

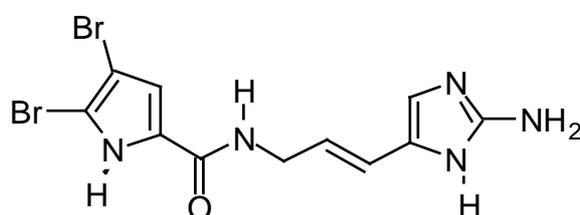
XCOCONPRO98: The Manual

2D NMR
molecular formula



XCOCONPRO 98

("press return")



plus **ALL** alternatives

Thomas Lindel

Jochen Junker

Matthias Köck

Heidelberg · Frankfurt 1998

Contents:

1. *What is XCOCONPRO98 for?*
2. *Constraints guiding the structure generation*
 - 2.1. *^{13}C NMR chemical shifts*
 - 2.2. *Geometry Requirements*
3. *Installation of XCOCONPRO98*
4. *The Input File*
 - 4.1. *General Calculation Options*
 - 4.2. *The Connectivity Table*
5. *Performing an XCOCONPRO98 Calculation*
 - 5.1 *Generating the COCONPRO98 Outputfile*
 - 5.2 *Visualization via XWindows*

1. *What is XCOCONPRO98 for?*

2D NMR experiments as COSY, HMQC, and HMBC have become standard tools in the structure elucidation of unknown organic molecules. They are applied worldwide on a routine basis, especially for the determination of unknown constitutions.

The computer software package XCOCONPRO98 (COCON: **C**onstitutions from **C**onnectivities") allows chemists to *comprehensively* discuss the entirety of molecular constitutions compatible with the *2D NMR data* of unknown compounds with known molecular formulas. With the help of a very efficient structure generator, XCOCONPRO98 extremely accelerates the process of reliable constitutional analysis. In particular, XCOCONPRO98 is a novel tool allowing the search for alternatives of structural proposals [1,2]. Assignment isomers representing identical constitutions with different assignments of ^{13}C NMR chemical shifts are treated as separate solutions.

There are two general applications for XCOCONPRO98:

A. Comprehensive discussion of structural proposals.

B. Structure elucidation of new compounds.

The calculated constitutions are visualized as 2D structural formulae under XWindows. In addition, "mdl" and "sdf" files are generated guaranteeing compatibility with most commercially available visualization programs.

2. Constraints guiding the generation of constitutions

It is immediately clear that structure generation without the use of connectivity information must lead to unmanageable numbers of structural proposals. Key feature of the structure generator *COCONPRO* being the heart of the XCOCONPRO98 software package is the evaluation of ambiguous HMBC-derived connectivity information. For a description of the modes of interpretation of connectivity information see chapter 5.

In addition to connectivity information, *COCONPRO* follows a few additional constraints to be briefly discussed.

2.1. ^{13}C NMR chemical shifts

COCONPRO does not apply any database. Nevertheless it takes into account the following, common sense rules:

a) with regard to the selection of hybridization states:

a carbon atom must not be the center of an allene if	δ_{C}	<	180 ppm,
a carbon atom must not be sp^2 -hybridized if	δ_{C}	<	75 ppm,
a carbon atom must not be sp -hybridized if	δ_{C}	>	115 ppm,
a carbon atom must not be sp^3 -hybridized if	δ_{C}	>	125 ppm.

b) with regard to bond formation:

C=S and C=O bonds are forbidden if	δ_{C}	<	150 ppm,
aliphatic-C-O bonds are forbidden if	δ_{C}	<	45 ppm,
olefinic-C-O bonds are forbidden if	δ_{C}	<	130 ppm,
olefinic-C-N bonds are forbidden if	δ_{C}	<	105 ppm,
Met-C bonds are forbidden if	δ_{CH_3}	>	35 ppm.

The only available ^{13}C NMR chemical shift rule referring to more than two atoms may optionally be activated by choosing the flag "multi_hetero = 1":

"An sp^3 -hybridized carbon atom with a ^{13}C NMR chemical shift above 90 ppm must be connected to at least two non-carbon hetero atoms."

2.2. Geometry Requirements

- a) *COCONPRO* always generates structural proposals consisting of a single molecule.
- b) Every structural proposal can be brought into an agreement with a multiple bond system.

