

PRELIMINARY COMMUNICATION

Open Access

Theoretical NMR correlations based Structure Discussion

Jochen Junker

Abstract

The constitutional assignment of natural products by NMR spectroscopy is usually based on 2D NMR experiments like COSY, HSQC, and HMBC. The actual difficulty of the structure elucidation problem depends more on the type of the investigated molecule than on its size. The moment HMBC data is involved in the process or a large number of heteroatoms is present, a possibility of multiple solutions fitting the same data set exists. A structure elucidation software can be used to find such alternative constitutional assignments and help in the discussion in order to find the correct solution. But this is rarely done. This article describes the use of theoretical NMR correlation data in the structure elucidation process with WEBCOCON, not for the initial constitutional assignments, but to define how well a suggested molecule could have been described by NMR correlation data. The results of this analysis can be used to decide on further steps needed to assure the correctness of the structural assignment. As first step the analysis of the deviation of carbon chemical shifts is performed, comparing chemical shifts predicted for each possible solution with the experimental data. The application of this technique to three well known compounds is shown. Using NMR correlation data alone for the description of the constitutions is not always enough, even when including ^{13}C chemical shift prediction.

Findings

Nuclear Magnetic Resonance allied with Elemental analysis or high resolution Mass Spectroscopy are the most common tools used for the structure elucidation of new compounds. The used 2D NMR experiments like COSY, HSQC, and ^{13}C -HMBC deliver correlation information between atoms that can be translated into connectivity information. Out of these, correlation information from COSY and HSQC experiments can be transcribed directly into connectivity between atoms. But the ^{13}C -HMBC correlations need more attention because of their ambiguity and complexity. Hence the difficulty of the structure elucidation problem depends more on the type of the investigated molecule than on its size [1]. Saturated compounds can usually be assigned unambiguously using mainly COSY and some ^{13}C -HMBC data, whereas condensed heterocycles are problematic due to their lack of protons that could show interatomic connectivities. This ambiguity has driven the development of different software packages to aid in the interpretation of the ^{13}C -HMBC correlation data [2-20] as

much as the development of additional correlation experiments [21,22].

Most of these approaches have in common that they work only based on experimental NMR correlation data. COCON [1,4,23,24] has recently been extended with the capability to create a theoretical NMR correlation data set, based on a molecule's suggested constitution. The theoretical data set is used as input data for the structure elucidation software COCON. The resulting set of constitutional assignments indicates how unambiguous NMR would have been able to describe the originally suggested molecule. The freely accessible online version of COCON (WEBCOCON at <http://cocon.nmr.de>) offers this analysis as "Alternative Constitutions".

The data derived from the NMR correlation spectra is the result of magnetization transfer via scalar coupling between the atoms in the molecule of interest. Since the scalar coupling is based on the interatomic bonds, the correlation data will reflect those bonds. Hence, a set of all feasible NMR correlation data (theoretical correlation data) can be derived from the molecular constitution. This is done by iteratively looking for all protons in the molecule, then building a list of their atoms in 2-bond and 3-bond distance. From each proton all

Correspondence: junker@cdfs.fiocruz.br
Fundação Oswaldo Cruz - CDTS, Rio de Janeiro - RJ, Brazil

connectivities are inspected recursively up to three bonds distance. If a carbon is found in a two bond distance, a 2J and a 1,1-ADEQUATE correlation are added to the list. If a carbon is found in a three bond distance, a HMBC correlation is added to the list, if a proton is found, a COSY correlation is added. In principle 4J correlations for COSY and HMBC could be generated, as sometimes they are observable in experiments as well. But, COCON can not handle 4J COSY correlations, therefore those are left out. The generation of 4J HMBC correlations is not used, because when the HMBC correlations are allowed to be 4J in the structure generation process, the process takes much more time and many more results are produced. Finally carbon chemical shifts are generated by table lookup, a table reverse generated based on the chemical shift rules that COCON uses. This values are not comparable to a chemical shift prediction, but enough to ensure that COCON will generate the starting structure.

