

Article

# The Advantage of Automatic Peer-Reviewing of $^{13}\text{C}$ -NMR Reference Data Using the CSEARCH-Protocol †

Wolfgang Robien 

Department of Organic Chemistry, University of Vienna, Währingerstrasse 38, A-1090 Vienna, Austria; Wolfgang.Robien@univie.ac.at

† Dedicated to Professor Dr. Mikhail Elyashberg on the occasion of his 85th birthday.

**Abstract:** A systematic investigation of the experimental  $^{13}\text{C}$ -NMR spectra published in *Molecules* during the period of 1996 to 2015 with respect to their quality using CSEARCH-technology is described. It is shown that the systematic application of the CSEARCH-Robot-Referee during the peer-reviewing process prohibits at least the most trivial assignment errors and wrong structure proposals. In many cases, the correction of the assignments/chemical shift values is possible by manual inspection of the published tables; in certain cases, reprocessing of the original experimental data might help to clarify the situation, showing the urgent need for a public domain repository. A comparison of the significant key numbers derived for *Molecules* against those of other important journals in the field of natural product chemistry shows a quite similar level of quality for all publishers responsible for the six journals under investigation. From the results of this study, general rules for data handling, data storage, and manuscript preparation can be derived, helping to increase the quality of published NMR-data and making these data available as validated reference material.

**Keywords:** NMR;  $^{13}\text{C}$ -NMR; computer-assisted peer-reviewing; spectrum prediction; structure generation; database



**Citation:** Robien, W. The Advantage of Automatic Peer-Reviewing of  $^{13}\text{C}$ -NMR Reference Data Using the CSEARCH-Protocol. *Molecules* **2021**, *26*, 3413. <https://doi.org/10.3390/molecules26113413>

Academic Editor: Antony J. Williams

Received: 7 May 2021

Accepted: 31 May 2021

Published: 4 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

NMR-spectroscopy is an important technique providing a massive amount of information during the structure elucidation process at the level of the constitution, configuration, and 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 a new bottleneck of spectrum interpretation. This has led to the effect of quite frequent misinterpretations of experimental data in terms of the 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] of computer-assisted peer reviewing and subsequent fully automatic structure revisions verified the tremendous effect of automatic quality control for spectral data.

The flood of experimental NMR data has stimulated the development of computer-software, helping spectroscopists 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 introduced in 1978 by Bremser [4]. Later on, neural networks [5] were used; this technology was 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 the constitutional properties of the

molecule (except for the easy case of cis-/trans-isomerism) and was later on expanded to also handle 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, many databases collecting the reference material published in the public domain chemical literature were built (ACD/NMR-Predictor [10], CAST/CNMR [11], CSEARCH [12], KnowItAll [13], NMRPREDICT [14], NMRSHIFTDB [15], and 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 for the assignment of 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 the 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 the chemical shift value—is not furthermore available then. Such an unassigned peak list only represents a fingerprint for a given structure sufficient for spectral similarity searches, but is 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 through a comparison of the new measurements with already published reference material taken from the public domain chemical literature. The main problem with the existing reference material is that there is usually no information available about the quality of these data [20]. During the last two decades, many journals have published so-called “Supplementary Material” for their articles—mainly as PDFs—showing, in principle, only pictures of the spectra, efficiently prohibiting the 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 for starting a necessary reprocessing and reinterpretation from the original datasets 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 using state-of-the-art 2D-techniques. A leading role with respect to this effect, seems to be HMBC-type spectra showing frequent misinterpretations of  $^2J$  (or  $^4J$ )–couplings, as  $^3J$ –couplings necessarily leading to incorrect structure proposals [25]. This kind of incorrect interpretation of 2D information is frequently preferred against the application of simple spectrum prediction, efficiently showing the inconsistency between the structure proposal and spectral data.

## 2. Results and Discussion

### 2.1. The CSEARCH-Robot-Referee—General Overview

The automatic structure verification based on  $^{13}\text{C}$ -NMR chemical shift data is the central procedure within the “CSEARCH-Robot-Referee” [3,26–28]. The knowledge base behind consists of some 340,000 curated  $^{13}\text{C}$ -NMR spectra taken from the public domain literature. A database holding 520 million predicted spectra was additionally used to allow for efficient structure dereplication. The workflow applied here consisted of the following steps:

- Formal check of the structure (valency, charge, and stereocenter);
- Create a table to link the structure to other databases (e.g., PUBCHEM [29,30], Chemspider [31], and eMolecules [32]);
- Check formal correctness of the supplied peak list (symmetry and 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 for the evaluation of the quality of the result;
- Assign signals if unassigned signals are given;
- Calculate and visualize the coincidence between the structure proposal and the given experimental data;
- Perform a search for identical structures contained in the underlying knowledge base;
- Perform a search for an identical spectral pattern associated with different structures;
- Detect positions with a large deviation between the experimental and predicted values;
- Start structure generator, which modifies the topology of the given structure exclusively at the positions with a large deviation;
- Perform dereplication based on the given peak list using the database of 520 million predicted spectra.

This workflow was applied to 10,039 spectra taken from *Molecules* between 1996 and 2015. The publications having assigned  $^{13}\text{C}$ -NMR data within this range of years were extracted and these data sets were used to build the CSEARCH database. The reason for selecting this period of time was simply that we were not able to extract more than approximately 25 K spectra per year from the literature, and we had to switch to other journals in order to cover as many journals as possible, with the goal of achieving a high structural diversity. When comparing *Molecules* with other prominent journals in the field of natural product chemistry, like *Chemical and Pharmaceutical Bulletin*, *Fitoterapia*, *Journal of Natural Products*, *Phytochemistry*, and *Planta Medica*, a similar number of compounds were found that seemed to be in error. It should be mentioned that the manually performed extraction of the data from the journal was already quite selective, because datasets that seemed to be in error at a first glance were completely ignored, 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  $^{13}\text{C}$ -NMR data published in these six above-mentioned journals (Table 1) came from compounds of a similar size with an average molecular weight between 431 and 520 amu; the average deviation between the experimental and predicted chemical shift values was also very similar (1.61–2.07 ppm). The number of compounds with at least one signal more than 20 ppm away from the predicted chemical shift value was between 1.34 and 3.40% of the entries available in the CSEARCH-collection. A nearly identical finding was given when the prediction was restricted by a partial structure search; the examples selected here contained either a chromone-fragment or a steroid skeleton. In both cases, the average deviation was again in a very narrow range, starting at 1.40 ppm and 1.27 ppm, respectively, with a maximum at 1.74 and 1.60 ppm, respectively. From the data compiled (Table 1), it can be concluded that all six journals showed a similar level of quality with respect to the  $^{13}\text{C}$ -NMR data contained therein.

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

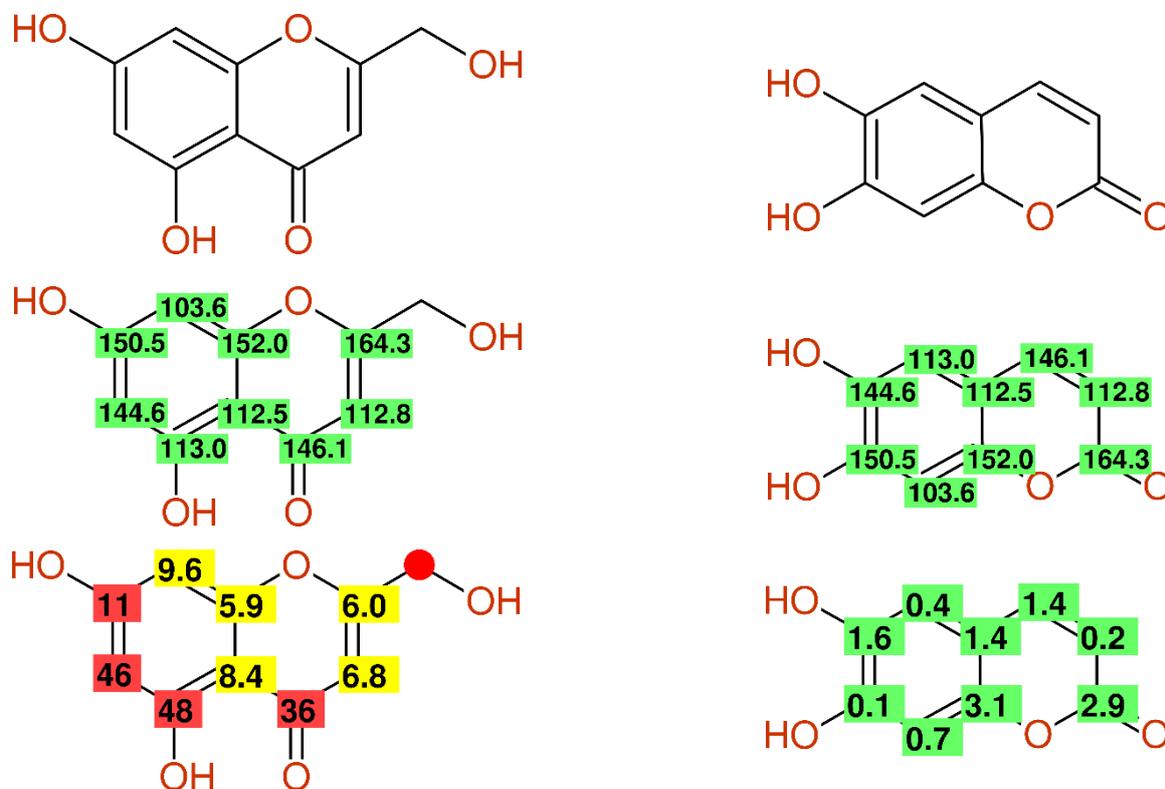
Journal	MOL	CPB	FT	JNP	PC	PM
Entries	10,039	17,863	1568	34,933	38,379	3621
Period	1996–2015	1977–2016	1998–2012	1979–2013	1976–2015	1977–2006
Average MWT (amu)	431	520	497	481	483	479
$\Delta\delta_{\text{C}}$ (ppm)	2.07	1.61	1.85	1.79	1.66	1.72
$\Delta\delta_{\text{C}} > 20$ ppm	341/3.40%	239/1.34%	42/2.68%	542/1.55%	533/1.39%	64/1.77%
$\Delta\delta_{\text{C}}$ (ppm) Chromones	1.59	1.40	1.74	1.68	1.45	1.57
$\Delta\delta_{\text{C}}$ (ppm) Steroids	1.60	1.27	1.54	1.47	1.37	1.36

