

## CoCON: From NMR Correlation Data to Molecular Constitutions<sup>§</sup>

Thomas Lindel<sup>1,\*</sup>, Jochen Junker<sup>2</sup>, and Matthias Köck<sup>2,\*</sup>

<sup>1</sup>Pharmazeutisch-Chemisches Institut der Universität, Im Neuenheimer Feld 364, D-69120 Heidelberg, Germany; Phone: +49-6221-54-4857; Fax: +49-6221-54-6430 (lindel@convex.phazc.uni-heidelberg.de)

<sup>2</sup>Institut für Organische Chemie, Johann Wolfgang Goethe Universität, Marie-Curie-Straße 11, D-60439 Frankfurt, Germany; Phone: +49-69-798-29143; Fax: +49-69-798-29128 (km@org.chemie.uni-frankfurt.de)

Received: 5 June 1997 / Accepted: 7 July 1997 / Published: 15 August 1997

### Abstract

The constitutional assignment of natural products by NMR spectroscopy is usually based on 2D NMR experiments like COSY, HSQC, and HMBC. In this paper, the resulting connectivity information is used as input for the new structure generating program CoCON which both improves and dramatically accelerates the process of constitutional assignment. CoCON allows to quantify the value of connectivity information (2D NMR correlation data) for structure elucidation problems. Applying CoCON, it is systematically evaluated to which degree the NMR experiments COSY, <sup>1</sup>H, <sup>13</sup>C-HMBC and 1,1-ADEQUATE constrain the number of constitutions compatible with the data sets of two secondary metabolites from marine sponges.

**Keywords:** Computer-assisted structure elucidation, Constitutional analysis, HMBC, ADEQUATE, NMR software

### Introduction

The structure elucidation of unknown natural products has always been a fascinating challenge to chemists. The identification of novel compounds from a variety of data becomes especially exciting if biological activity is involved. NMR experiments such as COSY [1], HMQC [2] or HSQC [3], and HMBC [4] have proven to be powerful in this respect and the development of novel NMR experimental tools is a continuing process. A recent example is the 1,1-ADEQUATE

experiment which allows the observation of two-bond correlations [5]. This new 2D NMR tool is extremely useful for distinguishing between <sup>2</sup>J<sub>CH</sub> and <sup>3</sup>J<sub>CH</sub> correlations obtained from HMBC spectra [5b].

### Method

The computer program CoCON (“CoNstitutions from CoNnectivities”) has been developed to serve natural product chemists in discussing the constitutions of unknown compounds for which crystal structures are not available [6]. It allows chemists to comprehensively discuss the entirety of molecular constitutions compatible with the given NMR data and the molecular formula.

<sup>§</sup> Presented at the 11. Molecular Modeling Workshop, 6 -7 May 1997, Darmstadt, Germany

\* To whom correspondence should be addressed

**Table 1.** Cocon results for 4,5-dibromopyrrol-2-carboxylic acid (**1**) and manzacidin (**2**).

NMR correlation Data	<b>1</b>	<b>1</b>	<b>2</b>
	(data set A) [a]	(data set B) [a]	
COSY, HMBC	10	6	190
COSY, HMBC, 1,1-ADEQUATE ( <b>non_adeq</b> = 0) [b]	3	2	40
COSY, HMBC, 1,1-ADEQUATE ( <b>non_adeq</b> = 1) [b]	3	2	24