For online use, the MarvinSketch applet from ChemAxon is available for drawing or loading of the molecule. The resulting MDL file contains all atoms, their connectivity and multiplicity information. Based on this file, the recently developed Module "Alternative Constitutions" in WEBCOCON generates atomtypes, theoretical correlation data and table-based carbon chemical shifts.

The actual magnitude of the scalar coupling, and therefore the observability of a correlation, depends on the atoms involved, their chemical environment and relative geometry. For 1J and 2J couplings mainly the atoms involved and their chemical environment are of importance, since the geometry varies little. That is different with 3J coupling, which depends on the dihedral angle, hence the actual molecular conformation decides on the magnitude of the coupling. The creation of theoretical correlation data disregards the molecule's real conformation, assuming that all correlations are observable. Hence the data set represents the upper limit of correlations that may be experimentally available for the constitution.

Calculations were run with three molecules (Figure 1) on the publicly available WEBCOCON server, running times varied from one to twelve minutes. All molecules were drawn in the "Alternative Constitutions" module and submitted to the server. The number of solutions suggested for Ascomycin **1** and Oroidin **2** in runs with theoretical and experimental data are shown in table 1. Also, a webpage allowing direct access to the results shown here has been set up on the WEBCOCON server at <http://cocon.nmr.de/StructureDiscussion/> (The results are mirrored at <http://science.jotjot.net/StructureDiscussion/>).

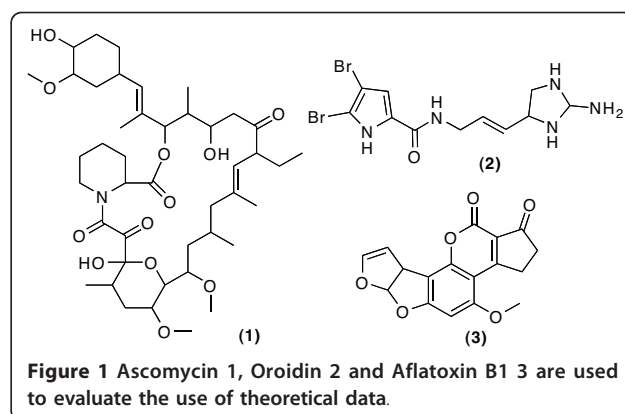


Figure 1 Ascomycin **1**, Oroidin **2** and Aflatoxin B1 **3** are used to evaluate the use of theoretical data.

Ascomycin **1** is a well known ethyl derivative of Tacrolimus, it serves as example of a large natural product, featuring 43 Carbon atoms. Using theoretical NMR correlation data (COSY and ^{13}C -HMBC correlations) COCON generates only one solution, independent of whether atom types are defined or not. Using experimental COSY and ^{13}C -HMBC correlation data the structure generator comes up with 100 structural assignments, which are reduced to one when the atom types are fixed as well. In this case NMR correlation data was able to define the constitution unambiguously.

Oroidin **2** has been frequently used for the demonstration of COCON. The use of theoretical COSY and ^{13}C -HMBC correlations leads to a total of 16 possible constitutional assignments, also predefining the atom types reduces this set to one constitutional assignment. The experimental data set leads to 252,566 structural assignments generated, which reduce to 1,486 when atom types are predefined as well. Hence the structure can not be safely determined by NMR alone. The original structure determination was carried out by chemical derivatization and total synthesis [25,26].

