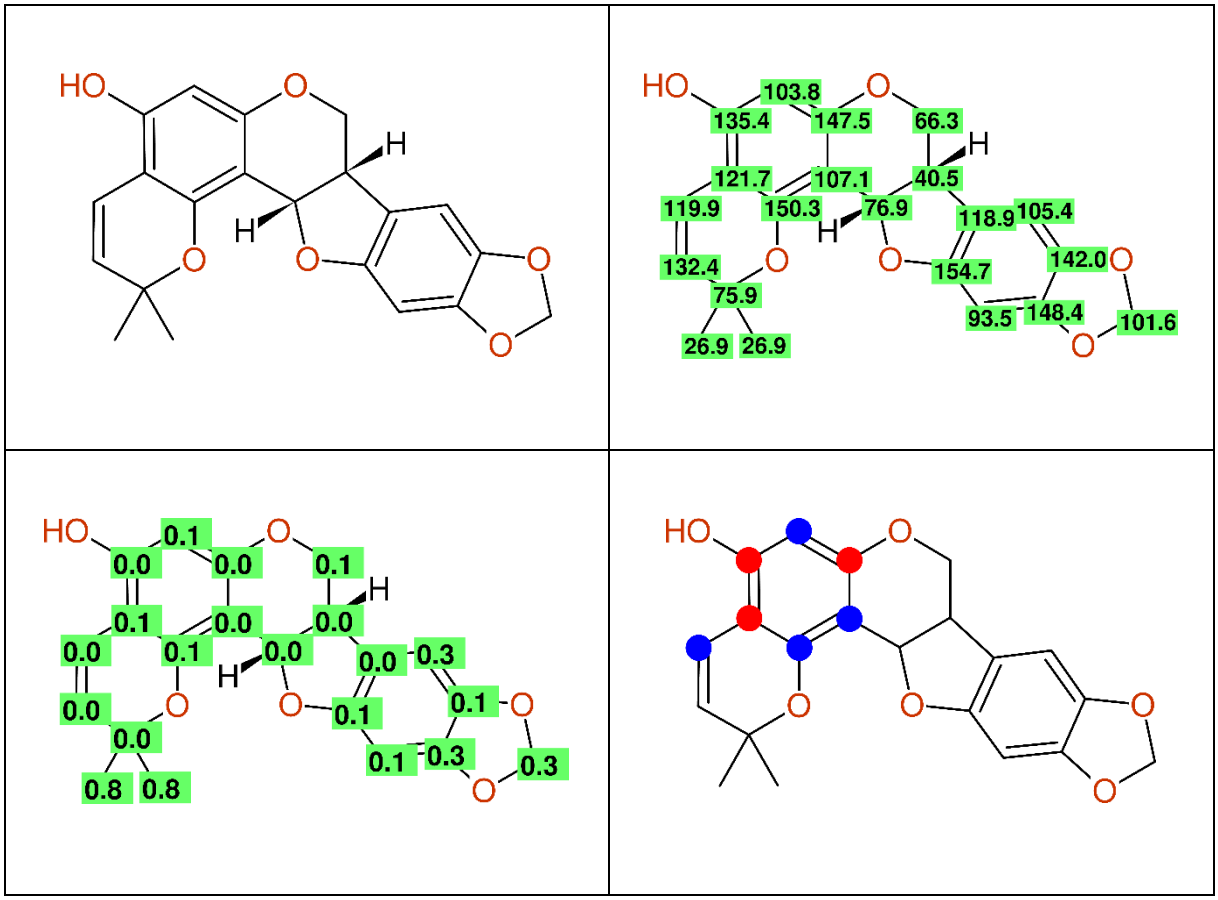
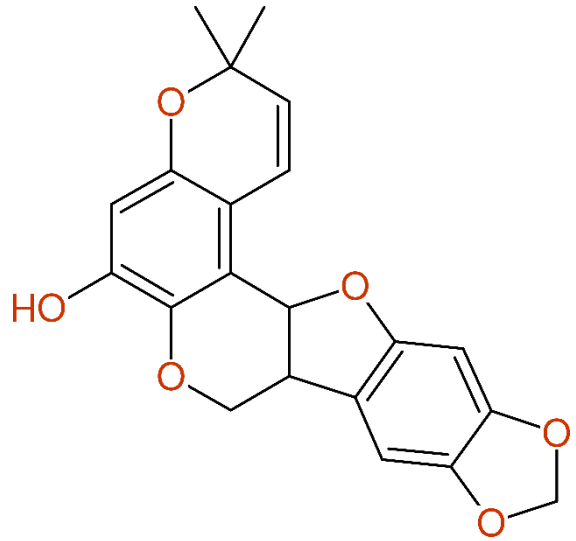
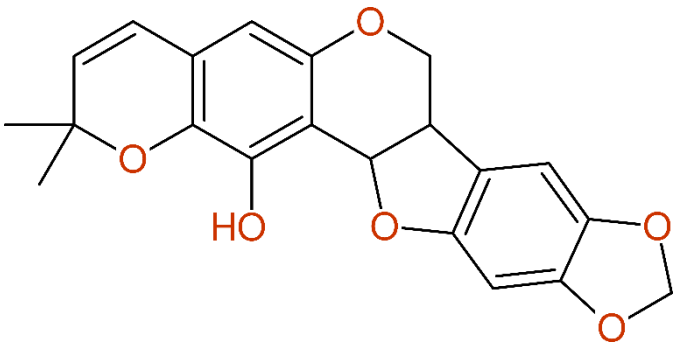
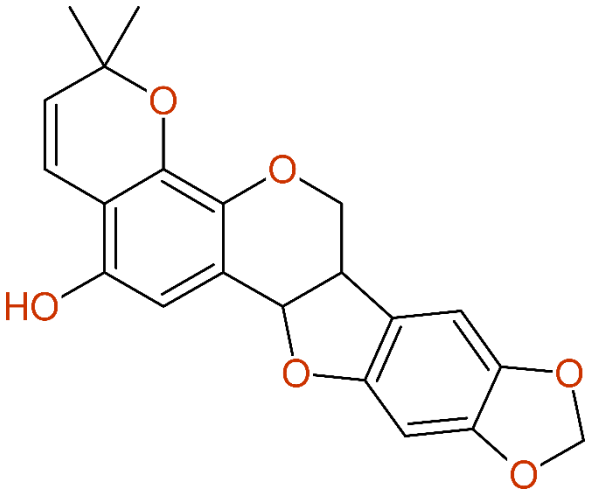
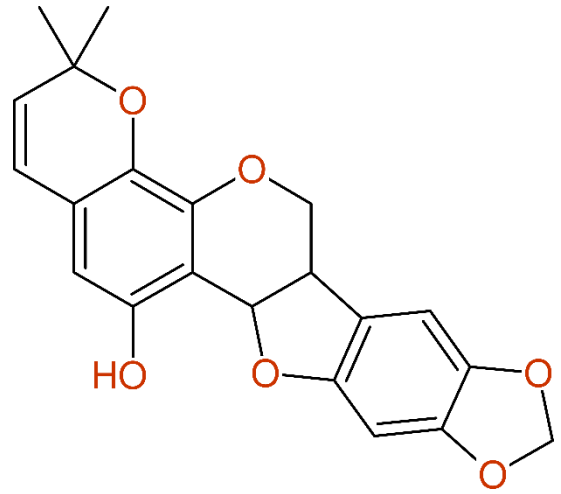
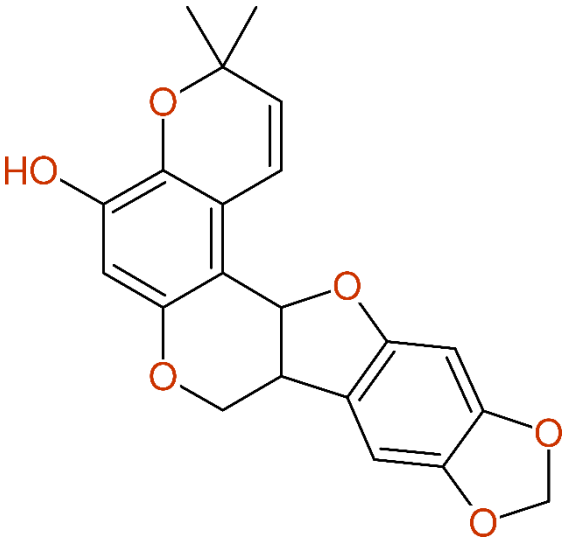
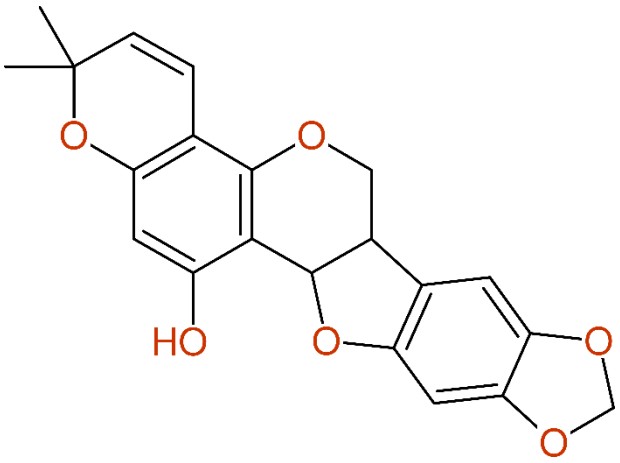


Table 8 Automatic structure revision starting from the data of Bolucarpan B [55]. Eight alternative structure proposals having the same structural motifs fit the ^{13}C-NMR data better than the published structure.

	Proposal 1 $\Delta\delta_C$=1.48ppm Unknown
	Proposal 26 $\Delta\delta_C$=1.70ppm Unknown
	Proposal 28 $\Delta\delta_C$=1.71ppm Unknown

	Proposal 29 $\Delta\delta_C$=1.71ppm Unknown
	Proposal 69 $\Delta\delta_C$=1.71ppm CAS-RN: 1160432-67-6 PUBCHEM-CID: 25,231,269 Alternative structure: Published in [56]
	Proposal 71 $\Delta\delta_C$=1.90ppm Unknown

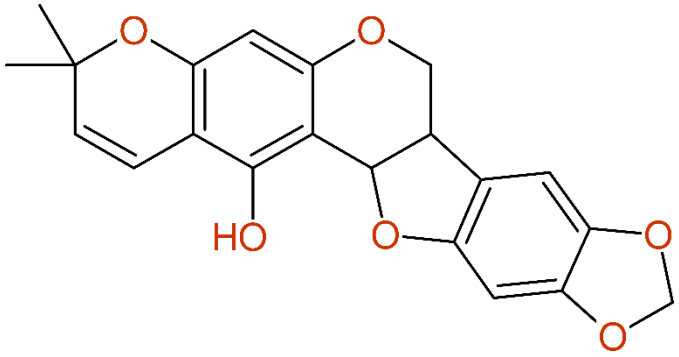
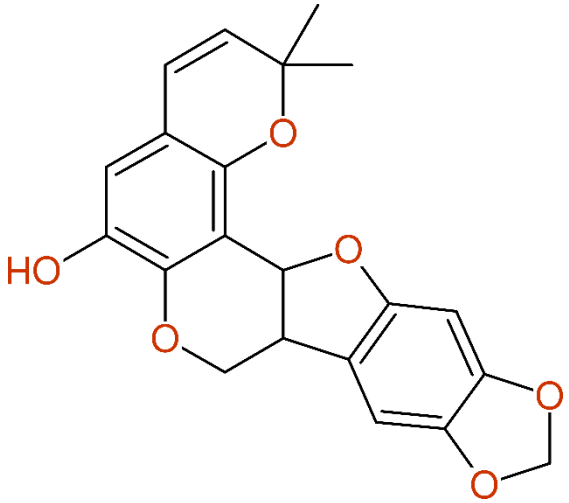
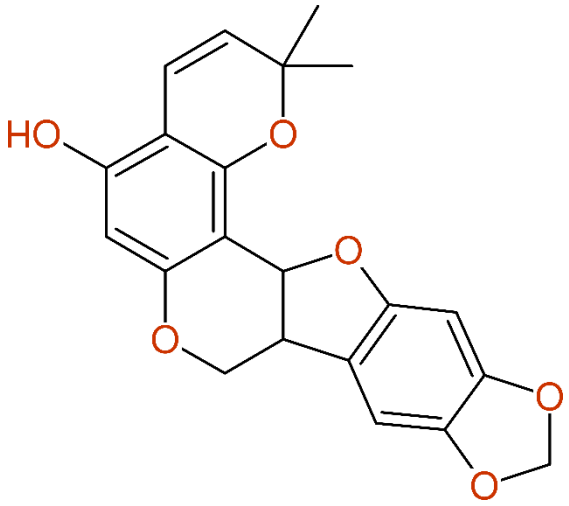
	Proposal 128 $\Delta\delta_C$=2.02ppm CAS-RN: 76175-39-8 PUBCHEM-CID: 44,257,467 Published in [57]
	Proposal 153 $\Delta\delta_C$=2.05ppm Unknown
	Proposal 167 $\Delta\delta_C$=2.07ppm CAS-RN: 467465-29-8 PUBCHEM-CID: 10,474,063 Proposed structure for Bolucarpan B Published in [55]

Table 9 ^{13}C-NMR data of Bolucarpan B (compound 3 in [55], proposal 167 (**Table 8**)) versus Maackiapterocarpan B (compound 5 in [56], proposal 69 (**Table 8**))

Compound 3 in [55] Acetone-d_6	**Compound 5 in [56] $CDCl_3$**	**Difference(ppm)**
S:	**S:**	
154.7	154.2	0.2
150.3	149.9	0.4
148.4	148.0	0.4
147.5	146.1	1.4
142.0	141.6	0.4
135.4	134.3	1.1
121.7	120.8	0.9
118.9	117.9	1.0
107.1	106.7	0.4
75.9	76.8	0.9
D:	**D:**	
132.4	132.0	0.4
119.9	119.2	0.7
105.4	104.7	0.7
103.8	103.1	0.7
93.5	93.8	0.3
76.9	77.2	0.3
40.5	40.1	0.4
T:	**T:**	
101.6	101.2	0.4
66.3	66.2	0.1
Q:	**Q:**	
26.9	27.6	0.7
26.9	27.7	0.8

Isoflavone from *Bituminaria morisiana*

The structure elucidation of compound 9 in [58] was performed by application of ^{1}H-, ^{13}C-, DQF-COSY, ROESY, HSQC and HMBC-spectra leading to an isoflavone-derivative (**Fig. 8**). The assigned experimental ^{13}C-NMR chemical shift values significantly differ from the predicted ones; obviously the 2D-NMR data have been wrongly interpreted ignoring basic chemical shift arguments. Using the given structure proposal together with the 1D-^{13}C-NMR peak list as a starting point for the structure generator implemented in the CSEARCH-Robot-Referee, a set of 4,751 alternative structures is obtained. After creating the ranked hitlist the published structure is found at position 77 having an average deviation $\Delta\delta_C$ of 2.72ppm. The best reasonable alternative – which is not contained in the CAS-Registry file - is found at position 1 having a $\Delta\delta_C$ of 1.80ppm (**Table 10**). Using the dihydrobenzofuran-moiety contained in the new structure proposal as query for a partial structure search retrieves compound 7 from [59] showing an excellent coincidence of ^{13}C-NMR chemical shift values in the dihydrofuran-fragment (**Fig. 9**).

Fig. 8 Compound 9 from [58]; the experimental ^{13}C-NMR chemical shift values and the differences between experimental and predicted data.

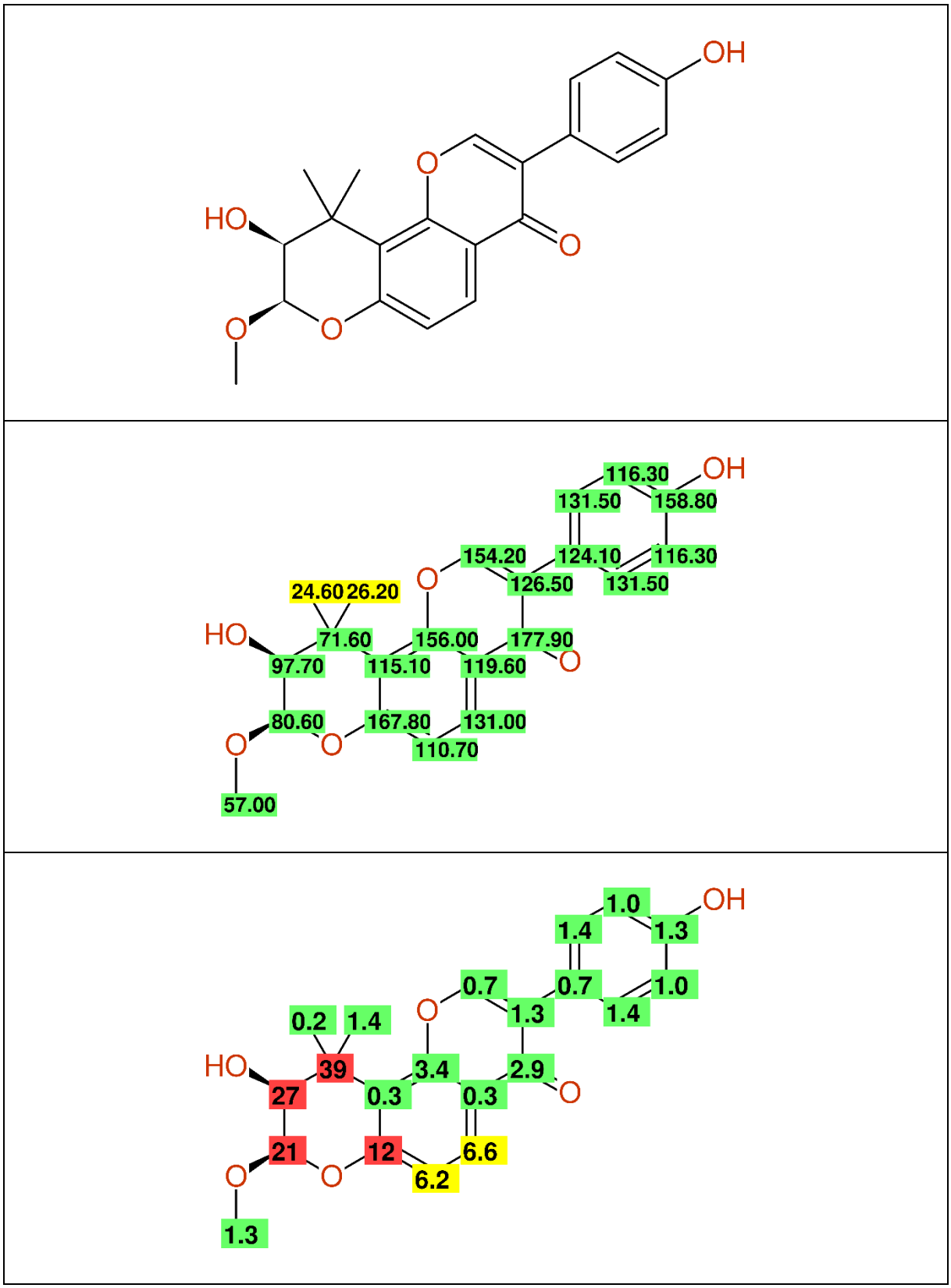


Table 10 Automatic structure revision of compound 9 from [58] giving the analogous benzofuran-derivative

