A COCON Analysis of Proton-Poor Heterocycles – Application of Carbon Chemical Shift Predictions for the Evaluation of Structural Proposals

Matthias Köck,*[a] Jochen Junker,[a] Walter Maier,[b] Martin Will,[b] and Thomas Lindel*[c]

Keywords: Structure elucidation / Computational method / Natural products / HMBC / Constitutional analysis

The application of the new computer program COCON (**Co**nstitutions from **Con**nectivities) to the 2D-NMR data sets of three different complex natural products is described. The investigated compounds are proton-poor and therefore underdetermined systems. For such molecules the number of possible constitutions and the computational speed of COCON are of interest. Our investigation is focused on how methods of ^{13}C-NMR chemical shift prediction can assist chemists with regard to refining the selection among the constitutions proposed by COCON.

Introduction

The constitution of a new non-crystalline natural product is frequently considered elucidated as soon as an assigned structure is in accordance with the molecular formula and the connectivity information derived from 2D-NMR experiments like COSY,[1] HSQC,[2] HMBC[3] and ADEQUATE.[4] Other methods such as IR and UV spectroscopy at most allow the distinction among alternatives in the process of the structure elucidation. John Faulkner characterized the most desirable approach to structure elucidations as follows: "Rather than defining a structure that can be shown to fit the data, it is best to examine many possible structures (we would say *every* possible structure) and to treat each proposed structure as a hypothesis that cannot be proved but can only be disproved by incompatible data."[5] We have developed the computer program COCON (**Co**nstitutions from **Con**nectivities) as a comprehensive method for generating the entirety of constitutions that are compatible with available 2D-NMR data and the molecular formula of an unknown compound.[6] In this article the use of ^{13}C-NMR chemical shift prediction with the computer program SpecEdit[7] is examined in assisting chemists with regard to the selection among several proposed constitutions.

The data sets of three well-known natural products, aflatoxin B_1 (**1**) from the microfungus *Aspergillus flavus*, 11-hydroxyrotenone (**2**) from the root *Derris elliptica*,[8] and haemoventosin (**3**) from the lichen *Ophioparma ventosa* were chosen as demonstrative examples to show the difficulties of systematic constitutional analyses (Figure 1). The structure elucidation of compounds **1**–**3** is a challenge because they are proton-poor and highly oxygenated. While alkaloids may be subjected to ^{15}N-NMR correlation spectroscopy (^{1}H,^{15}N HMBC[9]), NMR experiments sensitive to oxygen (^{17}O)[10] can usually not be applied to natural products.[11] Unlike **1**,[12] the structures of **2** and **3** have not been confirmed by X-ray analysis and have been determined solely using MS, NMR spectroscopy and chemical derivatization.

For each example, two different COCON calculations were performed. Analysis A (results see Table 2) considers the hybridization states of every atom as they are given by the constitutions **1**, **2**, and **3** (Figure 1). In practice, however, the hybridization states especially of heteroatoms are usually unknown. In the more in-depth analysis B (results see Table 5), the program starts with only the true molecular formula and the degree of protonation of each atom (available e. g. from DEPT-edited spectra for carbon atoms). Restricted by only a very limited number of ^{13}C-NMR chemical shift rules,[13] every possible combination of hybridization states is generated and afterwards analyzed independantly. Of course, the solutions generated in analysis A are a subgroup of those of analysis B. In both calculation modes, identical 2D-NMR-derived connectivity information was regarded. Computer-based ^{13}C-NMR chemical shift calculations for the entirety of proposed constitutions were performed as the final step of the analysis. In all analyses, the resulting structures have to be viewed with a "chemist's eye" to select the best candidates based e. g. on their chemical stability or reactivity. Furthermore, molecular mechanics calculations, biosynthetic arguments, the synthesis of derivatives, or chemical degradations may also be helpful.

Calculation Conditions

The computer program COCON was shortly introduced in ref.[6] A more detailed description of the computer program COCON is given in the preceding paper.[14] The discussion in this paper is focused on the NMR data used in COCON. Bonds between atoms not showing COSY or 1,1-ADEQUATE correlations can optionally be forbidden. For

---

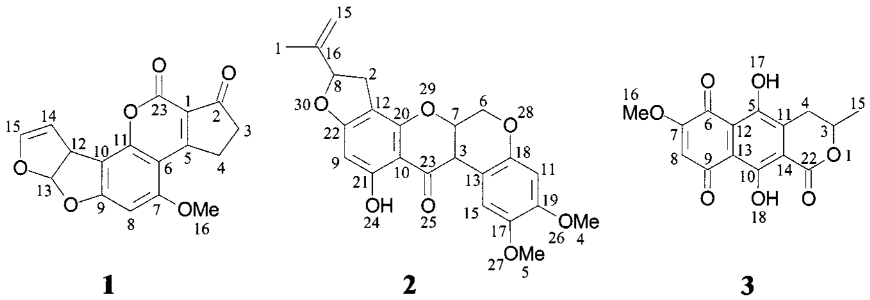
[a] Institut für Organische Chemie, J. W. Goethe-Universität,
Marie-Curie-Straße 11, D-60439 Frankfurt, Germany
Fax: (internat.) +49(0)69-798-29128
E-mail: km@org.chemie.uni.frankfurt.de
[b] ZHV-Molekülspektroskopie, BASF AG,
D-67056 Ludwigshafen, Germany
[c] Pharmazeutisch-Chemisches Institut der Universität,
Im Neuenheimer Feld 364, D-69120 Heidelberg, Germany
Fax: (internat.) +49(0)6221-54-6430
E-mail: Thomas.Lindel@urz.uni-heidelberg.de

1434–193X/99/0303–0579 $ 17.50+.50/0

Table 1. Quality of the input data of aflatoxin B_1 (**1**), 11-hydroxyrotenone (**2**) and haemoventosin (**3**)