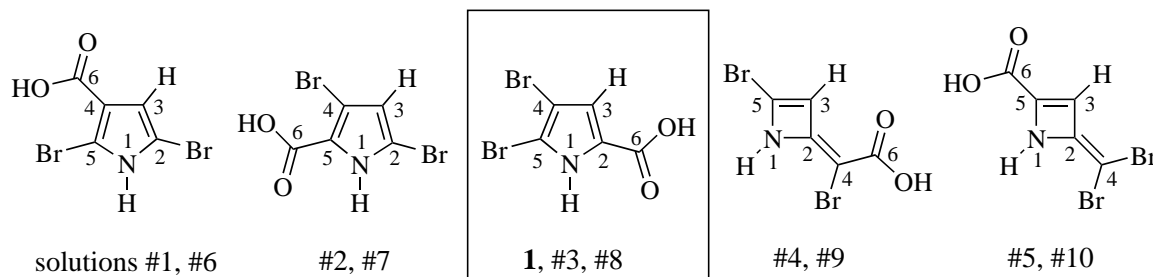
[a] The following correlations are given starting from H-3. Data set A: HMBC C-2, C-5; data set B: HMBC C-2, C-4, C-5 and C-6. The 1,1-ADEQUATE correlations to C-2 and C-4 are obtained in both cases. In carefully dried solutions, it is possible to observe correlations from the H-1 (NH) to C-2, C-3, C-4, and C-5 in the HMBC spectrum. The number of the generated structures was the same as for data set B.

[b] The explanation of the *non\_adeq* flag is given in the text.

A simple problem is the structure elucidation of 4,5-dibromopyrrol-2-carboxylic acid (**1**, see Figure 1), an immunosuppressive natural product from the marine sponge *Agelas* sp. [7]. COCON proposes 318 different constitutions for **1** if only the hybridization states of the involved nuclei and no 2D NMR correlations are taken into account. 157 solutions

remain, if three-membered rings with at least two sp<sup>2</sup>-hybridized atoms as well as cyclobutadienoid substructures are excluded. When applying very limited <sup>13</sup>C NMR chemical shift rules [a] and considering the experimentally observed HMBC correlations (data set A, correlations from H-3 to C-2 and C-5) only proposals #1 to #5 respectively their assignment isomers [b] were obtained (see Figure 1 and Table 1). On the basis of data set B (additional HMBC correlations from H-3 to C-4 and C-6) proposals #1, #3 and #5 and their assignment isomers were generated (see fig. 1 and table 1). All COCON calculations were carried out assuming any direct connection of two CH fragments to be forbidden if no COSY correlation was observed (*non\_cosy* = 1). This restriction can optionally be extended to heteroatoms (*non\_cosy* = 2).

The 1,1-ADEQUATE experiment [5a,b] allows the selective observation of two-bond correlations (pseudo <sup>2</sup>J<sub>CH</sub> couplings, which consists of a <sup>1</sup>J<sub>CH</sub> and a <sup>1</sup>J<sub>CC</sub> step). For **1**,



**Figure 1.** Ten different assignments representing five different constitutions were obtained for the data set of 4,5-dibromopyrrol-2-carboxylic acid (**1**, solution #3). In solutions #6 to #10, the assignments of the carbon atoms C-2 and C-5 are interchanged.