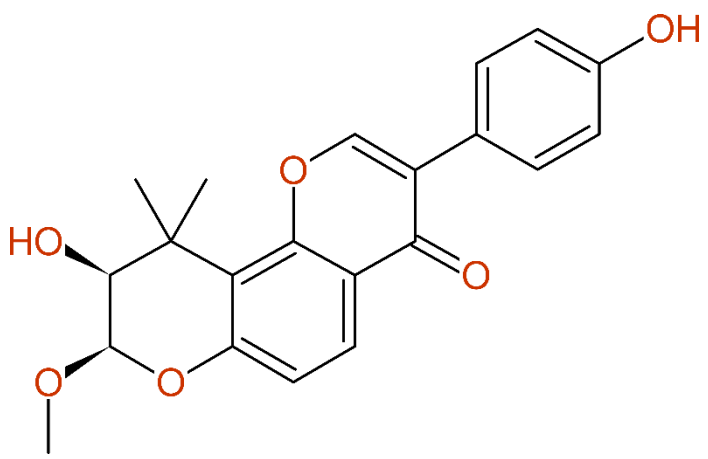
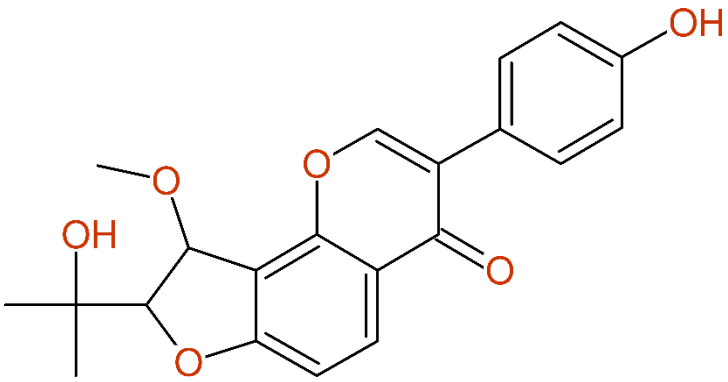
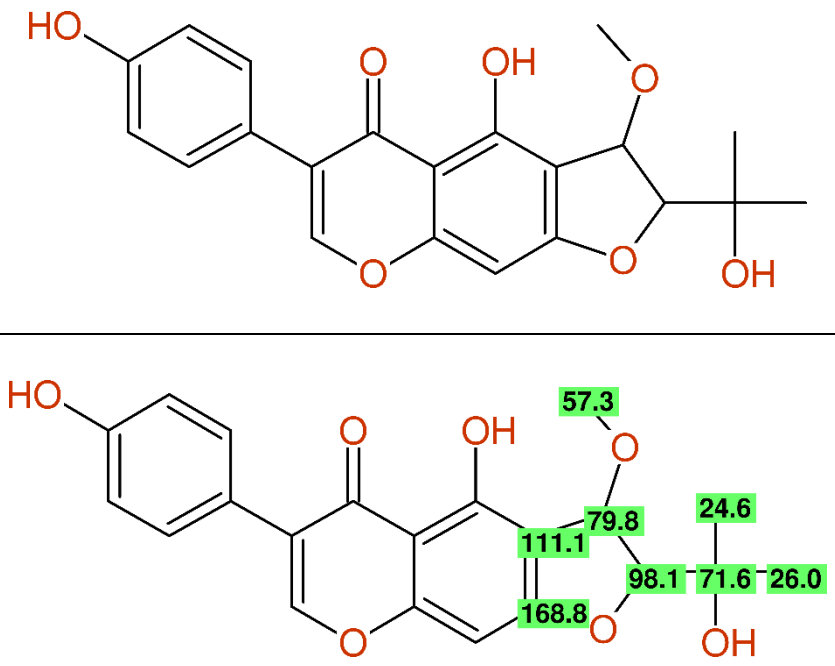

Published structure: Compound 9 from [58] CAS-RN: 857254-97-8 Position 77: $\Delta\delta_C$=2.72ppm

Best alternative proposal CAS-RN: unknown compound Position 1: $\Delta\delta_C$=1.80ppm

Fig. 9 The given ^{13}C-NMR chemical shift values of 24.6(Q), 26.2(Q), 57.0(Q), 71.6(S), 80.6(D), 97.7(D), 115.1(S) and 167.8(S) of compound 9 from [58] are in excellent agreement with the data for compound 7 from [59].



Compound 7 from [59]
CAS-RN: 221002-11-5
PUBCHEM-CID: 92,030,426

Flavonoids from *Psoralea corylifolia*

Corylin was described in [60] and identified by comparison with known literature data [61], the ^{13}C-NMR chemical shift values are summarized in the experimental section. The given data are consistent with respect to the chemical shift values with a previously published paper [62]. When using the given assignment and comparing the signal multiplicities derived from the H-count in the structure massive differences can be found (**Fig. 10**). Using the structure proposal together with the ^{13}C-NMR data and starting the structure generator thereof leads to 5,230 alternatives. The given structure is ranked at position 1 with an average deviation of 0.85ppm, two similar compounds can be found at positions 22 and 76, respectively, in the sorted hitlist (**Table 11**). From this analysis it can be concluded that the structure perfectly matches the ^{13}C-NMR data – only the signal assignment to the respective carbons is wrong. The signal assignment is obviously based on comparison with literature data [61], because no experimental details are given in [60]. The assignment given in the paper is compared with the fully automatic assignment done by the CSEARCH-Robot-Referee leading to a much better coincidence between experimental and predicted chemical shift values (**Fig. 11**).

It is interesting to note that a search for the substance identifier “corylin” using the CAS-Registry file reveals two different structures – the corylin investigated here (CAS-RN: 53947-92-5) and the alkaloid “corylin” (CAS-RN: 486-04-4) showing the confusing situation with trivial names in natural product chemistry.

Fig. 10 The differences between the ^{13}C-NMR experimental data [60] of corylin (CAS-RN: 53947-92-5 PUBCHEM-CID: 5,316,097).

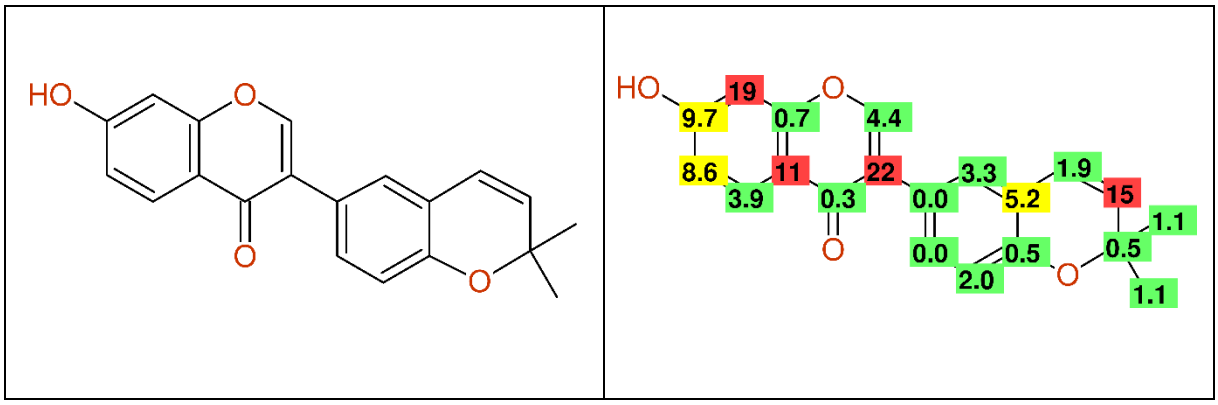


Table 11 Automatic structure revision using the ^{13}C-NMR data of corylin; the published structure is ranked at position 1 showing that 'only' an assignment error is present.

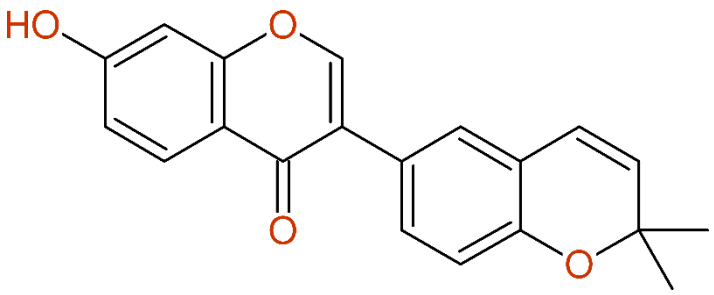
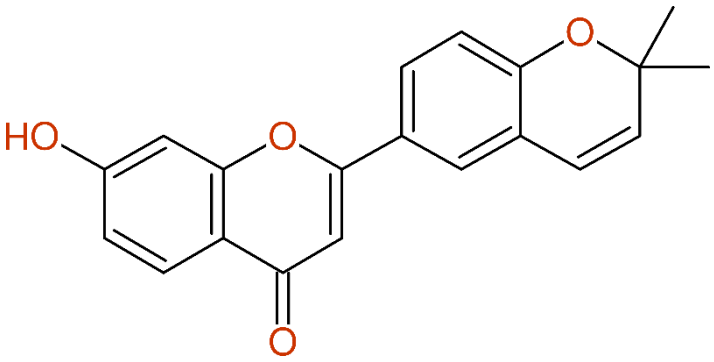
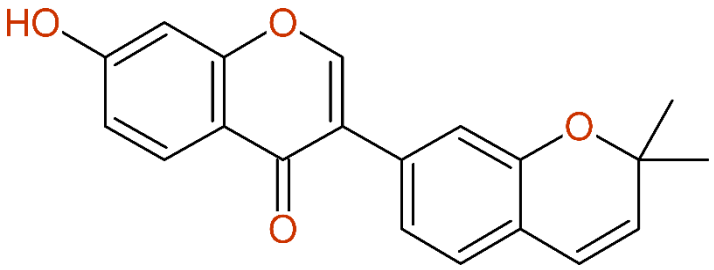
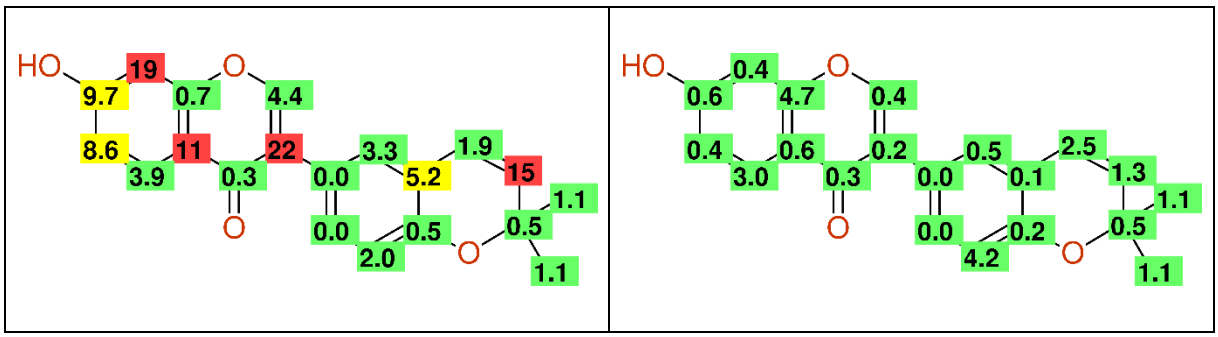

Corylin: Position 1 $\Delta\delta_C$=0.85ppm

Position 22 $\Delta\delta_C$=1.53ppm

Position 76 $\Delta\delta_C$=2.10ppm

Fig. 11 Left: δ_{exp}-δ_{pred} based on the assignment as given in [60]. Right: δ_{exp}-δ_{pred} when the signal assignment is done by the CSEARCH-Robot-Referee exclusively using the same chemical shift values.



Myriachrome from *Myriactis humilis*

Another extremely interesting example showing the necessity of reinspection of the experimental data starting at the level of the original data as proposed by Pauli [2] is the structure elucidation of Myriachrome, compound 1 from [63]. The strategy used relies on ^{1}H-, ^{13}C-NMR measurements as well as application of ^{1}H-^{1}H COSY, NOESY, HSQC and HMBC. The ring fusion is verified by a NOESY correlation between the aromatic proton at δ_H=6.36 ppm and the olefinic proton at δ_H=6.22ppm. The compatibility check of the published structure and the experimental data with the predicted ^{13}C-NMR data gives large differences (**Fig. 12**). The 1,2,3,4-tetrahydroxylated benzene ring shows singlets at 136.2, 149.3, 149.5 and 153.9ppm, whereas the prediction using the CSEARCH-Robot-Referee [26] gives 4 lines in the range between 134.5 and 140.9ppm. These values are verified by inspection of the spectra in the CAS-Registry file predicted with the ACD-software [40]. The given experimental chemical shift values are massively incompatible with a 1,2,3,4-tetrahydroxylated benzene (**Fig. 12**); this pattern points to a 1,2,3,5-tetrahydroxylated system. Starting the CSEARCH Robot-Referee with the probably wrong structure and the published ^{13}C-NMR chemical shift values a set of 1,918 structures is generated. 1,837 out of them are isomers and 1,462 have identical multiplicity information as necessary to explain the given structure proposal. Only the given structure is known in the PUBCHEM-collection. The resulting hitlist is post-processed by three partial structure searches: a) a chromene-skeleton; b) a methylendioxy-group and c) a methoxy-group must be present in the obtained structural proposals leading to a set of five reasonable alternative structures (**Table 12**). This set consists of two structures having 1,2,3,4-tetrahydroxylation at positions 146 and 372, two structures having 1,2,3,5-tetrahydroxylation (positions 87 and 187) and one structure with a 1,2,4,5-pattern at position 355 in the ranked hitlist. Only 2

structures at positions 146 and 372 are compatible with the NOESY-correlation described before. The structure at position 146 needs a signal around 60ppm for the methylgroup, the experimental line can be found at 56.0ppm pointing to a methoxy-group with at least one non-substituted ortho-position at the benzene-ring, therefore the argumentation based on 2D-NMR data leading to the structure of myriachromene ranked at position 372 seems to be straight forward and consistent. With respect to the ^{13}C-NMR chemical shift values there is a big discrepancy between experimental and predicted values which still needs further investigation. It should be mentioned that myriachromene occurs exclusively in the paper analyzed here [63], this publication has been cited by 20 other publications. The compound ranked at position 146 is also a known compound [64] again occurring exactly once in the literature and only characterized by its ^{1}H-NMR and MS-data. The same arguments based on the ^{13}C-NMR data with respect to the hydroxylation-pattern as already discussed before, are valid for this compound too.

This example shows the urgent need for a public domain data repository holding experimental data and their interpretation – this would allow the necessary re-interpretation of the original data starting from the 'Free Induction Decays'.

Fig. 12 Top: Myriachrome and experimental ^{13}C-NMR chemical shift values [63]. Bottom: Predicted ^{13}C-NMR chemical shift values [26] and differences between experiment and prediction.

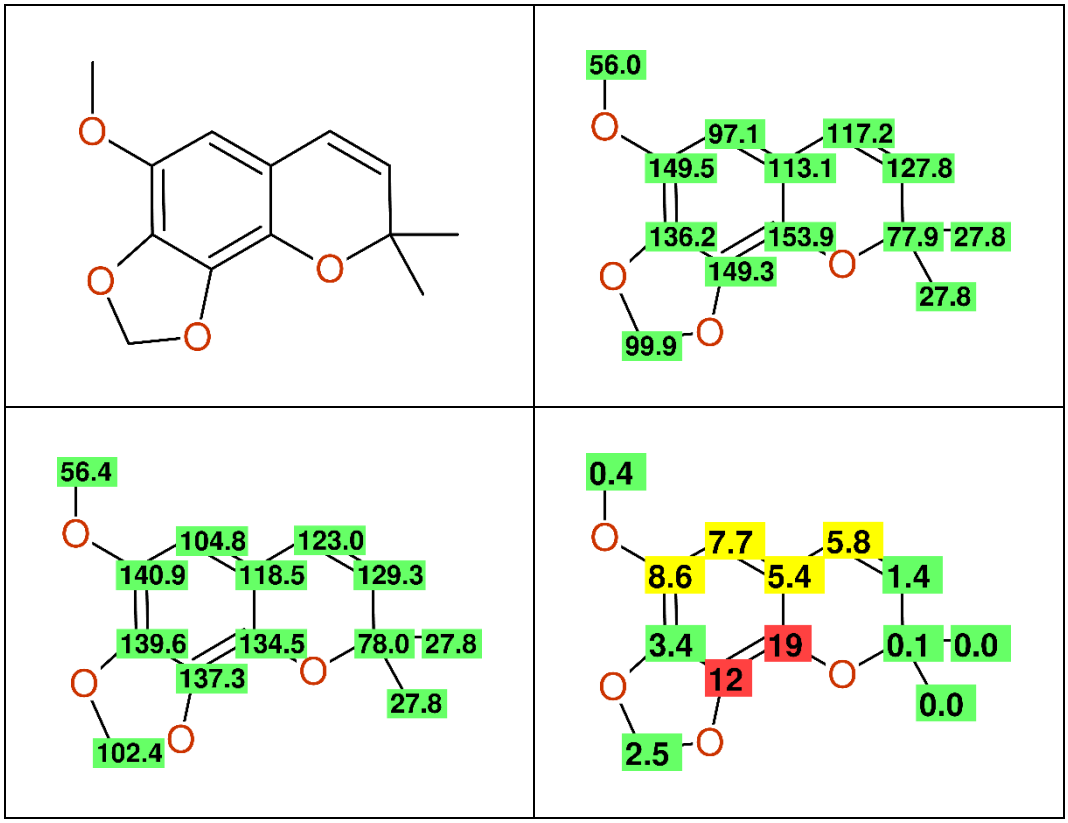
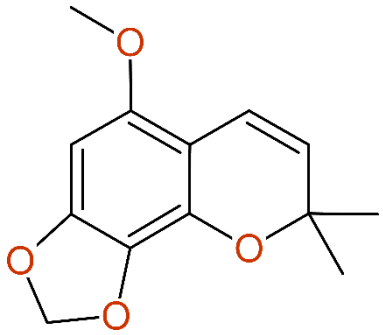
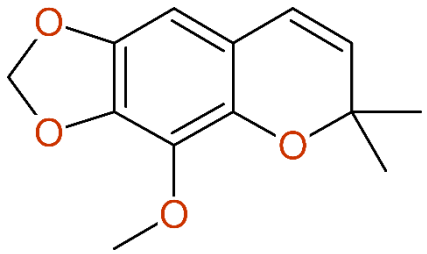
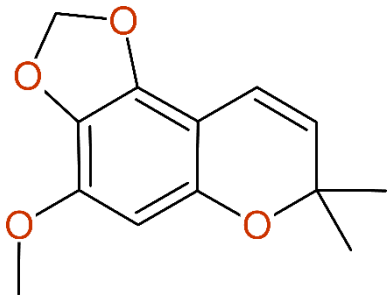
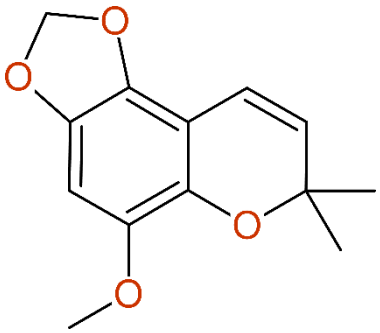
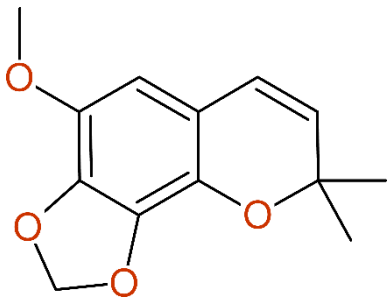


