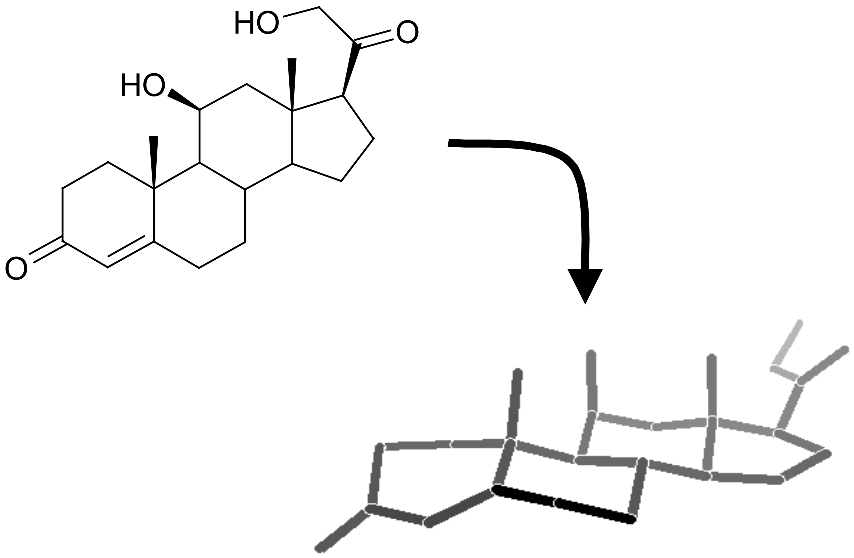
CORINA

3D Structure Generator

Version 3.0

Program Description



Jens Sadowski, Christof H. Schwab, and Johann Gasteiger

Molecular Networks GmbH Computerchemie
March 2003

http://www.mol-net.de

Molecular Networks GmbH
Computerchemie
Nägelsbachstr. 25
91052 Erlangen
Germany

Phone: +49-(0)9131-815668
Fax: +49-(0)9131-815669

Email: info@mol-net.de
WWW: www.mol-net.de

Contents

1 Program Installation

Since verion 2.6 CORINA is distributed on a CD-ROM, which contains the executable file(s) of CORINA, this program description in PDF format, and some example files of structure information (see section 4 on page 12).

The CD-ROM contains an ISO9660 file system and, thus, is readable by all common UNIX systems as well as by Microsoft Windows (win32) based platforms. The following directories and files are common for all hardware platforms.

name of directory	**description**	**name of file(s)**
examples	example files for structure data (ASCII)	*example.ctx* *example.sdf*
manual	this program description in PDF format	*corina30manual.pdf*

Please copy the example files *example.ctx* and *example.ctx* into your home directory. The program description *corina30manual.pdf* can be viewed and printed with a PDF document viewer, e.g. Adobe Acrobat Reader (http://www.adobe.com/acrobat).

In addition, the CD-ROM contains at least one the following directories, where the executable files of CORINA for the various hardware platforms reside.

name of directory	**executable file of CORINA for**	**name of file**
irix65	SGI workstations, IRIX 6.5	*corina.sgi*
linux22_redhat	x86 Linux workstations, Kernel 2.2, distribution by RedHat (7.0)	*corina.lnx*
linux24_redhat	x86 Linux workstations, Kernel 2.4, distribution by RedHat (8.0)	*corina.lnx*
linux24_suse	x86 Linux workstations, Kernel 2.4, distribution by SuSE (7.x)	*corina.lnx*
solaris26	Sun SPARC stations, Solaris 2.6	*corina.sun*
solaris8	Sun SPARC stations, Solaris8	*corina.sun*
tru64	DEC AlphaStations, Tru64 (OSF1)	*corina.dec*
win32	Microsoft Windows platforms (win32: NT4/95/98/2000/XP)	*corina.exe*

1.1 New Installation

1.1.1 UNIX Systems (SGI, Sun SPARC, x86 Linux, DEC AlphaStations)

1) Create a subdirectory, e.g., *corina* (for system administrators when installing software locally, e.g. */usr/local/bin/corina*).