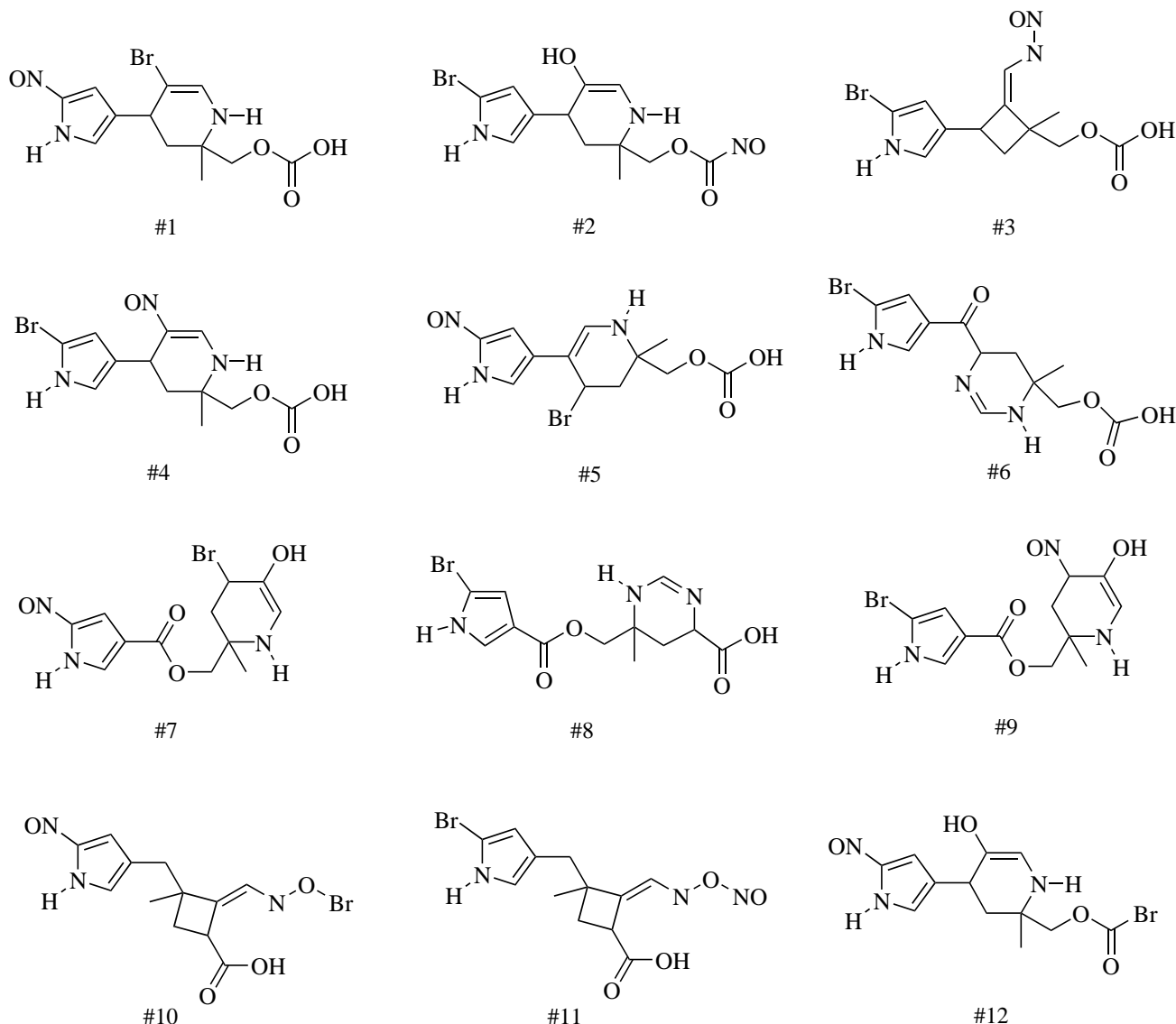
[a] COCON obeys the following <sup>13</sup>C NMR chemical shift rules:

- a carbon atom cannot be a C=S and C=O if  $\delta_C < 150$  ppm;
- an olefinic carbon atom bound to an oxygen is forbidden if  $\delta_C < 130$  ppm;
- an olefinic carbon atom bound to a nitrogen is forbidden if  $\delta_C < 100$  ppm;
- an aliphatic carbon atom bound to an oxygen is forbidden if  $\delta_C < 45$  ppm;
- a methyl group cannot be bound to a carbon atom if  $\delta_C > 35$  ppm.

correlations from H-3 to C-2 and C-4 were observed in the 1,1-ADEQUATE spectrum, leading to a reduction from 10 to 3 proposed structures (#1, #3 and #5) for data set A and to 2 proposed structures (#1 and #3) for data set B.