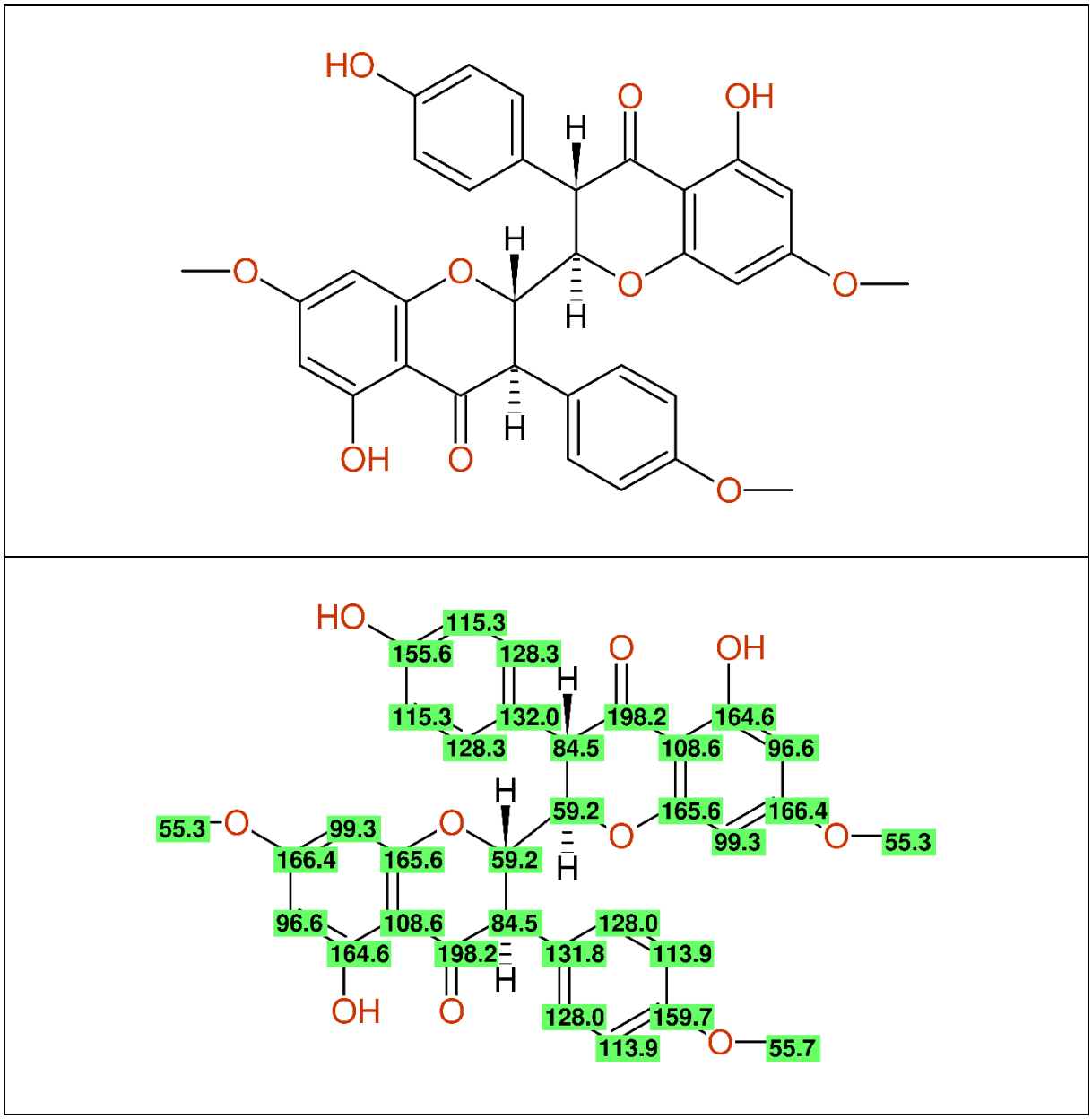
Table 12 Four alternative structures generated from the data published for myriachrome [63] – all four alternative proposals have some inconsistencies with respect to the measured 2D-data. The published structure is also inconsistent with the ^{13}C-NMR peaklist.

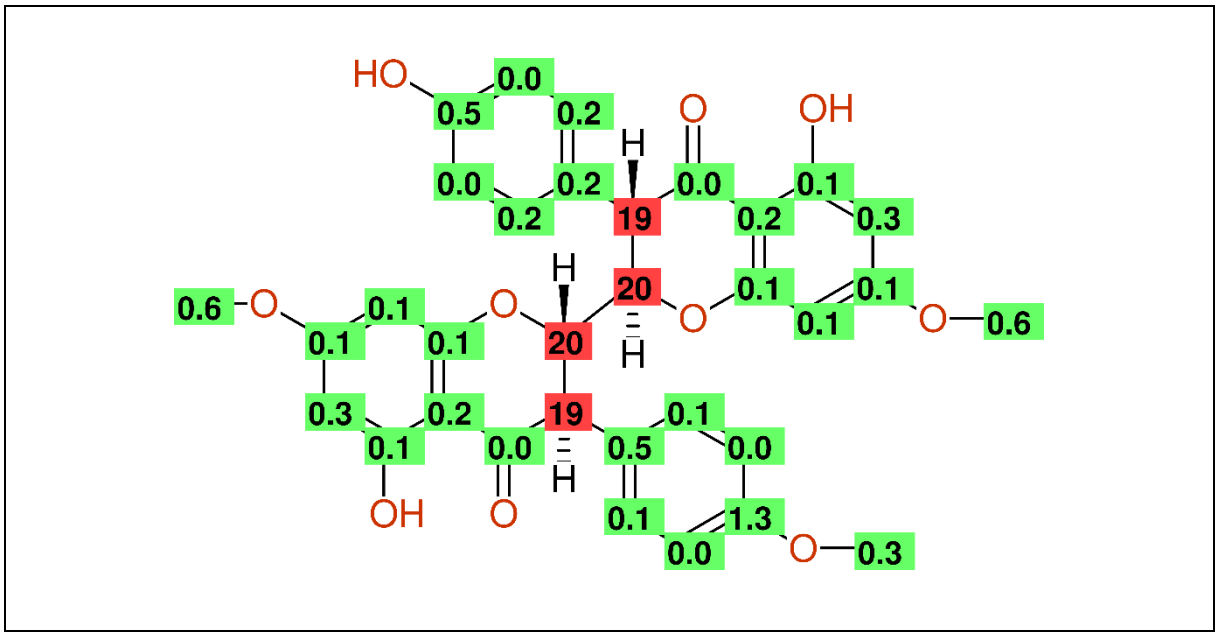
	
Position 87: $\Delta\delta_C$=2.83ppm	Position 146: $\Delta\delta_C$=3.15ppm CAS-RN: 128650-48-6
	
Position 187: $\Delta\delta_C$=3.32ppm	Position 355: $\Delta\delta_C$=3.80ppm
	Proposed structure in [63] Position 372 $\Delta\delta_C$=3.84ppm PUBCHEM-CID: 11,564,904 CAS-RN: 866560-84-1

Biflavonoids from *Ochna macrocalyx*

Hexaspermone C published as compound 2 in [65] shows massive deviations at four positions in both 4-chromanone moieties (**Fig. 13**). Detailed inspection of the chemical shift values and their assignment as given, immediately implies an assignment error and/or a type-setting error, which can be easily found. The given ^{13}C-NMR spectral data are in excellent agreement with an earlier publication [66]. The same assignment problem within the 4-chromanone system is also present in compound 1 named dehydroxyhexaspermone C.

Fig. 13 The chemical structure of hexaspermone C (CAS-RN: 155969-79-2, PUBCHEM-CID: 101,663,212); the assigned chemical shift values as given in [65] (middle) and the differences between experimental and predicted ^{13}C-NMR chemical shift values (bottom).

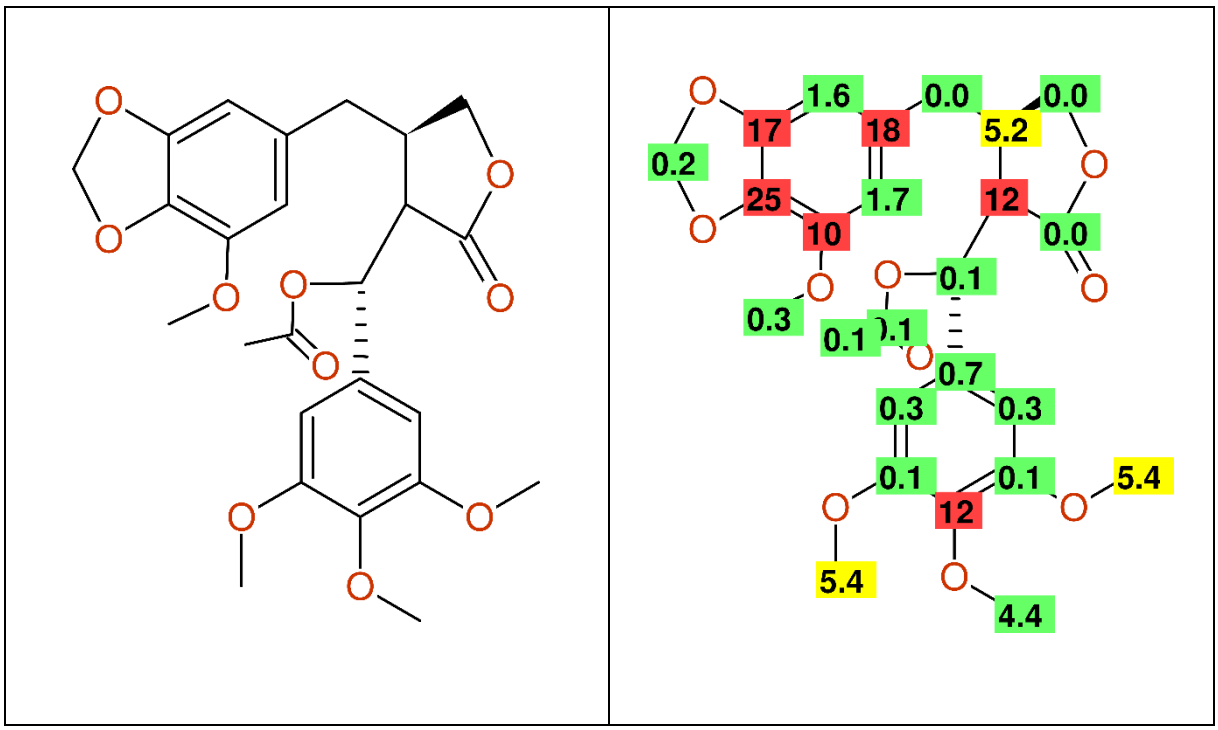



HO
0.0
0.5
0.2
O
OH
H
0.0
0.2
0.0
0.1
0.2
19
0.2
0.3
H
20
0.1
0.1
0.6
O
0.1
O
O
O
0.6
0.1
0.1
20
0.1
0.3
0.2
19
0.1
0.1
0.0
0.5
0.0
H
OH
O
0.1
1.3
O
0.3
0.0

Lignans from the Seeds of *Hernandia sonora*

Eight lignans were isolated from the seeds of *Hernandia sonora*; three out of these compounds were characterized using ^{13}C-NMR spectral data. The chemical shift values of 5-Methoxypodophyllotoxin and its acetate (compound 6 and 7 in [67]) differ by more than 11 ppm at C_5 far away from the substituted position, where the acetylation takes place. The quality of the presentation of the ^{1}H- and ^{13}C-NMR chemical shift values in table 1 in the original paper [67] suffers from many type-setting problems, e.g. according to the structural diagram there is a 4'-position available, but not shown in the table, furthermore the carbon-shift value of 149.2ppm is associated with a proton chemical shift value of 2.75ppm, which is also in error. When comparing the compounds 7 and 8 at position 4' (which is position 4 according to the table), two quite different ^{13}C-NMR chemical shift values of 135.4ppm (compound 7) and 149.2ppm (compound 8) are assigned to similar carbon positions in the well-separated spin-system of the 3,4,5-trimethoxy-benzene moiety. Compound 7 shows also a quite unusual chemical shift value for the acetate-carbonyl at 164.2ppm. Comparing the data given here for compound 7 with the chemical shift data of the identical compound 1 as published in [68] reveals differences larger than 2ppm at 10 carbon positions. Compound 8 shows a methoxy-group at 50.8ppm, the analogous carbons in compounds 6 and 7 resonate at 56.1ppm and 56.3ppm as to be expected. The structure of compound 8 from [67] named 5-methoxypodorhizol-acetate shows large deviations between the experimental ^{13}C-NMR chemical shift values and the expected ones (**Fig. 14**). These data sets published here were used (among others) to train an expert system named SISTEMAT [69] intended to predict the probability of the occurrence of certain lignan-type skeletons in natural products from their ^{1}H- and ^{13}C-NMR spectra.

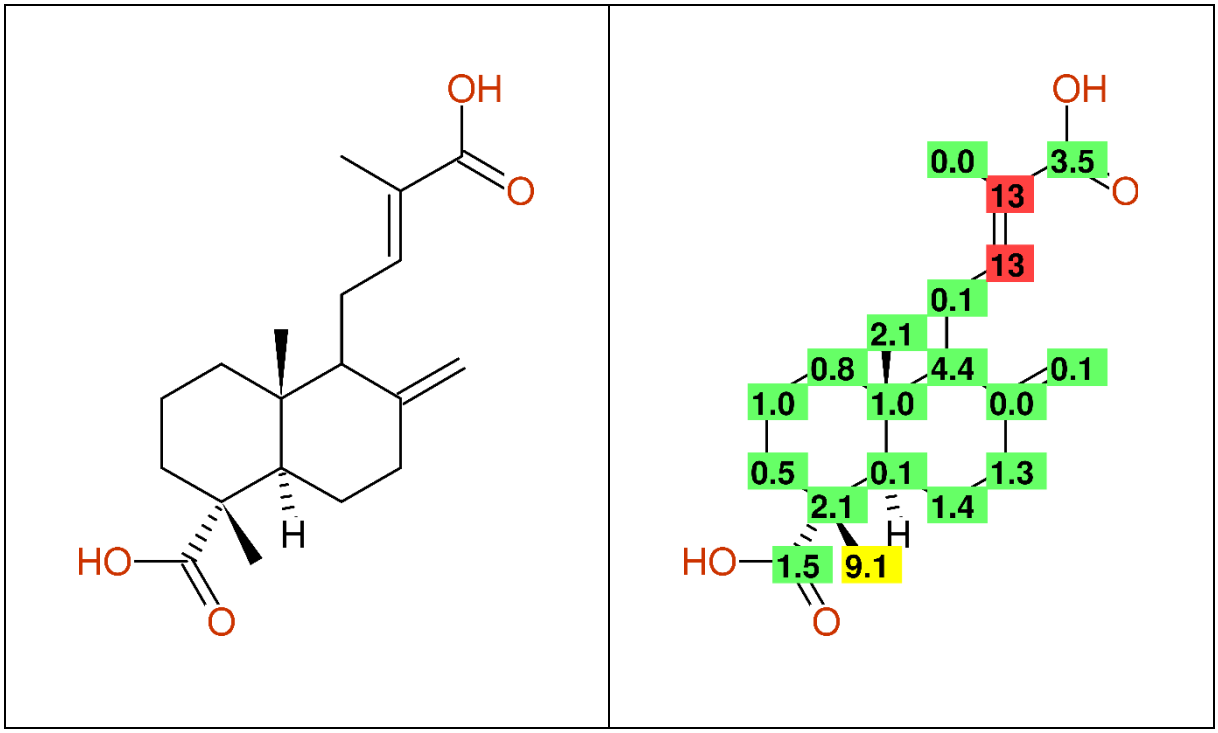
Fig. 14 Compound 8 from [67] (CAS-RN: 225669-53-4, PUBCHEM-CID: 1,009,361) showing large deviations between experimental and predicted ^{13}C-NMR chemical shift values.



A diterpenoid from *Cunninghamia lanceolata*

Compound 3 in [70] was elucidated as 15-nor-14-oxo-8(17),12-labdandiene-14,18-dioic acid using ^{1}H- and ^{13}C-DEPT measurements. The stereochemistry of the sidechain was defined as trans-configured based on a missing NOE between H-12 and the Me-group at position 16. This finding is shown in the figure 3 in the publication, whereas the structure holding the numbering for the signal assignment has a missing carboxylgroup and a cis-configured sidechain – this inconsistent drawing of the molecular structure leads to the effect, that compound 3 is not associated with this publication when searching the CAS-Registry file. Ignoring this confusing information and focusing on the ^{13}C-NMR data shows a trivial assignment error at positions C_{12} (CH) and C_{13} (Q_{quat}), which is quite surprising because of the DEPT-measurement. This assignment error is verified by another independent measurement published in [71] performed on a diastereomer having the inverted configuration at C_4 (CAS-RN: 477878-45-8). The ^{13}C-NMR data of both diastereomers significantly differ in the sidechain because of the assignment error and with respect to the pattern of the signals of the 3 methyl groups because of the different 1,3-diaxial interactions for Me-19 and Me-18, respectively. The difference between the experimental ^{13}C-NMR chemical shift values and the predicted values clearly points to an assignment error of the resonances in the sidechain (**Fig. 15**). The deviation at the methyl group at position 18 can be attributed to the presence of both diastereomers (among other similar compounds) as reference material in the database creating therefore a weighted average chemical shift value from reference data having either R- or S-configuration at C_4.