3. Installation of XCOCONPRO98

The software package "XCOCONPRO98" contains the programs *xcocon*, *coconpro*, *cocon2mdl*, *molcoord* [2] and the shell script *run_molcoord*. All binary codes have been compiled on an SGI workstation (mips2 or mips 4). Guided by connectivity information, *coconpro* generates the entirety of constitutions compatible with a given molecular formula. *cocon2mdl* converts the connectivity strings generated by *coconpro* into 2D plain structures in mdl-file format (2D coordinates generation by *molcoord* which is automatically started by *cocon2mdl* resp. *run_molcoord*). The mdl files can be visualized using the implemented XWindows window (*xcocon*).

a) Sent by e-mail

You will receive a tar file of the software package (*cocon.tar.gz.uue*) with all programs and files you need to run and to test the program. This tar file was zipped and uuencoded. First, this file has to be decoded using the UNIX command *uudecode cocon.tar.gz.uue*. In a next step this file has to be unzipped using the command *gzip -d cocon.tar.gz*. The last command is *tar xvf cocon.tar* which generates the directory *cocon* containing all files.

b) Sent on a floppy disk or CD Rom

On this floppy disk or CD Rom you will see the tar file (see a) or just all files. In the latter case you have just to copy the files onto your hard disk.

Files: *xcocon*, *coconpro*, *cocon2mdl*, *molcoord*, *run_molcoord* and *oroidin.in*.

4. The Input File

The input file of *coconpro* should be created in the vi-mode (UNIX version) or any other available editor. Its name may be chosen according to the UNIX rules using the extension ".in" and it will appear in the *coconpro* start command line (see section 5.1.).

It is absolutely crucial for running *coconpro* that the input file is created very carefully without any typing errors! Blank lines (frequently occurring at the end of the input file if conventional word processors are used) are not allowed and would lead to a core dump error while executing *coconpro*.

For an example of an input file see Appendix 2 (data set of aflatoxin B1, basis of the calculation of structural alternatives). Every line of additional *comment* within the input file has to start with the hash symbol ("#").

4.1. General Calculation Options

XCOCONPRO98 calculations can optionally be guided by constraints regarding allowed substructures and interpretation of connectivity information. These are set by the option flags:

a) charge_

The flag "charge_" refers to the way XCOCONPRO98 selects possible hybridization state combinations compatible with the molecular formula. On default, (charge_ = 0) only neutral (non-radical!) and cationic hybridization states are considered. This selection may either be restricted to exclusively neutral, non-radical hybridization states (charge_ = 1) or be extended to all hybridization states (charge_ = 3). If charge_ is set to 2, only neutral and radical hybridization states are regarded.

b) cy_propene

If "cy_propene" is set to 1, XCOCONPRO98 excludes every constitution containing a cyclopropenoid structure, such as cyclopropenes, azirines, and oxirenes. Three-ring radialenes are also excluded.

c) cy_butadiene

If "cy_butadiene" is set to 1, XCOCONPRO98 excludes every constitution containing cyclobutadienoid structures only a few examples of which are known at all. Cyclobutadienoid radialenes are still allowed.

d) multi_hetero

If the flag "multi_hetero" is set to 1, carbon atoms will never be connected to more than one protonated oxygen substituent (e. g. -OH groups). In addition, sp³-hybridized carbon atoms with a chemical shift of more than 90 ppm will always be connected to two or more heteroatoms.

e) non_cosy

There are three possibilities of setting the non_cosy flag. non_cosy = 0 means that two proton bearing atoms may be directly connected to each other, even if there is no $^1\text{H},^1\text{H}$ -COSY correlation observed. This may be the case, if the dihedral angle between the protons is close to zero degrees. Setting non_cosy = 1 forbids every connection between proton bearing carbon atoms for which no $^1\text{H},^1\text{H}$ -COSY correlation is observed. Setting non_cosy = 2 forbids every connection between any proton bearing atoms for which no $^1\text{H},^1\text{H}$ -COSY correlation is observed. This selection can only be made, if the COSY data were obtained in aprotic NMR solvents.

f) non_adeq

Setting the flag "non_adeq" to 1 forbids any direct connection between two carbon atoms of which at least one bears a proton, as soon as no 1,1-ADEQUATE correlation is defined between them.

4.2. The Connectivity Table

Every heteroatom (=non-hydrogen atom) has to be assigned an atom number. The kinds and numbers of heteroatoms can easily be derived from the molecular formula. The currently manageable maximal number of heteroatoms is 100.