The second example discussed here is manzacidin A (**2**), an alkaloid from the Okinawan sponge *Hymeniacidon* sp. [8]. In addition to its 4-bromopyrrol-2-carbonyl unit, it possesses a tetrahydropyrimidine moiety. Based on the COSY and HMBC data and on the hybridization states reported in

[b] The carbon atoms C-2 and C-5 of the solutions #1 to #5 may be interchanged without violation of any connectivity or other constraints. The resulting solutions #6 to #10 represent identical constitutions, but different <sup>13</sup>C NMR assignments which are of importance for subsequent <sup>13</sup>C NMR chemical shift calculations.



**Figure 2a.** Constitutions #1 to #12 calculated by the COCON computer program based on the reported data and on 1,1-ADEQUATE correlations to be expected for manzacidin A (2).

the original publication [8], 190 different constitutional assignments are still possible (see Table 1). Of course, these COCON proposals include a large share of constitutions which do not agree with the rules of chemical stability or of  $^{13}\text{C}$  NMR chemical shifts.

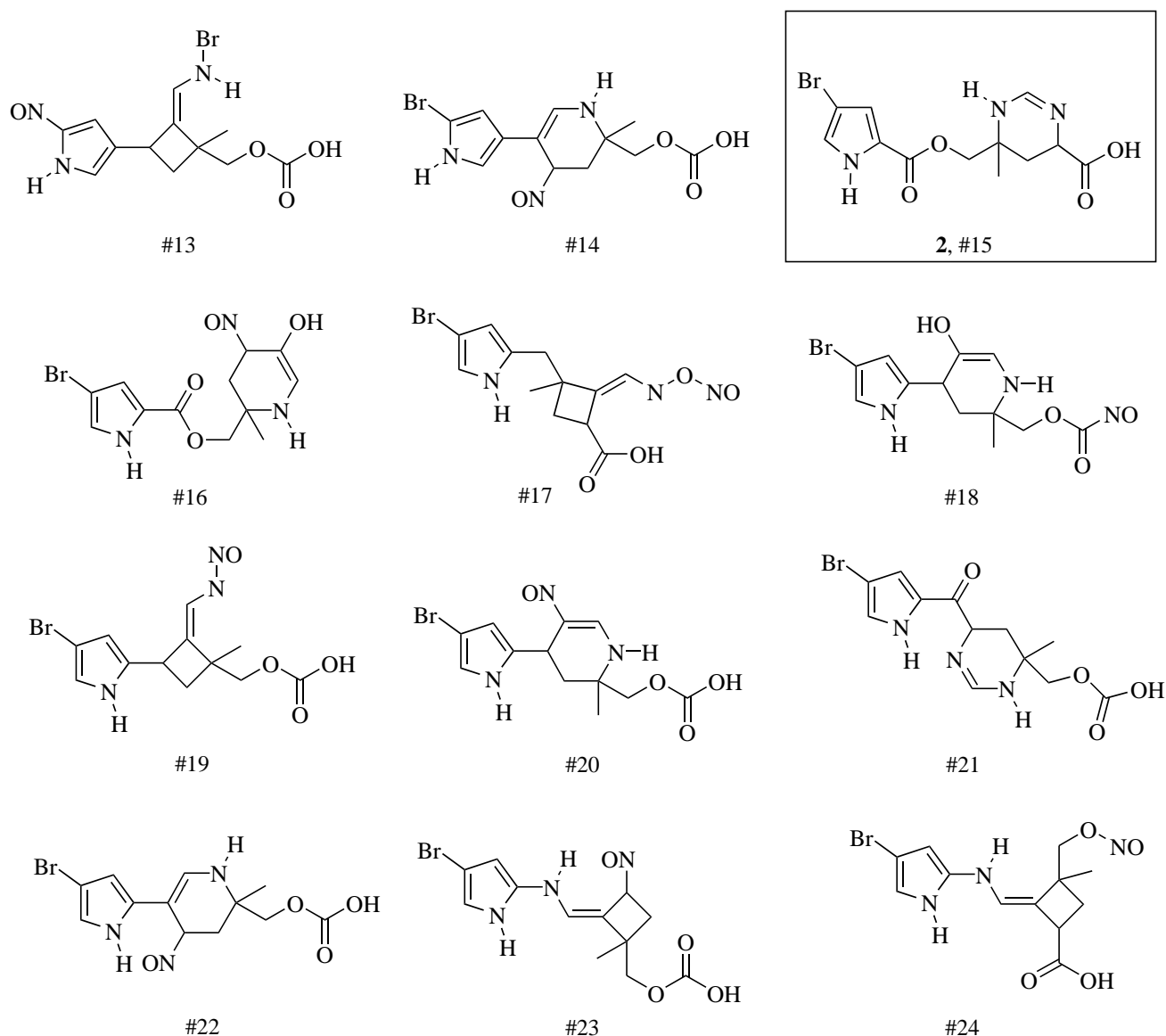
The number of compatible solutions is substantially diminished if 1,1-ADEQUATE correlations are taken into consideration. If all correlations to be theoretically expected for the correct structure #15 of manzacidin A (2, see Figure 2B) are assumed, 40 proposals are generated (see Table 1). As soon as bonds between two carbon atoms are prohibited, if at least one of them is proton-bearing and no 1,1-ADEQUATE correlation is observed between them (non\_adeq = 1), only 24 different constitutions are proposed by Cocon (see Fig-