2) Copy the executable file of CORINA *corina.sgi/sun/lnx/dec* from the CD-ROM to the subdirectory *corina* and rename the file *corina. sgi/sun/lnx/dec* to *corina*. **Please note:** *corina.sgi/sun/lnx/dec* is a binary file.

3) Add the *corina* subdirectory name to the environment variable *PATH* in your *.login* or *.cshrc* files (*.profile* or *.bashrc*).

1.1.2 Microsoft Windows Platforms (win32: NT4/95/98/2000/XP)

The directory *win32* on the CD-ROM contains the win32 executable file.

1) Create a subdirectory, e.g., *corina* (for system administrators when installing the software locally, e.g. *X:\programs\corina*).

2) Copy the file *corina.exe* from the CD-ROM to the subdirectory *corina*. **Please note:** *corina.exe* is a binary file.

3) Add the *corina* executable file (*corina.exe*) and the path where the program resides (e.g. *X:\programs\ corina*) to your environment variables of your system settings (variable: *corina*; value: *X:\programs\ corina*).

1.2 Program Updates

1) Before installing the new version, please copy the old executable and configuration files to a new directory, e.g. *corinaxxx* (*xxx* = *old-version-number*, e.g., *corina24*).

2) According to the hardware platform install the new version following the instructions given in section 1.1 on page 2.

3) **Please note:** Since CORINA version 2.4, the data files *stdval.ctx* and *rings.ctx* are no longer part of the distribution. All data has been included in the binary file of CORINA (see section 3.1.6)

2 Problems and Help!

If you have any difficulties with the installation of CORINA or if any problems occur while running CORINA, please send all your inquiries to the following address:

Molecular Networks GmbH Computerchemie
Nägelsbachstr. 25
91052 Erlangen
Germany

or contact us by email support@mol-net.de,
or by Fax +49-(0)9131-815669.

Please include your input file, the output file, and the CORINA trace file *corina.trc* generated by CORINA on an MS/DOS diskette (3½") or send it to us by email. These files will help us to analyze the problem; if your system displays any error messages, please add them to your report. Thank you!

You can also use the report form at the end of this manual.

3 Release Notes

3.1 CORINA (Full Version)

3.1.1 Version 1.6

CORINA version 1.6 represents a substantial improvement of version 1.5. Both the quality of the results became higher and the program became more flexible. There are five major changes in version 1.6 compared to version 1.5.

1) The input file format SMILES linear notation was added [1].
2) The output file formats SYBYL MOL/MOL2 [2] and the Brookhaven Protein Databank PDB [3] were added.
3) The algorithm, which refines atom overlap and close contacts was improved by implementing of a set of rules obtained from a statistical analysis of the conformational preferences of open-chain portions in small molecule crystal structures contained in the Cambridge Structural Datafile (CSD) [4], [5].
4) A substantial speed-up of almost a factor of 2 was achieved by optimizing the the algorithm.
5) The command line options now follow the UNIX command syntax standard.

The quality and speed improvements are illustrated in detail in section 6.6 on page 29 of this manual. A side-effect of the quality improvements is of course that the resulting 3D structures for a number of structural classes might have changed.

The changes in the command syntax might cause some portability inconveniences for the user but gave more flexibility for the addition of new options as, e.g., the new input and output file specifications. The old options are no longer valid—the program exits with an error message when recognizing the use of the old syntax.

3.1.2 Version 1.7

CORINA version 1.7 was tailored especially to the database business:

1) The two new driver options **-d flapn** and **sc** were added for generating multiple ring conformations.
2) The additional PDB output options **-o pdbludi** and **pdbludilabel** allow the generation of fragments for databases interfacing to the *de novo*-design program Ludi [6].

An exhaustive study on the effect of multiple ring conformations on the performance of flexible 3D pharmacophor searches was performed (see section 9 on page 47).

3.1.3 Version 2.0

CORINA 2.0 is now able to interact with the ligand docking program FlexX [7] as a conformer generator for ring systems (see section 7.8 on page 39). Thus, CORINA ring conformations can be used for flexible ligand docking into a receptor pocket. Changes were mainly made concerning the file format interfaces and in the ring conformation options.

1) Two new input file formats SYBYL MOL/MOL2 [2] (**-i t=mol** and **mol2**) as required by FlexX were added.

