ORGANIC LETTERS

1999
Vol. 1, No. 5
737−740

# Computer-Assisted Constitutional Assignment of Large Molecules: COCON Analysis of Ascomycin

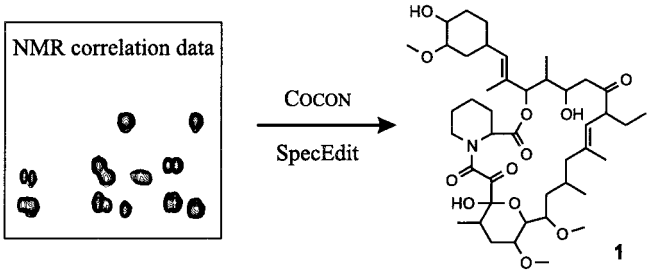
**Jochen Junker,[†] Walter Maier,[‡] Thomas Lindel,[§] and Matthias Köck*,[†]**

*Institut für Organische Chemie, Johann Wolfgang Goethe-Universität, Marie-Curie-Strasse 11, D-60439 Frankfurt, Germany, ZHV-Molekülspektroskopie, BASF AG, D-67056 Ludwigshafen, Germany, and Pharmazeutisch-chemisches Institut der Universität, Im Neuenheimer Feld 364, D-69120 Heidelberg, Germany*

*km@org.chemie.uni-frankfurt.de*

**Received June 9, 1999**

## ABSTRACT




**The computer-assisted constitutional assignment of ascomycin (1) is discussed. This example demonstrates that the NMR-based structure generator COCON is able to analyze data sets of organic molecules covering the full known range in size and complexity. The structural proposals of ascomycin were validated by calculating the $^{13}$C NMR chemical shifts using the computer program SpecEdit. The enhanced calculation time of COCON achieved by preorganizing the input data is also demonstrated for the data sets of aflatoxin $B_1$ (2), 11-hydroxyrotenone (3), and haemoventosin (4).**

Recently, the computer program COCON ("COnstitutions from CONnectivities")[1] was introduced as a tool for the computer-assisted structure elucidation[2] of unknown compounds, in particular of complex natural products. COCON is of special value for the comprehensive constitutional analysis of proton-poor compounds.[1c] Although the calculation speed is secondary to the comprehensiveness of the constitutional analysis itself, it becomes an important property of COCON when data sets of large molecules are discussed, such as those of ascomycin (**1**). A new, accelerated version of COCON is presented. The effect on the calculation time is also demonstrated for aflatoxin $B_1$ (**2**), 11-hydroxyrotenone (**3**), and haemoventosin (**4**).

The natural product ascomycin (**1**, $C_{43}H_{69}NO_{12}$, 56 heavy atoms) is a close structural analogue to the potent immunosuppressive agent FK-506.[3] The experimental NMR data set of **1** includes 52 $^1$H,$^1$H-COSY[4] correlations (90% of those correlations that could theoretically be expected for the constitution of **1**) and 86 $^1$H,$^{13}$C-HMBC[5] correlations (53%).[6] The NMR correlation data of **1** is given in the Supporting Information.

[†] Johann Wolfgang Goethe-Universität.
[‡] BASF AG.
[§] Pharmazeutisch-chemisches Institut der Universität.

(1) (a) Lindel, T.; Junker, J.; Köck, M. *J. Mol. Model.* **1997**, *3*, 364−368. (b) Lindel, T.; Junker, J.; Köck, M. *Eur. J. Org. Chem.* **1999**, 573−577. (c) Köck, M.; Junker, J.; Maier, W.; Will, M.; Lindel, T. *Eur. J. Org. Chem.* **1999**, 579−586.

(2) Jaspers, M. *Nat. Prod. Rep.* **1999**, *16*, 241−248, and references cited therein.

(3) Or, Y. S.; Clark, R. F.; Xie, Q.; McAlpine, J.; Whittern, D. N.; Henry, R.; Luly, J. R. *Tetrahedron* **1993**, *49*, 8771−8786.