ures 2a and 2b) [c]. The calculation results include a real alternative (solution #8) to manzacidin A (2, solution #15). The average deviations of the experimental  $^{13}\text{C}$  NMR chemical shifts of the four pyrrole ring carbon atoms from calculated values [d] were 9.2 ppm for solution #8 and 4.2 ppm for solution #15 (manzacidin A (2)), favoring the correct structure. Most other proposed structures represent carbonic acid hemiesters or otherwise chemically unstable constitutions.

The calculation time for the manzacidin problem on a personal computer (Pentium 133 MHz) was 6 min. 24 seconds, if only the hybridization states, COSY, and HMBC correlations were considered, and 2 seconds if theoretical 1,1-ADEQUATE correlations were also included. The high calculation speed of the COCON program is mainly based on

[c] Cyclobutadienoid and cyclopropenoid constitutions were excluded by default.

[d] The increment calculation started from the pyrrole chemical shifts and used the increments of the vinyl system. See also: Foley, L. H.; Habgood, G. I.; Gallagher, K. S. *Magn. Res. Chem.* **1988**, 26, 1037-1038.



**Figure 2b.** Constitutions #13 to #24 calculated by the COCON computer program based on the reported data and on 1,1-ADEQUATE correlations to be expected for manzacidin A (2). The correct structure of manzacidin A (2, solution #15) is highlighted.

the integrated evaluation of HMBC information simultaneously with the construction of constitutions. The detection of constitutionally equivalent isomers at the earliest possible stage of their generation proved to be another very efficient step, especially for natural products containing several groups of equivalent heteroatoms.

Although it may seem overwhelming to obtain several hundreds or thousands of solutions, each proposed structure should be treated as a hypothesis [9] and *e. g.* be subject to a subsequent, accurate calculation of NMR chemical shifts and of molecular mechanics. The final decision on the constitu-

tion of a natural product of course relies on additional information obtained from other spectroscopic methods or on chemical or biosynthetic considerations. COCON may also be applied after a structure of a compound was proposed. This is especially important for proton-poor compounds to validate a structural proposal, because these systems are not very well defined by NMR correlation data. For many alkaloids and other proton-poor compounds, spectroscopic underdetermination is responsible for the combinatorial explosion of solutions and not the difficulties of evaluating ambiguous or poorly resolved NMR data.

## Conclusion

In summary, the computer program COCON is presented as a powerful tool for the evaluation of NMR-derived connectivity data. For the marine alkaloid manzacidin A (2), it was

demonstrated how the information available from the 1,1-ADEQUATE experiment can greatly reduce the number of constitutions compatible with given data sets. COCON should enable the chemical community to automate the process of comprehensive constitutional assignment. To the interested reader, the program will be made available on the PC platform.