2) A number of new options were introduced for ring conformations (**-d de**, **timeout** and **flexx**) for tailoring the results for FlexX.

3.1.4 Version 2.1

The following changes and improvements were made:

1) The SMILES interface was made more stable (many thanks to the people at Oxford Molecular and Dr. Peter Ertl, Novartis for useful hints).

2) Three new options **-d ow**, **amide**, and **-i sdfict** related to the handling of stereochemical information for MDL SDFiles [8] were added (see section 5.2 on page 14).

3) The most important change concerns the handling of the configuration of amide bonds. In earlier versions, the configuration (*cis* or *trans*) was taken from the 2D drawing in the input file. This behavior must now be switched on explicitly. By default, now the most suitable configuration is taken—in most cases *trans*. Thus, cases with unexpected *cis* amides will be no longer generated.

3.1.5 Version 2.3

The following changes and improvements were made:

1) A new option **-d no3d** allows to use CORINA as a file format converter for the supported file formats without generating 3D coordinates.

2) The FlexX interface, the SMILES interpreter and the MDL SDFile were made more stable.

3) Additional ring conformation patterns for cyclo-octa-1,3-diene were added to the template data file *rings.ctx*.

3.1.6 Version 2.4

The following changes and improvements were made:

1) The data files *stdval.ctx* and *rings.ctx* are now inline–easier installation, less mistakes with different versions.

2) The new option **-d 3dst** forces the use of a given 3D configuration instead of the atomic stereodescriptors. This might be useful if the stereodescriptors are not specified properly but the 3D structure is correct.

3) The new option **-d neu** neutralizes formal charges at acids, alcoholates, and basic nitrogens by adding or removing protons. Often it is useful to have all molecules of a database in the same protonation state. This option can be used with the option **-d rs** in order to remove counter-ions from salts.

4) The new option **-d ori** orients the generated 3D structure according to the moments of inertia. This might be useful when the structure is directly forwarded to a graphical viewer. The molecule then appears more often in an orientation that shows as much of it as possible on one sight.

5) Some minor problems in the FlexX and the MDL interfaces with no influence on the 3D generation process were fixed.

3.1.7 Version 2.6

The following improvements and changes were implemented:

1) The file format MDL RDFile [8] was added to the read and write functions of CORINA.

2) In order to provide interfaces to the protein crystallographic and NMR program packages CCP4 [9] and X-PLOR [10] the output file formats CCP4 dictionary file format (**-o dic**), X-PLOR topology (**-o top**), and X-PLOR parameter file format (**-o par**) were added. These features allow in conjunction with the additional options **-o resnam**, **typchr**, **dicid** the generation of input files for the CCP4 and X-PLOR program suites.

3) Atoms with isotopic mass are now defined for MDL SDFile, SMILES linear code, and Gasteiger ClearText format [11].

4) The SMILES reader and interpreter is now more general: SMILES strings containing heteroaromatic rings without explicitly defined hydrogen atoms at the hetero atoms are now tolerated. For example, pyrrole compounds can now be inputted also as the "incorrect" SMILES n1cccc1 according to the SMILES language definition (correct coding: [nH]1cccc1).

5) The SMILES reader now accepts only one SMILES linear code per line. The SMILES code is expected to be the first string in the line. With the input option **-i smilesname**, all following strings are interpreted as compound name and copied into the corresponding field of the output file. Thus, white or blank spaces within the compound name are now allowed.

6) Non-element symbols, dummy atom types or groups like X, R, Du, Lp, D, T, and * are defined for the file formats MDL SDFile, SMILES linear code, and SYBYL MOL/MOL2. For SMILES linear code the interpretation of dummy atom types or groups has to be specified explicitly by using the new input option **-i dummies**.

7) With the new input option **-i csdmol2** specific extensions and information in SYBYL MOL/MOL2 input files, which were generated by the Cambridge Structural Database (CSD) software [5], are written to the output file.

8) The new output option **-o m2l** ("mass to label") copies isotopic mass labels given in the input file into the corresponding atom name field in SYBYL MOL/MOL2 files. Atoms without given mass label remain untouched.The atom name has the format <symbol><mass>. If the corresponding atom is a non-element symbol, the atom name has the format R<mass>. This can be used to create extension points for virtual combinatorial library, e.g., as input files for FlexX.