	Sum formula	Degrees of unsaturation	No. of COSY correlations[a][b]	No. of HMBC correlations[a]	No. of ADEQUATE correlations[c]	No. of fixed bonds
Aflatoxin B_1 (**1**)	$C_{17}H_{12}O_6$	12	8	32	13	-
Rotenone (**2**)	$C_{23}H_{22}O_7$	13	6 (100%)	34 (72%)	15	6[d]
Haemoventosin (**3**)	$C_{15}H_{12}O_7$	10	4 (100%)	12 (55%)	7	5[e]

[a] The number in parenthesis gives the percentage of the obtained correlations to those correlations (COSY or HMBC) that could theoretically be expected for the finally proposed constitution. – [b] The COSY correlations were counted on both sides of the diagonal. Therefore only half of the correlations contain the connectivity information. – [c] The ADEQUATE correlations are theoretical 1,1-ADEQUATE correlations which would be expected for the finally proposed constitution. – [d] Six bonds were set fixed: 4–26 (C–O), 5–27 (C–O), 6–28 (C–O), 7–29 (C–O), 8–30 (C–O), 23–25 (C=O) (see also text). – [e] In conclusion from additional considerations and experiments Huneck et al.[36] assumed the bonds a) 1–3, b) 5–17, c) 6–12, d) 7–8, and e) 10–18 which are considered by us accordingly.



1 **2** **3**

Figure 1. Constitutions of aflatoxin B_1 (**1**), 11-hydroxyrotenone (**2**) and haemoventosin (**3**); the structures are numbered as used for the COCON calculations

COSY correlations this rule can be extended from proton bearing carbon atoms to all proton bearing atoms. Of course, the application of this rule can be problematical because a COSY correlation between two protons might not be observed if the dihedral angle ϕ is close to 90 degrees.[15] Further difficulties can result from long range correlations ($^4J_{HH}$ or $^5J_{HH}$) or from relaxation allowed coherence transfer[16] due to cross correlations.[17] The latter effect leads to cross peaks in the COSY spectrum in spite of the fact that the two spins are not scalar coupled. The 1,1-ADEQUATE experiment allows the observation of every H–C–C two-bond correlation present in a molecule.[18] The application of the non-1,1-ADEQUATE option implies that every carbon–carbon bond involving at least one proton bearing carbon atom will be forbidden if no 1,1-ADEQUATE correlation is observed.[19] *Whenever there are doubts that the additional restrictions on the COSY or the 1,1-ADEQUATE data influence the results of a COCON calculation, it should be repeated under conditions as unrestricted as possible.* The HMBC correlations are interpreted by COCON as two- or three-bond distances. A detailed description of the implementation of the use of HMBC correlations in the COCON algorithm is given in the preceding paper.[14] Besides two- and three-bond correlation also four-bond correlations may be observed in a HMBC spectrum. To avoid the problem of interpreting ^4J HMBC correlations as two- or three-bond interactions, it is recommended to run a COCON calculation in an iterative way and omit the HMBC cross peaks with low intensity in the first calculation. If COCON does not generate any solution in the next calculations, the new correlations are not two- or three-bond distances. COCON optionally obeys rules on ^{13}C-NMR chemical shifts[20] which are intentionally defined very coarsely. With one exception,[21] rules refering to more than two nuclei have not been implemented.

The ^{13}C-NMR chemical shift calculation by SpecEdit[7] is based on a matching of the spherical environment of every carbon atom (HOSE code)[22] of all query constitutions with the most similar spherical substructures of reference library compounds. The used SpecEdit library consists of 3.2×10^6 six sphere HOSE codes. The chemical-shift deviations of all carbon atoms allow to rank constitutions generated by COCON in a hitlist. The smallest deviations are expected to correspond to the most likely constitutions.

Calculation Results

For the COCON evaluation of the data sets of **1**, **2**, and **3**, any direct connection between proton-bearing carbon atoms not showing a COSY correlation as well as every carbon–carbon bond involving at least one protonated carbon (non-1,1-ADEQUATE option) were forbidden.[23] The ^{13}C-chemical shift rules of COCON[20] were used. To reference the structures generated by COCON the following nomenclature is used: (a) the bold number indicates the data set (**1**, **2** or **3**) and (b) the number after the hyphen is the numbering of the constitution as generated by COCON, e. g. **2**-212 would be data set **2** and structural proposal 212.

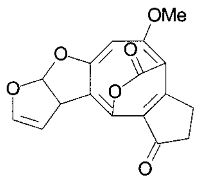
Aflatoxin B_1 (**1**): For aflatoxin B_1 (**1**) theoretical COSY and HMBC data sets from the proposed constitution (see Figure 1) were generated.[24] The ^{13}C-chemical shifts [δ(^{13}C)] were taken from ref.[25] COCON generated 1004 constitutions, if exactly the hybridization states present in aflatoxin B_1 (**1**) were used as input (analysis A). The number of possible structures was dramatically reduced by more than 98% (17 constitutions, see Table 2) when additionally using the 1,1-ADEQUATE correlations.

Table 2. Results of the COCON calculation assuming the fixed hybridization states of aflatoxin B_1 (**1**), 11-hydroxyrotenone (**2**), and haemoventosin (**3**)

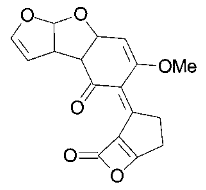
		COSY, HMBC[b]		COSY, HMBC, 1,1-ADEQ.[b]	
1	No. of structures	1004	134	17	5
	Calculation time[a]	17 min 54 s	3 min 14 s	0.4 s	0.4 s
2	No. of structures	492		4	
	Calculation time[a]	28.9 s		< 0.1 s	
3	No. of structures	49		12	
	Calculation time[a]	1.3 s		0.2 s	

[a] The calculations were carried on a SGI R10000, 195 MHz processor, the source code of COCON was compiled for a 64 Bit computer. – [b] For **1** two columns are shown, the left refers to the normal calculation, while for the right the ^{13}C chemical shift rule from ref.[21] was applied. For the data sets of **2** and **3**, some bonds were set fixed (for details see Table 1).