*Acknowledgments:* T. L. wishes to thank the Deutsche Forschungsgemeinschaft for a Habilitandenstipendium (Li 597/2-1). M. K. thanks the Fonds der chemischen Industrie for a Liebig-Stipendium (Li 149/4). J. J. and M. K. acknowledge the support of Prof. Dr. C. Griesinger. We further thank Dr. J. P. Marino for careful reading of the manuscript.

*Supplementary material:* A demo version for SGI of the program COCON (works with up to 15 heavy atoms) and the manual are included.

## References

1. Aue, W. P.; Bartholdi, E.; Ernst, R. R. *J. Chem. Phys.* **1976**, *64*, 2229-2246.
2. (a) Müller, L. *J. Am. Chem. Soc.* **1979**, *101*, 2093-2094; (b) Bendall, M. R.; Pegg, D. T.; Doddrell, D. M. *J. Magn. Reson.* **1983**, *52*, 81-117; (c) Bax, A.; Griffey, R. H.; Hawkins, B. L. *J. Am. Chem. Soc.* **1983**, *105*, 7188-7190.
3. Bodenhausen, G.; Ruben, D. J. *Chem. Phys. Lett.* **1980**, *69*, 185-189.
4. Bax, A.; Summers, M. F. *J. Am. Chem. Soc.* **1986**, *108*, 2093-2094.
5. (a) Reif, B.; Köck, M.; Kerssebaum, R.; Kang, H.; Fenical, W.; Griesinger, C. *J. Magn. Reson.* **1996**, *A118*, 282-285; (b) Köck, M.; Reif, B.; Fenical, W.; Griesinger, C. *Tetrahedron Lett.* **1996**, *37*, 363-366; (c) Köck, M.; Reif, B.; Gerlach, M.; Reggelin, M. *Molecules* **1996**, *1*, 41-45; (d) Reif, B.; Köck, M.; Kerssebaum, R.; Schleucher, J.; Griesinger, C. *J. Magn. Reson.* **1996**, *B112*, 295-301.
6. Computer programs systematically evaluating 2D NMR spectroscopical data: SESAMI: (a) Christie, B. D.; Munk, M. E. *J. Am. Chem. Soc.* **1991**, *113*, 3750-3757. LSD: (b) Nuzillard, J.-M.; Massiot, G. *Tetrahedron* **1991**, *47*, 3655-3664. (c) Ley, S. V.; Doherty, K.; Massiot, G.; Nuzillard, J.-M. *Tetrahedron* **1994**, *42*, 12267-12280. (d) LSD/CSP: Nuzillard, J. M.; Naanaa, W.; Pimont, S. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1068-1073. CISOC-SES: (e) Peng, C.; Yuan, S.; Zheng, C.; Hui, Y. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 805-813. (f) Peng, C.; Yuan, S.; Zheng, C.; Hui, Y.; Wu, H.; Ma, K.; Han, X. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 814-819. (g) Peng, C.; Yuan, S.; Zheng, C.; Shi, Z.; Wu, H. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 539-546. LUCY: (h) Steinbeck, C. *Angew. Chem.* **1996**, *108*, 2108-2110. (i) for a review see: Massiot, G.; Nuzillard, J.-M. *Phytochem. Anal.* **1992**, *3*, 153-159.
7. Forenza, S.; Minale, L.; Riccio, R.; Fattorusso, E. *J. Chem. Soc., Chem. Commun.* **1971**, 1129-1130.
8. Kobayashi, J.; Kanda, F.; Ishibashi, M.; Shigemori, H. *J. Org. Chem.* **1991**, *56*, 4574-4576.
9. Faulkner, D. J. in *Marine Biotechnology, Vol. 1*; Attaway, D. H., Zaborsky, O. R., Eds.; Plenum Press: New York 1993; pp. 459-474.