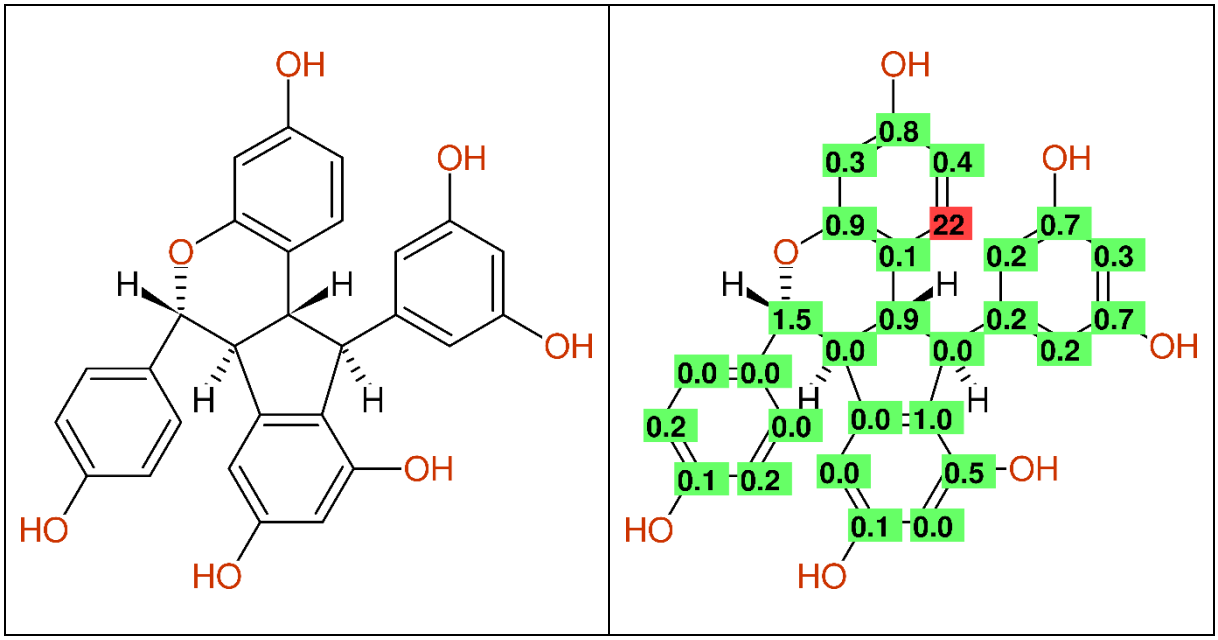
Fig. 15 Compound 3 from [70] labeled with the difference between experimental and predicted ^{13}C-NMR chemical shift values showing massive deviations in the sidechain. This stereoisomer is not contained in the CAS-Registry file (Diastereomer at C_4: CAS-RN: 477878-45-8 PUBCHEM-CID: 22,297,541).



Gnetuhainin S from *Gnetum hainanense*

The published ^{13}C-NMR spectral data and the structure of Gnetuhainin S derived thereof (compound 2 in [72]) are in perfect agreement except at one position in a dihydroxylated benzene ring system (**Fig. 16**); the structure of Gnetuhainin S was also proven by XRAY-analysis. From the data given in the publication it can be concluded that there is simply a transmission error, because the line at 158.49ppm is used twice – hydroxylation at this position is not a reasonable alternative, because the proton-NMR data show all characteristics of an ABC-spin system. This example shows that the process of creating publishable data sets has to be done in a more straightforward and mainly automatic way to avoid this type of transmission errors.

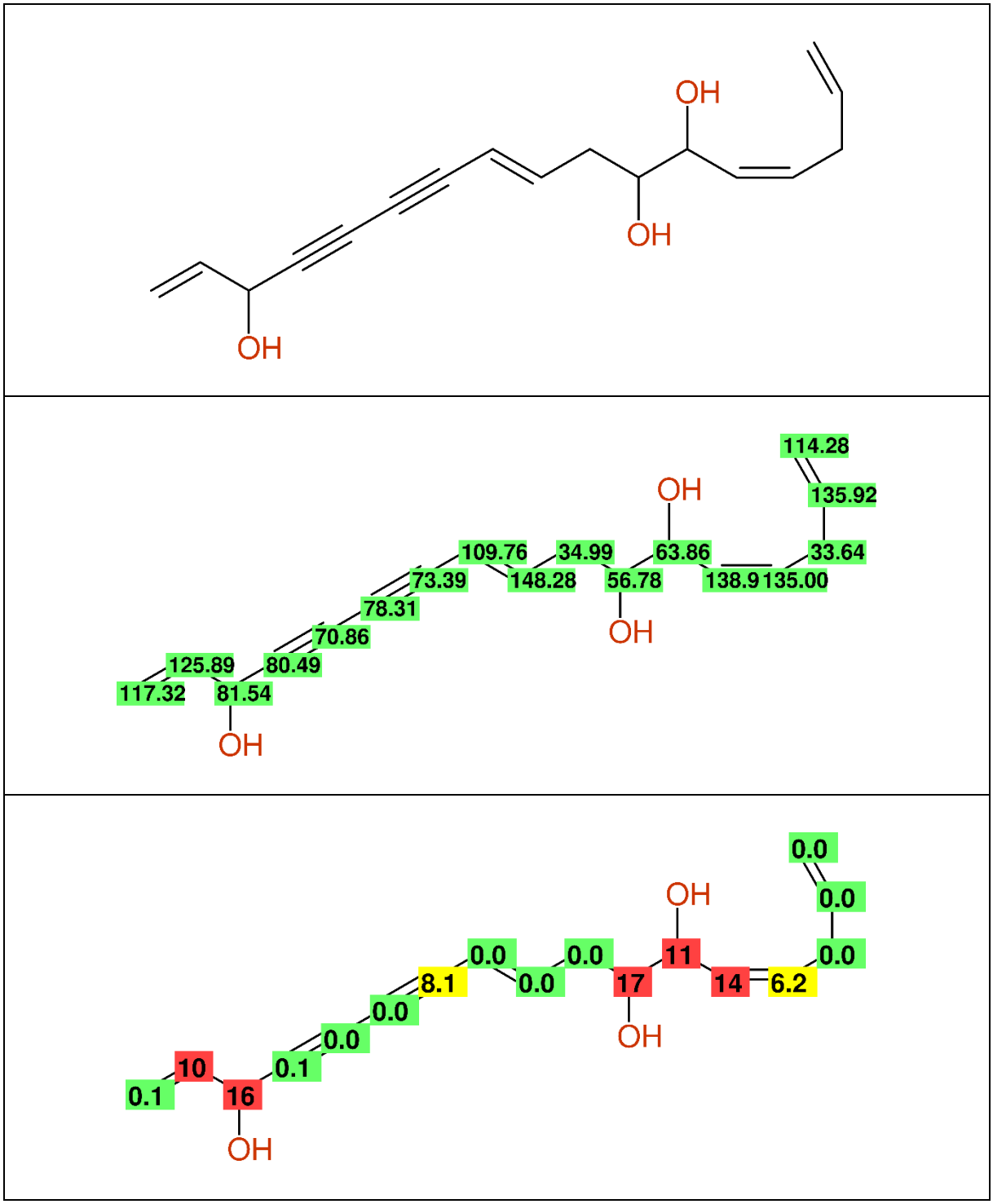
Fig. 16 The structure of gnetuhainin S (CAS-RN: 361444-40-8, PUBCHEM-CID: 11,016,150) from [72] showing the differences between experimental and predicted ^{13}C-NMR chemical shift values.



Polyacetylenes from *Artemisia eriopoda*

Two new polyacetylenes were isolated from *Artemisia eriopoda* and characterized by spectral methods [73]. The measured ^{13}C-NMR spectral data significantly differ from the predicted ones (**Fig. 17**). Two lines assigned to the sp^2-carbons seem to be exchanged; the assignment of the hydroxylated CH-groups is also different from the assignments already published for similar compounds [74,75,76,77]. The chemical shift value of 56.78ppm is also not in full agreement with the given structure - this value might be compatible with an epoxide; furthermore the line at 81.54ppm does not fit to the CH connected to the acetylene unit because of the anisotropy of the triple bond. Comparison of compound 1 with the similar polyacetylene described as compound 2 in the publication is impossible because there are no ^{13}C-NMR data given in table 1 of [73]. The experimental part describing both compounds holds appropriate hints to the table, but the ^{13}C-NMR data of the second polyacetylene are missing there. The ^{1}H-NMR data are also not very helpful, because of severe signal overlap despite the measurements were performed at 400MHz. This example reflects the typical case proving the necessity of depositing the experimental data in a public repository for later inspection, even when this later evaluation leads to the result of repeating the structural characterization starting from the substance itself. The CSEARCH-Robot-Referee generates a few isomeric trihydroxylated polyacetylenes, but the lack of data as well as the same arguments discussed above prohibit a more reasonable structure proposal for this example. The only conclusion from this evaluation is, that the published proposal is incompatible with the given data with a high degree of confidence.

Fig. 17 ^{13}C-NMR data of compound 1 from [73] (CAS-RN: 209066-87-5, PUBCHEM-CID: 101,944,746) together with the differences between experiment and prediction.



Phenylpropanes from *Acorus tatarinowii*

The tabulated ^{13}C-NMR chemical shift data [78] of erythro-1',2'-dihydroxyasarone and threo-1',2'-dihydroxyasarone are compatible with the given structures (**Fig. 18**). The 1',2'-epoxypropane derivative named cis-epoxyasarone has massive deviations at both epoxy-carbons. The CSEARCH-Robot-Referee generated 559 alternative structure proposals. At position 15 in the ranked hitlist an alternative structure was found having an average deviation of 3.06ppm, which might be compatible with the two signals around 79ppm and having the identical substitution pattern at the aromatic ring system (**Fig. 19**). The need of more experimental data in order to clarify the situation is obvious - the only conclusion withdrawn from this analysis is, that an epoxy-moiety can be clearly excluded by the published ^{13}C-NMR chemical shift data.

Fig. 18 Compounds 11 and 12 from [78] showing small differences between experiment and prediction.

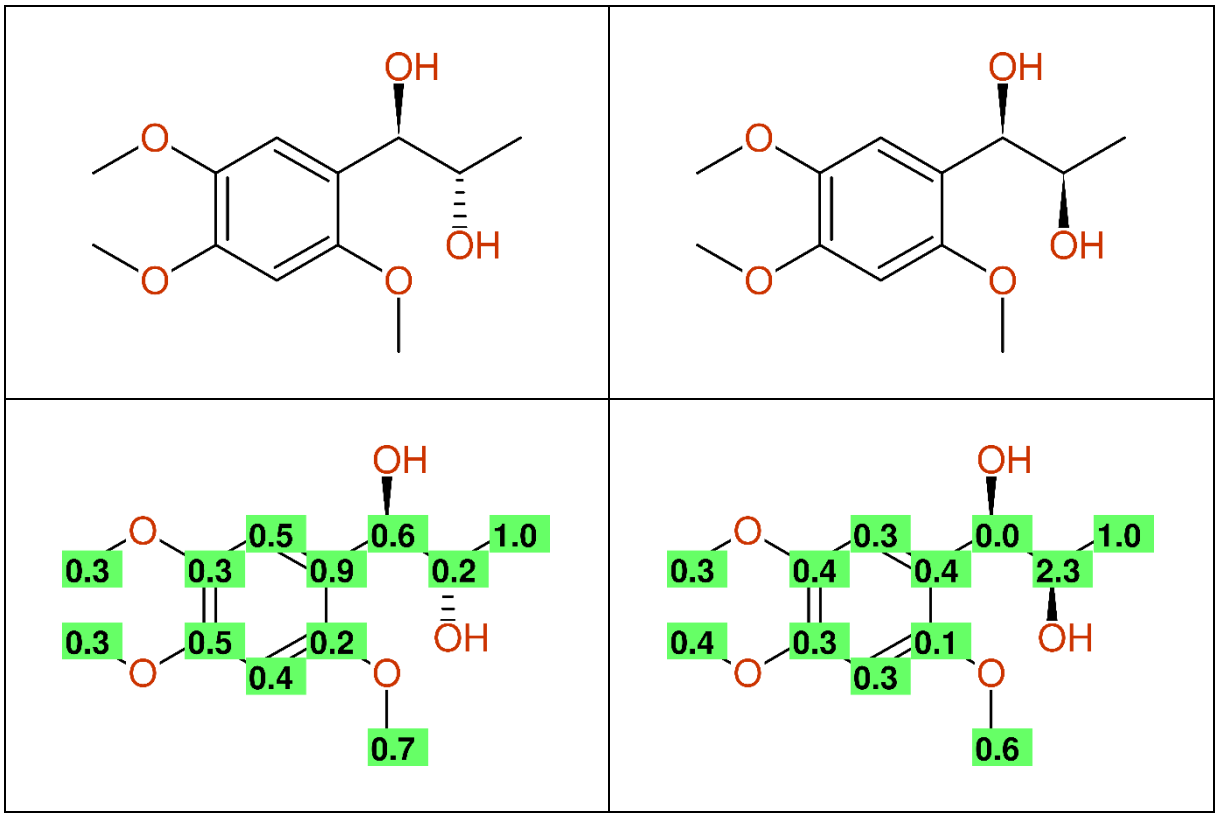
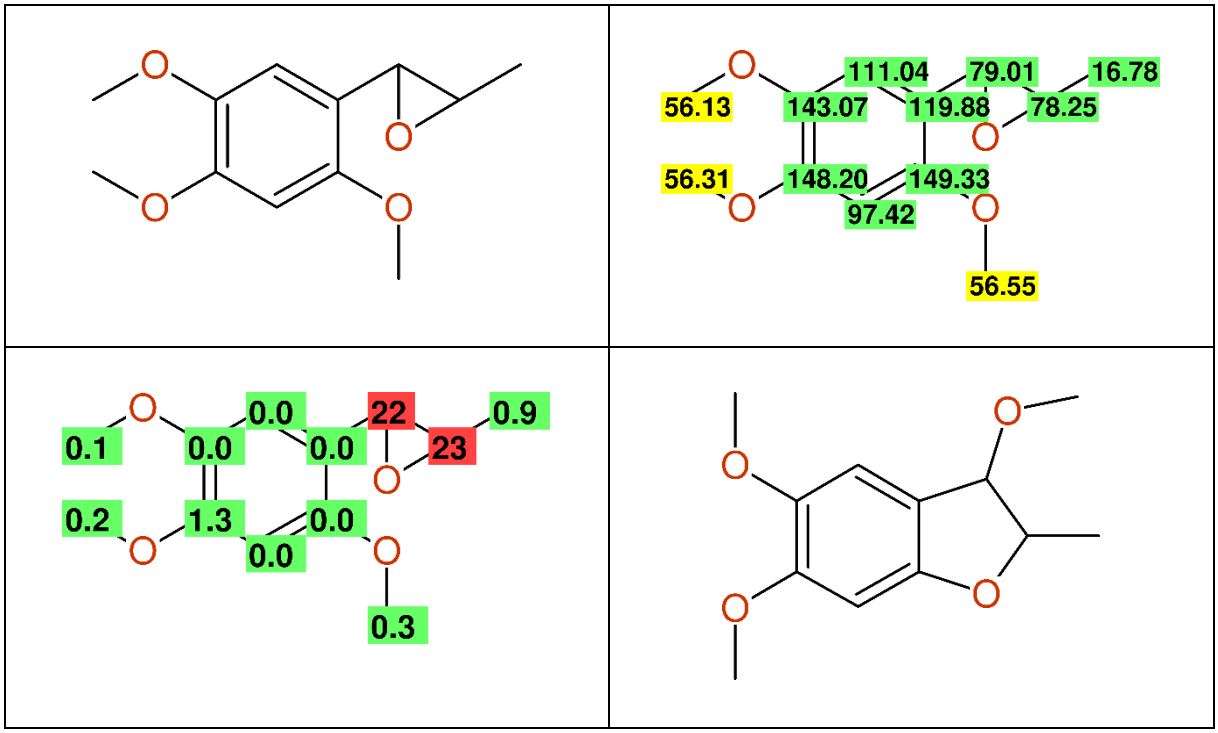


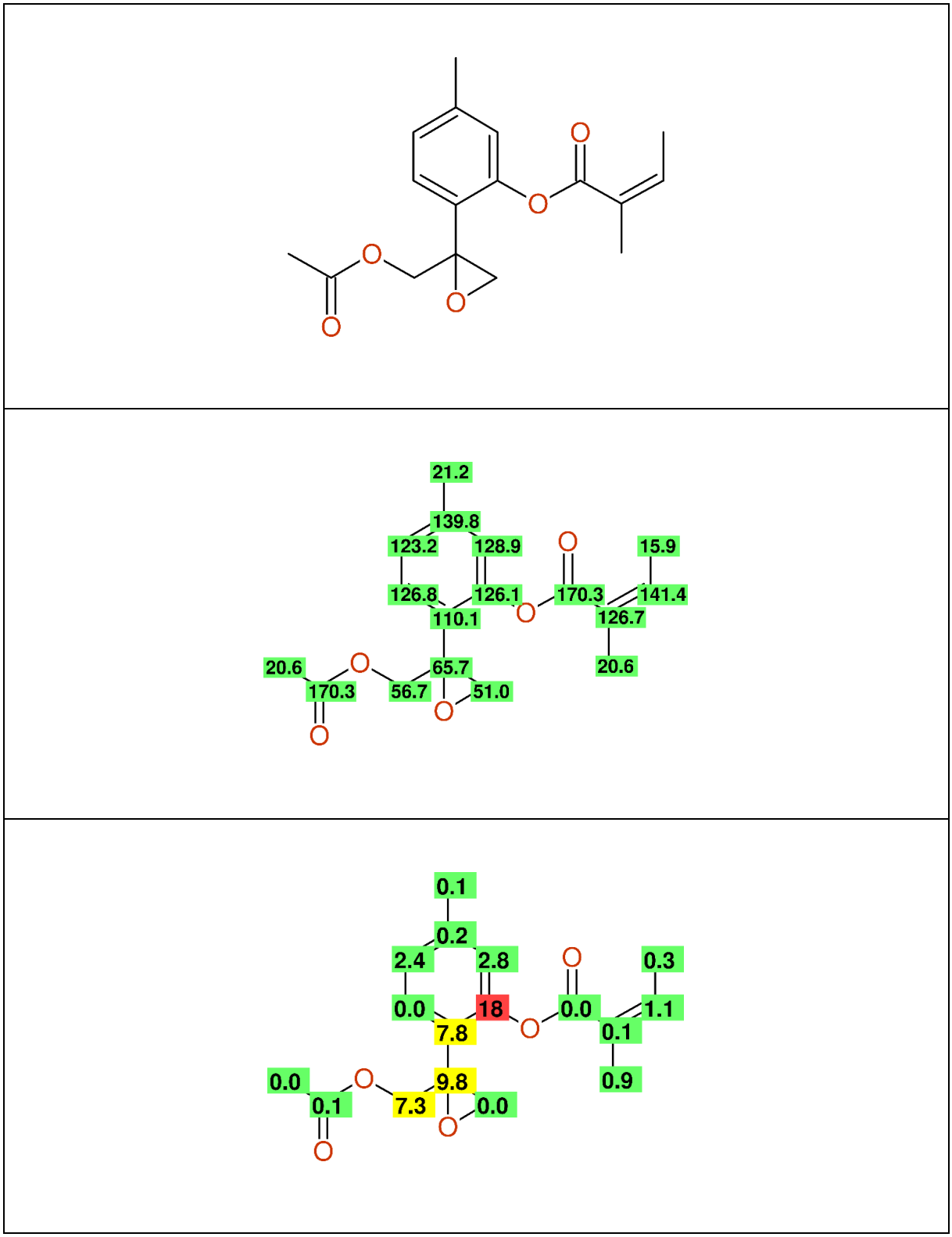
Fig. 19 ^{13}C-NMR chemical shift values of cis-epoxyasarone (CAS-RN: 321365-66-6; compound 10 in [78]) located at position 93 in the ranked hitlist having $\Delta\delta_C$=4.45ppm showing large deviations at the oxirane-moiety (bottom/left). A reasonable alternative ($\Delta\delta_C$=3.06ppm; position 15) – still unknown with CAS and PUBCHEM (bottom/right).