(4) Aue, W. P.; Bartholdi, E.; Ernst, R. R. *J. Chem. Phys.* **1976**, *64*, 2229−2246.

(5) Bax, A.; Summers, M. F. *J. Am. Chem. Soc.* **1986**, *108*, 2093−2094.

10.1021/ol990725b CCC: $18.00 
**Published on Web 08/12/1999**

**Table 1.** Results of the COCON Calculation Assuming Predefined (Fixed) Hybridization States (fh) and with Permutation of Hybridization States (oh) According to the Molecular Formula and the Degrees of Protonation of Aflatoxin $B_1$ (**2**), 11-Hydroxyrotenone (**3**), and Haemoventosin (**4**)

| compd[a] | | COSY, HMBC[f] | | COSY, HMBC, 1,1-ADEQ[f] | |
|---|---|---|---|---|---|
| **2** (fh) | no. of structures | 1004 | 134 | 17 | 5 |
| | calculation time[b] | 17.2 s | 1.9 s | 0.1 s | <0.1 s |
| **2** (oh) | no. of ATCs[c] | 1932 | 1932 (34) | 1932 (35) | 1932 (3) |
| | no. of structures | >6.4 million | 2636 | 1341 | 15 |
| | calculation time[b] | >10000 min | 4 min 48 s | 38 min 24 s | 3.8 s |
| **3** (fh) | no. of structures | 492 | | 4 | |
| | calculation time[b] | 26.7 s | | <0.1 s | |
| **3** (oh) | no. of ATCs[c,d] | 155648 (6) | | 155648 (2) | |
| | no. of structures | 2996 | | 20 | |
| | calculation time[b] | 3h3m | | 4m22s | |
| **4** (fh) | no. of structures | 49 | | 12 | |
| | calculation time[b] | 0.8 s | | 0.2 s | |
| **4** (oh) | no. of ATCs[c] | 320 (15) | | 320 (8) | |
| | no. of structures[e] | 936 | | 204 | |
| | calculation time[b] | 13.6 s | | 3.2 s | |

[a] COCON calculations with predefined (fixed) hybridization states (fh) and COCON calculations with open hybridization states (oh). [b] The calculations were carried out on a SGI R10000, 195 MHz processor, the source code of COCON was 64-bit optimized. [c] The number in parentheses is the number of atom type combinations (ATCs) with COCON solutions. [d] The number of ATCs is reduced in comparison to ref 1c because the lower limit for the chemical shift of a carbon atom involved in a triple bond was set to 45 ppm. [e] One ATC (two constitutions) is eliminated due to an improved treatment of triple bonds and allene systems in comparison to ref 1c. [f] For **2** two columns are shown, the left refers to the normal calculation, while the right includes the extra $^{13}C$ chemical shift rule that carbon atoms connected to two oxygens must have a chemical shift larger than 90 ppm. For the data sets of **3** and **4**, some bonds were set fixed (see ref 1c).

To the best of our knowledge, ascomycin (**1**) is one of the compounds with the highest number of heavy atoms reported as investigated by a computer program. Nuzillard et al.[7] described the application of their program LSD to the constitutional analysis of azadirachtin ($C_{35}H_{44}O_{16}$, 51 heavy atoms) and Zheng et al.[8] of CISOC-SES to alborixin ($C_{48}H_{84}O_{14}$, 62 heavy atoms). COCON is able to calculate molecules with more than 150 heavy atoms as demonstrated for palytoxin ($C_{129}H_{223}N_3O_{54}$, 186 heavy atoms).

COCON allows the comprehensive discussion of the entirety of molecular constitutions compatible with a 2D NMR data set, required *respectively* forbidden substructures, and the molecular formula. The high calculation speed of the COCON program is based on the integrated evaluation of HMBC information simultaneously with the construction of constitutions and on the detection of constitutionally equivalent but combinatorially different isomers at the earliest possible stage of their generation. For an optimized calculation speed it is preferable to sort the atoms of a given data set according to the number of HMBC correlations in which they are involved.