Appendix 2: 17 carbon atoms, 6 oxygen atoms, leading to 23 lines.

Column 1 (No.):

Atom numbers assigned to each non-hydrogen atom by the operator. Atoms must be listed consecutively without omitting a number.

Appendix 2: Atom number 1 to 23.

Column 2 (OZ):

Order number (periodic table of elements) of every atom of column 1.

Appendix 2: No. 1-16, 23: carbon, OZ 6; No. 17-22: oxygen, OZ 8.

Column 3 (numH):

Number of hydrogen substituents attached to the heteroatom of column 1. This information is derived from DEPT experiments for carbon atoms and from ^1H NMR spectra for most heteroatom-bound hydrogen atoms. In rare cases, the number of hydrogen substituents have to be assumed for certain heteratoms in order to comply with the molecular formula.

Appendix 2: No. 1, 2, 4, 5, ,7, 9, 10, 11, 17-23: numH = 0; No. 8, 12-15: numH = 1; No. 3, 4: numH = 2.

Column 4 (Cshift):

^{13}C NMR chemical shifts (integers!) of every atom of column 1. For heteratoms other than carbon, a hyphen ("-") has to be entered.

Appendix 1: hyphens in the columns of the (oxygen) atoms 17 to 22, otherwise shifts.

Column 5 (Atype):

Optionally, heteroatom types (see Appendix 2) may be defined. Otherwise, a hyphen ("-") has to be entered. The contents of column 5 are disregarded as soon as "-a" is omitted on starting the *coconpro98* calculation.

Columns 6 to 11: Input of connectivity information. Entry of the partners of the atoms of column 1. As soon as no connectivity information is available for a certain atom, a hyphen ("-") has to be entered in the corresponding column.

If an atom has several entries in any of the connectivity columns 4-9, the correlating partner numbers have to be separated by comma and blank, *e. g.* (6, 12). *There must be no comma following the last entry of a column.*

Column 6 (COSY):

Connectivity information obtained from $^1\text{H},^1\text{H}$ -COSY experiments. Although $^1\text{H},^1\text{H}$ -COSY correlations are observed for both involved hydrogen atoms, it is allowed to enter the connectivity information for only one partner. The interpretation of COSY information is defined by the flag "non_cosy" in the option list (see below). Attention: connectivity information in column 6 is strictly interpreted as a *direct connection* of the heteroatoms!

Appendix 2: *e. g.*, "4" in line "3": "based on COSY considerations, atom 3 is connected to atom 4."

Column 7 (HMBC):

Connectivity information obtained from $^1\text{H},^{13}\text{C}$ -HMBC experiments. The unsymmetry of HMBC spectra should be reflected in the input file and only the observed direction (^1H to ^{13}C) should be entered. Attention: connectivity information in column 7 is strictly interpreted as a *direct connection or a two-bond distance* of the heteroatoms!

Appendix 2: *e. g.*, "1, 2, 4, 5" in line "3": "at least one hydrogen atom attached to the heteroatom 3 is either 2 or 3 bonds distant from the heteratoms 1, 2, 4, and 5".

Column 8 (ADEQ):

Connectivity information obtained from 1,1-ADEQUATE experiments. The unsymmetry of 1,1-ADEQUATE spectra should be reflected in the input file and only one direction (from ^1H) should be entered. The interpretation of 1,1-ADEQUATE information is defined by the flag "non_adeq" in the option list (see below). Attention: connectivity information in column 8 is strictly interpreted as a *direct connection* of the heteroatoms!

Column 9 (NHMBC):

Connectivity information obtained from $^1\text{H},^{15}\text{N}$ -HMBC experiments. The unsymmetry of HMBC spectra should be reflected in the input file and only one direction (^1H to ^{15}N) should be entered. Attention: connectivity information in column 7 is strictly interpreted as a *direct connection or a two-bond distance* of the heteroatoms!

Column 10 (fixed):

Those atoms which are defined by the operator to be directly connected to the atom number in column 1.

Column 11 (forbidden):

Those atoms which are defined by the operator *not* to be directly connected to the atom number in column 1.