The following examples were taken from articles published in *Molecules* between 2010 and 2015, and were exactly reproduced for input into the CSEARCH-Robot-Referee—any later correction or erratum was ignored, because the intention of this summary is only to show what can be avoided by fully automatic peer-reviewing at the time of uploading a manuscript to the editorial office. The examples are grouped according to the kind of error and only typical cases are shown, far away from being a complete analysis of all published  $^{13}\text{C}$ -NMR data in *Molecules*. Furthermore, it should be mentioned that this computer-assisted peer-reviewing could be regarded as an effective tool to support the usual peer-reviewing process with respect to quality and time-consumption, going into the details of the structure proofs based on  $^{13}\text{C}$ -NMR spectroscopy. Throughout the figures, the following coloring scheme is applied in order to increase readability. The first row shows the chemical structure, and in the second row, the structure together with the experimental chemical shift values, as published, are shown, where green highlighting means that signals were assigned by authors and yellow highlighting points to exchangeable signal assignment. The third row shows the differences between the experimental and predicted chemical shift values; deviations between 5 and 10 ppm are highlighted in yellow. Smaller deviations are given in green, whereas larger deviations are shown with red.

## 2.2. Examples of Wrong/Useless $^{13}\text{C}$ -NMR Data

### 2.2.1. Using the Same Data Twice

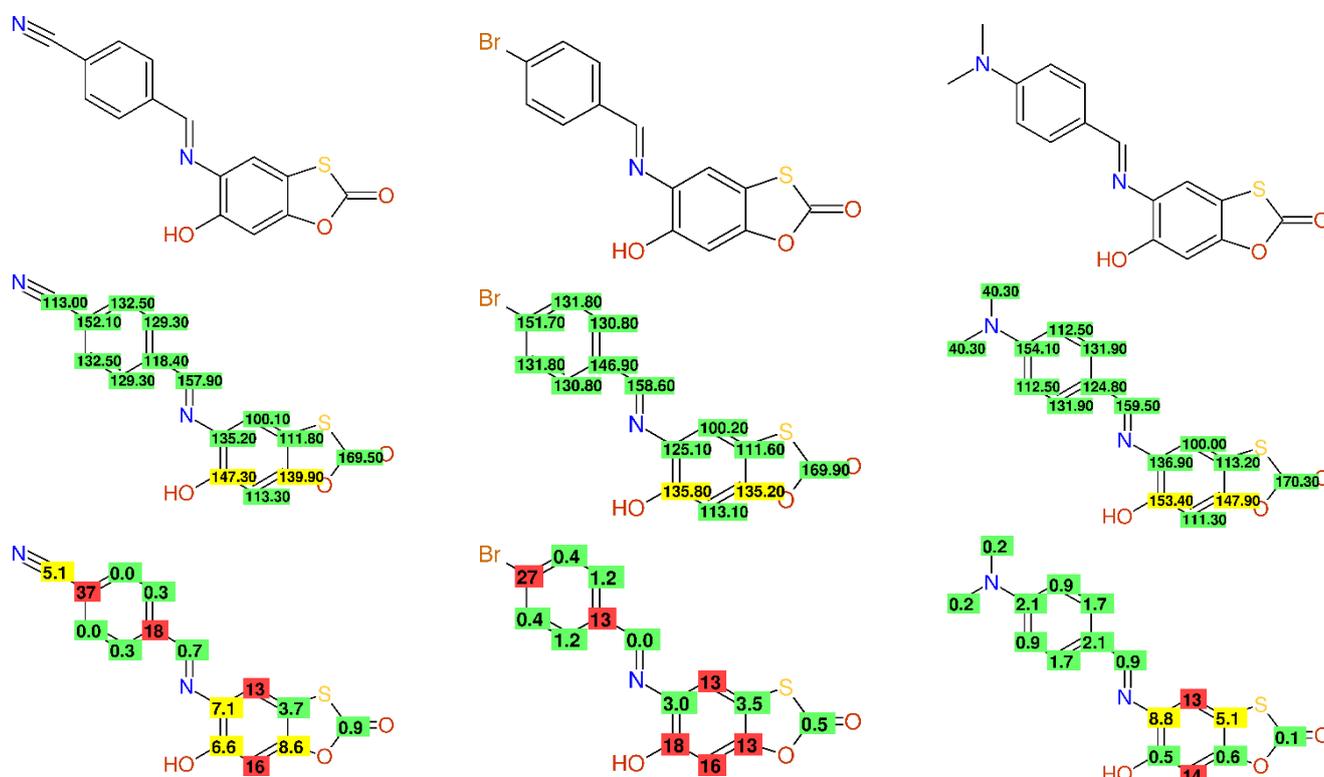
In [33], compounds 4 (esculetin; 6,7-dihydroxycoumarin) and 5 (5,7-dihydroxy-2-hydroxymethylchromone) showed different  $^1\text{H}$ -NMR data, but identical  $^{13}\text{C}$ -NMR chemical shift values were published, as shown in Figure 1. The given values fit to esculetin, but were incomplete and wrong for the proposed chromone-derivative. The structure elucidation was done by comparison with the literature data; therefore it should be mentioned that obviously correct  $^{13}\text{C}$ -NMR data for compound 5 were already published in [34].



**Figure 1.** The structures of compound 5 (left) and compound 4 (right) from [33] together with the published  $^{13}\text{C}$ -NMR data (green, signal assigned by author, and yellow, signal exchangeable assigned) and the differences between experimental and predicted values (red > 10 ppm, yellow < 10 ppm and > 5 ppm, green < 5 ppm).

### 2.2.2. Wrong Values—Strange Substituent Effects

Bromine-substitution is known to induce a high-field shift of approximately 6 ppm in benzene-derivatives, a more pronounced effect can be observed for a nitrile-moiety (16 ppm); for this reason, the given chemical shift data for the carbon 4' in compounds **4a** (152.1 ppm) and **4d** (151.7 ppm) published in [35] are far away from the expected values, as can be seen in Figure 2. It is interesting to note that the carbons 6 and 7a in the benzo[d][1,3]-oxathiol-2-one fragment are exchangeable, assigned with values ranging from 135.2/135.8 ppm (compound **4d**) to 153.4/147.9 (compound **4i**); this difference in chemical shift values was obviously attributed, according to the research, to the exchange of a bromo-substituent with a dimethylamino-moiety seven, respectively nine bonds away! The compounds described in this publication contain quite interesting structural features, but the carbon NMR-data are unusable as a high-quality reference material.



**Figure 2.** The structures of compound **4a** (left), compound **4d** (middle), and compound **4i** (right) from [35], together with the published  $^{13}\text{C}$ -NMR data and the differences between the experimental and predicted values.

In Figure 3, the published NMR-data of ligand **L1** from [36] are summarized, showing a methoxy-group at 30.12 ppm, which is known to resonate within an extremely narrow range of around 56 ppm when one ortho-position is unsubstituted; additionally the chemical shift values of the three  $\text{CH}_2$ -groups are far away from the predicted values, and the spectroscopic characterization of the aromatic ring system is incomplete.

In [37], a series of benzimidazole derivatives was published and the relevant data were compiled in the Supplementary Information. A few compounds contained fluorine, but no C-F couplings in the  $^{13}\text{C}$ -NMR data were given. In compound **26**, which has two different names (5-Methyl- according to the header, and 5-Methoxy according to the experimental procedure), the aromatic  $-\text{OCH}_3$ -group resonated at 55.35 ppm, whereas in compound **34**, a chemical shift value of 33.26 ppm was given. The signal assignment was completely inconsistent within the presented series of compounds—some carbons with a quite small influence to be expected on their chemical shift value covered a range of more than 40 ppm (e.g., C-2). Many compounds showed differences in the range of 40 to

60 ppm between the experimental and predicted values, as given in Figure 4. The overall impression coming from this research was that they were interesting compounds, with  $^{13}\text{C}$ -NMR data beyond repair.

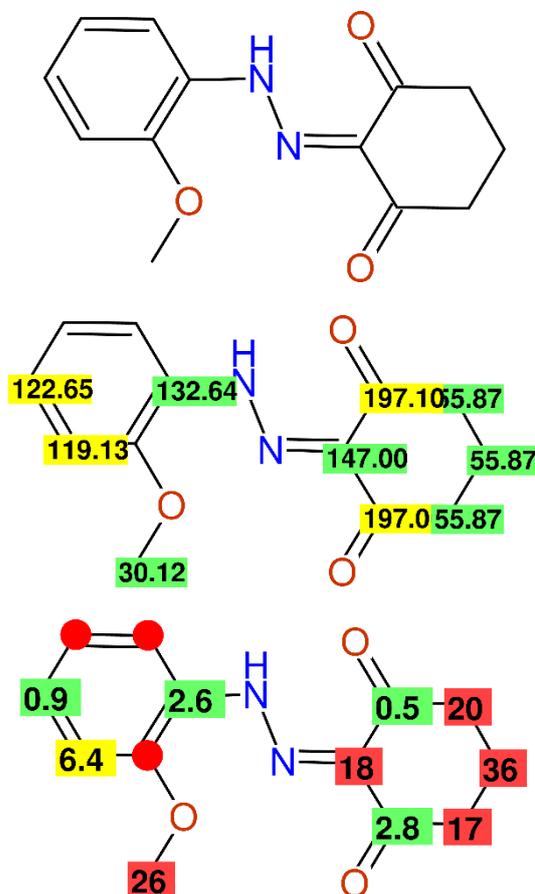


Figure 3. The  $^{13}\text{C}$ -NMR data of compound L1 as published in [36], together with the differences between the experimental and predicted values.