9) The new output option **-o mdldb** creates the additional data fields <MODEL.SOURCE>, containing information about the program version of CORINA, and <MODEL.CCRATIO>, giving the close contact ratio of the CORINA generated 3D molecular model. This option has been added for compatibility reasons with databases distributed by MDL Information Systems, Inc.

10) The new output option **–o noccat** switches off the automatic conversion of the carbon atom in amidinium-like structures ($[NH_2^+]$=CN) to the carbo-cation type SYBYL atom type C.cat (N$[C^+]$N). The conversion to this atom type, which is the default, is still strongly recommended.

11) The conformational analysis package for small and medium sized ring systems has been improved: CORINA is now able to generate and to output different ring geometries for ring systems consisting of up to nine ring atoms. In lower program versions, this was limited to a ring size up to eight atoms.

12) The conformational analysis package has been extended to a set of over 900 rules to avoide or eliminate close contacts of non-bonded atom pairs in 3D molecular models. These rules have been derived from a statistical analysis of the conformational preferences of open chain portions in small molecule crystal structures contained in the Cambridge Structural Database (CSD) [4],[5],[12].

13) The new driver option **-d sanpyr** allows the generation of pyramidal nitrogen atoms in sulfonamide groups. The default, which is strongly recommended, is the generation of a planar configuration of the nitrogen atom. The sampling of the "out-of-plane" distances of 1216 sulfonamide nitrogen atoms as found in the Cambridge Structural Database (CSD) [5] has shown, that the in majority of cases (901 of 1216 sulfonamides – 74%) an "out-of-plane" distance of less than 0.3 Å is exihibited. Thus, the planar configuration is the preferred geometry compared to the pyramidal configuration.

14) The new driver option **-d newtypes** forces CORINA to generate new atom types for the output file by discarding any given input types plus aromaticity information. This allows the use of CORINA for, e.g., correct retyping of aromatic groups in corrupted input records.

3.1.8 Version 3.0

The following improvements, changes, and new features were implemented:

1) The functionality of the stereoisomer generator STERGEN [13] has been integrated in CORINA. The driver option **-d stergen** forces CORINA to determine all stereocenters in a given input structure and to generate the 3D structures of all possible, but unique stereoisomers. Configurational isomers at tetrahedral coordinated centers as well as at double bonds (*cis*/*trans*) are considered. Duplicate configurations, such as *meso*-compounds are identified and removed. By default (if the driver option **-d stergen** is set), a maximum number of four stereocenters are processed and a maximum number of 16 stereoisomeric compounds are generated. However, the driver options **-d msc** and **msi** allow to set a user defined number of stereocenters which should be processed (**msc=<value>**) and to restrict the total number of generated stereoisomers (**msi=<value>**). Stereocenters which have a defined stereochemistry (stereodescriptor) are also processed, unless the driver option **-d preserve** is set which prevents from processing those centers which have a defined stereochemistry, i.e., a stereodescriptor is given in the input structure.

2) In order to provide interfaces to the molecular modeling package MacroModel [14], CORINA now supports the uncompressed MacroModel structure file format (input option **-i t=mmod**) as well as the Maestro file format (input option **-i t=mae**) [15] as new input and output file formats.

3) In addition, the file format CIF (Crystallographic Information File, **-o cif**) [16] supported by a variety of crystallographic program packages, the file format ODB (O Database file format, **-o odb**) [17] to interface to the crystallographic modeling tool O, and the file format of the NMR structure calculation program DYANA (**-o dyana**) [18],[19] were added.

4) The input option **-i expandapo** forces CORINA to expand attachment points defined in MDL SDFiles ("`M APO`" field in the properties block) into 3D space. The attachment points are added as "artificial" atoms to the connection table (both to the atom and bond list) and 3D coordinates are calculated. Dummy atom types are assigned to the "artificial" atoms, i.e. "Du" in SYBYL MOL/MOL2

files, "*" (first attachment point) and "**" (second attachment point), respectively, and "X" in PDB files. In addition, the atom names of the attachment point atoms are set to "R1" (first attachment point) and "R2" (second attachment point), respectively, in the output file for formats which support atom names (e.g., SYBYL MOL2).