5. Performing an XCOCONPRO98 Calculation

5.1 Generating the COCONPRO98 Outputfile

The command line starting a *coconpro98* calculation is

```
coconpro inputfile [six-digit binary code] -a > outputfile
```

The six-digit binary code specifies the consideration of input connectivity information. *E. g.*, the code "111000" (= "column 6, column 7, column 8, not column 9, not column 10, not column 11") would tell *coconpro* to include $^1\text{H},^1\text{H}$ COSY data, $^1\text{H},^{13}\text{C}$ HMBC data, and 1,1-ADEQUATE data. The columns corresponding to $^1\text{H},^{15}\text{N}$ HMBC data, to fixed bonds, and to forbidden bonds would be disregarded. In that way, a single input file may be used for

different calculations regarding different sets of information. If the six-digit binary code is not given in the command line the defined value of the flag *nmr_corr* in the input file is used.

As *outputfile.out*, e. g. *oroidin.out* may be chosen. For identification purposes, it is advised to use the six-digit binary code even as part of the outputfile, e. g.,

```
coconpro oroidin.in 111000 -a > aflatoxin111000.out
```

If "-a" is used, the atom types (column 5) are taken into account. A hyphen will be interpreted as an undefined atom type.

5.2. Visualization via XWindows

To start XWindows visualization tool type: *xcocon*. If the working directory already contains mol files they will be shown in the left column if you click on *Run* under *Menue*. If you click on the structure name it will be visualized on the right side of the Xwindow (Quickview). If there are no mol files in the working directory but a COCONPRO output file (must be named as *coconpro.out*) the program automatically generates the mol files after click on *Run* under *Menue*.

The structures can be viewed with protons (button *Protons on/off*), with carbons (button *Carbons on/off*) and with atom No. (button *Atom No. on/off*).

To start a COCONPRO calculation in the XWindow you click on *Input Infos* under *Menue*. Two things have to be typed in a) Name of the input file (w or w/o options, e.g. 111010 and -a) and b) the project name. Quit this window with *OK*. The calculation is started by *Run* under *Menue*.

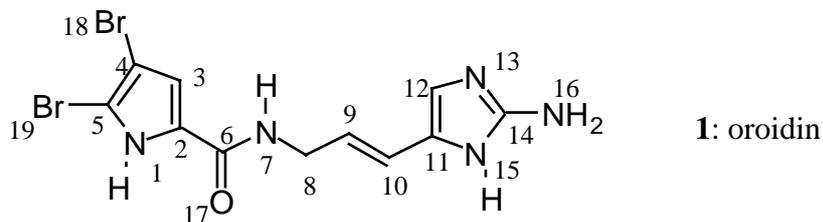
To edit the input from the XWindow use the *Edit* command under *Menue* (xedit editor). Type in the name of the input file and click on *Load* to view the input file.

Under *Menue* it is possible to define if the mol files should be deleted or not. Furthermore, an sdf file (all mol files in one file) can be generated.

To quit the program use *Quit* under *Menue* (mol and sdf files are usually deleted if not specified in the *Menue*!!!).

The calculation itself can only be interrupted by pressing "ctrl-c".

Appendix 1: Example of an XCOCONPRO Input File, based on the data of oroidin (1).



Which molecules should be accepted?

```
single_mol [0 1]          1
multiple_bond [0 1]       1
charge_ [0 default, 1 only neutral, 2 neutral and radicals, 3 all] 1
#
```

Exclusion of certain structural elements

```
cy_propene [0 1]         1
cy_butadiene [0 1]       1
hydrate_ex [0 1]         1
no_peroxy [0 1]         1
#
```

Inclusion of very coarse chemical shift rules

```
shift_ch [0 1]          1
#
```

Inclusion of NMR correlations

```
non_cosy [0 1 2]        2
non_adeq [0 1]          1
#
```

Which kind of NMR correlations should be included?