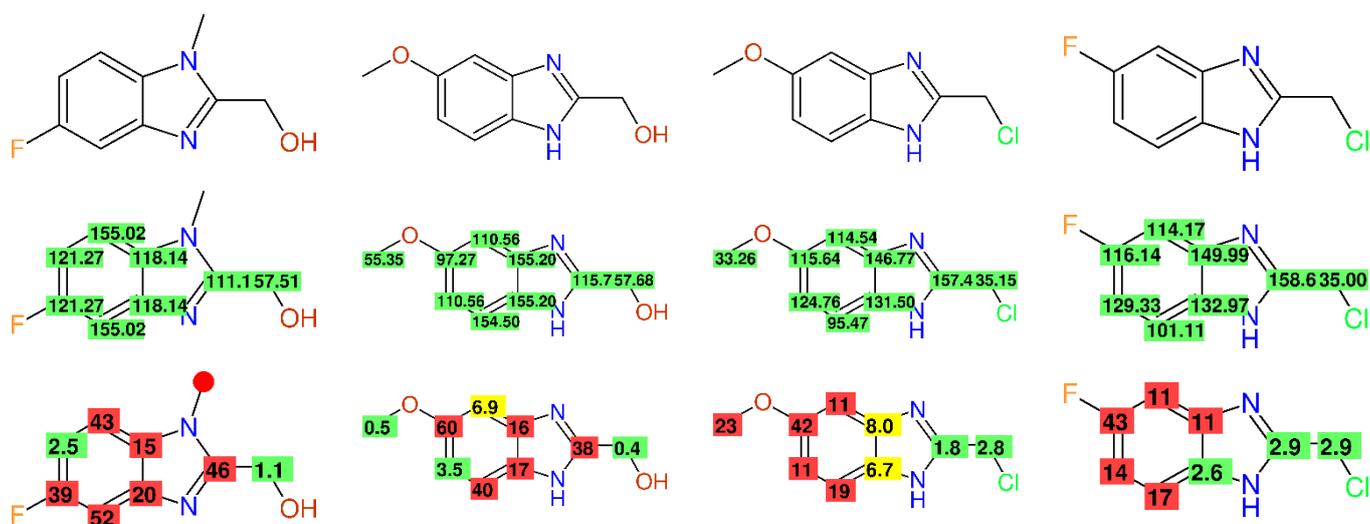
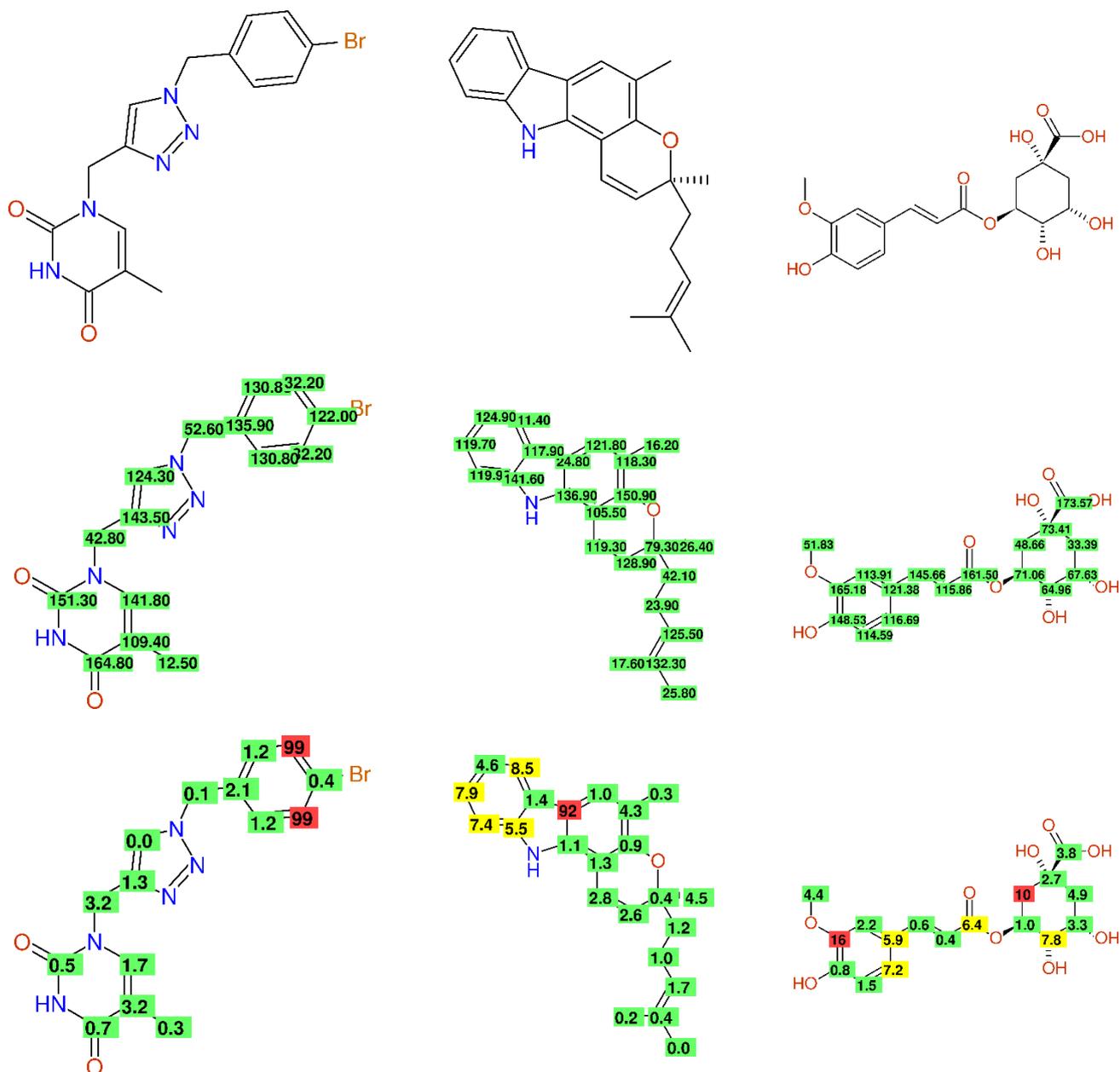


Figure 4.  $^{13}\text{C}$ -NMR data of compounds 51, 26, 34, and 37 (columns from left to right) as published in [37], together with the published chemical shift values and the differences between the experimental and predicted chemical shift values.

### 2.2.3. Typos and Transmission Errors

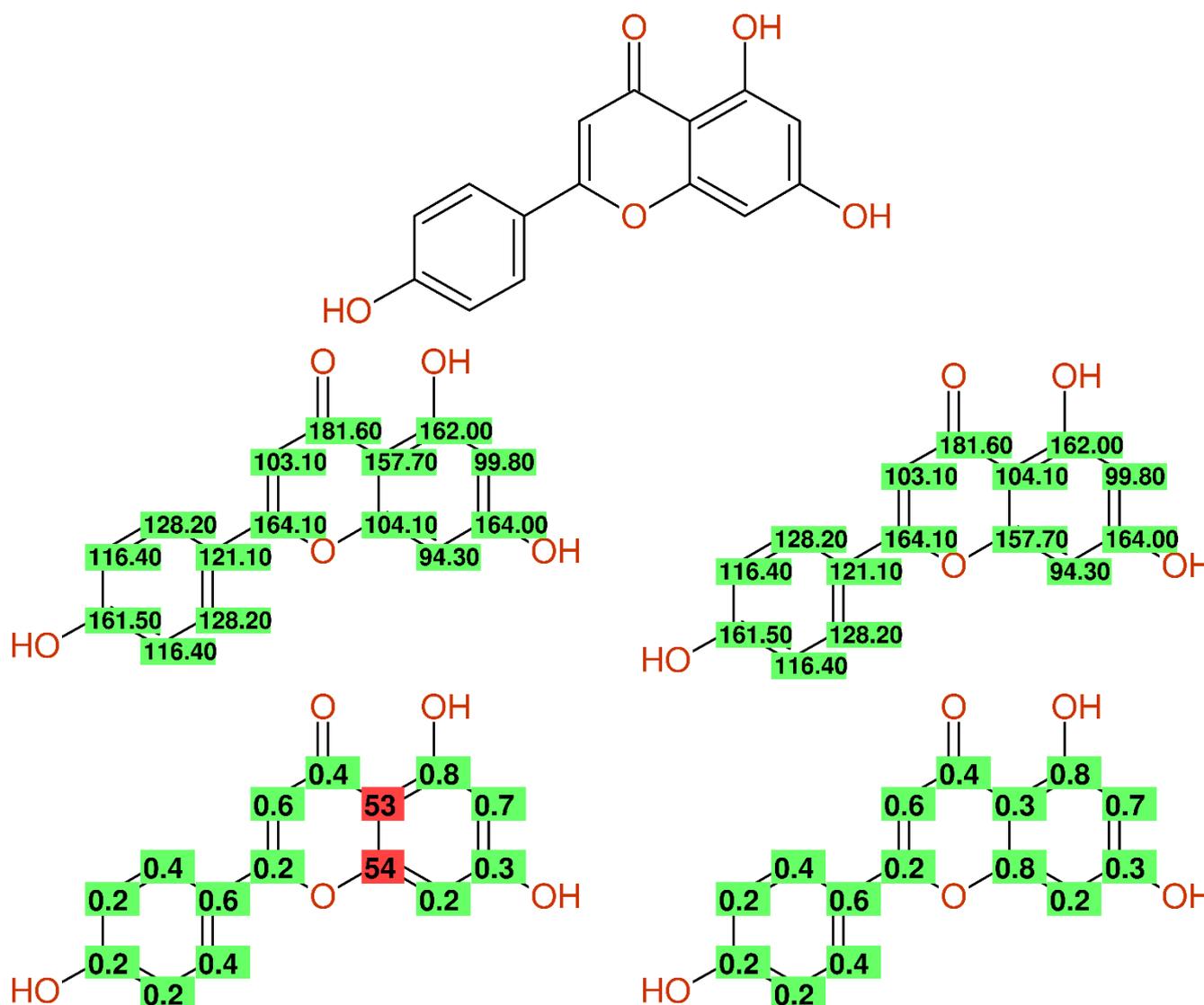
The examples compiled in Figure 5 showed some significant deviations between the predicted and experimental  $^{13}\text{C}$ -NMR chemical shift values, which could be attributed to typos and transmission errors, because the other shift values were mainly within their expectation ranges. In the case of compound 3 (mahanimbine) from [38] and compound 5 from [39], additional assignment errors could be present, as can be seen in Figure 5 (middle and right column). Compound 5 from [39] shows a quite unusual chemical shift value for an aromatic  $-\text{O}-\text{CH}_3$  group (51.83 ppm) making reinspection of the underlying experimental data necessary in order to verify the structural proposal. In the case of compound 13 from [40], a value of 32.2 ppm is given there in Table 3, whereas the experimental part shows a value of 132.2 ppm, as to be expected.