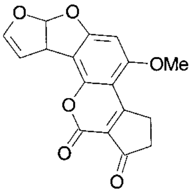
When the hybridization states of the atoms were left undefined and only the molecular formula was used (analysis B), 1932 atom type combinations (ATCs) were proposed by COCON. Only 35 of 1932 ATCs yielded in total 1341 constitutions when using COSY, HMBC and 1,1-ADEQUATE data (see Table 5). A closer inspection of the structures revealed that most of them contained an sp^3-carbon atom with δ(^{13}C) of 114 ppm (C-13) that was connected to one oxygen atom. By applying the rule that carbons connected to only one oxygen atom are not allowed with δ(^{13}C) over 90 ppm,[21] the number of constitutions was reduced from 1341 to 15 (see Table 5).[26] Only three of the 1932 atom type combinations (ATCs) generated for **1** yielded valid constitutions (#201: 6 constitutions, #1041: 4 constitutions, and #1167: 5 constitutions, see Table 5). It becomes clear from this data how strongly the connectivity information restricts the selection of hybridization states. For further analysis, the well-developed methodology of ^{13}C-NMR chemical shift prediction was applied. The program SpecEdit[7] was used to calculate all ^{13}C-NMR chemical shifts to be expected for every constitution generated by COCON. The structures of ATC #201 show an averaged chemical shift deviation (average over the deviations of all atoms) calculated by SpecEdit of 10.05 ppm (rmsd 3.3 ppm), ATC #1041 of 9.78 ppm (3.4) and ATC #1167 of 5.26 ppm (3.1). The SpecEdit results clearly prefer ATC #1167 which represents the atom types used in analysis A. The five structures corresponding to ATC #1167 are discussed in detail (see Figure 2). The constitutions **1**-11, **1**-12 and **1**-14 can be disregarded because they violate Bredt's rule. Bredt's rule is not automatically checked in the COCON program.[27]



1-12



1-13



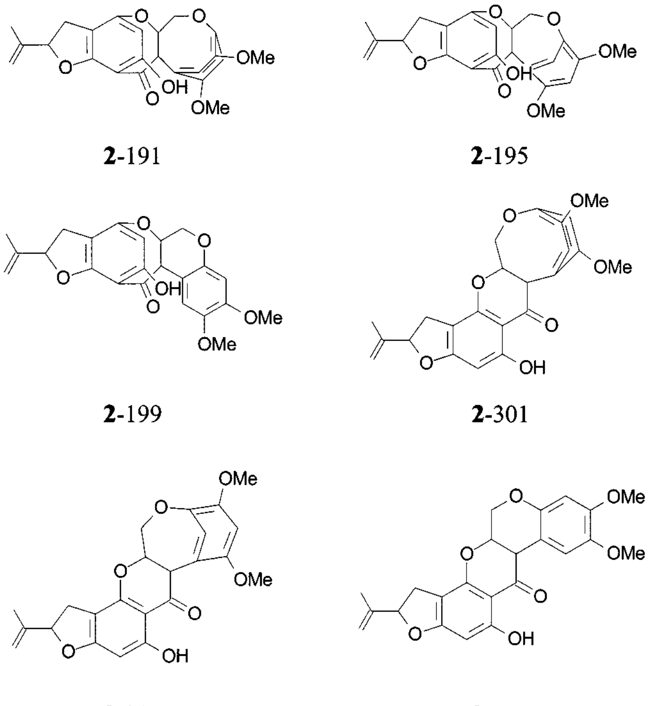
1-15

Figure 2. Selected constitutions generated by COCON (with COSY, HMBC and 1,1-ADEQUATE data) on the basis of the data set of aflatoxin B_1 (**1**); these three structures are discussed in the text

Constitution **1**-12, a cyclophane with a brigded cyclooctatetraene, is shown as an example for a structure which violates Bredt's rule (the total energy is about twice as large as for **1**-15). Structure **1**-13 consists of an α,β-unsaturated β-lactone moiety. Beside the chemical shifts for the carbonyl carbons there is one very deshielded carbon atom (C-5) at 177 ppm which is not in accordance with **1**-13. The β-position of the α,β-unsaturated cyclopentenone is expected with a ^{13}C-chemical shift of about 165 ppm. The chemical shift of 177 ppm is only fulfilled in structural proposal **1**-15 because there is another carbonyl group bound to this double bond (the –CO–OR substituent in α-position leads to a downfield shift of about 7 ppm[28]). Constitution **1**-15 is the correct structure of aflatoxin B_1 (**1**) and also prefered by SpecEdit with an averaged chemical shift deviation over all carbon atoms of 1.18 ppm.

11-Hydroxyrotenone (**2**):[29] The COSY- and HMBC-correlation data for 11-hydroxyrotenone (**2**)[30] were obtained experimentally.[31] COCON generated 24994 possible constitutions when using the COSY and HMBC data and the hybridization states shown in Figure 1.[32] To reduce the number of solutions, fixed bonds for **2** were introduced. First, the C=O bond was set fixed (23–25) according to the IR data[33] and the carbon chemical shifts (see Table 1). With this restriction the number of possible constitutions was still 5148.[34] Therefore further fixed bonds were introduced. Since the proton at δ(^{1}H) = 12.4 ppm is the only exchangable proton and the IR does not give any hint for further carbonyl groups, the other oxygen atoms were assumed as –O–. The following bonds were set fixed because of the carbon chemical shifts which indicate a neighbouring oxygen atom: C4–O26 (57 ppm, $-OCH_3$), C5–O27 (58 ppm, $-OCH_3$), C6–O28 (67 ppm, $-OCH_2-$), C7–O29 (72 ppm, –OCH<) and C8–O30 (89 ppm, –OCH<). With this input data COCON generated 492 possible constitutions. For all generated constitutions ^{13}C-chemical shifts were calculated using the software SpecEdit. 24 of the 492 solutions showed an averaged deviation of ^{13}C-chemical shifts from 1.48 to 2.91 ppm, while the constitution ranked 25th in

SpecEdit calculation already showed an averaged deviation of 4.78 ppm. Out of the 24 structures, 18 are assignment isomers[35] which ultimately represent six different constitutions (**2**-191, **2**-195, **2**-199, **2**-301, **2**-305 and **2**-309, see Figure 3). Besides proposal **2**-309 all constitutions are strained cyclophanes and violate Bredt's rule. The total energies are 2.5 to 5 times higher than **2**-309 (see Table 3). Structure **2**-309 is the correct constitution of 11-hydroxyrotenone (**2**), it also showed the smallest averaged deviation (1.48 ppm) for all δ(^{13}C) (see Table 3). By applying theoretical 1,1-ADEQUATE correlation data, the 492 structural proposals would be dramatically reduced to four which demonstrates the usefulness of this experiment (see Table 2).