[COSY HMBC ADEQ NHMBC]

```
nmr_corr                1110
#
```

# No.	OZ	numH	Cshift	Atype	COSY	HMBC	ADEQ	NHMBC
1	7	1	-	66	-	2, 3, 4	-	-
2	6	0	128	11	-	-	-	-
3	6	1	113	74	-	2, 5	2, 4	1
4	6	0	98	11	-	-	-	-
5	6	0	105	11	-	-	-	-
6	6	0	159	11	-	-	-	-
7	7	1	-	66	8	6, 8	-	-
8	6	2	40	130	7, 9	6, 9, 10	9	7
9	6	1	127	74	8, 10	8, 10, 11	8, 10	7
10	6	1	116	74	9	8, 9, 11, 12	9, 11	15
11	6	0	125	11	-	-	-	-
12	6	1	111	74	-	10, 11, 14	11	13, 15
13	7	0	-	10	-	-	-	-
14	6	0	148	11	-	-	-	-
15	7	1	-	66	-	11, 12, 14	-	-
16	7	2	-	129	-	-	-	13, 15
17	8	0	-	9	-	-	-	-
18	35	0	-	1	-	-	-	-
19	35	0	-	1	-	-	-	-

Appendix 2: List of Available Atom Types (for Column 5 ("Atype") of the Input File)

Li: OZ 3

-Li: 1; Li(+): 257;

Be: OZ 4 (analogous: Mg: OZ 12)

>Be: 2;

B: OZ 5 (analogous Al: OZ 13, Ga: OZ 31, In: OZ 49, Tl: OZ 81)

-BH2: 129; >BH: 66; >B-: 3; =BH: 73; =B-: 10; ≡B: 33;

-BH3(-): 705; >BH2(-): 642; >BH(-): 579; >B<(-): 516; =BH2(-): 649; =BH(-): 586;

=B<(-): 523; =B(-): 530; ≡BH(-): 609; ≡B(-): 546;

C: OZ 6 (analogous Si: OZ 14, Ge: OZ 32, Sn: OZ 50, Pb: OZ 82)

-CH3: 193; >CH2: 130; >CH-: 67; >C<: 4; =CH2: 137; =CH-: 74; =C<: 11; =C=: 18;

≡CH: 97; ≡C-: 34;

-CH2(·): 897; >CH(·): 834; >C(·)-: 771; =CH(·): 841; =C(·)-: 778; ≡C(·): 801; -CH2(+): 385;

>CH(+): 322; >C-(+): 259; =CH(+): 329; =C-(+): 266; +C(+): 289; -CH2(-): 641; >CH(-):

578; >C(-): 515; =CH(-): 585; =C(-): 523; +C(-): 545; -C-: 1026; =C: 1033.

N: OZ 7

-NH2: 129; >NH: 66; >N-: 3; =NH: 73; =N-: 10; ≡N: 33;

-NH3(+): 449; >NH2(+): 386; >NH-(+): 323; >N<(+) : 260; =NH2(+): 393; =NH-(+): 330;

=N<(+) : 267; =N=(+) : 274; ≡NH(+): 353; ≡N-(+) : 290;

-NH(·): 833; >N(·): 770; =N(·): 777; -NH(-): 577; >N(-): 514; =N(-): 521; -N: 1025.

P: OZ 15 (analogous As: OZ 33, Sb: OZ 51, Bi: OZ 83)

-PH2: 129; >PH: 66; >P-: 3; =PH: 73; =P-: 10; +≡P: 33; >P<-: 5; =P<-: 14; =P=: 19;

O: OZ 8

-OH: 65; >O: 2; =O: 9;

-OH2(+): 385; >OH(+): 322; >O(+): 259; =OH(+): 329; =O(+): 266; +O(+): 289; -O(·): 769;

-O(-): 513.

S: OZ 16 (analogous Se: OZ 34, Te: OZ 52)

>>S<: 6; >>S=: 13; =>S=: 20; ==S=: 27; >S<: 4; =SH-: 74; =S<: 11; =S=: 18; -SH: 65;

-S-: 2; =S: 9; >>S(·)-: 773; >-S=(·): 780; =S(·)=: 787; >S-(·): 771; =S-(·): 778; -S(·): 769;

>>S(+)-: 261; >-S=(+): 268; =S(+)=: 275; >S-(+): 259; =S-(+): 266; -S(+): 257; =S(-)-: 522;

-S(-): 513;

F: OZ 9 (analogous Cl: OZ 17, Br: OZ 35)

-F: 1.

I: OZ 53

-I: 1; >I-: 3; =I-: 10; >>I-: =5; =>I-: 12; ==I-: 19; >>I<-: 7;=>I<-: 14; ==I<-: 21; ==I=-: 28;

[1] T. Lindel, J. Junker, M. Köck, *J. Mol. Model.* **1997**, *3*, 364-368.

[2] Parts of the visualization software (function "molcoord") were written by Dr. Eric Fontain, TU München, Germany.