The higher the number of HMBC correlations of an atom, the earlier it has to appear as an entry in the data set. The idea behind this optimization is given in a short example. Take a molecule with two distinct regions (A and B) without interactions. If an input file contains, for example, a poorly defined region (A) first and a well-defined region (B) second COCON will generate all possible permutations of the connectivities for the poorly defined region A and then, for each of these, all possible permutations for the second region B. As only the second region contains the correlations, it will decide whether an overall permutation is a valid solution or not. If the input file is resorted to contain first the well-defined region B and then the poorly defined region A, a permutation of region A is only calculated, when the actual permutation of region B seems promising for an overall solution. The number of generated structures is the same for both cases but not the number of permutations. Therefore, the input data set is automatically sorted.[9,10]

The improvements associated with the preorganization of the input data set become apparent when the calculation times obtained for the optimized data sets of aflatoxin $B_1$ (**2**, Figure 1), 11-hydroxyrotenone (**3**, Figure 1), and haemoventosin (**4**, Figure 1)[11] are compared with those of ref 1c (see Table 1). For the data set of aflatoxin $B_1$ (**2**) the calculation time

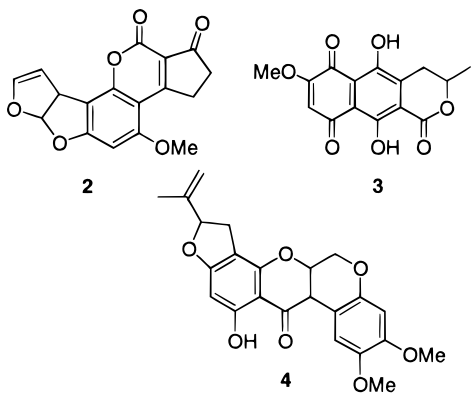



**Figure 1.** Structures of aflatoxin $B_1$ (**2**), 11-hydroxyrotenone (**3**), and haemoventosin (**4**).

using COSY and HMBC correlations and open hybridization states is reduced by a factor of almost 300 from 23 h 31 min to 4 min 47 s.

The COCON calculation results for the data set of ascomycin (**1**) are shown in Table 2. Despite 52 COSY and 86 HMBC correlations COCON generated 350 possible structures with predefined hybridization states for all atoms. With open hybridization states COCON generated 6389 atom type combinations (ATCs). Only 4 of them yielded in total 2015 constitutions. Of course, the number of constitutions is reduced if theoretical 1,1-ADEQUATE correlations are used additionally, in this case from 2015 to 273 (see Table 2). Finally, the theoretical data set for ascomycin containing all correlations due to $^3J_{HH}$, $^2J_{CH}$, and $^3J_{CH}$ couplings led to 6389 ATCs and 23 constitutions. In contrast to proton-poor compounds analyzed earlier, ascomycin (**1**) is well-defined by its complete correlation data set. The major problem with larger systems is the overlap of signals in the spectra and therefore usually only a reduced correlation data set is available. This is especially a problem for ascomycin (**1**) because it exists as two rotamers in solution. We have shown that methods of $^{13}$C NMR chemical shift prediction (SpecEdit[12]) can assist chemists with regard to refining the selection among the constitutions proposed by COCON.[1c] This approach allows a selection or ranking of the generated structures as well as a ranking of the ATCs.[1c] A SpecEdit calculation[12] was carried out for the resulting COCON structures of **1** obtained with the experimental data set. The averaged $^{13}$C NMR chemical shifts of all four possible ATCs of **1** are very similar and no distinction is possible (see Table 3). When considering only the best 10 structures of each ATC, the correct ATC (no. 1354, 350 constitutions) is clearly favored (see Table 3). Out of the 350 constitutions corresponding to the correct ATC no. 1354, ascomycin (**1**) is ranked at position 10 in the SpecEdit calculations with an averaged $\delta(^{13}C)$ deviation of 4.35 ppm ($\langle\Delta\delta(^{13}C)\rangle$). The best 20 constitutional proposals remaining after the SpecEdit analysis are discussed here (all 20 structures are shown in the Supporting Information).