Thymol Derivative from *Eupatorium stoechadosmum*

Compound 2 from [79] shows significant deviations between experimental and predicted ^{13}C-NMR chemical shift values around the epoxid ring which can attributed to a signal misassignment of the lines at 65.70 and 56.72 ppm, respectively. The chemical shift value of 126.10ppm assigned to a hydroxylated carbon in a benzene ring is also far away from the expected value. In the paper no information about the measuring conditions are given nor the experimental techniques used for multiplicity determination are available. The possible misassignment as well as the suspected typing error could be easily corrected by inspection of the original experimental data. The proposed structure and the published chemical shift values are incompatible at four positions (**Fig. 20**).

Fig. 20 Compound 2 (CAS-RN: 76475-33-7; PUBCHEM-CID: 56,680,552) from [79] with experimental ^{13}C-NMR chemical shift data and deviation between predicted and measured values.

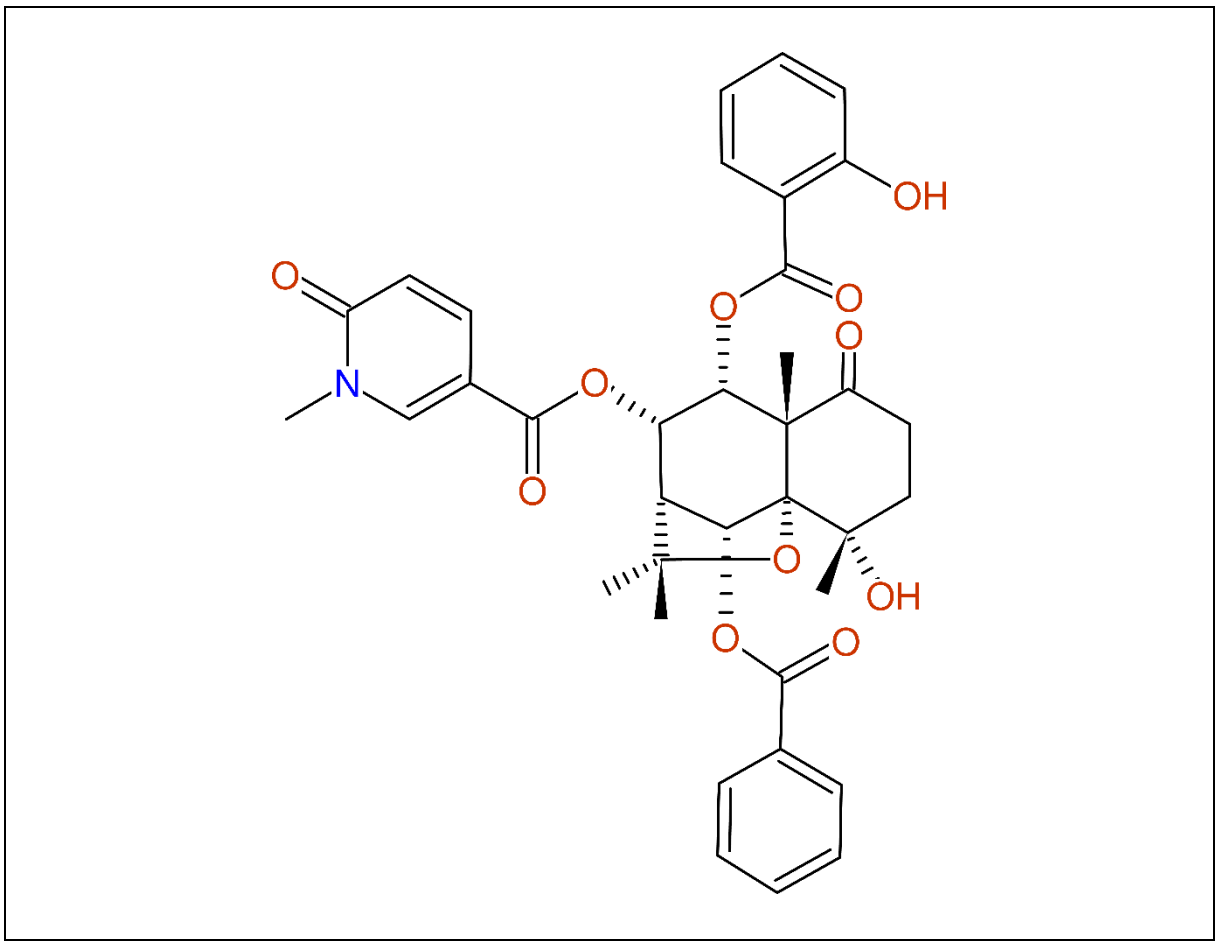


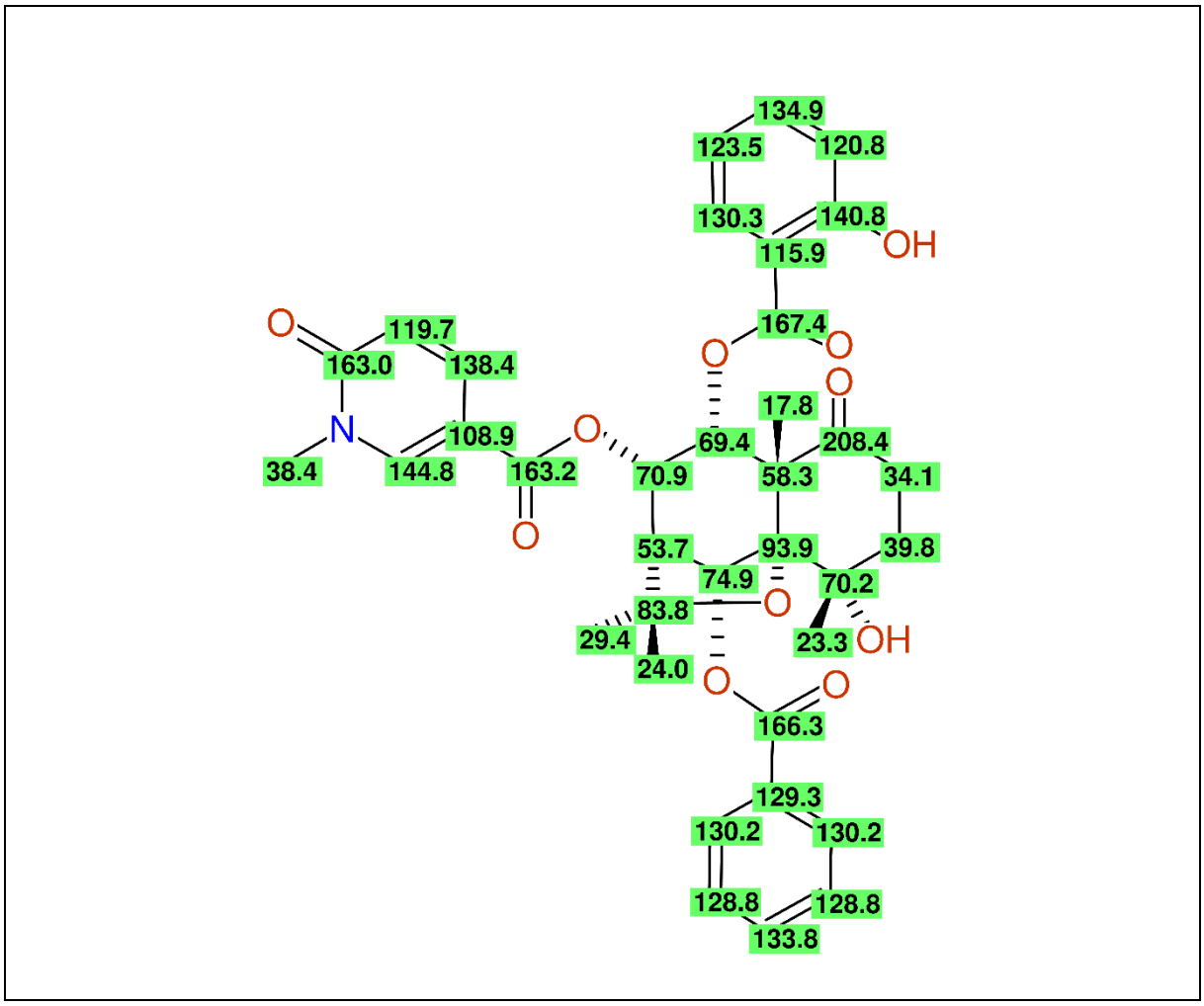
Typos and transmission errors

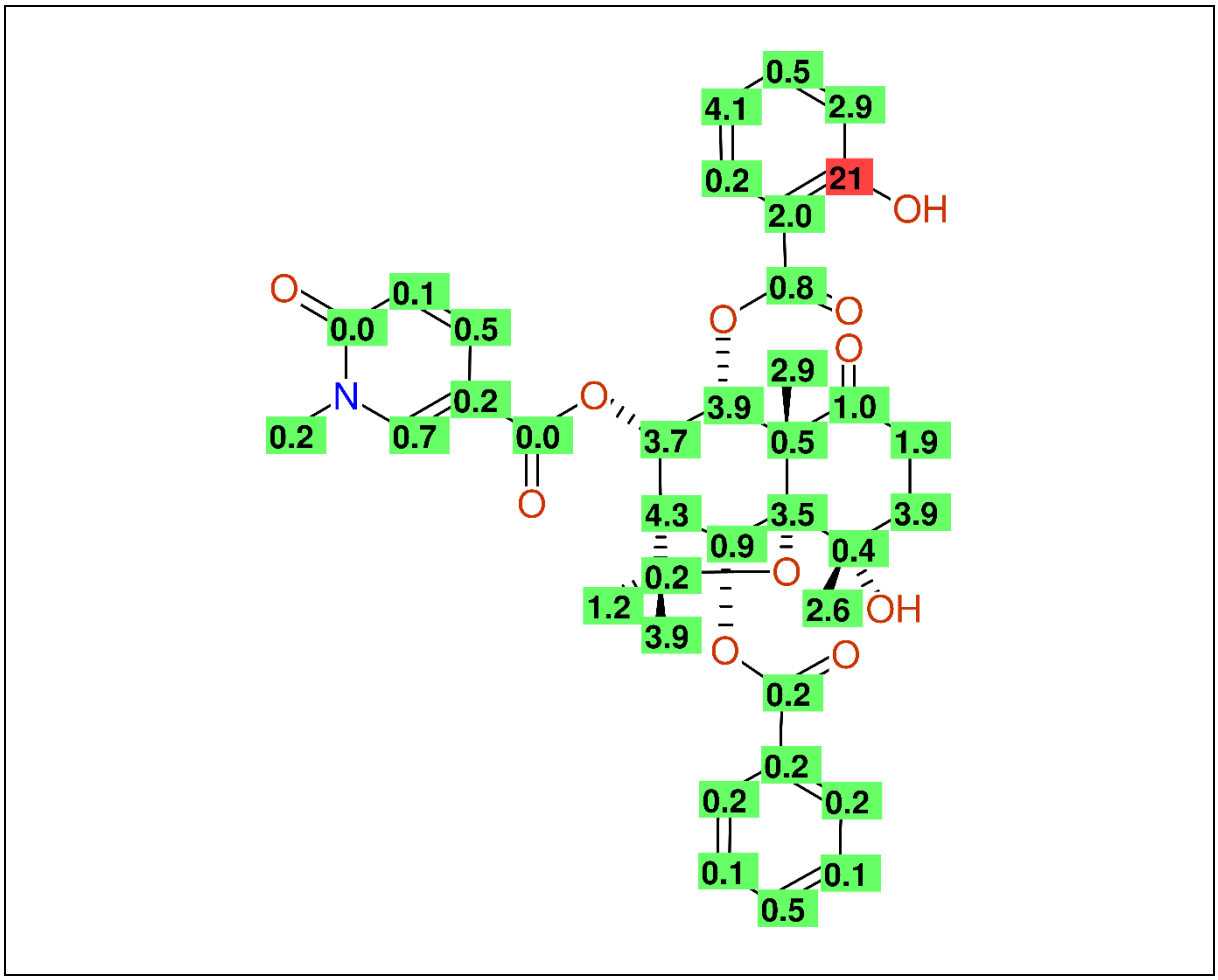
Many steps in the process of publishing the structures of natural products and their associated data have to be done manually leading to a good chance of introducing errors at every step. As long as the original experimental data – best would be the FIDs – are available, re-processing them will clarify this inconsistency, furthermore massive application of spectrum prediction tools together with automatic peer-reviewing of the interpretation of experimental data would help to avoid this type of error.

Reissantin H, a new agarofuran sesquiterpene from *Reissantia buchananii* [80] shows excellent agreement between experimental and predicted ^{13}C-NMR chemical shift values except at the hydroxylated carbon in the 2-hydroxybenzoic acid fragment. The prediction proposes a chemical shift value for this carbon around 160ppm – a typing error, 140.8ppm instead 160.8ppm is therefore assumed (**Fig. 21**).

Fig. 21 Reissantin H (compound 3 from [80], CAS-RN: 910651-22-8) showing a large deviation at C_2 in the 2-hydroxy-benzoate moiety – 160.8ppm instead of 140.8ppm is assumed.




134.9
123.5
120.8
130.3
140.8
OH
115.9
167.4
O
O
O
17.8
O
119.7
163.0
138.4
69.4
208.4
N
108.9
O
38.4
144.8
163.2
70.9
58.3
34.1
O
53.7
93.9
39.8
74.9
70.2
O
83.8
29.4
23.3
OH
24.0
O
O
166.3
129.3
130.2
130.2
128.8
128.8
133.8


0.5
4.1
2.9
0.2
21
OH
2.0
0.8
O
O
O
O
2.9
3.9
1.0
0.5
1.9
3.7
4.3
3.5
3.9
0.9
0.4
0.2
O
1.2
2.6
OH
3.9
O
O
0.2
0.2
0.2
0.2
0.1
0.1
0.5
O
0.1
0.0
0.5
N
0.2
0.2
0.7
0.0
O
O

A similar problem as described for Reissanthin H can be found in [81] for the three 8-O-methylingol-esters described therein. The methoxy-group at position 8 has chemical shift values at 80.32ppm (compound 5), 80.43ppm (compound 6), and 80.34ppm (compound 7) assigned. The spectrum prediction expects values around 56ppm for this functional group as usual. The signal assignment of the very similar esters is done by combined application of ^{1}H-, ^{13}C-, COSY-45, HOHAHA, APT, DEPT, HMQC and NOESY-measurements showing nearly identical chemical shift values, however this trivial error has not been detected, neither by the authors nor during the peer-reviewing process.

An additional problem occurs in all six compounds, which have been isolated for the very first time (compound 1 and 3-7). The assignment of the ^{13}C-NMR signals of the methyl-group in position 4' in the 2-methylbutanoyl-moiety is also wrong; it seems to be exchanged with a methyl group contained in the two acetate units present in the structure. The numbering of the CH_3-groups is also correct as to be seen from the multiplicity information provided in the published table holding the ^{1}H-NMR spectral data [81]. The analogous problem as with compound 7 (**Fig. 22**) can also be found in compounds 1 and 3-6.

Fig. 22 Deviation between experimental and predicted ^{13}C-NMR chemical shift values of compound 7 (CAS-RN: 928298-80-0) from [81] showing the typing error at the -OCH_3 group, the missing signals for the acetate-residues and the misassignment in the 2-methylbutanoyl-moiety.