5) The combined input and output option **-i/-o xelement** only has an impact if dummy atom types ("Du") or element symbols which are unknown SYBYL atom types are defined in SYBYL MOL2 input files. The new input option **-i xelement** forces CORINA to derive–if possible–SYBYL atom types either from the atom names or from the element symbol, or to interpret element symbols in order to internally set appropriate atom types for the 3D structure generation process. By default, CORINA then outputs dummy atom types ("Du") for these atoms. In addition, the new output option **-o xelement** allows to write the derived SYBYL atom types or the element symbols ("artificial" SYBYL atom types) to the output file. Please use these options carefully and manually check the results, since ambiguous definitions in the input file might lead to misinterpretations or false assignment of atom types.

6) The new output option **-o mdlcompact** restricts the number of output fields in the atom lines of the atom block in MDL SDFiles (RDFiles) to the x-, y-, and z-coordinates, the atom type (symbol), the mass difference, the atom charge, and the stereochemical atom parity (columns 1 through 7 of the atom block). All other fields in the atom lines are omitted, since they contain no data which is mandatory for 3D structure information. The goal is to save disk space (up to 40%) in case of large files containing hundreds of thousand compounds.

7) If stereo information is missing in the input file CORINA assumes reasonable stereodescriptors following some implemented rules (see section 0 on page 42). The new output options **-o mdl3dparity** forces CORINA to output the stereochemical atom parities in MDL SDF and RDF files which were calculated and used by CORINA during the 3D structure generation process.

8) If the output file format is set to SYBYL MOL2 the new output option **-o gold** forces CORINA to assign the atom types and the bond orders according to the atom and bond types conventions of the docking program GOLD [20],[21] for difficult groups, i.e., functional groups which have more than one canonical form (e.g., guanidinium groups).

9) Furthermore, the new output option **-o fcharges** has only an impact if the output file format is set to SYBYL MOL2 format. Formal atom charges which are given in the input structure are then written to the charge column (column 9) in the corresponding `@<TRIPOS>ATOM` data lines of the SYBYL MOL2 output file. In addition, the keyword `USER_CHARGES` is set in the `@<TRIPOS>MOLECULE` field.

10) Usually, if the atom name and the atom type differs, CORINA tries to derive a reosonable atom name from the atom type for PDB, MacroModel, and Maestro output files. The new output option **-o keepnames** forces CORINA to keep and to output atom names which are defined in the input file.

11) The driver option **-d names** consecutively numbers the different conformations (**-d rc**) or stereoisomers (**-d stergen**) in ensembles which have been generated by CORINA. A counter is added to the compound name of each conformation (*compoundname*_c00*n*, *n*=1,2,3,...*N*; where *N* is the total number of generated conformations) or stereoisomer (*compoundname*_i00*n*, *n*=1,2,3,...*N*; where *N* is the total number of generated stereoisomers). Thus, the different conformations or stereoisomers of the same input structure are named uniquely and can therefore be easily distinguished by other program systems or any post-processing software.

12) Several problems in the interfaces to the various input and output file formats as well as in the 3D structure generation routines were fixed. The authors would like to take the opportunity to thank all CORINA users who made us aware of these insufficiencies in our software and, thus, helped us to make CORINA a more reliable and sophisticated piece of software.

3.2 CORINA-F (Restricted Version)

CORINA-F is a restricted version of CORINA interfacing to the flexible ligand docking program FlexX [7]. The same interface functionality for FlexX is contained in the full version of CORINA since version 2.0 (**-d flexx**, *vide supra* and see section 5.2 on page 14). During the docking process, CORINA generates low-energy conformations for the ring systems of the ligand. The only difference between CORINA and CORINA-F is that the latter runs only if a number of restrictions are fulfilled:

1) Only one input structure is allowed.

2) The structure has to contain one but only one ring system (fused, bridged and spiro systems are regarded as one ring system!).

3) Ring systems having up to nine atoms are allowed.

4) Exocyclic parts exceeding two bonds are not allowed.