2-191 **2**-195

2-199 **2**-301

2-305 **2**-309

Figure 3. Selected structures for 11-hydroxyrotenone (**2**) which were obtained by COCON with COSY, HMBC and fixed bonds; out of the generated 492 structures only six have to be discussed (for details, see text); the structures are shown in this figure as generated by COCON

If the atom types are not defined (analysis B), 226304 atom type combinations (ATCs) were generated when disregarding all connectivity information. Only six ATCs finally yielded at least one constitution when the connectivity information was regarded (see Table 5). If theoretical 1,1-ADEQUATE correlations were included, only two ATCs remained which yielded in 20 constitutions (ATC #17473 resulted in 16 structures, ATC #22225 in 4). These two groups of constitutions can easily be distinguished by their averaged chemical shift deviations calculated by SpecEdit which are 9.70 ppm (0.53 ppm) for ATC #17473 and 2.68 ppm (0.11 ppm) for ATC #22225. The results for 11-hydroxyrotenone (**2**) show how dramatically the correlation data reduce the number of atom type combinations which finally result in constitutions.

Haemoventosin (**3**): Haemoventosin (**3**), a pigment from the lichen *Ophioparma ventosa*, has been subjected to several constitutional assignments. Huneck et al. forwarded their proposal in 1995 on the basis of 2D-NMR-spectral data including HMBC and of chemical derivatizations.[36] Using the reported COSY and HMBC correlations as well as fixed bonds[37] we have carried out a COCON calculation. COCON generated 320 ATCs for the data set of **3**. Only 15 of them yielded in total 938 possible constitutions. The correct ATC #24 (atom types present in the constitution proposed by Huneck et al.[36]) can be distinguished by their averaged ^{13}C-chemical shift deviation of 8.50 ppm (see Table 6) in the SpecEdit calculations (see Figure 4). After introducing 1,1-ADEQUATE correlations, only eight of the 320 ATCs yielded COCON solutions, resulting in total in 206 possible constitutions. The SpecEdit results are given in Table 6.

The ATC #24 includes 49 constitutions,[38] among which 36 can be disregarded due to a violation of Bredt's rule. The dihydrofuran substructure proposed in earlier studies of 1971 and 1990 can be excluded on the basis of the correlation data obtained by Huneck et al.[36] In their paper the

Table 3. Calculation of the ^{13}C-chemical shift deviations of selected structures for **1**, **2** and **3**; for the structures of **2** also an energy minimization[a] was carried out

1	<Δ[δ(^{13}C)]>[b] [ppm]	**2**	<Δ[δ(^{13}C)]>[b] [ppm]	*E* [kcal/mol]	**3**	<Δ[δ(^{13}C)]>[b] [ppm]
1-11	6.65	**2**-191	2.65	189.18	**3**-3	5.00
1-12	3.18	**2**-195	2.57	163.37	**3**-4	5.67
1-13	6.18	**2**-199	2.48	105.75	**3**-6	9.60
1-14	9.12	**2**-301	1.83	97.89	**3**-10	9.87
1-15	1.18	**2**-305	1.74	87.21	**3**-14	6.71
-	-	**2**-309	1.48	36.49	**3**-15	7.27
-	-	-	-	-	**3**-18	8.33
-	-	-	-	-	**3**-19	10.20
-	-	-	-	-	**3**-20	11.40
-	-	-	-	-	**3**-22	10.20
-	-	-	-	-	**3**-24	10.87
-	-	-	-	-	**3**-40	11.27
-	-	-	-	-	**3**-43	10.33

[a] The energy minmization was carried out with the HyperChem program package using a modified MM2 force field (MM+). – [b] Averaged ^{13}C chemical shift deviation calculated by SpecEdit.

Table 4. Possible atom types proposed by COCON for **1**, **2** and **3**

	Aflatoxin (**1**)		11-Hydroxyrotenone (**2**)		Haemoventosin (**3**)	
Atom no.[a]	$\delta(^{13}C)$ [ppm]	Possible atom types	$\delta(^{13}C)$ [ppm]	Possible atom types	$\delta(^{13}C)$ [ppm]	Possible atom types
1	117	>C<, =C<	18	$-CH_3$	-	>O, =O
2	201	=C<, =C=	31	$>CH_2$	-	>O, =O
3	35	$>CH_2$	44	>CH−, ≡CH	75	>CH−, =CH−, ≡CH
4	29	$>CH_2$	57	$-CH_3$	29	$>CH_2$
5	177	=C<	58	$-CH_3$	156	=C<
6	104	>C<, =C<, ≡C−	67	$>CH_2$	185	=C<, =C=
7	161	=C<	72	>CH−, ≡CH	158	=C<
8	91	>CH−, =CH−, ≡CH	89	>CH−, =CH−, ≡CH	113	>CH−, =CH−, ≡CH
9	166	=C<	93	>CH−, =CH−, ≡CH	181	=C<, =C=
10	108	>C<, =C<, ≡C−	101	>C<, =C<, ≡C−	142	=C<
11	153	=C<	102	>CH−, =CH−, ≡CH	138	=C<
12	48	>CH−, =CH	105	>C<, =C<, ≡C−	116	>C<, =C<
13	114	>CH−, =CH−, ≡CH	105	>C<, =C<, ≡C−	122	>C<, =C<
14	103	>CH−, =CH−, ≡CH	111	>CH−, =CH−, ≡CH	26	=C<
15	145	=CH−	114	$>CH_2$, $=CH_2$	51	$-CH_3$
16	57	$-CH_3$	143	=C<	57	$-CH_3$
17	-	>O, =O	144	=C<	-	−OH
18	-	>O, =O	149	=C<	-	−OH
19	-	>O, =O	150	=C<	-	>O, =O
20	-	>O, =O	157	=C<	-	>O, =O
21	-	>O, =O	166	=C<	-	>O, =O
22	-	>O, =O	170	=C<	-	>O, =O
23	155	=C<	194	=C<, =C=	-	-
24	-	-	-	−OH	-	-
25	-	-	-	>O, =O	-	-
26	-	-	-	>O, =O	-	-
27	-	-	-	>O, =O	-	-
28	-	-	-	>O, =O	-	-
29	-	-	-	>O, =O	-	-
30	-	-	-	>O, =O	-	-