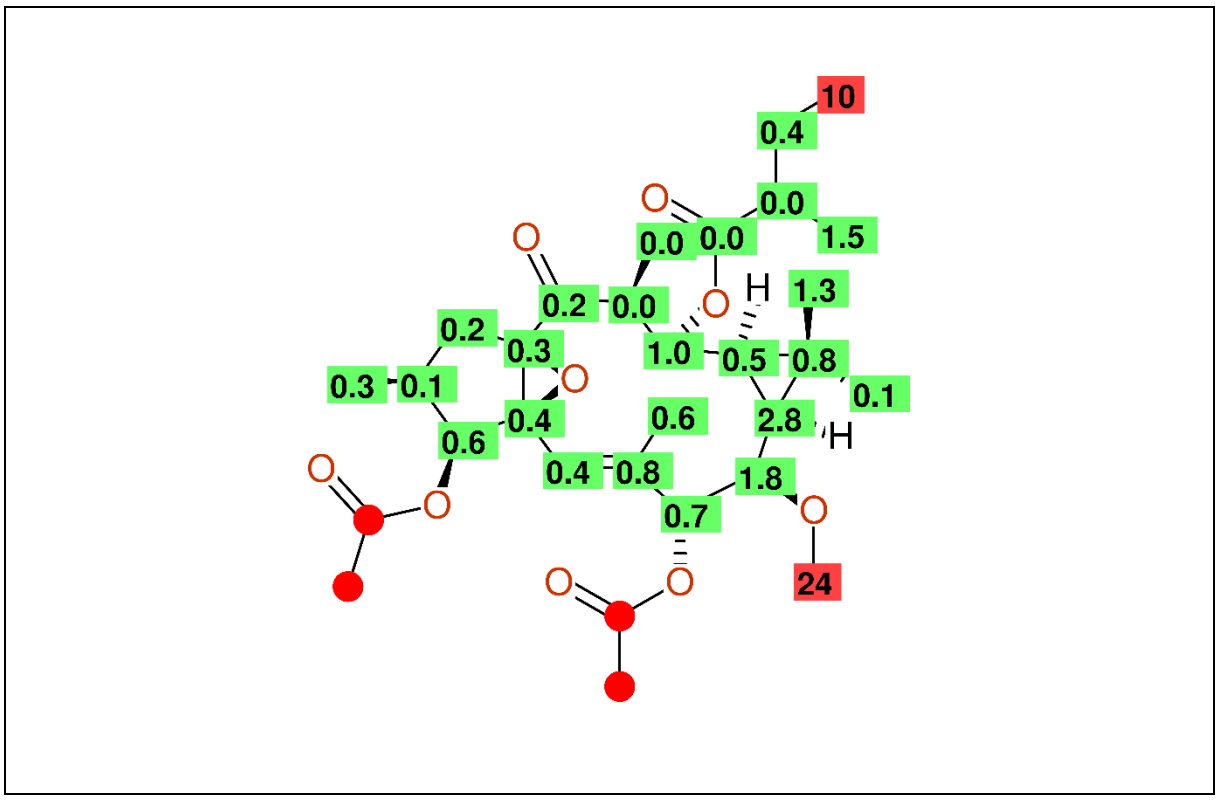
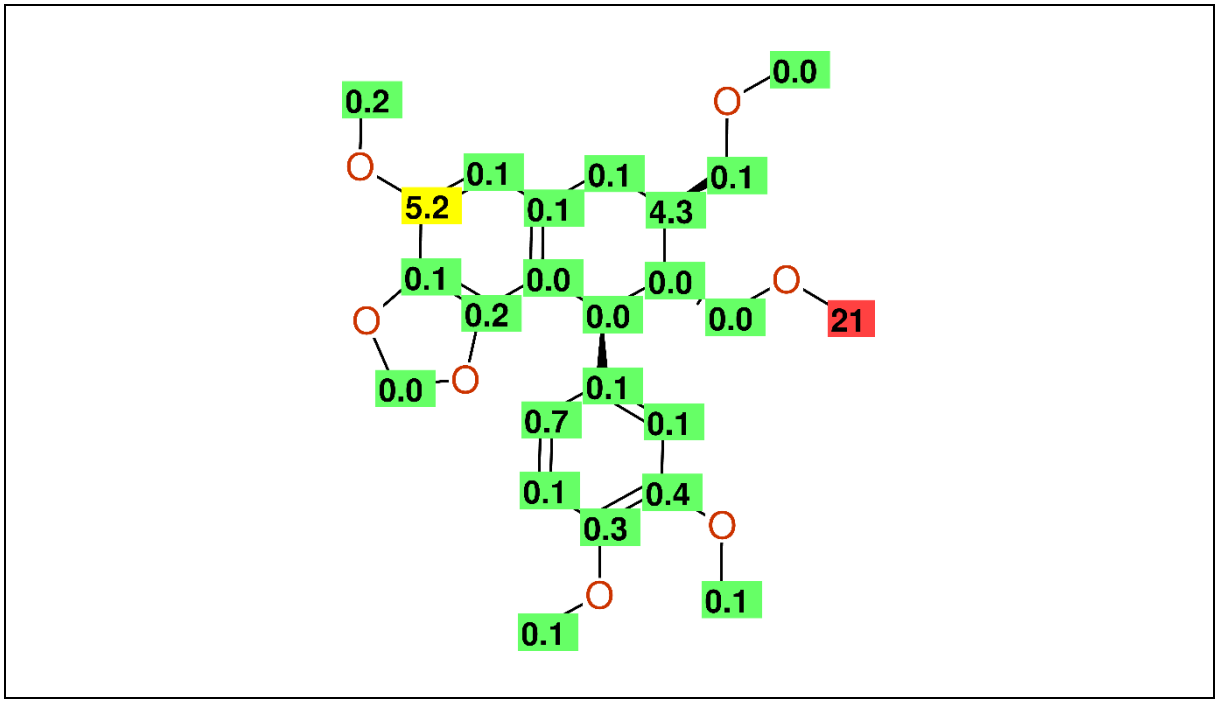
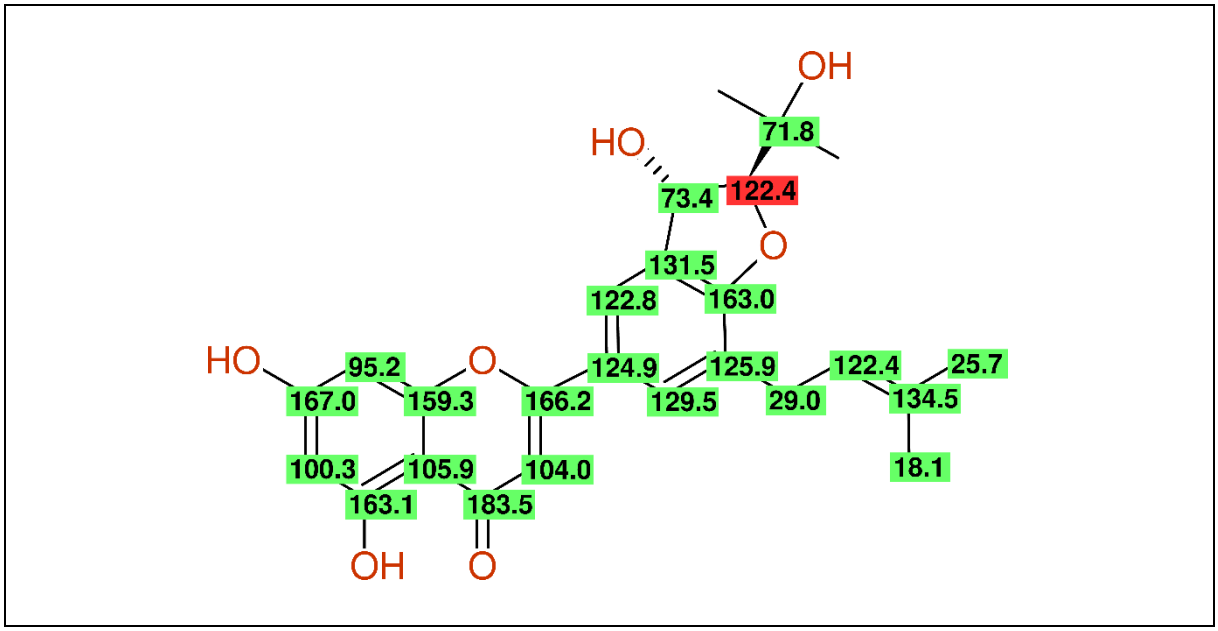


Fig. 23 One methoxy-group of hypophyllanthin (compound 2 in [82], CAS-RN: 33676-00-5; PUBCHEM-CID: 92,466,660) is given at 38.16ppm, the prediction gives a value around 58 ppm, therefore a typing error is assumed.



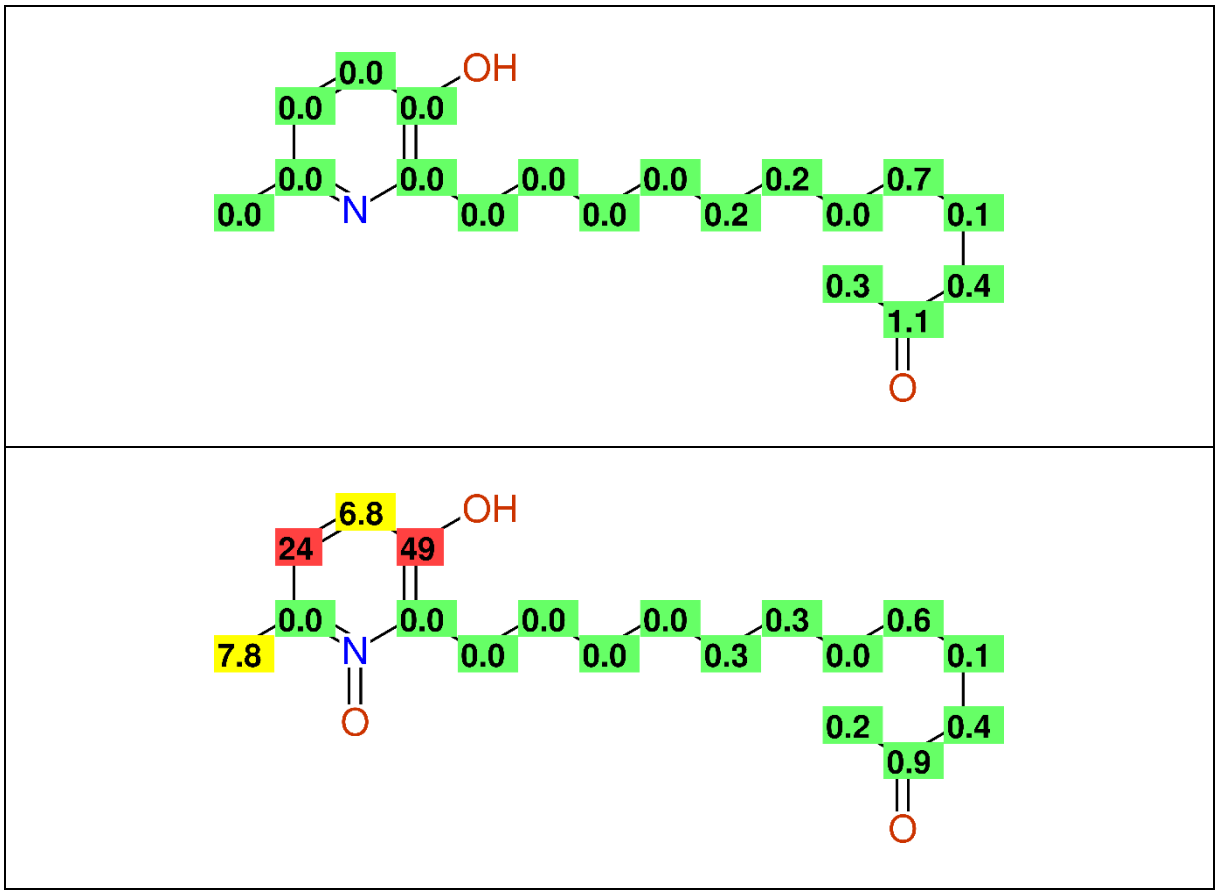
Breviflavone B, published in [83], is another example for a typing error in the given assignment table. The carbon at 95.2ppm is directly attached to a proton resonating at 6.43ppm having a meta-coupling of 2.1Hz to another proton at 6.20ppm connected to a carbon at 100.3ppm. This group of lines is attributed to the positions C_6 and C_8 of a 5,7-dihydroxylated benzopyran-4-one fragment. The benzofuran-moiety needs two more signals between 70 and 75ppm as well as one signal between 90 and 100 ppm in order to verify the necessary partial structures in breviflavone B (**Fig. 24**). As already discussed before the line around 95ppm is the most important one to prove the existence of the benzofuran-unit, which is missing here. Instead the line at 122.4ppm assigned to the CH-group in the sidechain is used twice (**Fig. 24**). In the publication the carbon resonating at 122.4ppm is connected to another carbon at 73.4ppm via their directly bond protons showing a $^3J_{HH}$-coupling in between.

Fig. 24 Structure of Breviflavon B (CAS-RN: 856900-06-6, PUBCHEM-CID: 9,845,980) together with the experimental ^{13}C-NMR chemical shift values from [83].



Three new alkaloids have been isolated from the flowers of *Senna spectabilis* [84] and structurally characterized by 1D- and 2D-NMR techniques. Compounds 4 and 5 in this paper differ by N-oxidation in the pyridine moiety. Whereas compound 4 – a 5-hydroxy-2-methyl-6-(11'-oxododecyl)-pyridine – gives the expected ^{13}C-NMR spectral pattern, shows the corresponding N-oxide significant deviations between predicted and measured ^{13}C-NMR data (**Fig. 25**). Interestingly the overall spectral pattern of this compound seems to be correct pointing to an assignment error; when going into the details it looks like a simple inconsistency between the numbering given for the structure and the numbering used in the table. The carbons in both structures are numbered clockwise for the pyridine ring, whereas the chemical shift values seem to be given counterclockwise in the table for this fragment for compound 5. This situation can be easily clarified with the help of the information on experimentally determined signal multiplicity, which is unfortunately missing in this paper.

Fig. 25 High probability of inconsistent numbering of the carbons in the structure (compound 5 in [84]) resulting in an assignment error.



Flavones from *Artemisia giraldii*

Two new flavones – compound 2 and 3 in [85] were isolated from *Artemisia giraldii* and their structures were identified using spectroscopic methods. According to the paper both structures and the signal assignments were supported by ^{1}H-^{13}C-COSY and HMBC. The large deviations between the experimental and the predicted ^{13}C-NMR chemical shift values point to wrong structure proposals and/or wrong signal assignments. The problem with compound 2 seems to be a combination of a massive assignment error based on misinterpretation of the HMBC-spectrum for the 3,5-dimethoxy-4-hydroxybenzene fragment and an error in the substitution pattern of the chromone system (**Fig. 26**).

When starting the structure generator of the CSEARCH-Robot-Referee with the data given for compound 2 the best already known structure is tricin (CAS-RN: 520-32-1). The best-case scenario is, that simply the spectral data of tricin – compound 1 in the paper – were associated by mistake with structure 2. The situation with compound 3 from [85] is even more complicated, because the published structure as defined in this paper is not associated with the literature citation in the CAS-registry file. Using the spectral data of compound 3 as query for a spectral similarity search a high coincidence with jaceosidin [86] can be detected, furthermore a similarity with different quercetin-derivatives is found. The final conclusion from this analysis of the data of compound 3 is that the structure proposal and the ^{13}C-NMR spectral data are incompatible or at least the signals are wrongly assigned. The missing experimental data efficiently prohibit any further clarification.

The scope of the problem can be easily shown when analyzing the underlying data available for tricin, which is used as reference throughout this analysis.

Within the CSEARCH NMR-database tricin can be found in seven literature citations. The best [87] and the worst [88] coincidence between experimental and predicted ^{13}C-NMR chemical shift values show, that the same structure is elucidated from quite different sets of ^{13}C-NMR spectral data (**Fig. 27**).

Fig. 26 Compounds 2 (left) and 3 (right) from [85] labeled with their ^{13}C-NMR chemical shift values and the deviation between experimental and predicted values.