**Figure 5.**  $^{13}\text{C}$ -NMR data of compound 13 from [40] (left), compound 3 (mahanimbine) from [38] (middle), and compound 5 from [39] (right), together with the differences between the experimental and predicted chemical shift values. Differences larger than 99 ppm are always shown as 99 ppm.

### 2.2.4. Trivial Assignment Errors

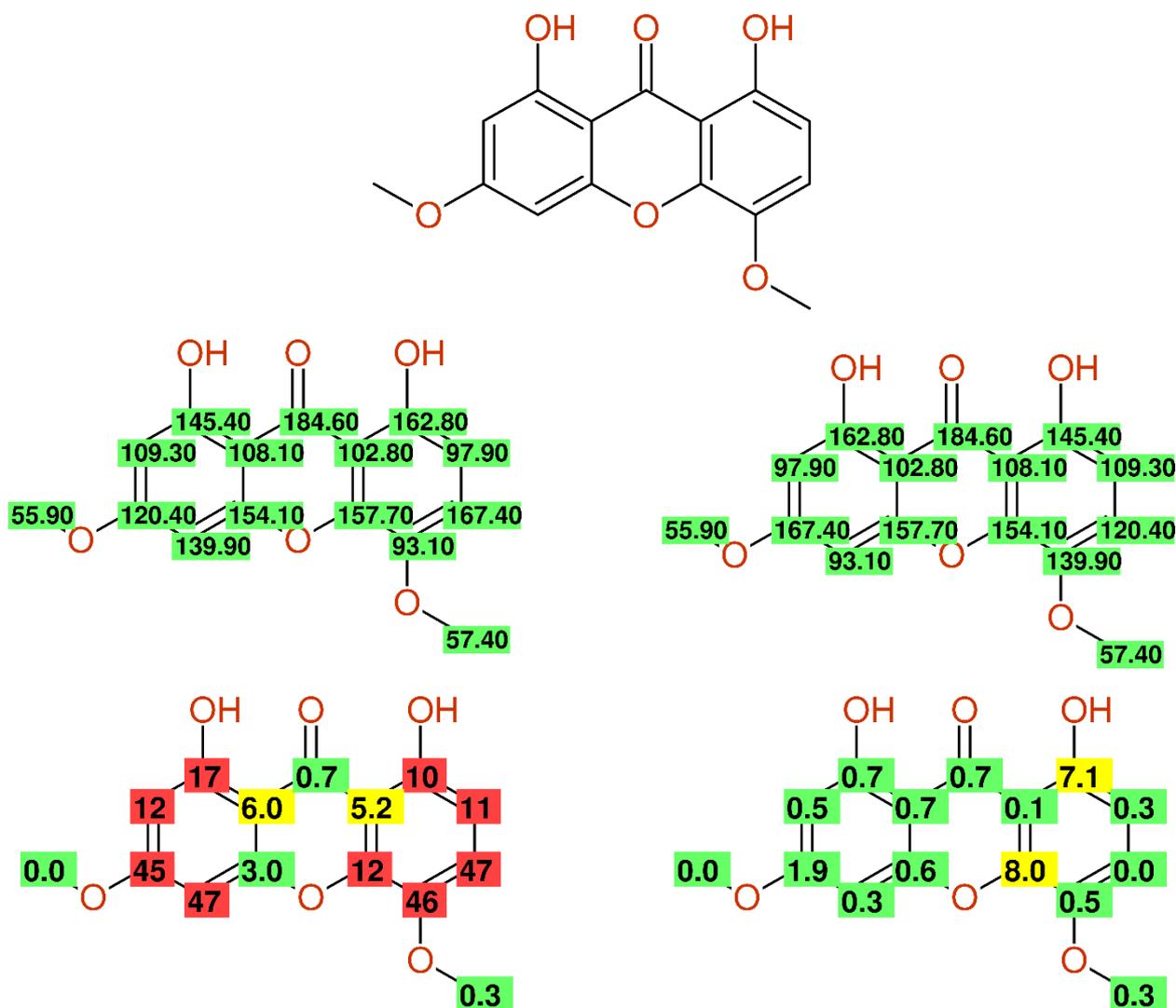
The data of compound 4 (apigenin) from [41] show a typical error based on interchanging the numbering of the structure and the sequence of chemical shift values according to the structure drawing. After this easy correction, an excellent agreement between the predicted and experimental chemical shift values could be achieved, as summarized in Figure 6. The identical problem could be observed with compound D1 from [42] and compound 12 from [43].



**Figure 6.** <sup>13</sup>C-NMR data for compound 4 in [41] as published (left) and after correction (right), together with the differences between the experimental and predicted chemical shift values.

Other trivial assignment errors could be found with compound 3 published in [44]; compound 3 from [45]; compound 3 from [46]; compound 2 in [47], despite excellent Supplementary Information; compound 1 from [48]; and compound 3 from [49].

A more sophisticated error occurred when the numbering schemes of two complete ring systems were intermixed [50], leading to massive deviations between the experimental and predicted chemical shift values at nearly all positions; in this particular case, another possible assignment error (carbon 1 versus 4a) increased the complexity of the correction, as can be seen in Figure 7.



**Figure 7.**  $^{13}\text{C}$ -NMR data of compound 1 from [50] as published (left) and after correction (right), together with the differences between the experimental and predicted chemical shift values. The deviations remaining after the first step of correction (7.1/8.0 ppm) point out an additional assignment problem.

### 2.2.5. Alkyl-Chains

It is known from basic textbooks on NMR-spectroscopy that an unbranched alkyl chain is characterized by a quartet around 14 ppm, 2 triplets around 22 and 32 ppm, and all of the other more centered  $\text{CH}_2$ -groups resonate around 29 ppm. In the chemical literature, there are frequently examples with a different sequence of signals occurring, either caused by typos or wrong assignments. Compound 11 (Figure 8) and compound 12 from [51] showed this misassignment, whereas other compounds (e.g., compounds 4, 5, and 9) from the same paper were correctly assigned.

The identical problem with the wrong assignment of an alkyl chain can be observed at compound OH-1 in [52]; the published data of (3S,8S)-faltarindiol are compiled in Figure 9.

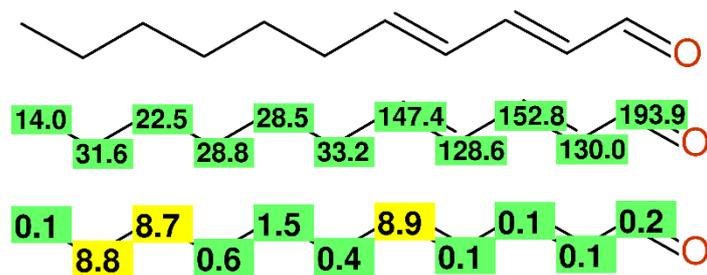


Figure 8. Misassigned alkyl-chain of compound 11 from [51]. Experimental chemical shift values (middle) and differences between experimental and predicted chemical shift values (bottom).