The FlexX-CORINA interface is described in more detail in section 7.8 on page 39. The method implemented in CORINA and CORINA-F, respectively, for generating multiple ring conformations is briefly described in section 9.1 on page 47. It is absolutely identical for both program versions.

4 Getting Started

CORINA is a command line oriented program. Therefore, CORINA has to be executed in a UNIX (Linux) shell (e.g., csh, tcsh, or bash) or, on MS Windows platforms, in a MS DOS shell (DOS prompt).

The example file *example.sdf* provided with the distribution contains the structure information of three molecules in MDL SDFile format [8], which is the default file format for input and output of CORINA.

Please, copy this example file into your working directory and type the following command at the command line prompt:

```
corina example.sdf out.sdf
```

CORINA now creates the output file *out.sdf* containing the input information and the generated 3D coordinates. Figure 1 shows the generated 3D structures.

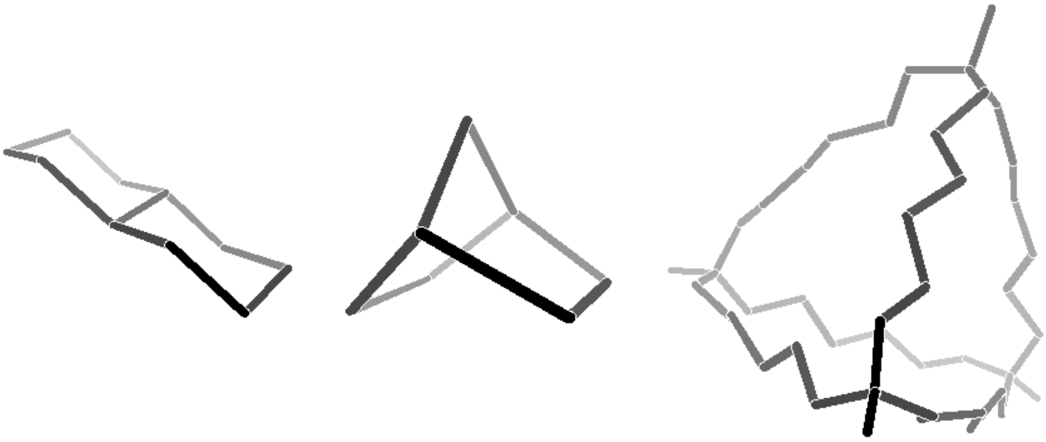


Figure 1 3D models of the structures of the example files.

Furthermore, a trace file (log file) named *corina.trc* which contains information on the CORINA run, such as used parameters, defined options, computation times, details on the 3D structure generation process, as well as warning and error messages (see section 10 on page 53 and section 11 on page 59), is created in your working directory. The additional trace file option **-t s** redirects this trace information to your standard output device. With the command

```
corina -t s example.sdf out.sdf
```

the following output appears on the screen:

```
corina 3.00 <serial number - compilation date - user - date - time - host>

Input file type MDL SDFile
Output file type MDL SDFile

Options:

Standard values, version 3.0, March 18, 2001
Ring templates, version 3.0, March 18, 2001
Torsion angle library, version December 1999 (1088 patterns)
                       all rights CCDC, Cambridge, UK

*** RECORD no.:  1 read ***********************
   Ident 1
   Name Decaline
   10 atoms
   Elapsed time: 5 ms

*** RECORD no.:  2 read ***********************
   Ident 2
   Name Norbornane
   7 atoms
   Elapsed time: 10 ms

*** RECORD no.:  3 read ***********************
   Ident 3
   Name Trimacrocyclus
   44 atoms
   Elapsed time: 110 ms

3 record(s) read, 3 converted
   Totally elapsed time: 0 sec
```

5 Program Use

5.1 Synopsis

The general synopsis for using CORINA is:

```
corina [ -option(s) [suboption[=value],...] ] [ infile ] [ outfile ]
```

Infile and outfile are the input and output file names. If no file names are given, the program reads from standard input and writes to standard output. If only one file name is given, this file will be read as input file and the output will be written to standard output. By default, a minimum of trace output is by written to the file *corina.trc*.