The pictures change with Aflatoxin B1 **3** with 17 Carbon atoms. Using theoretical COSY and ^{13}C -HMBC data alone, COCON generates 1,048 structures, compared to 1,932 solutions using experimental data. When the atom types are predefined, COCON generates 55 constitutional assignments, compared to 108 with experimental data. The molecule set generated contains constitutions with the element cyclobutadiene, a

Table 1 Number of constitutional assignments suggested for **1** and **2**.

	open atom types		fixed atom types	
	theo	exp	theo	exp
1	1	100	1	1
2	16	252.566	1	1486

structural element that is very uncommon in natural products. COCON has several built-in rules that eliminate certain constitutional elements, like cyclobutadiene, cyclopropene and peroxides. By default these rules are not used, but in this special case we observed a substantial difference in the number of results.

When these rules are activated the number of solutions drops to 58 for the experimental correlation data set and 33 for the theoretical data set. All planar molecules suggested are shown in Figure 2, the correct constitution and starting point of the analysis is **6**. For the small number of interesting constitutions a back-calculation on the carbon chemical shifts was made (ChemDraw v11), that were compared to the experimental values (see table 2). The last line in the table contains the sum of the absolute chemical shift differences for all carbons, exposing molecule **6** as the one that best fits the experimental data [24,27,28].

The theoretical NMR correlation dataset is the upper limit of number of correlations that are possible with a given constitution. Therefore all alternative constitutions generated with this data are "NMR-identical" with regard to correlation data. A careful analysis of this alternatives might be used to direct further investigations needed to confirm the proposed constitution. Whilst Ascomycin's structure can be confirmed by NMR correlations, Oroidin's structure can not. The results obtained would direct further work towards chemical derivatization and synthesis [25,26] or x-ray crystallography. The results obtained for Aflatoxin B1 show nicely how carbon chemical

Table 2 Experimental and predicted ^{13}C chemical shifts for the different constitutions suggested for Aflatoxin B1.

	exp	^{13}C shifts for molecule							
		4	5	6	7	8	9	10	
C	201	170	167	200	203	202	190	190	
C	177	167	166	177	161	167	175	166	
C	166	163	163	161	155	154	169	162	
C	161	152	154	159	150	145	158	160	
C	155	149	149	159	137	144	153	154	
C	153	140	142	151	136	141	152	127	
C	117	128	125	123	133	126	140	122	
C	108	117	116	111	129	122	131	121	
C	104	112	113	104	128	107	104	114	
CH	145	149	149	149	149	149	149	149	
CH	114	106	106	100	105	105	105	107	
CH	103	105	105	100	100	94	93	101	
CH	91	93	93	97	100	88	93	100	
CH	48	43	46	45	48	50	43	45	
CH ₂	35	35	35	34	30	32	33	33	
CH ₂	29	24	21	28	25	31	14	14	
CH ₃	57	53	53	56	56	59	58	52	
$\sum \Delta\delta $		130	127	56	171	122	116	129	

shift prediction can be used as tool for the structure discussion, exposing one suggested constitutional assignment as best fitting.

Availability

The WEBCOCON server is freely accessible via <http://cocon.nmr.de>.