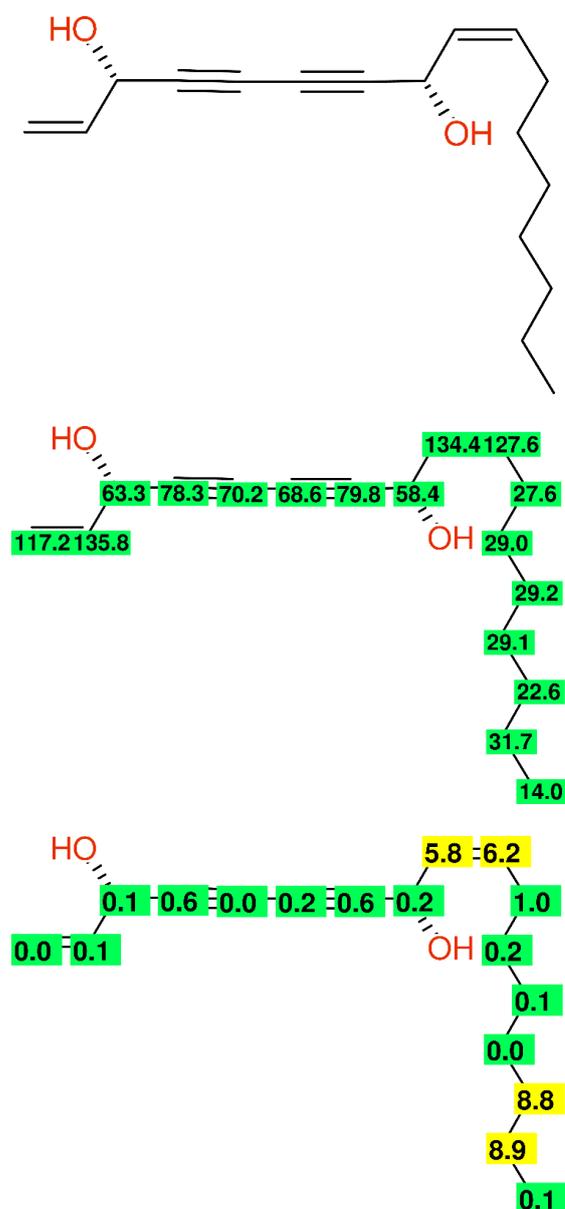
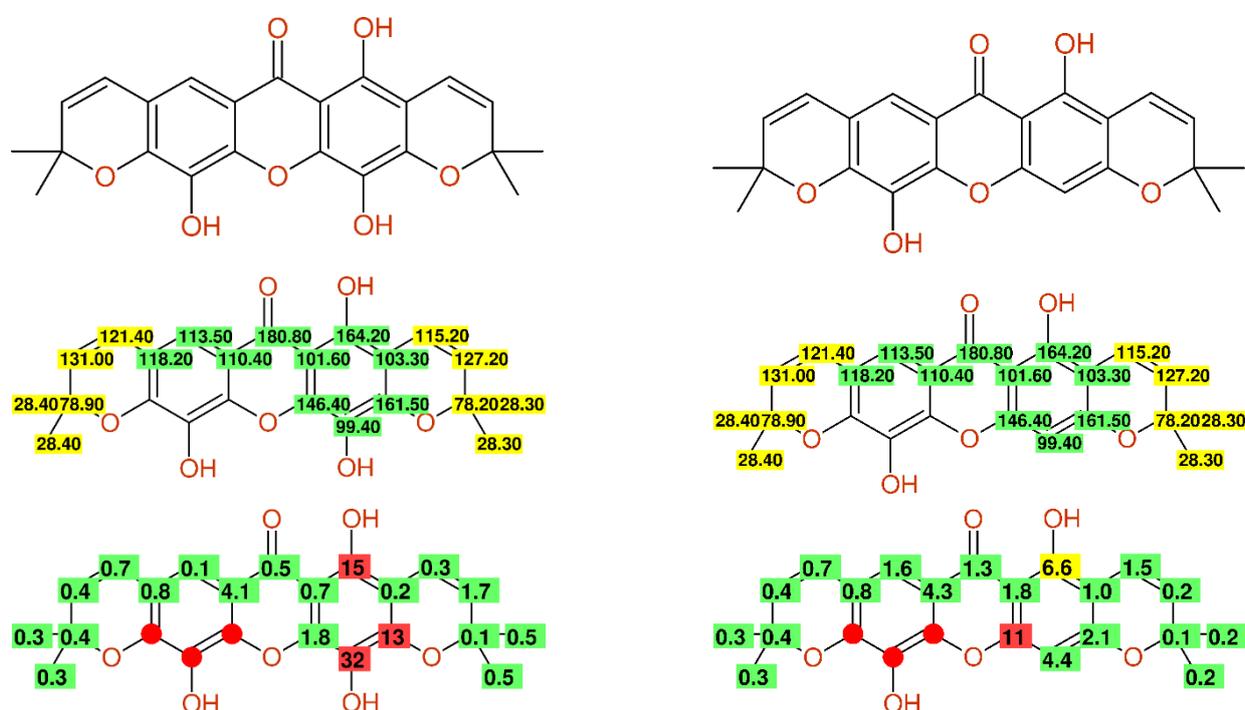


Figure 9. Compound OH-1 from [52] as another example of a misassigned alkyl chain. Experimental chemical shift values (middle) and the differences between the experimental and predicted chemical shift values (bottom).

### 2.2.6. Wrong Structure Drawing

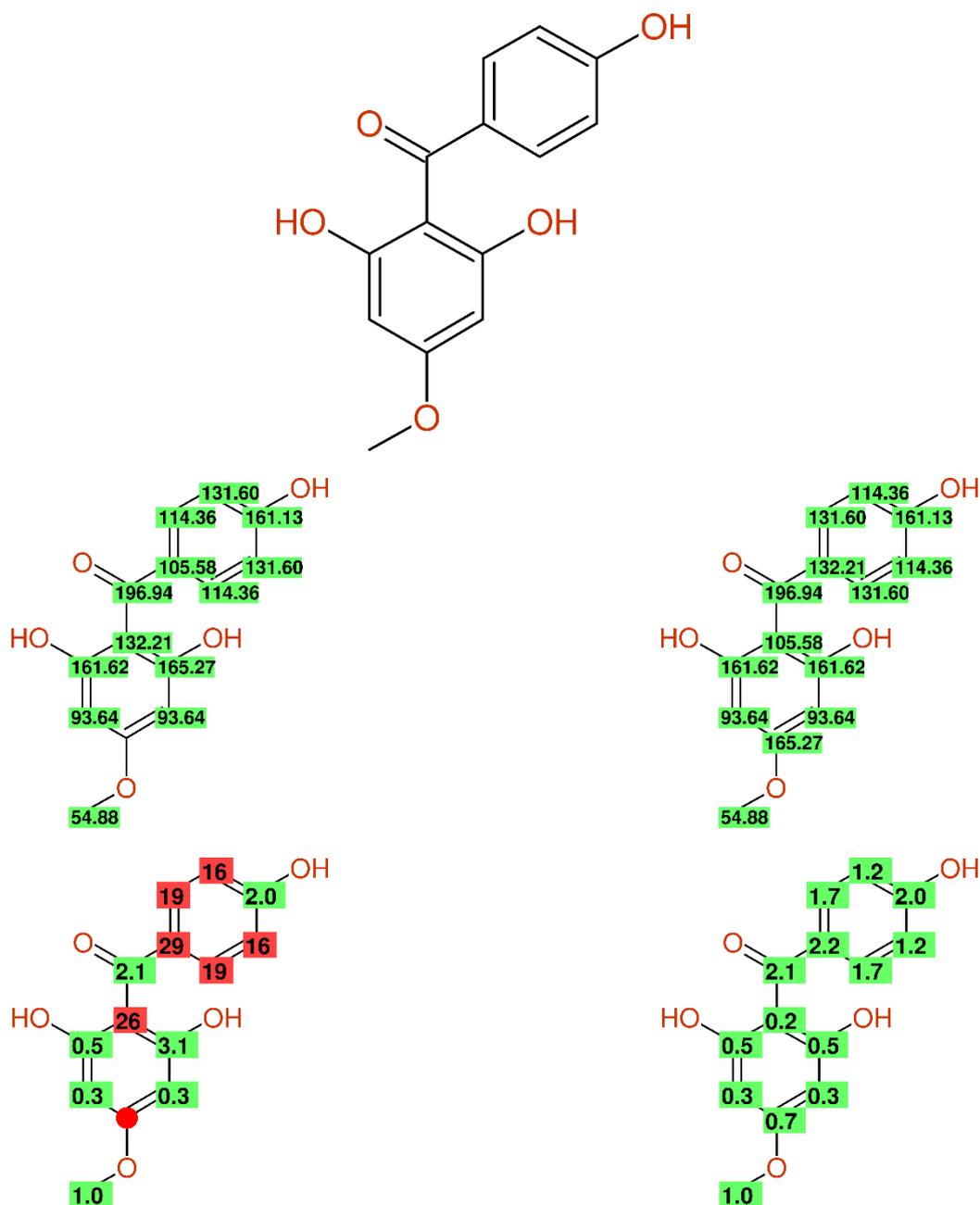
The  $^{13}\text{C}$ -NMR data of pyrojacareubine (compound 3 in [53]) show severe deviations between the experimental and predicted chemical shift values; the differences were focused into one aromatic ring system, as given in Figure 10. This asymmetric  $\text{C}_{23}$ -compound was characterized by only 20 distinct carbon chemical shift values. The molecular formula was explicitly given as  $\text{C}_{23}\text{H}_{20}\text{O}_6$ , whereas the structural diagram held seven oxygens. Removing the 4-hydroxy group from the structure resulted in the well-known compound pyranojacareubin, here named “pyrojacareubine”. According to the  $^1\text{H}$ -NMR data, a chemical shift value of 6.27 ppm for H-4 was given, supporting the assumption of a drawing error with respect to an additional hydroxy-group in position 4. Independently of this correction, the problem with three missing signals remained and a minor assignment error was still present. The wrong structure from this publication was used later in an investigation on antimalarial QSAR analysis [54], influencing the results published there.



**Figure 10.** The data of compound 3 from [53] named pyrojacareubine, as published (left) and after correction of the structure (right), still holding additional assignment errors.