[a] The atom numbering is given in Figure 1 for all three compounds.

Table 5. Results of the COCON calculation with permutation of hybridization states according to the molecular formula and the degrees of protonation of **1**, **2**, and **3**

		COSY, HMBC[b]	COSY, HMBC, 1,1-ADEQ.[b]	
1	No. of ATCs[a]	1932 (34)	1932 (35)	1932 (3)
	No. of structures	2636	1341	15
	Calculation time[b]	23 h 31 min	1 h 45 min	2 min 38 s
2	No. of ATCs[a]	226304 (6)	226304 (2)	
	No. of structures	2996	20	
	Calculation time[b]	9 h 55 min	8 min 42 s	
3	No. of ATCs[a]	320 (15)	320 (8)	
	No. of structures	938	206	
	Calculation time[b]	10 min 13 s	1 min 32 s	

[a] The number in parenthesis represents the number of atom type combinations (ATCs) with COCON solutions. − [b] The calculations were carried out on a SGI R10000, 195 MHz processor, the source code of COCON was compiled for a 64-Bit computer. − [c] For **1** two columns are shown, the left refers to the normal calculation, while the right includes the ^{13}C chemical shift rule from ref.[21] For the data sets of **2** and **3**, some bonds were set fixed (for details see Table 1).

authors already excluded the former constitutional proposals because of the IR band at 1750 cm^{-1}.[36] The remaining 13 proposals correspond to only nine different constitutions (see Figure 5) because of assignment isomers.[35] The structural proposals **3-**3, **3-**4, **3-**14, and **3-**15 (see Figure 5) are clearly favored by their $\delta(^{13}C)$ deviation (see Table 3). These are exactly the four constitutions which have been thoroughly analyzed by Huneck et al.[36] leading to **3-**3 as the final constitution of haemoventosin.[39] Structural proposal **3-**3 was not considered best in the SpecEdit calculations, while three constitutions violating Bredt's rule are ranked first to third (**3-**29 with 3.27 ppm, **3-**5 with 3.40 ppm and **3-**26 with 3.80 ppm).

Summary and Conclusion

The constitutional analysis of natural products will continue to play an important role in chemistry. With the introduction of high field NMR spectrometers into many laboratories, the compounds to be analyzed will show increasing complexity and size. Therefore, NMR-data analysis has to be accelerated and to become more accurate. To achieve this goal, the new computer program COCON was developed as a method to generate constitutions on the basis of connectivity information. The chemist is forced to first accept all constitutions which are in accordance with the NMR-correlation data and to find further arguments why certain structures cannot be the correct constitution. In contrast to pure structure generators which disregard connectivity information, the approach followed by COCON makes both the constitutional space and the calculation times managable for chemists.

Proton-poor heterocycles are expected to be among the most challenging subjects of NMR-based structure elucidations. The COCON analysis of the investigated compounds confirmed that assumption by generating a large number of

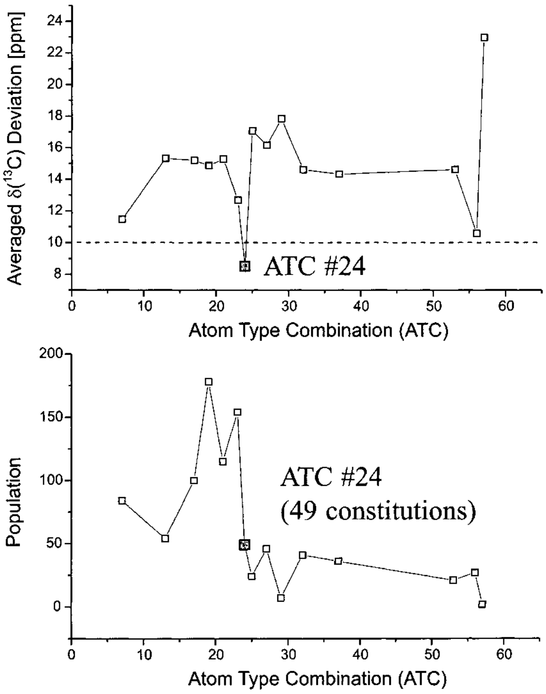


Figure 4. Without fixed hybridization states for haemoventosin (**3**), Cocon generated 320 ATCs from which 15 yield constitutions (altogether 936); for all of them the ^{13}C chemical shifts were calculated using SpecEdit; the averaged ^{13}C chemical shift deviation for all 15 ATCs is shown on the left; the correct ATC (#24) shows the smallest deviation; on the right side the number of structures per ATC is given

constitutions, even if almost the entire theoretically available connectivity information was used.

The Cocon calculations show that the consideration of connectivity information strongly reduces not only the number of constitutions compatible with a given atom type combination (ATC), but also the number of ATCs yielding constitutions at all. The inclusion of 1,1-ADEQUATE correlations has a dramatic influence on the number of generated structures as well as on the calculation time as it was demonstrated for all three examples.