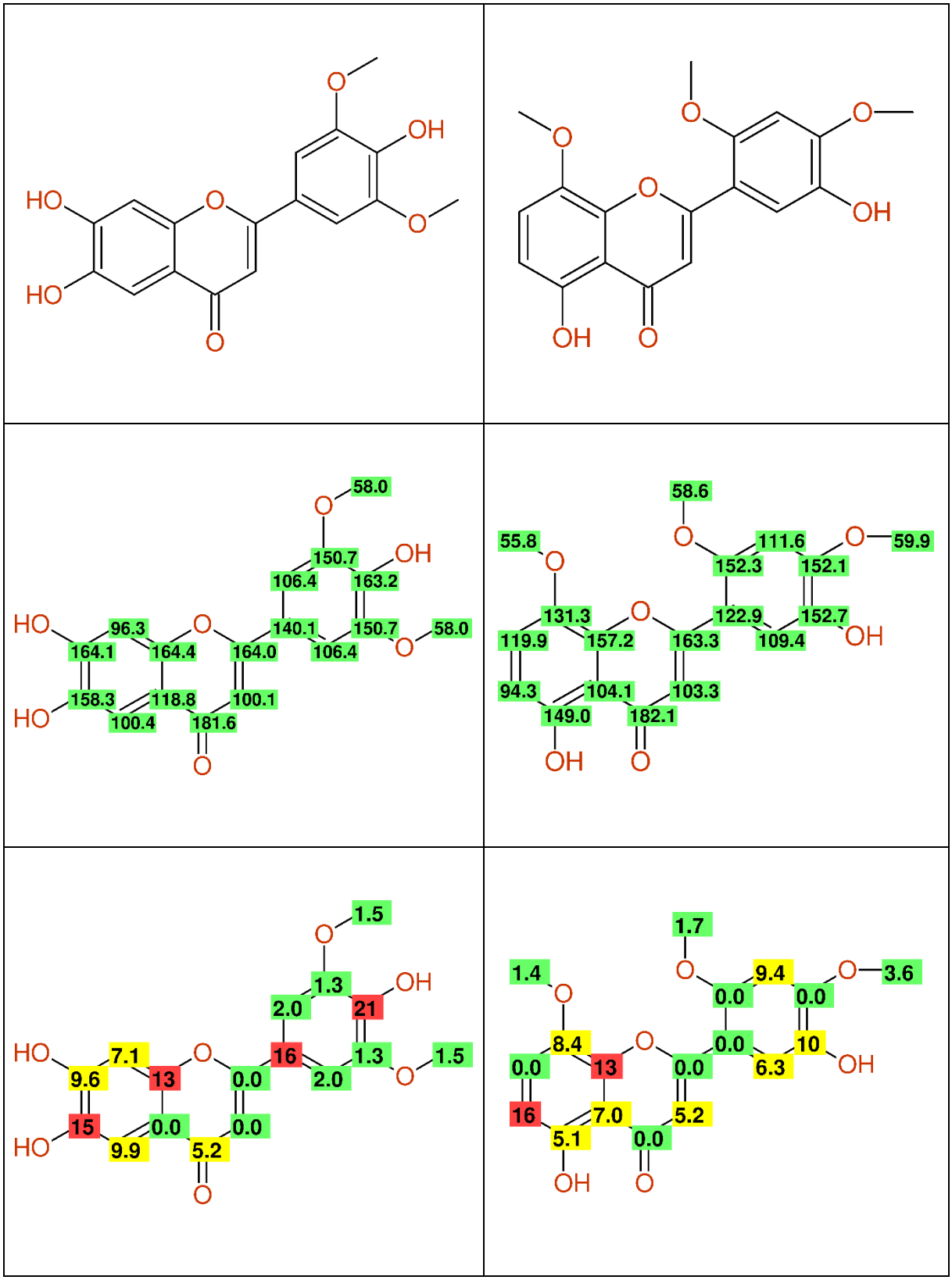
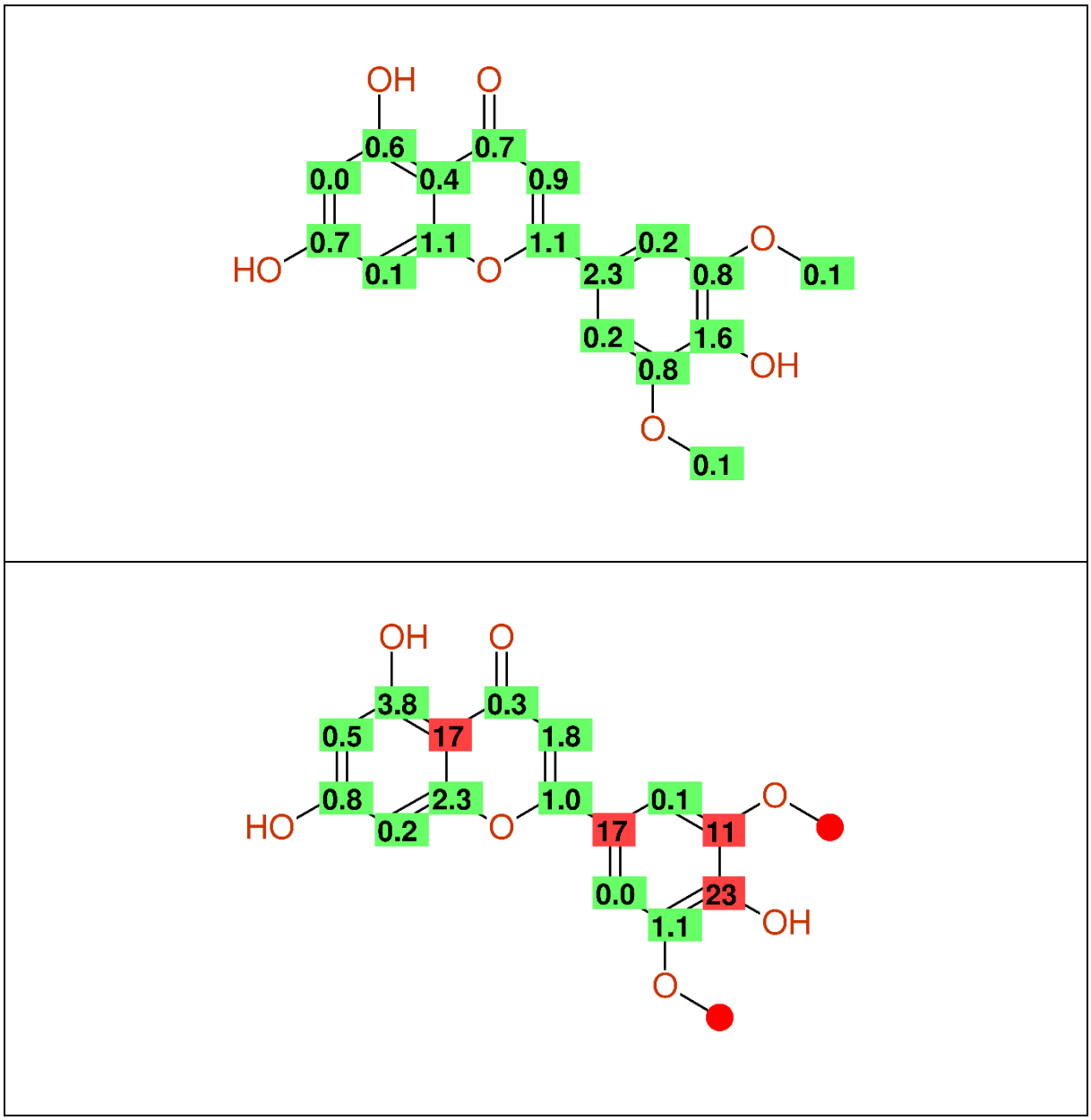


Fig. 27 The identical structure elucidated from quite different sets of ^{13}C-NMR spectral data showing the large inconsistency found in the chemical literature. Top: Excellent consistency between structure proposal and ^{13}C-NMR data [87]. Bottom: Large deviations between assigned and predicted ^{13}C-NMR chemical shift values [88].

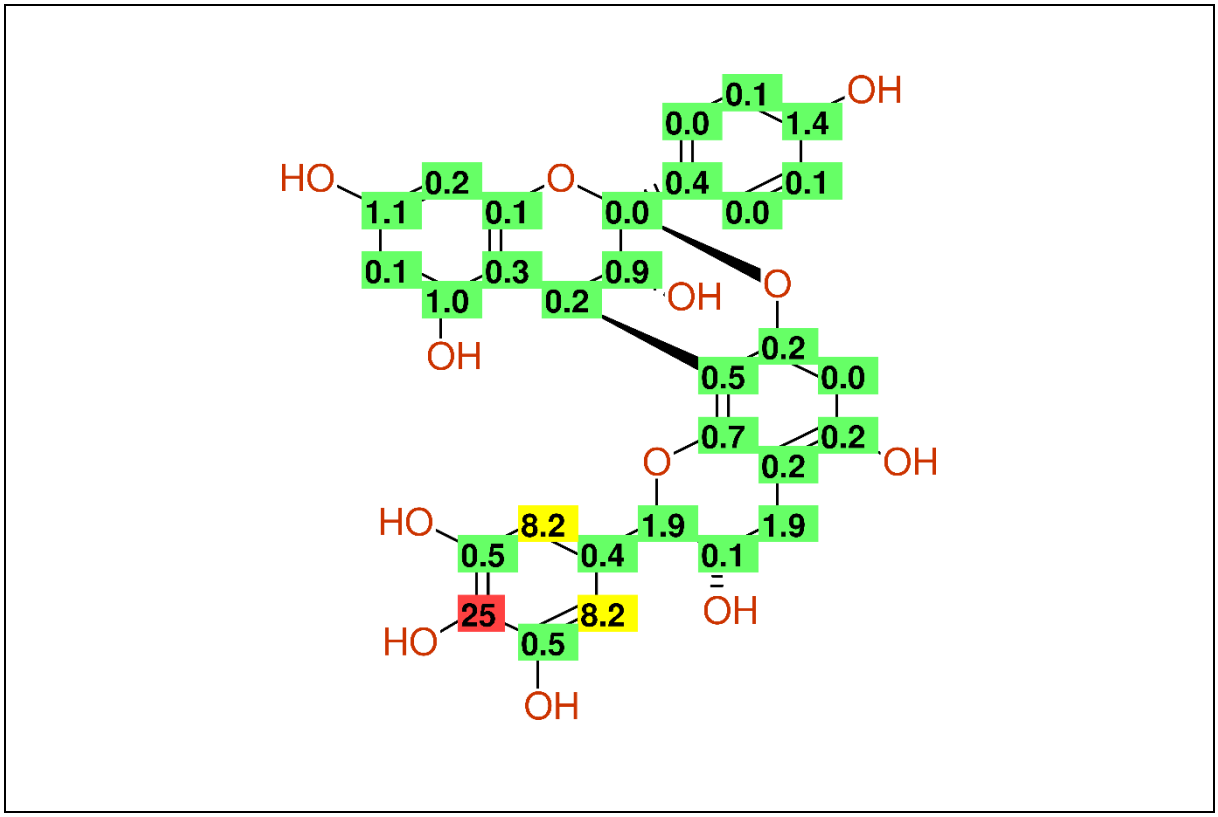


Inconsistency of unknown origin between ^{13}C-NMR data and the chemical structure

This chapter holds a few examples of data inconsistent with the given structural proposal. Neither comparison with other literature citations nor application of the structure generator, which creates only a small subset of all possible structures was able to clarify the situation. From spectrum prediction extremely large deviations are obtained showing this inconsistency but no alternative or explanation could be derived. In such cases it would be best to start from scratch redoing all measurements, which implies the availability of the compound under investigation. The original experimental data are usually also not available anymore, because we missed to store them together with the publication as supplementary material. The excuse of high-priced storage devices is no longer valid as it was some fifteen years ago.

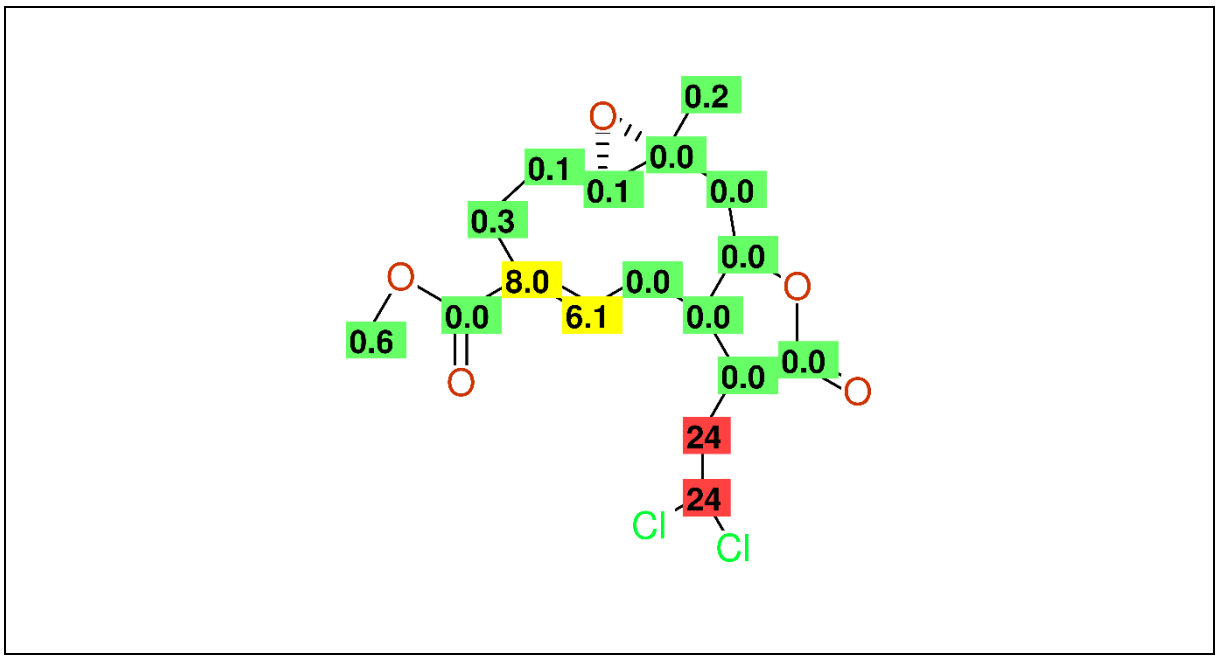
An example which allows to claim an inconsistency between the proposed structure and the given ^{13}C-NMR spectral data without offering any alternative structure proposal or any other explanation is geranin C [89] (**Fig. 28**).

Fig. 28 Geranin C (compound 1 from [89], CAS-RN: 385366-99-4; PUBCHEM-CID: 10,415,768) having massive deviations in its 3,4,5-trihydroxybenzene-moiety. The CSEARCH-Robot-Referee does not create reasonable alternative structures, the reason for this discrepancy remains unidentified.



The isolation and structure elucidation of a natural product named dichloromethyldihydrohirsutolide was published as compound 3 in [90]; the data show some severe inconsistency with respect to ^{13}C-NMR chemical shift values in the dichlorinated sidechain (**Fig. 29**). Comparison with 1,1-dichlorinated n-alkanes supports this finding [91]. When studying the paper [90] in detail, the given isotope ratio in the MS used to verify the existence of a dichlorinated species is wrong, the signal assignment of compound 2 named dihydrohirsutolide is disturbed by a typo in the numbering of the carbons (C-2 given twice) increasing the confusion when using this published data as reference material for future research.

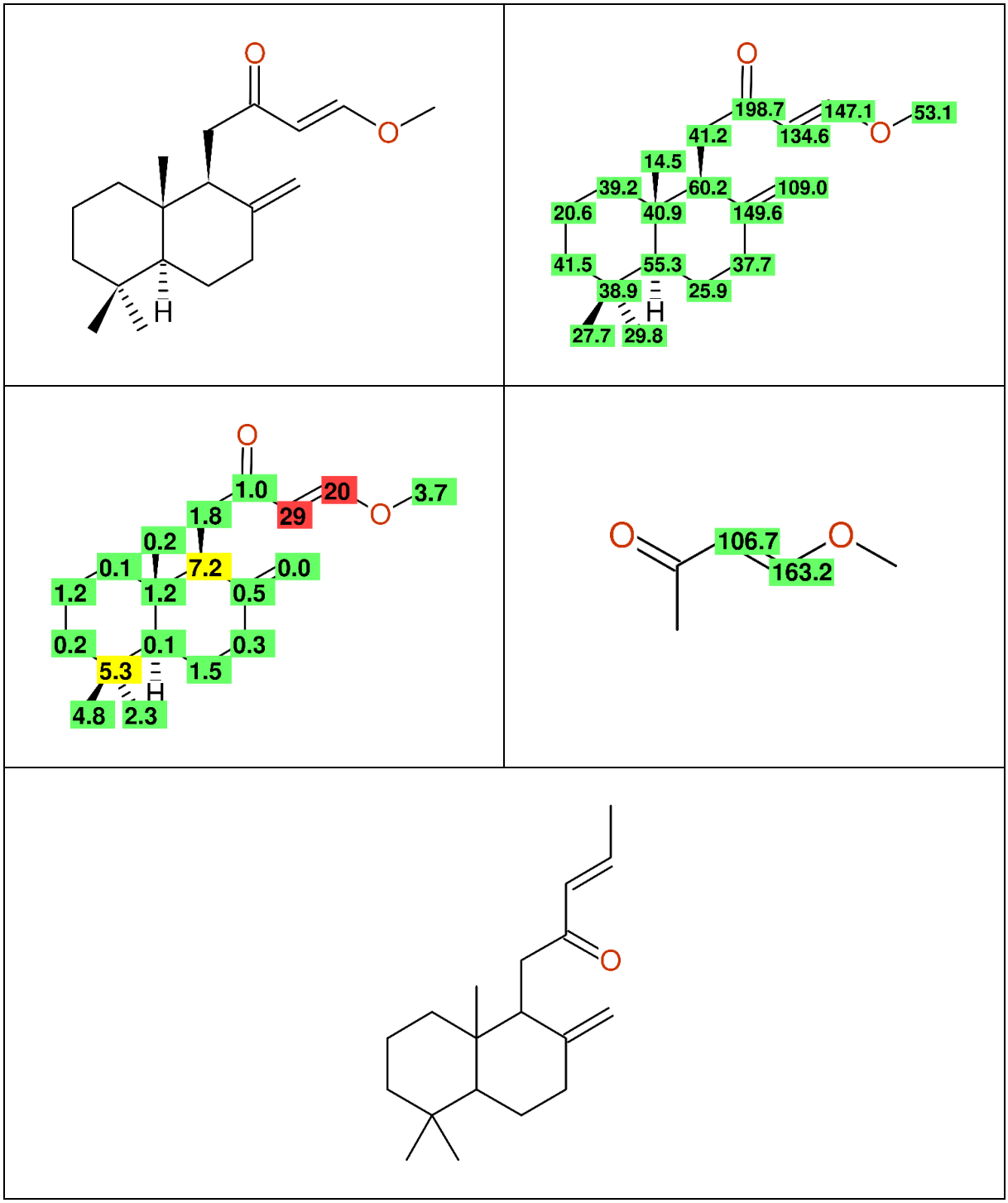
Fig. 29 Dichloromethyldihydrohirsutolide (compound 3 from [90], CAS-RN: 302580-18-3, PUBCHEM-CID: 101,078,307) shows large deviations between experimental and predicted ^{13}C-NMR data in the dichlorinated sidechain.



The structure of the alkaloid coryximine, compound 9 in [92], is shown with five oxygens according to the molecular drawing, whereas in the experimental part a molecular formula of $C_{20}H_{19}NO_6$ is given. This molecular formula is supported by the fact that two -O-CH_2-O- groups with signals at 101.22 and 101.68ppm respectively, are given and furthermore four signals around 147ppm are present. Even the simple strategy of checking the given molecular formula against the calculated one from the structure drawing might avoid such a trivial oversight.

Another example is a bioactive terpene from the roots of *Chloranthus henryi*, compound 2 in [93], showing massive differences between experimental and predicted ^{13}C-NMR chemical shift values in compatible with the data of a similar model compound [94] (**Fig. 30**). One possible explanation might be a drawing error of the structural diagram leading to an alternative structure, which is again incompatible with the given HMBC-data in the paper.

Fig. 30 Compound 2 from [93] (CAS-RN: 929684-45-7, PUBCHEM-CID: 15,984,079) labeled with the experimental ^{13}C-NMR chemical shift values (top/right) showing large deviations between experimental and predicted data (middle/left). A reference compound for the methoxylated enone-fragment (middle/right). A known non-isomeric structural alternative (CAS-RN: 879497-69-5, PUBCHEM-CID: 12,160,803) still incompatible with the given 2D-data.



Summary and Conclusion:

The analysis of some 3,600 published ^{13}C-NMR data in Planta Medica in the period from 1977 to 2006 reveals a set of approximately 100 questionable entries having assignment errors and/or wrong structure proposals. A few classes of compounds are frequently and systematically intermixed in the literature. The error rate in Planta Medica is of comparable order of magnitude as it is in other relevant journals in the field of natural product chemistry despite intense peer-reviewing. The CSEARCH-Robot-Referee has been shown to be a very efficient tool to detect these inconsistencies between the proposed structure and the assigned ^{13}C-NMR signals. In all cases analyzed here, the wrong proposal was contained in the underlying knowledge base used for spectrum prediction showing the efficiency when using (at least) a second technology for spectrum estimation. Even trivial facts like the compatibility of the drawn structure and the molecular formula (or molecular weight) are worth to be checked. It should be mentioned that data sets, which seem to be wrong at a first glance, will be already excluded before data input into the CSEARCH-database; therefore the findings presented here are already biased by the process of data-selection itself – this effect is valid for all journals, because of the journal-independent workflow implemented.

The main problem associated with wrong assignments and possibly wrong structure proposals is the inaccessibility of the raw data used during the structure elucidation process for later reinspection in case of doubt. Blind trust in 2D-NMR spectroscopy paired with neglection of basic rules describing ^{13}C-NMR chemical shift values seems to be the main source for wrong proposals in modern structure elucidation. The final recommendation for authors, reviewers and publishers is, that the publishing process must be dramatically streamlined

and as many manual steps as possible must be avoided. Every fact which can be checked by appropriate computer software should be checked already by the author(s) and later on during peer-reviewing. The raw data and the structures derived thereof must be deposited in searchable repositories together with a measure of their quality in order to have validated reference material on a long-term perspective.