### 2.2.7. Multiple Inconsistencies

In [55], a benzophenone-derivative was published, together with its  $^1\text{H}$ - and  $^{13}\text{C}$ -NMR data, as a structure proof. The following problems were related with this example. Only the facts freely available in the PDF are summarized here. The numbering scheme was wrong (1 versus 1'), leading to a misassignment of the carbon signals at 132.21 and 105.58 ppm. The compound name was inconsistent with the drawing (drawing had four hydroxy-groups, the name was given as “trihydroxy-”). The signals at 131.60 and 114.36 ppm, as well as the associated  $^1\text{H}$  chemical shift values, were interchanged. According to the  $^1\text{H}$  and  $^{13}\text{C}$  chemical shift values in positions 3' and 5', the symmetry of the aryl-moiety was correctly represented, whereas the carbon chemical shift values in positions 2' and 6' reflected non-equivalence and the resonance line of C-4' was missing. The summary for the presentation of this compound, as shown in Figure 11, is as follows: wrong structure drawing, compound name inconsistent with the structure drawing, wrong numbering, wrong signal assignment, missing line caused by wrong assignment—it is a fact that this paper successfully passed the peer-reviewing procedure.

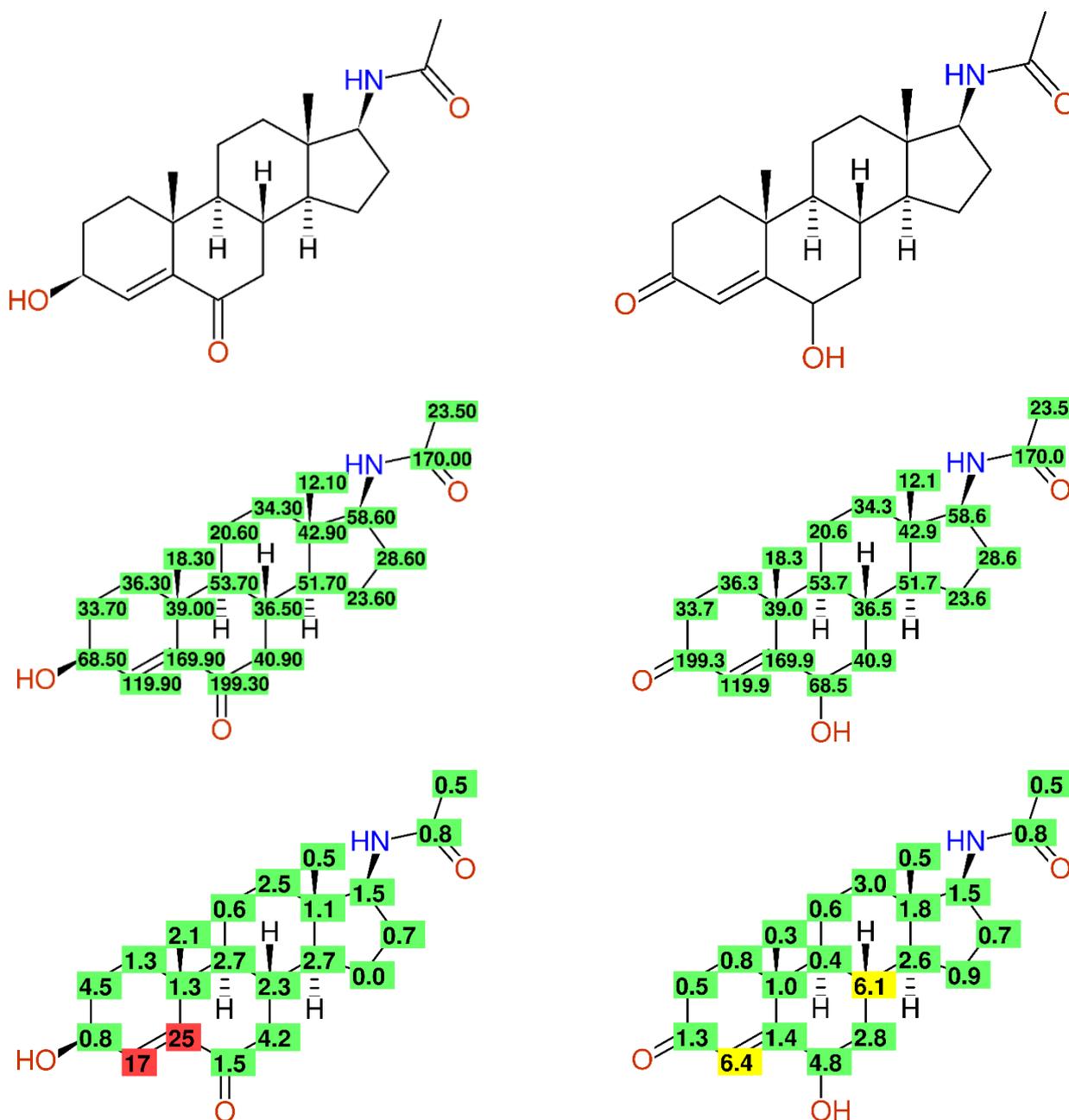


**Figure 11.**  $^{13}\text{C}$ -NMR data of a benzophenone-derivate (already corrected with respect to the erroneously drawn hydroxy-group in position 2) taken from [55], as published (left) and the corrected version (right).

### 2.2.8. Fully Automatic Structure Revisions

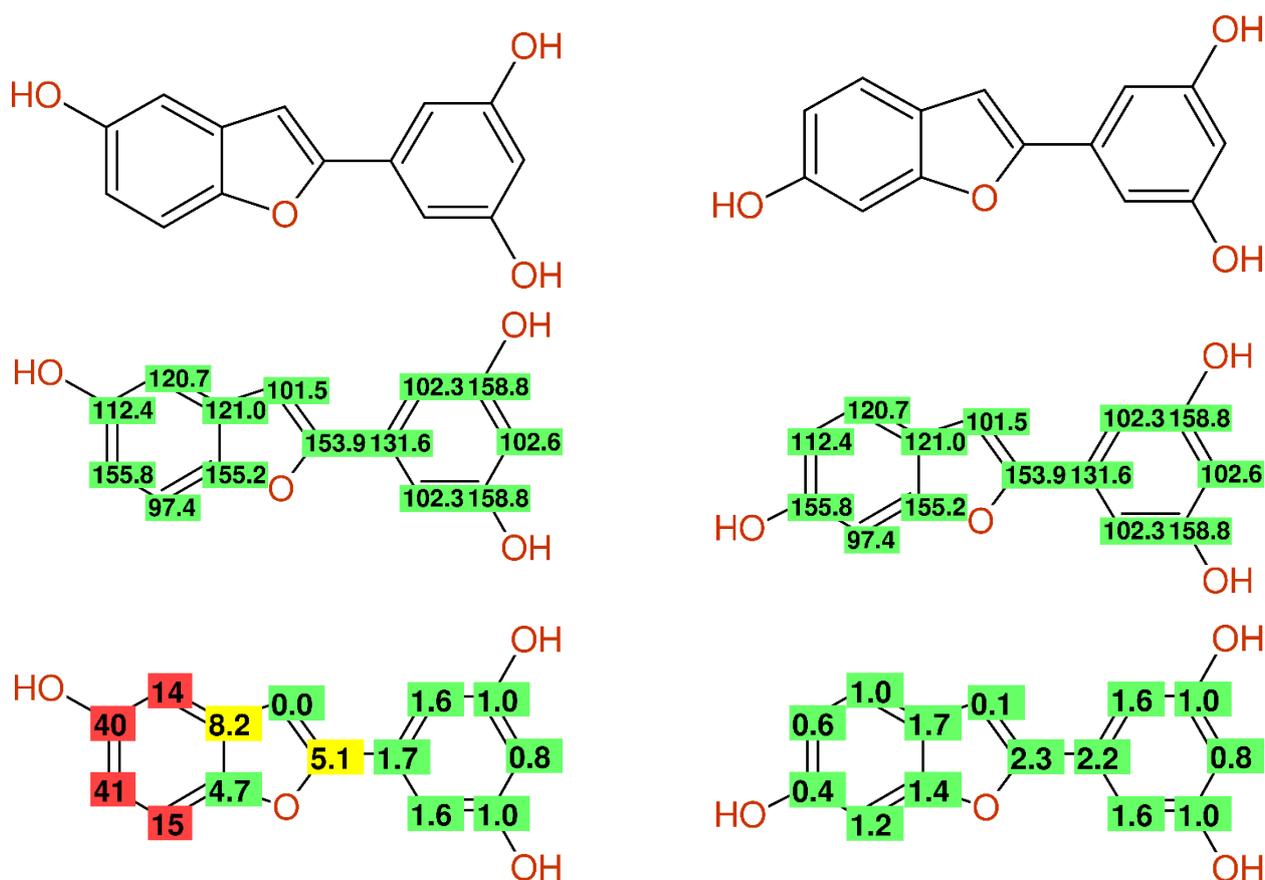
In [56], the selective reduction of 17-acetamidoandrost-4-en-3,6-dione (compound 6) using  $\text{NaBH}_4/\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$  was published. The product was elucidated as  $3\beta$ -hydroxy-17-acetamidoandrost-4-en-6-one (compound 8), showing the  $^{13}\text{C}$ -NMR chemical shift values as given in Figure 12. The prediction of the spectral data led to significant deviations for both carbons of the 4,5-double bond, showing some errors during the structure elucidation process. Starting the structure generator with the obviously wrong structure proposal and the given experimental  $^{13}\text{C}$ -NMR data produced 1591 alternative structures. The given proposal was found at position 358, with an average deviation of 5.20 ppm, whereas the corresponding 6-hydroxy-3-one derivative was ranked at position 2, having an average deviation of only 1.49 ppm. The  $^{13}\text{C}$ -NMR data of only one stereoisomer of this alternative

structure are known [57]—these data slightly differ from those published in [56], showing the necessity of going back to the original measurements to clarify these inconsistencies.