From our point of view it appears much more efficient to perform a thorough ^{13}C-NMR chemical shift calculation *after* constitutions have been generated by a Cocon calculation and not during the generation process. If the generation of constitutions is unbiased by experience, methods of ^{13}C-NMR shift prediction and assignment should be challenged. The computer program SpecEdit offers in most cases a powerful filter to break even several thousands of generated Cocon constitutions down to only a few candidate structures. SpecEdit showed a reasonable performance by ranking the correct constitutions of aflatoxin B_1 (**1**), 11-hydroxyrotenone (**2**), and haemoventosin (**3**) always among the first choices. The limit of this data based filter is clearly the quality of chemical shift assignment and similarity of reference substructures of the ^{13}C-chemical shift library. But, the absolute value of the chemical shift deviation is quite large in some cases. Therefore, a larger data base is needed in order to cover the whole area of natural products. The SpecEdit results allow the selection of the correct atom type combination on the basis of their chemical shift deviations.

The Cocon calculations discussed in this paper demonstrate the usefulness of the program for problems of structure elucidation. The calculations were fast and comprehensive. There are two general applications: (a) analysis of structural proposals which have already been made and (b) de novo structure elucidation. A demo version and documentation of the Cocon software can be ordered from the authors. It is also possible to run a Cocon calculation over the internet (http://cocon.org.chemie.uni-frankfurt.de).

Experimental Section

Calculations Conditions. – Aflatoxin B_1 (1): The Cocon calculations used the following connectivity data: (a) direct connections

Table 6. Cocon results for haemoventosin (**3**) with the averaged ^{13}C-chemical shift deviation from SpecEdit[a]

COSY, HMBC[b]				COSY, HMBC, 1,1-ADEQUATE[b]			
ATC #	No. of structures	<Δ[δ(^{13}C)]> [ppm]	rmsd [ppm]	ATC #	No. of structures	<Δ[δ(^{13}C)]> [ppm]	rmsd [ppm]
7	84	11.46	1.25	-	-	-	-
13	54	15.32	1.80	-	-	-	-
17	100	15.20	1.85	17	68	15.00	1.91
19	178	14.86	2.27	19	34	15.52	1.21
21	115	15.29	2.26	21	60	16.35	2.26
23	154	12.68	1.81	23	5	15.91	0.90
24	**49**	**8.50**	**2.65**	**24**	**12**	**7.83**	**2.30**
25	24	17.07	2.51	-	-	-	-
27	46	16.16	3.51	27	4	15.08	1.58
29	7	17.84	3.72	-	-	-	-
32	41	14.61	3.61	-	-	-	-
37	36	14.32	1.72	-	-	-	-
53	21	14.61	2.10	53	21	14.61	2.10
56	27	10.58	1.20	-	-	-	-
57	2	22.97	1.08	57	2	22.97	1.08

[a] In the first column the number of the ATCs with solutions as generated by Cocon is given. The second column contains the number of Cocon solutions for each ATC. The third column includes the average ^{13}C chemical shift deviation over all constitutions of each ATC (over all carbon atoms for each structure) calculated by SpecEdit. The fourth column gives the root mean square deviation of column three. The bold printed numbers represent the correct atom type combination for haemoventosin which is for both correlation data sets the preferred solution. – [b] Also fixed bonds were used (for details see Table 1).

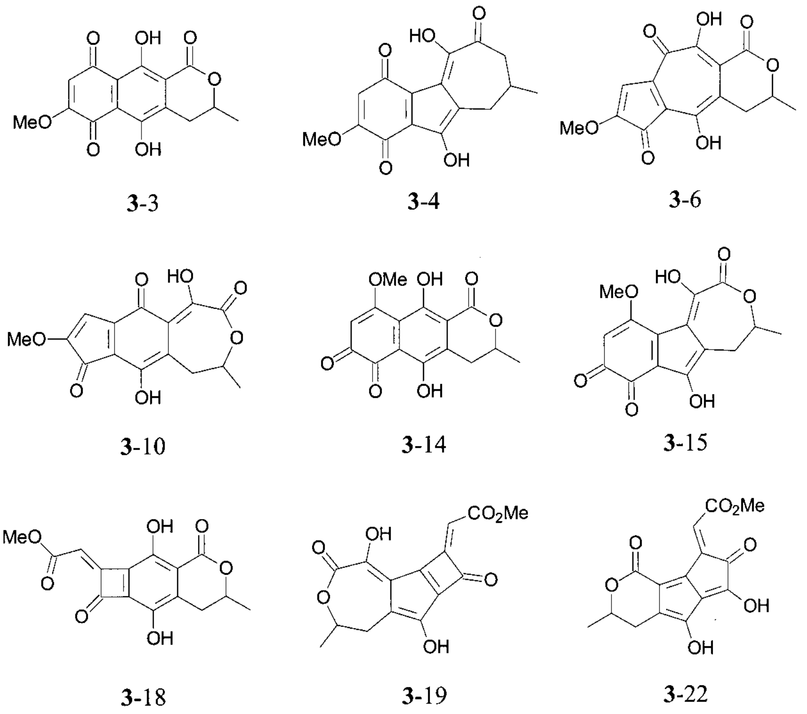


Figure 5. The nine COCON proposals for haemoventosin (3) are discussed in this investigation. Out of 49 structures generated by COCON with COSY, HMBC and fixed bonds, 36 were disregarded because of a violation of Bredt's rule. The number of 13 structures is reduced to nine because of assignment isomers (*e. g.* **3**-18, **3**-20 and **3**-40 have the same constitutions, but not the same assignment of ^{13}C chemical shifts)

as derived from the ^{1}H,^{1}H-COSY experiment: C-3 and C-4, C-12 and C-13, C-12 and C-14, C-14 and C-15; (b) two- or three-bond distances as derived from the ^{1}H,^{13}C-HMBC experiment: C-3 and C-1, C-2, C-4, C-5; C-4 and C-1, C-2, C-3, C-5, C-6; C-8 and C-6, C-7, C-9, C-10; C-12 and C-9, C-10, C-11, C-13, C-14, C-15; C-13 and C-9, C-10, C-12, C-14, C-15; C-14 and C-10, C-12, C-13, C-15; C-15 and C-12, C-13, C-14; C-16 and C-7; (c) direct connections as derived from the 1,1-ADEQUATE experiment: C-3 and C-2, C-4; C-4 and C-5; C-8 and C-7, C-9; C-12 and C-10, C-13, C-14; C-14 and C-15.