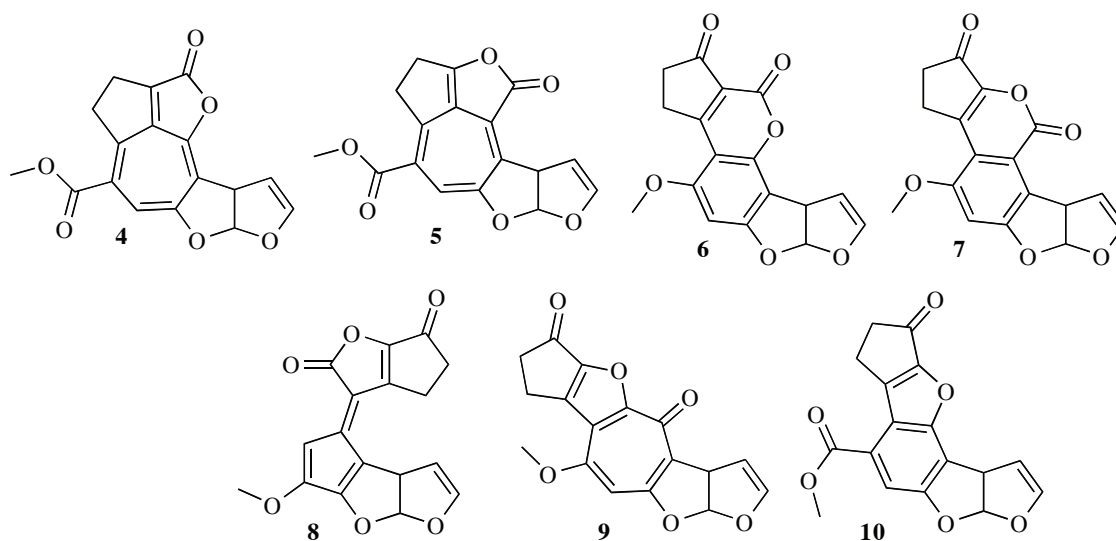


Figure 2 Planar constitutions suggested for Aflatoxin B1. Suggestions 4 - 6 are obtained using theoretical data, 5 - 10 using experimental data. Constitution 6 is the correct one.

Acknowledgements

The author wishes to acknowledge Rainer Haessner and the Technische Universität München for providing the Hardware for the WEBCOCON Server.

Authors' contributions

JJ maintains the WEBCOCON software and has run all the calculations shown.

Competing interests

The author declares that they have no competing interests.