**Figure 12.** 3 $\beta$ -Hydroxy-17-acetamidoandrost-4-en-6-one as published in [56] for compound 8 (left), together with the differences between the experimental and predicted chemical shift values. The automatic structure revision proposes the 6-hydroxy-3-one derivative (right) leading to a much better, but not perfect coincidence, with the experimental data.

Another fully automatic structure revision started from the <sup>13</sup>C-NMR data and the obviously wrong structure proposal (compound 7 in [41]) given for Moracin M, as summarized in Figure 13. The structure generator created 1794 alternative structures, which were sorted by the difference between the experimental and the predicted chemical shift values. The published proposal was ranked at position 66, with an average deviation of 2.46 ppm; however, the correct structure of Moracin M is found at position 1, with an average deviation of only 1.42 ppm. This automatic structure correction was further supported when retrieving the known structure of Moracin M using CAS-Scifinder.



**Figure 13.** (left): Published structure of Moracin M (compound 7 in [41]) together with the given  $^{13}\text{C}$ -NMR data and the deviations between the experiment and prediction. (right): Fully automatic structure revision leading to excellent coincidence between the experimental and predicted chemical shift values.

### 3. Conclusions

A similar number of examples of problematic or unusable  $^{13}\text{C}$ -NMR data can be found in nearly every journal—this is simply an indication of the “publish or perish” mentality concomitant with the preference of quantity instead of quality. From the examples summarized here, the following conclusions can be drawn with the intention of improving the quality of published NMR reference data.

- Structures must be deposited as computer-readable files (e.g., MOLfile)—every structure drawing must be derived thereof in order to avoid drawing errors.
- Every structure must be accompanied by a unique identifier (e.g., INCHIKEYS), avoiding transmission errors and allowing “identical structure search” via text-based search-engines.
- Every NMR-dataset must be validated (e.g., using CSEARCH [26], MNova [18], and ACD [10]).
- Every NMR-dataset that is uploaded (e.g., NMRDATA [21,22]) must be stored in a searchable, public domain, open access, and curated repository—allowing for the automatic detection of “reusing” already known NMR-data in order to verify another structure proposal [58].
- It is highly advisable to combine a database-oriented approach, as described here, with DFT calculations [59,60].
- As many steps as possible in the process of publishing results must be done in a software-supported way. It must be mandatory for authors to provide all of the necessary experimental data (free induction decays for NMR) so that every conclusion can be reproduced, and it must be mandatory for the publishers to allow for the

uploading of such data. Furthermore, these data must be made searchable and downloadable for later use.

- Improvements in the actual “peer-reviewing” workflow, including massive computer-supported technologies, every set of spectral data has to be automatically checked during upload, and the associated protocol as described here must be an integral part of the manuscript available to the reviewer(s) and, upon publication, to the readers.

**Funding:** Open Access Funding by the University of Vienna.

**Data Availability Statement:** Data available from <https://c13nmr.at/journal/molecules/>.

**Acknowledgments:** Open Access Funding by the University of Vienna.

**Conflicts of Interest:** The author declares no conflict of interest.

**Sample Availability:** No samples available.

## References

1. Nicolaou, K.C.; Snyder, S.A. 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. [CrossRef]
2. McAlpine, J.B.; Chen, S.N.; Kutateladze, A.; MacMillan, J.B.; Appendino, G.; Barison, A.; Beniddir, M.A.; Biavatti, M.W.; Bluml, S.; Boufridi, A.; et al. The value of universally available raw NMR data for transparency, reproducibility, and integrity in natural product research. *Nat. Prod. Rep.* **2019**, *36*, 35–107. [CrossRef]
3. Robien, W. A Critical Evaluation of the Quality of Published  $^{13}\text{C}$  NMR Data in Natural Product Chemistry. In *Progress in the Chemistry of Organic Natural Products*; Kinghorn, A.D., Falk, H., Gibbons, S., Kobayashi, J., Eds.; Springer International Publishing: Cham, Switzerland, 2017; Volume 105, pp. 137–215. [CrossRef]
4. Bremser, W. HOSE—A novel substructure code. *Anal. Chim. Acta* **1978**, *103*, 355–365. [CrossRef]
5. Purtuc, V. Abschätzung von  $^{13}\text{C}$ -NMR-Spektren Mittels Neuronaler Netze. Master’s Thesis, University of Vienna, Vienna, Austria, 1997.
6. Jonas, E.; Kuhn, S. Rapid prediction of NMR spectral properties with quantified uncertainty. *J. Cheminform* **2019**, *11*, 50. [CrossRef]
7. Schütz, V.; Purtuc, V.; Felsing, 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. [CrossRef]
8. Technical Note: ACD/Labs NMR Predictors. Advanced Chemistry Development, Inc.: Toronto, ON, Canada, 21 November 2019. Available online: <https://www.acdlabs.com/download/technotes/2019/Technical-Note-Drawing-with-NMR-Predictors.pdf> (accessed on 5 May 2021).
9. Kuhn, S.; Johnson, S.R. Stereo-aware extension of HOSE-codes. *ACS Omega* **2019**, *4*, 7323–7329. [CrossRef] [PubMed]
10. NMR Predictor Software. Available online: [https://www.acdlabs.com/products/adh/nmr/nmr\\_pred/](https://www.acdlabs.com/products/adh/nmr/nmr_pred/) (accessed on 5 May 2021).
11. Satoh, H.; Koshino, H.; Uno, T.; Koichi, S.; Iwata, S.; Nakata, T. Effective consideration of ring structures in CAST/CNMR for highly accurate  $^{13}\text{C}$ -NMR chemical shift prediction. *Tetrahedron* **2005**, *61*, 7431–7437. [CrossRef]
12. Kalchauer, 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. [CrossRef]
13. KnowItAll NMR Spectral Database Collection. Available online: <https://sciencesolutions.wiley.com/solutions/technique/nmr/knowitall-nmr-collection/> (accessed on 5 May 2021).
14. NMRPredict Overview. Available online: [http://www.modgraph.co.uk/product\\_nmr.htm](http://www.modgraph.co.uk/product_nmr.htm) (accessed on 5 May 2021).
15. 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. [CrossRef]
16. SpectraBase<sup>TM</sup>. Available online: <https://spectrabase.com/> (accessed on 5 May 2021).
17. IR, UV/VIS and NMR Spectra Predictions. Available online: <http://insideinformatics.cambridgesoft.com/videosanddemos/143/ir-uvvis-and-nmr-spectra-predictions> (accessed on 5 May 2021).
18. Accurate Prediction of  $^1\text{H}$  and  $^{13}\text{C}$  NMR Spectra from a Chemical Structure. Available online: <https://mestrelab.com/software/mnova/nmr-predict/> (accessed on 5 May 2021).
19. NMR Software. Available online: <https://www.bruker.com/products/mr/nmr/software.html> (accessed on 5 May 2021).
20. Robien, W. Do high-quality  $^{13}\text{C}$ -NMR spectral data really come from journals with high impact factors? *TrAC Trends Anal. Chem.* **2009**, *28*, 914–922. [CrossRef]
21. Pupier, M.; Nuzillard, J.M.; Wist, J.; Schlörer, N.E.; Kuhn, S.; Erdelyi, M.; Steinbeck, C.; Williams, A.J.; Butts, C.; Claridge, T.D.W.; et al. NMRReDATA, a standard to report the NMR-assignment and parameters of organic compounds. *Magn. Reson. Chem.* **2018**, *56*, 703–715. [CrossRef]
22. NMRReDATA Initiative. Available online: <http://nmrredata.org/> (accessed on 5 May 2021).
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. [CrossRef]