11-Hydroxyrotenone (2): The COCON calculations used the following connectivity data: (a) direct connections as derived from the ^{1}H,^{1}H-COSY experiment: C-2 and C-8, C-3 and C-7, C-6 and C-7; (b) two- or three-bond distances as derived from the ^{1}H,^{13}C-HMBC experiment: C-1 and C-8, C-15, C-16; C-2 and C-8, C-12, C-16, C-20, C-22; C-3 and C-13, C-14, C-23; C-4 and C-19; C-5 and C-17; C-6 and C-3, C-7, C-18; C-8 and C-1, C-15; C-9 and C-10, C-12, C-21; C-11 and C-13, C-17, C-18; C-14 and C-3, C-13, C-17, C-18; C-15 and C-1, C-8, C-16; C-24 and C-9, C-10, C-21; (c) direct connections as derived from the 1,1-ADEQUATE experiment: C-1 and C-16; C-2 and C-8, C-12; C-3 and C-13, C-23; C-6 and C-7; C-8 and C-16; C-9 and C-20, C-21; C-11 and C-17, C-18; C-14 and C-13, C-19; C-15 and C-16.

Haemoventosin (3): The COCON calculations used the following connectivity data: (a) direct connections as derived from the ^{1}H,^{1}H-COSY experiment: C-3 and C-4, C-15; (b) two- or three-bond distances as derived from the ^{1}H,^{13}C-HMBC experiment: C-3 and C-22; C-4 and C-5, C-11, C-14; C-8 and C-6, C-7, C-9, C-13; C-16 and C-7; C-17 and C-5, C-11, C-12; (c) direct connections as derived from the 1,1-ADEQUATE experiment: C-3 and C-4, C-15; C-4 and C-11; C-8 and C-7, C-9

Acknowledgments

T. L. wishes to thank the Deutsche Forschungsgemeinschaft (Li 597/2-1 and 2-3) and Professor Dr. Richard Neidlein for his generous support. M. K. thanks the Fonds der Chemischen Industrie (Li 149/4) and the Deutsche Forschungsgemeinschaft (Ko 1314/3-1 and 3-2). J. J. and M. K. acknowledge the support of Professor Dr. Christian Griesinger. We thank Dr. John P. Marino for careful reading of the manuscript.

---

[1] W. P. Aue, E. Bartholdi, R. R. Ernst, *J. Chem. Phys.* **1976**, *64*, 2229−2246.

[2] G. Bodenhausen, D. J. Ruben, *Chem. Phys. Lett.* **1980**, *69*, 185−189.

[3] A. Bax, M. F. Summers, *J. Am. Chem. Soc.* **1986**, *108*, 2093−2094.

[4] [4a] B. Reif, M. Köck, R. Kerssebaum, H. Kang, W. Fenical, C. Griesinger, *J. Magn. Reson.* **1996**, *A118*, 282−285. − [4b] M. Köck, B. Reif, W. Fenical, C. Griesinger, *Tetrahedron Lett.* **1996**, *37*, 363−366.

[5] D. J. Faulkner in *Marine Biotechnology*, vol. 1 (Hrsg.: D. H. Attaway, O. R. Zaborsky), Plenum Press, New York, **1993**, pp. 459−474.

[6] T. Lindel, J. Junker, M. Köck, *J. Mol. Mod.* **1997**, *3*, 364−368. COCON is written in the C programming language.

[7] W. Maier in *Computer - Enhanced Analytical Spectroscopy*, vol. 4, (Ed.: C. L. Wilkins), Plenum Press, NY and London, **1993**, 37−55.

[8] Isolated from an extract found in the BASF screening.

[9] M. Köck, J. Junker, T. Lindel, manuscript in preparation.

[10] S. Berger, S. Braun, H.-O. Kalinowski, *NMR-Spektroskopie von Nichtmetallen; Grundlagen ^{17}O, ^{33}S- und ^{129}Xe-Spektroskopie*, Thieme Verlag, Stuttgart, **1992**.

[11] Since ^{17}O relaxes very fast, no 2D correlation experiments can be obtained. For natural product samples (1 to 20 mg) it is almost impossible to even get a 1D ^{17}O spectrum.

[12] A. C. Pavao, L. A. S. Neto, J. F. Neto, M. B. C. Leao, *Theochem. J. Mol. Struct.* **1995**, *337*, 57−60.

[13] Cocon selects possible hybridization states (part of analysis B) on the basis of the following rules: (a) the total numbers of single, double, and triple bond valencies represented by a hybridization state combination must be even; (b) ^{13}C-NMR chemical shift criteria (not identical to ref.[20]): $\delta_C < 180$ ppm: the carbon atom must not be an allene carbon atom; $\delta_C > 125$ ppm: the carbon atom must not be sp^3-hybridized; $\delta_C > 115$ ppm: the carbon atom must not be sp-hybridized; $\delta_C < 75$ ppm: the carbon atom must not be sp^2-hybridized.

[14] T. Lindel, J. Junker, M. Köck, *Eur. J. Org. Chem.* **1999**, 573−577.

[15] The Karplus equation correlates the ^3J coupling constant with the involved dihedral angle ϕ. The function has a minimum for $\phi = 90$ degrees where the coupling constant will be close to zero. M. Karplus, *J. Chem. Phys.* **1959**, *30*, 11−15 and *J. Am. Chem. Soc.* **1963**, *85*, 2870−2871.

[16] S. Wimperis, G. Bodenhausen, *Chem. Phys. Lett.* **1987**, *140*, 41−45 and *Mol. Phys.* **1989**, *66*, 897−919.

[17] T. E. Bull, *J. Magn. Reson.* **1987**, *72*, 397−413.