Several of the 20 structures can be excluded because of their chemical instability under laboratory conditions, e.g., carbamic acids. In Figure 2 the final two proposals (**1**-282 and **1**-286) are shown as well as the two structures which were rated first (**1**-222, $\langle\Delta\delta(^{13}C)\rangle = 3.51$ ppm) and second (**1**-271, $\langle\Delta\delta(^{13}C)\rangle = 3.56$ ppm) with regard to the deviation of their $^{13}$C NMR chemical shift from the calculated values (SpecEdit). The structural proposal **1**-286 is the correct constitution of ascomycin.

**Table 2.** COCON Results for Ascomycin (**1**)

| *a* | *b* | | COSY, HMBC | COSY, HMBC, 1,1-ADEQ |
|---|---|---|---|---|
| **1** (oh) | exp | no. of ATCs[c] | 6389 (4) | 6389 (4) |
| | | no. of structures | 2015 | 273 |
| | | calculation time[d] | 9 h 32 min (89 h 11 min) | 2 h 16 min (5 h 14 min) |
| **1** (fh) | exp | no. of structures | 350 | 76 |
| | | calculation time[d] | 48 min 6 s (8 h 27 min) | 11 min 26s (26 min 51 s) |
| **1** (oh) | theo | no. of ATCs[c] | 6389 (4) | 6389 (3) |
| | | no. of structures | 23 | 6 |
| | | calculation time[d] | 5 min 43 s (18 min 50 s) | 4 min 19 s (7 min 50 s) |
| **1** (fh) | theo | no. of structures | 8 | 2 |
| | | calculation time[d] | 13 s | 4 s |

[a] COCON calculations with predefined (fixed) hybridization states (fh) and COCON calculations with open hybridization states (oh). [b] The abbreviation "exp" stands for experimental data set, "theo" for theoretical data set. [c] The number in parentheses is the number of atom type combinations (ATCs) with COCON solutions. [d] The calculations were carried out on a SGI R10000, 195 MHz processor, the source code of COCON was 64-bit optimized. The number in parentheses is the calculation time obtained with the previous COCON version (see results for **2**−**4** in ref 1c).

**Table 3.** SpecEdit Results for Ascomycin (**1**)

| ATC | no. of structures | averaged $^{13}$C chemical shift deviation[a] for all constitutions | averaged $^{13}$C chemical shift deviation[a] for the best 10 |
|---|---|---|---|
| 1353 | 1109 | 10.63 (2.90) | 5.62 (0.09) |
| 1354 | 350 | 9.78 (3.32) | 3.88 (0.30) |
| 1358 | 543 | 9.78 (2.77) | 5.48 (0.12) |
| 1359 | 13 | 8.96 (2.66) | 7.77 (1.29) |

[a] The rms deviation is given in parentheses.