24. Schlörer, N.; Kuhn, S. Moleküle und Strukturbeweise teilen: NMR-Spektroskopie-Datenbank für organische Moleküle. *Chem. Unserer Zeit.* **2020**, *54*, 6–13. [CrossRef]
25. Burns, D.C.; Reynolds, W.F. Minimizing the risk of deducing wrong natural product structures from NMR data. *Magn. Reson. Chem.* **2021**, *59*, 500–533. [CrossRef]
26. CSEARCH Robot Referee. Available online: <https://c13nmr.at/molecules/robot.php> (accessed on 5 May 2021).
27. Robien, W. Computer-assisted peer reviewing of spectral data: The CSEARCH-protocol. *Mh. Chem.* **2019**, *150*, 927–932. [CrossRef]
28. Haider, N.; Robien, W. Automatisierte Qualitätskontrolle von <sup>13</sup>C-NMR-Daten. *Nachr. Chem.* **2016**, *64*, 196–198. [CrossRef]
29. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Res.* **2019**, *47*, D1102–D1109. [CrossRef]
30. Index of /Pubchem/Compound/CURRENT-Full/SDF. Available online: <https://ftp.ncbi.nih.gov/pubchem/Compound/CURRENT-Full/SDF/> (accessed on 5 May 2021).
31. ChemSpider: Search and Share Chemistry. Available online: <https://www.chemspider.com> (accessed on 5 May 2021).
32. eMolecules: Empowering Drug Discovery. Available online: <https://www.emolecules.com> (accessed on 5 May 2021).
33. Huang, Y.-H.; Zeng, W.-M.; Li, G.-Y.; Liu, G.-Q.; Zhao, D.-D.; Wang, J.; Zhang, Y.-L. Characterization of a New Sesquiterpene and Antifungal Activities of Chemical Constituents from *Dryopteris fragrans* (L.) Schott. *Molecules* **2014**, *19*, 507–513. [CrossRef] [PubMed]
34. Baba, K.; Kawanishi, H.; Taniguchi, M.; Kozawa, M. Chromones from *Cnidium monnieri*. *Phytochemistry* **1992**, *31*, 1367–1370. [CrossRef]
35. Chazin, E.D.L.; Sanches, P.D.S.; Lindgren, E.B.; Vellasco Júnior, W.T.; Pinto, L.C.; Burbano, R.M.R.; Yoneda, J.D.; Leal, K.Z.; Gomes, C.R.B.; Wardell, J.L.; et al. Synthesis and Biological Evaluation of Novel 6-Hydroxy-benzo[d][1,3]oxathiol-2-one Schiff Bases as Potential Anticancer Agents. *Molecules* **2015**, *20*, 1968–1983. [CrossRef]
36. Turan, N.; Körkoca, H.; Adigüzel, R.; Çolak, N.; Buldurun, K. Synthesis, Structural Characterization and Biological Activity of Novel Cyclohexane-1,3-dione Ligands and Their Metal Complexes. *Molecules* **2015**, *20*, 9309–9325. [CrossRef]
37. Alasmay, F.A.S.; Snelling, A.M.; Zain, M.E.; Alafeefy, A.M.; Awaad, A.S.; Karodia, N. Synthesis and Evaluation of Selected Benzimidazole Derivatives as Potential Antimicrobial Agents. *Molecules* **2015**, *20*, 15206–15223. [CrossRef] [PubMed]
38. Nagappan, T.; Segaran, T.C.; Wahid, M.E.A.; Ramasamy, P.; Vairappan, C.S. Efficacy of Carbazole Alkaloids, Essential Oil and Extract of *Murraya koenigii* in Enhancing Subcutaneous Wound Healing in Rats. *Molecules* **2012**, *17*, 14449–14463. [CrossRef] [PubMed]
39. Estork, D.M.; Gusmão, D.F.; Paciencia, M.L.B.; Díaz, I.E.C.; Varella, A.D.; Younes, R.N.; Reis, L.F.L.; Montero, E.F.S.; Bernardi, M.M.; Suffredini, I.B. First Chemical Evaluation and Toxicity of *Casingsa-cheirosa* to Balb-c Male Mice. *Molecules* **2014**, *19*, 3973–3987. [CrossRef] [PubMed]
40. González-Olvera, R.; Espinoza-Vázquez, A.; Negrón-Silva, G.E.; Palomar-Pardavé, M.E.; Romero-Romo, M.A.; Santillan, R. Multicomponent Click Synthesis of New 1,2,3-Triazole Derivatives of Pyrimidine Nucleobases: Promising Acidic Corrosion Inhibitors for Steel. *Molecules* **2013**, *18*, 15064–15079. [CrossRef] [PubMed]
41. Sun, J.; He, X.-M.; Zhao, M.-M.; Li, L.; Li, C.-B.; Dong, Y. Antioxidant and Nitrite-Scavenging Capacities of Phenolic Compounds from Sugarcane (*Saccharum officinarum* L.) Tops. *Molecules* **2014**, *19*, 13147–13160. [CrossRef] [PubMed]
42. Martínez-Martínez, F.J.; Razo-Hernández, R.S.; Peraza-Campos, A.L.; Villanueva-García, M.; Sumaya-Martínez, M.T.; Cano, D.J.; Gómez-Sandoval, Z. Synthesis and in Vitro Antioxidant Activity Evaluation of 3-Carboxycoumarin Derivatives and QSAR Study of Their DPPH• Radical Scavenging Activity. *Molecules* **2012**, *17*, 14882–14898. [CrossRef]
43. Zhou, Y.; Yang, B.; Jiang, Y.; Liu, Y.; Wang, X.; Kuang, H. Studies on Cytotoxic Activity against HepG-2 Cells of Naphthoquinones from Green Walnut Husks of *Juglans mandshurica* Maxim. *Molecules* **2015**, *20*, 15572–15588. [CrossRef]
44. Wang, C.F.; Yang, K.; Zhang, H.M.; Cao, J.; Fang, R.; Liu, Z.L.; Du, S.S.; Wang, Y.Y.; Deng, Z.W.; Zhou, L. Components and Insecticidal Activity against the Maize Weevils of *Zanthoxylum schinifolium* Fruits and Leaves. *Molecules* **2011**, *16*, 3077–3088. [CrossRef] [PubMed]
45. Zhao, X.; Mei, W.; Gong, M.; Zuo, W.; Bai, H.; Dai, H. Antibacterial Activity of the Flavonoids from *Dalbergia odorifera* on *Ralstonia solanacearum*. *Molecules* **2011**, *16*, 9775–9782. [CrossRef]
46. Han, L.; Ji, L.; Boakye-Yiadom, M.; Li, W.; Song, X.; Gao, X. Preparative Isolation and Purification of Four Compounds from *Cistanches deserticola* Y.C. Ma by High-Speed Counter-Current Chromatography. *Molecules* **2012**, *17*, 8276–8284. [CrossRef] [PubMed]
47. Kwiecień, A.; Ciunik, Z. Stable Hemiaminals: 2-Aminopyrimidine Derivatives. *Molecules* **2015**, *20*, 14365–14376. [CrossRef] [PubMed]
48. Li, H.Q.; Liu, Q.Z.; Liu, Z.L.; Du, S.S.; Deng, Z.W. Chemical Composition and Nematicidal Activity of Essential Oil of *Agastache rugosa* against *Meloidogyne incognita*. *Molecules* **2013**, *18*, 4170–4180. [CrossRef]
49. Popoola, O.K.; Marnewick, J.L.; Rautenbach, F.; Iwuoha, E.I.; Hussein, A.A. Acylphloroglucinol Derivatives from the South African *Helichrysum niveum* and Their Biological Activities. *Molecules* **2015**, *20*, 17309–17324. [CrossRef] [PubMed]
50. Filho, H.D.S.R.; Pacheco, L.C.; Andrade, E.D.S.; Correa, M.J.C.; Araújo, R.N.M.; Guilhon, G.M.S.P.; Silva, J.K.R.d.; Santos, L.S. Xanthones from the Roots of *Moutabea guianensis* Aubl. *Molecules* **2015**, *20*, 127–134. [CrossRef]
51. Petroski, R.J.; Vermillion, K.; Cossé, A.A. Two-Carbon Homologation of Aldehydes and Ketones to  $\alpha,\beta$ -Unsaturated Aldehydes. *Molecules* **2011**, *16*, 5062–5078. [CrossRef]

52. Meng, L.-Z.; Huang, W.-H.; Wang, C.-Z.; Yuan, C.-S.; Li, S.-P. Anticancer Activities of Polyynes from the Root Bark of *Oplonanax horridus* and Their Acetylated Derivatives. *Molecules* **2014**, *19*, 6142–6162. [[CrossRef](#)] [[PubMed](#)]
53. Marti, G.; Eparvier, V.; Litaudon, M.; Grellier, P.; Guéritte, F. A New Xanthone from the Bark Extract of *Rheedia acuminata* and Antiplasmodial Activity of Its Major Compounds. *Molecules* **2010**, *15*, 7106–7114. [[CrossRef](#)] [[PubMed](#)]
54. Syahri, J.; Yuanita, E.; Nurohmah, B.A.; Wathon, M.H.; Syafri, R.; Armunanto, R.; Purwono, B. Xanthone as Antimalarial: QSAR Analysis, Synthesis, Molecular Docking and In-vitro Antimalarial Evaluation. *Orient J. Chem.* **2017**, *33*, 29–40. [[CrossRef](#)]
55. Andriani, Y.; Tengku-Muhammad, T.S.; Mohamad, H.; Saidin, J.; Syamsumir, D.F.; Chew, G.-S.; Abdul Wahid, M.E. *Phaleria macrocarpa* Boerl. (*Thymelaeaceae*) Leaves Increase SR-BI Expression and Reduce Cholesterol Levels in Rats Fed a High Cholesterol Diet. *Molecules* **2015**, *20*, 4410–4429. [[CrossRef](#)] [[PubMed](#)]
56. Huang, Y.; Cui, J.; Jia, L.; Gan, C.; Song, H.; Zeng, C.; Zhou, A. Synthesis and Evaluation of Some 17-Acetamidoandrostane and *N,N*-Dimethyl-7-deoxycholic Amide Derivatives as Cytotoxic Agents: Structure/Activity Studies. *Molecules* **2013**, *18*, 7436–7447. [[CrossRef](#)] [[PubMed](#)]
57. Holland, H.L.; Lakshmaiah, G.; Ruddock, P.L. Microbial hydroxylation of acetylamino steroids. *Steroids* **1998**, *63*, 484–495. [[CrossRef](#)]
58. Dieltiens, N.; Claeys, D.D.; Zhdankin, V.V.; Nemykin, V.N.; Allaert, B.; Verpoort, F.; Stevens, C.V. The Pyroglutamate Hydantoin Rearrangement. *Eur. J. Org. Chem.* **2018**, 3372. [[CrossRef](#)]
59. Krivdin, L.B. Computational protocols for calculating  $^{13}\text{C}$  NMR chemical shifts. *Prog. Nucl. Magn. Reson. Spectrosc.* **2019**, *112–113*, 103–156. [[CrossRef](#)] [[PubMed](#)]
60. Semenov, V.A.; Krivdin, L.B. Computational  $^1\text{H}$  and  $^{13}\text{C}$  NMR of strychnobaillonine: On the way to larger molecules calculated at lower computational costs. *Magn. Reson. Chem.* **2021**, *59*, 108–116. [[CrossRef](#)]