[18] For a H(1)−C(1)−C(2) fragment the 1,1-ADEQUATE correlation from H(1) to C(2) defines the connection between C(1) and C(2). The same information can be obtained from a COSY experiment if both carbons bear a proton. For two non-protonated carbon atoms this information can only be derived from the INADEQUATE experiment (A. Bax, R. Freeman, T. A. Frankiel, *J. Am. Chem. Soc.* **1981**, *103*, 2102−2104) which is very insensitive.

[19] This may be critical if the chemical shifts of two carbon atoms involved in the double quantum coherence are identical. Then the approximation of a weakly coupled spin system ($\Delta\delta \gg J$) does not hold anymore. As a result, the corresponding cross peak cannot be observed (as it is already known from the INADEQUATE experiment). This case should be rare, because the chemical shift dispersion in the ^{13}C dimension is quite large.

[20] Cocon chemical shift rules concerning the consideration of connectivities: a) C=S and C=O bonds are forbidden if $\delta_C < 150$ ppm, b) aliphatic C−O bonds are forbidden if $\delta_C < 45$ ppm, c) olefinic C−O bonds are forbidden if $\delta_C < 130$ ppm, d) olefinic C−N bonds are forbidden if $\delta_C < 105$ ppm and e) Met−C bonds are forbidden if $\delta_{CH3} > 35$ ppm. These rules are based on E. Pretsch, T. Clerc, J. Seibl, W. Simon, *Tables of Spectral Data for Structure Determination of Organic Compounds*, 2nd ed.; Springer-Verlag, New York **1989**.

[21] Carbon atoms connected to two oxygen atoms must have a chemical shift larger than 90 ppm [see also the discussion of aflatoxin B$_1$ (**1**)].

[22] HOSE stands for *H*irarchically *O*rdered *S*perical Description of *E*nvironment and was introduced by W. Bremser, *Anal. Chim. Acta* **1978**, *103*, 151−162 and 355−365.

[23] The following restrictions were obeyed additionally: every generated solution had to represent a single molecule and to be compatible with a multiple bond system; cyclopropenoids, cyclobutadienoids, and geminal diols were excluded.

[24] There is no report about HMBC data of pure aflatoxin B$_1$. This kind of data was published for an adenine adduct of aflatoxin B$_1$ epoxide (R. S. Iyer, M. W. Voehler, T. M. Harris, *J. Am. Chem. Soc.* **1994**, *116*, 8863−8869).

[25] K. G. R. Pachler, P. S. Steyn, R. Vleggaar, P. L. Wessels, D. B. Scott, *J. Chem. Soc., Perkin Trans. 1* **1976**, 1182−1189.

[26] The calculation time was dramatically reduced by a factor of 40.

[27] A simple method to introduce Bredt's rule into a structure generator was published by J.-M. Nuzillard, *J. Chem. Inf. Comp. Sci.* **1994**, *34*, 723−724.

[28] H.-O. Kalinowski, S. Berger, S. Braun, *^{13}C-NMR-Spektroskopie*, Thieme Verlag, Stuttgart, **1984**.

[29] L. Crombie, *Nat. Prod. Rep.* **1984**, 3−19.

[30] 11-Hydroxyrotenone is unknown but ^{13}C-NMR data of rotenone was already published: L. Crombie, G. W. Kilbee, D. A. Whiting, *J. Chem. Soc., Perkin Trans. 1* **1975**, 1497−1499; G. Blaskó, H.-L. Shieh, J. M. Pezzuto, G. A. Cordell, *J. Nat. Prod.* **1989**, *52*, 1363−1366.

[31] All measurements were carried out with an appr. 2-mg sample of **2** in 190 µl using a Shigemi NMR tube on a Bruker AC400 and a Varian INOVA500. The following NMR experiments were carried out: 1D ^{1}H NMR, 1D ^{13}C NMR, 1D DEPT (90 and 135), COSY, TOCSY, ROESY, HSQC and HMBC.

[32] The calculation time was 2 h 9 min 57 s. By using the 1,1-ADEQUATE data additionally the number of possible structures is reduced to 78 (calculation time of 0.8 s).

[33] The wave number of 1648 cm^{-1} in the IR spectrum indicates a carbonyl and especially because of the low field shift a β-hydroxy α,β-unsaturated carbonyl. This is in accordance with an exchangeable proton at $\delta(^1H) = 12.4$ ppm which indicates a hydrogen bond.

[34] The calculation time was 1 h 8 min 7 s. By using the 1,1-ADEQUATE data additionally the number of possible structures is reduced to 38 (calculation time of 0.5 s).

[35] Assignment isomers are isomers in which two carbon atoms in two constitutions may be interchanged without violation of any connectivity or other constraints. The resulting solutions represent identical constitutions, but different ^{13}C NMR assignments which are of importance for subsequent ^{13}C chemical shift calculations.

[36] D. S. Rycroft, J. D. Connolly, S. Huneck, U. Himmelreich, *Z. Naturforsch.* **1995**, *50b*, 1557−1563.

[37] Huneck et al. used 19 NMR arguments from which they deduced bonds between two atoms. Their arguments were: two COSY correlations, four chemical shift arguments, one ROESY correlation and twelve HMBC correlations, where they derived bond information. We interpreted the HMBC correlations observed between H-8 and C-6 resp. C-9 as corresponding to $^2J_{CH}$ or $^3J_{CH}$, while Huneck et al. considered those correlations as exclusive $^3J_{CH}$ resp. $^2J_{CH}$ couplings.

[38] The defined fixed bonds for the data set of **3** have a dramatic influence on the number of structures and the calculation time. Cocon generated 29175 possible constitutions using COSY, HMBC data and fixed bonds (calculation time of 2 h 8 min 34 s). The number of structures is reduced to 219 if 1,1-ADEQUATE data is used (calculation time 45.5 s).

[39] The constitutions **3**-4 and **3**-15 were excluded by Huneck et al. by $\delta(^1H)$ of the hydroxy group involved in a hydrogen bond.[36] The final distinction between the *o*-quinone and the *p*-quinone structure was carried out by Huneck et al.[36] by a dithionite reduction.

Received May 29, 1998
[O98246]