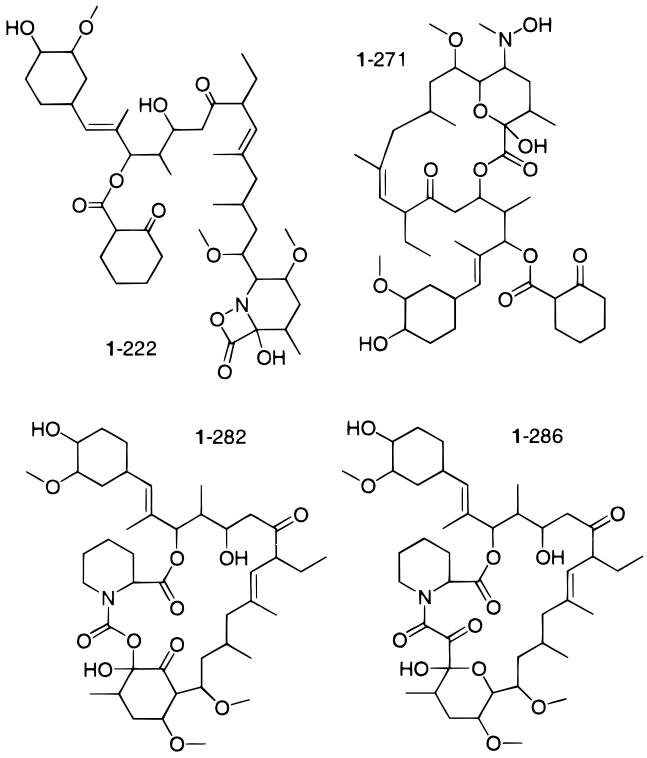



**Figure 2.** Constitutional proposals for ascomycin (**1**).

Our short investigation demonstrates that the computer program COCON is able to cope with large systems as shown for ascomycin (**1**) with more than 50 heavy atoms and even very large systems such as palytoxin with more than 150 heavy atoms. COCON generated only one constitutional proposal for palytoxin using the complete COSY and HMBC data set.[13] For ascomycin (**1**), it was shown that the subsequent calculation of the $^{13}C$ chemical shifts using SpecEdit is a very valuable tool for validating structural proposals generated by COCON. The accelerated calculation speed of the new COCON version was demonstrated for ascomycin (**1**), aflatoxin $B_1$ (**2**), 11-hydroxyrotenone (**3**), and haemoventosin (**4**).

**Acknowledgment.** This research was supported by the Deutsche Forschungsgemeinschaft (Ko 1314/3-1 to 3-4 and Li 597/2-3). T.L. thanks Professor Dr. Richard Neidlein for his generous support. J.J. and M.K. acknowledge the support of Professor Dr. Christian Griesinger.

**Supporting Information Available:** Table with the NMR correlation data obtained for ascomycin (including a numbered structure) and a figure with 20 constitutional proposals for ascomycin which were rated best after the SpecEdit calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

OL990725B

(6) All NMR correlations of ascomycin (**1**) are given in the Supporting Information. All NMR measurements of **1** were carried out in a 5-mm tube with 8 mg in 0.7 mL of $CDCl_3$ on a Bruker AMX360 and a Varian Unity Inova 600. The following experiments were carried out: 1D $^1H$ and $^{13}C$ NMR, COSY, HSQC, and HMBC. The delay for the heteronuclear long-range couplings was optimized to 8 Hz (62.5 ms) in the HMBC experiment (256 acquisitions and 128 complex increments which represent 256 FIDs). Because of the very good *S*/*N* ratio of the HMBC spectrum, linear prediction was applied in $F_1$ (factor of 2).

(7) Ley, S. V.; Doherty, K.; Massiot, G.; Nuzillard, J.-M. *Tetrahedron* **1994**, *50*, 12267−12280.

(8) Peng, C.; Yuan, S.; Zheng, C.; Shi, Z.; Wu, H. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 539−546.

(9) This allows the user to number the atoms, for example, first by the order number and then by the decreasing carbon chemical shift. Since the HMBC correlations are given from protons to carbons, the first entries are usually nonprotonated carbons and therefore without HMBC correlations. For a known compound a given numbering might be used and therefore the resulting input file is unpredictable.

(10) The presorting is also advantageous for the detection of constitutionally equivalent isomers because identical atoms with identical connectivities are automatically set in one group.

(11) The correlation data used for the COCON analysis of haemoventosin were originally obtained for the *O*-acetyl derivative.

(12) Maier, W. In *Computer − Enhanced Analytical Spectroscopy*; Wilkins, C. L., Ed.; Plenum Press: New York, London, 1993; Vol. 4, pp 37−55.

(13) The calculation time was 15 min on a SGI R10000 (195 MHz).