Received: 19 April 2011 Accepted: 28 July 2011 Published: 28 July 2011

References

- Junker J, Maier W, Lindel T, Kock M: Computer-assisted constitutional assignment of large molecules: COCON analysis of ascomycin. *Org Lett* 1999, **1**:737-740.
- Elyashberg M, Williams A, Martin G: Computer-assisted structure verification and elucidation tools in NMR-based structure elucidation. *Prog Nucl Mag Res Sp* 2008, **53**(1-2):1-104.
- Peng C, Bodenhausen G, Qiu S, Fong H, Farnsworth N, Yuan S, Zheng C: Computer-assisted structure elucidation: Application of CISOC-SES to the resonance assignment and structure generation of betulinic acid. *Magn Reson Chem* 1998, **36**(4):267-278.
- Lindel T, Junker J, Kock M: COCON: From NMR correlation data to molecular constitutions. *J Mol Model* 1997, **3**:364-368.
- Stefani R, Nascimento P, Costa F: Computer-aided structure elucidation of organic compounds: Recent advances. *Quim Nova* 2007, **30**(5):1347-1356.
- Elyashberg M, Blinov K, Molodtsov S, Williams A, Martin G: Fuzzy structure generation: A new efficient tool for computer-aided structure elucidation (CASE). *J Chem Inf Model* 2007, **47**(3):1053-1066.
- Smurnyy Y, Elyashberg M, Blinov K, Lefebvre B, Martin G, Williams A: Computer-aided determination of relative stereochemistry and 3D models of complex organic molecules from 2D NMR spectra. *Tetrahedron* 2005, **61**(42):9980-9989.
- Sharman G, Jones I, Parnell M, Willis M, Mahon M, Carlson D, Williams A, Elyashberg M, Blinov K, Molodtsov S: Automated structure elucidation of two unexpected products in a reaction of an alpha, beta-unsaturated pyruvate. *Magn Reson Chem* 2004, **42**(7):567-572.
- Steinbeck C: Recent developments in automated structure elucidation of natural products. *Nat Prod Rep* 2004, **21**(4):512-518.
- Schulz K, Korytko A, Munk M: Applications of a HOUDINI-based structure elucidation system. *J Chem Inf Comp Sci* 2003, **43**(5):1447-1456.
- Steinbeck C: SENECA: A platform-independent, distributed, and parallel system for computer-assisted structure elucidation in organic chemistry. *J Chem Inf Comp Sci* 2001, **41**(6):1500-1507.
- Steinbeck C: Recent advancements in the development of SENECA, a computer program for Computer Assisted Structure Elucidation based on a stochastic algorithm. *Abstr Pap Am Chem S* 1999, **218**:U360-U360.
- Strokov I, Lebedev K: Computer aided method for chemical structure elucidation using spectral databases and C-13 NMR correlation tables. *J Chem Inf Comp Sci* 1999, **39**(4):659-665.
- Madison M, Schulz K, Korytko A, Munk M: SESAMI: An integrated desktop structure elucidation tool. *Internet J Chem* 1998, **1**(34):CP1-U22.
- Steinbeck C: LUCY - A program for structure elucidation from NMR correlation experiments. *Angew Chem Int Edit* 1996, **35**(17):1984-1986.
- Bangov I, Laude I, Cabrolbass D: Combinatorial Problems in the Treatment of fuzzy C-13 NMR Spectral Information in the Process of Computer-Aided Structure Elucidation - Estimation of the Carbon-Atom Hybridization and Alpha-Environment States. *Anal Chim Acta* 1994, **298**:33-52.
- Funatsu K: Computer-Assisted Structure Elucidation for Organic-Compound. *J Syn Org Chem Jpn* 1993, **51**(6):516-528.
- Lebedev K, Nekhoroshev S, Kirshansky S, Derendjaev B: Computer Method of Fragmentary Formula Prediction of an unknown by its Mass and NMR-Spectra. *Sibirskii Khim Zh+* 1992, **3**: 72-79.
- Guzowskaswider B, Hippe Z: Structure Elucidation of organic-compounds aided by the Computer-Program System Scannet. *J Mol Struct* 1992, **275**:225-234.
- Nuzillard J, Massiot G: Computer-Aided Spectral Assignment in NMR Spectroscopy. *Anal Chim Acta* 1991, **242**:37-41.
- Reif B, Kock M, Kerssebaum R, Kang H, Fenical W, Griesinger C: ADEQUATE, a new set of experiments to determine the constitution of small molecules at natural abundance. *J Magn Reson Ser A* 1996, **118**(2):282-285.
- Kock M, Junker J, Lindel T: Impact of the H-1, N-15-HMBC experiment on the constitutional analysis of alkaloids. *Org Lett* 1999, **1**:2041-2044.
- Lindel T, Junker J, Kock M: 2D-NMR-guided constitutional analysis of organic compounds employing the computer program COCON. *Eur J Org Chem* 1999, **573**-577.
- Kock M, Junker J, Maier W, Will M, Lindel T: A COCON analysis of proton-poor heterocycles - Application of carbon chemical shift predictions for the evaluation of structural proposals. *Eur J Org Chem* 1999, **579**-586.
- Garcia E, Benjamin L, Fryer R: Reinvestigation into structure of Oroidin, a bromopyrrole derivative from marine sponge. *J Chem Soc Chem Comm* 1973, **3**: 78-79.
- Forenza S, Minale L, Riccio R: New bromo-pyrrole derivatives from sponge Agelas-Oroides. *J Chem Soc Chem Comm* 1971, **18**: 1129-1130.
- Meiler J, Sanli E, Junker J, Meusinger R, Lindel T, Will M, Maier W, Kock M: Validation of structural proposals by substructure analysis and C-13 NMR chemical shift prediction. *J Chem Inf Comp Sci* 2002, **42**:241-248.
- Meiler J, Kock M: Novel methods of automated structure elucidation based on C-13 NMR spectroscopy. *Magn Reson Chem* 2004, **42**(12):1042-1045.

doi:10.1186/1758-2946-3-27

Cite this article as: Junker: Theoretical NMR correlations based Structure Discussion. *Journal of Cheminformatics* 2011 **3**:27.

Publish with **ChemistryCentral** and every scientist can read your work free of charge

"Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge."

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.chemistrycentral.com/manuscript/