References

[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-Chãvez 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 ^{13}C NMR Data in Natural Product Chemistry. *Progress in the Chemistry of Organic Natural Products*, Kinghorn AD, Falk H, Gibbons S, Kobayashi J (eds), Springer International Publishing, Cham, 2017; 105:137–215

[4] Bremser W. HOSE – a novel substructure code. Analytica Chimica Acta 1978; 103:355-365

[5] Purtuc V. Abschätzung von ^{13}C-NMR-Spektren mittels neuronaler Netze. Work for diploma 1997; University of Vienna.

[6] Jonas E, Kuhn S. Rapid prediction of NMR spectral properties with quantified uncertainty. J Cheminform 2019; 11:50

[7] Schütz V, Purtuc V, Felsinger S, Robien W. CSEARCH-STEREO: A new generation of NMR database system allowing three-dimensional spectrum prediction. Fresenius J Anal Chem 1997; 359:33-41

[8] Technical Note: ACD/Labs NMR Predictors. Advanced Chemistry Development, Inc., Toronto, ON, Canada. November 21, 2019. https://www.acdlabs.com/download/technotes/2019/Technical-Note-Drawing-with-NMR-Predictors.pdf accessed July 10th,2020

[9] Kuhn S, Johnson SR. Stereo-aware extension of HOSE-codes. ACS Omega,2019; 4:7323-7329

[10] https://www.acdlabs.com/products/adh/nmr/nmr_pred/ accessed July 10th, 2020

[11] Satoh H, Koshino H, Uno T, Koichi S, Iwata S, Nakata T. Effective consideration of ring structures in CAST/CNMR for highly accurate ^{13}C-NMR chemical shift prediction. Tetrahedron 2005; 61:7431-7437

[12] Kalchhauser H, Robien W. CSEARCH: A computer program for identification of organic compounds and fully automated assignment of carbon-13 nuclear magnetic resonance spectra. J Chem Inform Comput Sci 1985; 25:103-108

[13] http://www.modgraph.co.uk/product_nmr.htm accessed July 10th, 2020

[14] Steinbeck C, Krause S, Kuhn S. NMRShiftDB – Constructing a free chemical information system with open-source components. J Chem Inf Comput Sci 2003; 43:1733–1739

[15] https://application.wiley-vch.de/stmdata/specinfo.php accessed July 10th, 2020

[16] https://spectrabase.com/ accessed July 10th,2020

[17] http://insideinformatics.cambridgesoft.com/videosanddemos/143/ir-uvvis-and-nmr-spectra-predictions accessed July 10th, 2020

[18] https://mestrelab.com/software/mnova/nmr-predict/ accessed July 10th, 2020

[19] https://www.bruker.com/products/mr/nmr/software.html accessed July 10th, 2020

[20] Robien W. Do high-quality ^{13}C-NMR spectral data really come from journals with high impact factors? TrAC – Trends Analytical Chemistry 2009; 28:914-922

[21] Pupier M, Nuzillard JM, Wist J, Schlörer NE, Kuhn S, Erdelyi M, Steinbeck C, Williams AJ, Butts C, Claridge TDW, Mikhova B, Robien W, Dashti H, Eghbalnia HR, Farès C, Adam C, Kessler P, Moriaud F, Elyashberg M, Argyropoulos D, Pérez M, Giraudeau P, Gil RR, Trevorrow P, Jeannerat D. NMReDATA, a standard to report the NMR-assignment and parameters of organic compounds. Magn Reson Chem 2018; 56:703-715

[22] http://nmredata.org/ accessed July 10th, 2020

[23] Urbas A, Schönberger T, Corbett C, Lippa K, Rudolphi F, Robien W. NPS-Data hub: a web-based community driven analytical data repository for new psychoactive substances. Forensic Chem 2018; 9:76-81

[24] Schlörer N, Kuhn S. Moleküle und Strukturbeweise teilen: NMR-Spektroskopie-Datenbank für organische Moleküle. Chem Unserer Zeit 2020; 54:6-13

[25] Burns DC, Reynolds WF. Minimizing the risk of deducing wrong natural product structures from NMR data. Magn Reson Chem 2019; 1-34

[26] https://nmrpredict.orc.univie.ac.at/planta_medica/robot.php accessed July 10th, 2020

[27] Robien W. Computer-assisted peer reviewing of spectral data: the CSEARCH-protocol. Mh Chem 2019; 150:927-932

[28] Haider N, Robien W. Automatisierte Qualitätskontrolle von ^{13}C-NMR-Daten. Nachr Chem, 2016; 64:196-198

[29] Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE. PubChem 2019 update: improved access to chemical data. Nucleic acids research 2019; 47(D1):D1102–D1109

[30] ftp://ftp.ncbi.nih.gov/pubchem/Compound/CURRENT-Full/SDF/ accessed July 10th, 2020

[31] http://www.chemspider.com/Default.aspx accessed July 10th, 2020

[32] https://www.emolecules.com/ accessed July 10th, 2020

[33] Rakotoson JH, Fabre N, Jacquemond-Collet I, Hannedouche S, Fourasté I, Moulis C. Alkaloids from *Galipea officinalis*. Planta Med 1998; 64:762-763

[34] Houghton PJ, Woldemariam TZ, Watanabe Y, Yates M. Activity against Mycobacterium tuberculosis of alkaloid constituents of angostura bark, *Galipea officinalis*. Planta Med 1999; 65:250-254

[35] Njar VCO, Alao TO, Okugun JI, Holland HL. 2-Methoxycanthin-6-one: A new alkaloid from the stem wood *Quassia amara*. Planta Med 1993; 59:259-261

[36] Pettit GR, Singh SB, Goswami A, Nieman RA. Antineoplastic agents. 157. *Quassia kerstingli*. Tetrahedron 1988; 44:3349-3354

[37] Ohmoto T, Tanaka R, Nikaido T. Studies on the constituents of *Ailanthus altissima Swingle*. On the alkaloidal constituents. Chem Pharm Bull 1976; 24:1532-1536

[38] Koike K, Ohmoto T. Carbon-13 nuclear magnetic resonance study of canthin-6-one alkaloids. Chem Pharm Bull 1985; 33:5239-5244

[39] Zhang G, Rücker G, Breitmaier E, Mayer R, Steinbeck C. Alkaloids from *Thalictrum przewalskii*. Planta Med 1998; 64:165-171

[40] Advanced Chemistry Development Inc. (ACD/Labs) Software V11.01

[41] Rios MY, Aguilar-Guadarrama AB, Navarro V. Two new benzofuranes from *Eupatorium aschenbornianum* and their antimicrobial activity. Planta Med 2003; 69:967-970

Erratum: Title changed to: Two new benzofurans from *Ageratina pichinensis* var. *bustamenta* and their antimicrobial activity. Planta Med 2004; 70(4):384

[42] Gomez F, Quijano L, Calderon JS, Perales A, Rios T. 2,2-Dimethylchromenes from Eupatorium aschenbornianum. Phytochemistry 1982; 21:2095-2097

[43] Sigstad EE, Catalan CAN, Diaz JG, Herz W. Chromanones, benzofurans and other constituents from *Ophryosporus lorentzii*. Phytochemistry 1996; 42:1443-1445

[44] Harmala P, Kaltia S, Vuorela H, Hiltunen R. A furanocoumarin from *Angelica archangelica*. Planta Medica 1992; 58:287-289

[45] Akagah B, Estour F, Vérite P, Seguin E, Tillequin F, Lafont O. Synthesis of models of metabolites: Oxidation of variously substituted chromenes including acronycine, by a porphyrin catalytic system. J Heterocyclic Chem 2005; 42:1267-1272

[46] Machida K, Trifonov LS, Ayer WA, Lu ZX, Laroche A, Huang HC, Cheng KJ, Zantige JL. 3(2H)-Benzofuranones and chromanes fromoquid cultures of the mycoparasitic fungus *Coniothyrium minitans*. Phytochemistry 2001; 58:173-177

[47] Castaneda P, Gomez L, Mata R, Lotina-Hennson B, Anaya AL,Bye R. Phytogrowth-inhibitory and antifungal constituents of *Helianthella quinquenervis*. J Nat Prod 1996; 59:323-326

[48] Ulubelen A, Mericli AH, Mericli F, Tan N. Two new coumarins from *Haplophyllum ptilostylum*. J Nat Prod 1993; 56:1184-1186

[49] Huang YL, Chen CC, Chen YJ, Huang RL, Shieh BJ. Three xanthones and a benzophenone from *Garcinia mangostana*. J Nat Prod 2001; 64:903-906

[50] Radwan MM, ElSohly MA, Slade D, Ahmed SA, Khan IA, Ross SA. Biologically active cannabinoids from high-potency *Cannabis sativa*. J Nat Prod 2009; 72:906-911

[51] Kassem M, Mosharrafa SA, Saleh NAM, Abdel-Wahab SM. Two new flavonoids from *Retama raetam*. Fitoterapia 2000; 71:649-654

[52] Tan YX, Yang Y, Zhang T, Chen RY, Yu DQ. Bioactive 2-arylbenzofuran derivatives from *Morus wittiorum*. Fitoterapia 2010; 81:742-746

[53] Kapche GDWF, Fozing CD, Donfack JH, Fotso GW, Amadou D, Tchana AN, Bezabih M, Moundipa PF, Ngadjui BT, Abegaz BM. Prenylated arylbenzofuran derivatives from *Morus mesozygia* with antioxidant activity. Phytochemistry 2009; 70:216-221

[54] Ryu HW, Cho JK, Curtis-Long MJ, Yuk HJ, Young S, Jung S, Kim YS, Lee BW, Park KH. α-Glucosidase inhibition and antihyperglycemic activity of prenylated xanthones from *Garcinia mangostana*. Phytochemistry 2011; 72:2148-2154

[55] Bojase G, Majinda RRT, Gashe BA, Wanjala CCW. Antimicrobial Flavonoids from *Bolusanthus speciosus*. Planta Med 2002; 68:615-620

[56] Li X, Wang D, Xia M, Wang Z, Wang W, Cui Z. Cytotoxic prenylated flavonoids from the stem bark of *Maackia amurensis*. Chem Pharm Bull 2009; 57:302-306

[57] Breytenbach JC, Rall GJH. Structure and synthesis of isoflavonoid analogs from *Neorautanenia amboensis* Schinz. J Chem Soc Perkin Trans 1 1980; 1980:1804-1809

[58] Cottiglia F, Casu L, Bonsignore L, Casu M, Floris C, Leonti M, Gertsch J, Heilmann J. New cytotoxic prenylated isoflavonoids from *Bituminaria morisiana*. Planta Med 2005; 71:254-260

[59] Lee SJ, Wood AR, Maier CGA, Dixon RA, Mabry TJ. Prenylated flavonoids from *Maclura pomifera*. Phytochemistry 1998; 2573-2577

[60] Wang D, Li F, Jiang Z. Osteoblastic proliferation stimulating activity of *Psoralea corylifolia* extracts and two of its flavonoids. Planta Med 2001; 67:748-749

[61] Peng G, Guan P, Li H, Yuan Y. Study on chemical constituents of *Psoralea corylifolia*. Chinese Traditional and Herbal Drugs 1996; 19:563-565

[62] Nkengfack AE, Kouam J, Vouffo TW, Meyer M, Tempesta MS, Fomum ZT. An isoflavanone and a coumestan from *Erythrina sigmoidea*. Phytochemistry 1994; 35:521-526

[63] Chen JJ, Duh CY, Chen IS. Cytotoxic Chromenes from *Myriactis humilis*. Planta Med 2005; 71:370-372

[64] Dallacker F, Holschbach M, Konings AWT. 1,3-Benzodioxole derivatives as radioprotective agents. Part 4. Preparation of 6,6-disubstituted 1,3-dioxolo[4,5-g]chromans and -chromenes. Chemiker-Zeitung 1990; 114:105-109

[65] Tang S, Bremner P, Kortenkamp A, Schlage C, Gray AI, Gibbons S, Heinrich M. Biflavonoids with cytotoxic and Antibacterial activity from *Ochna macrocalyx*. Planta Med 2003; 69:247-253

[66] Moreira IC, Sobrinho DC, De Carvalho MG, Braz-Filho R. Isoflavanone dimers hexaspermone A, B and C from *Ouratea hexasperma*. Phytochemistry 1994; 35:1567-1572

[67] Udino L, Abaul J, Bourgeois P, Gorrichon L, Duran H, Zedde C. Lignans from the seeds of *Hernandia sonora*. Planta Med 1999; 65:279-281

[68] Gu JQ, Park EJ, Totura S, Riswan S, Fong HHS, Pezzuto JM, Kinghorn AD. Constituents of the twigs of *Hernandia ovigera* that inhibit the transformation of JB6 murine epidermal cells. J Nat Prod 2002; 65:1065-1068

[69] Costantin MB, Ferreira MJP, Rodrigues GV, Emerenciano VP, Computer-assisted approach to structure elucidation of lignans. Spectroscopy Letters 2008; 41:405-421

[70] Du J, Wang ML, Chen RY, Yu DQ. Two new bislabdane-type diterpenoids and three new diterpenoids from the roots of *Cunninghamia lanceolate*. Planta Med 2001; 67:542-547

[71] Minami T, Wada S, Tokuda H, Tanabe G, Muraoka O, Tanaka R. Potential antitumor-promoting diterpenes from the cones of *Pinus luchuensis*. J Nat Prod 2002; 65:1921-1923

[72] Huang KS, Zhou S, Lin M, Wang YH. An isorhapontigenin and a novel stilbene dimer from *Gnetum hainanense*. Planta Med 2002; 68:916-920

[73] Hu J, Jia Z, Bai S. Two new polyacetylenes from *Artemisia eriopoda*. Planta Med 1998; 64:378-379

[74] Hu CQ, Chang JJ, Lee KH. Antitumor agents, 115. Seselidiol, a new cytotoxic polyacetylene from *Seseli mairei*. J Nat Prod 1990; 53:932-935

[75] Ruperto G, Cannizzo S, Amico V, Bizzini M, Piattelli M. Chemical constituents of *Ferulago nodosa*. J Nat Prod 1994; 57:1731-1733

[76] Bernart MW, Cardellina JH II, Balaschak MS, Alexander M, ShoemakerRH, Boyd MR. Cytotoxic falcarinol oxylipins from *Dendropanax arboreus*. J Nat Prod 1996; 59:748-753

[77] Kobayashi M, Mahmud T, Umezome T, Wang W, Murakami N, Kitagawa I. The absolute stereostructures of the polyacetylenic constituents of ginseng *Radix rubra*, Tetrahedron 1997; 53:15691-15700

[78] Hu J, Feng X. Phenylpropanes from *Acorus tatarinowii*. Planta Med 2000; 66:662-664

[79] Trang NTD, Wanner MJ, Koomen GJ, Dung NX. New acetophenone and thymol from *Eupatorium stoechadosmum*. Planta Med 1993; 59:480-481

[80] Chang FR, Chen IH, Liao SC, Issa HH, Hayashi K, Nozaki H, Wu YC, Lee, KH. Three new agarofuran sesquiterpenes reissantins F - H from *Reissantia buchananii*. Planta Med 2006; 72:92-96

[81] Baloch IB, Baloch MK, Saqib QNU. Cytotoxic macrocyclic diterpenoids esters from *Euphorbia cornigera*. Planta Med 2006; 72:830-834

[82] Murugaiyah V, Chan KL. Antihyperuricemic lignans from the leaves of *Phyllanthus niruri*. Planta Med 2006; 72:1262-1267

[83] Yap SP, Shen P, Butler MS, Gong Y, Loy CJ, Yong EL. New estrogenic prenylflavones from *Epimedium brevicornum* inhibits the growth of breast cancer cells. Planta Med 2005; 71:114-119

[84] Sriphong L, Sotanaphun U, Limsirichaikul S, Wetwitayaklung P, Chaicvhantipyuth C, Pummangura S. Cytotoxic alkaloids from the flowers of *Senna spectabilis*. Planta Med 2003; 69:1054-1056

[85] Zheng WF, Tan RX, Yang L, Liu ZL. Two flavones from *Artemisia giraldii* and their antimicrobial activity. Planta Med 1996; 62:160-162

[86] Hajdu Z, Martins A, Orban-Gyapai O, Forgo P, Jedlinszki N, Mathe I, Hohmann J. Xanthine oxidase-inhibitory activity and antioxidant properties of the methanol extract and flavonoids of *Artemisia asiatica*. Records of Natural Products 2014; 8:299-302

[87] Li H, Zhou CX, Pan Y, Gao X, Wu X, Bai H, Zhou L, Chen Z, Zhang S, Shi S, Luo J, Xu J, Chen L, Zheng X, Zhao Y. Evaluation of antiviral activity of compounds isolated from *Ranunculus sieboldin* and *Ranunculus sceleratus*. Planta Medica 2005; 71:1128-1133

[88] Awaad AS, Mohamed NH, Maitland DJ, Soliman GA. Anti-ulcerogenic Activity of Extract and Some Isolated Flavonoids from *Desmostachia bipinnata* (L.) Stapf. Rec Nat Prod 2008; 2:76-82

[89] Calzada F, Cedillo-Rivera R, Bye R, Mata R. Geranins C and D, additional new antiprotozoal A-type proanthocyanidins from *Geranium niveum*. Planta Med 2001; 67:677-680

[90] Ali AA, El-Emary NA, Khalifa AA, Volk FJ, Frahm AW. The heliangolide dihydrohirsutoide from *Arctotis fastuosa* and dichloromethyldihydrohirsutolide. Planta Med 2000; 66:585-587

[91] Aitken RA, Hodgson PKG, Morrison JJ, Oyewale AO. Flash vacuum pyrolysis over magnesium. Part 1. Pyrolysis of benzylic, other aryl/alkyl and aliphatic halides. J Chem Soc, Perkin Trans 1, 2002; 402-415

[92] Zhou J, Tong X, Lian W, Fang Q. Chemical studies on the alkaloids of *Corydalis hsuchowensis*. Planta Med 1991; 57:156-158

[93] Wu B, He S, Wu X, Pan Y. Bioactive terpenes from the roots of *Chloranthus henryi*. Planta Med 2006; 72:1334-1338

[94] Weingarten LR, Levin RH. Application of carbon-13 nuclear magnetic resonance spectroscopy to the analysis of charge distribution patterns in unsaturated carbonyl-containing compounds. J Am Chem Soc 1976; 98:4571–4577