5.2 Options

The command line options follow the rules of the UNIX command syntax standard.

-i Input file

t=<value>	Set the input file type to <value>. Allowed values (file formats) are:	
	sdf	MDL SDFile [8] (default),
	rdf	MDL RDFile [8],
	smiles	SMILES linear code [1],
	mol	SYBYL MOL file [2],
	mol2	SYBYL MOL2 file [2],
	ctx	Gasteiger ClearText file format [11],
	mmod	MacroModel structure file format [14], and
	mae	Maestro file format [15].
csdmol2	Allow the CSD specific extensions in SYBYL MOL/MOL2 input files.	
dummies	Allow the interpretation of dummy atom types in SYBYL MOL/MOL2 ("Du") and SMILES linear code ([*])input files.	

sdfi2n=<value>	Copy the data item <value> to the compound name field (1st line) in MDL SDF input files ("sdf item to name").
sdfict	Ignore *cis*/*trans* configuration of double bonds in MDL SDF input files.
expandapo	Expand attachment points into 3D. The input option **-i expandapo** forces CORINA to expand attachment points defined in MDL SDFiles ("`M  APO`" field in the properties block) into 3D space. The attachment points are added as "artificial" atoms to the connection table (both to the atom and bond list) and 3D coordinates are calculated. Dummy atom types are assigned to the "artificial" atoms, i.e. "Du" in SYBYL MOL/MOL2 files, "*" (first attachment point) and "**" (second attachment point), respectively, and "X" in PDB files. In addition, the atom names of the attachment point atoms are set to "R1" (first attachment point) and "R2" (second attachment point), respectively, in the output file for formats which support atom names (e.g., SYBYL MOL2).
xelement	Allow extra elements in SYBYL MOL2 input files. If dummy atom types ("Du") or element symbols which are unknown SYBYL atom types (e.g., "Ni" for a nickel atom) are defined in SYBYL MOL2 input files this option forces CORINA to derive–if possible–SYBYL atom types either from the atom names or from the element symbol, or to interpret element symbols in order to internally set appropriate atom types for the 3D structure generation process. By default, CORINA then outputs dummy atom types ("Du") for these atoms (see below: output option **-o xelement**).

-o Output file

t=<value>	Set the output file type. Allowed values <value> are:

sdf	MDL SDFile [8] (default),
rdf	MDL RDFile [8],
mol	SYBYL MOL file [2],
mol2	SYBYL MOL2 file [2],
pdb	Brookhaven Protein Data Bank format [3],
ctx	Gasteiger ClearText file format [11],
dic	CCP4 dictionary file [9],
top	X-PLOR topology file [10],
par	X-PLOR parameter file [10],
mmod	MacroModel structure file format [14],
mae	Maestro file format [15],
cif	Crystallographic Information File format [16],
odb	O Database file format [17], and
dyana	DYANA file format [19].

a	Append the output to the input file instead of creating a new output file.
pdbatom	Write the keyword `ATOM` instead of `HETATM` in PDB output files.
pdbnoconect	Skip all `CONECT` statements in PDB output files.
pdbludi	Create a PDB input file for a Ludi [6] fragment database.
pdbludilabel	Generate unique three-character labels for a Ludi [6] fragment database in PDB output files.

nodummies	Suppress writing of unknown (dummy) atom types in SYBYL MOL/MOL2 output files. If an unknown SYBYL atom type or a dummy ("Du") atom type is encountered the record is discarded from the output file.
noccat	Suppress the automatic conversion of the carbon atom in amidinium-like structures ([NH$_2^+$]=CN) to the carbocation type SYBYL atom type C.cat (N[C$^+$]N) in SYBYL MOL/MOL2 output files. The conversion to this atom type, which is default, is still strongly recommended.
m2l	Copy the given isotopic mass labels in the input file into the corresponding atom name field in the SYBYL MOL/MOL2 output file ("mass to label"). Atoms without given mass label remain untouched. The atom name has the format <symbol><mass>. Non-element symbols are replaced by "R". Thus, an atom [8*] would get the atom name "R8". This can be used for the preparation of combinatorial libraries for FlexX.
mdldb	Add the additional data fields `<MODEL.SOURCE>` and `<MODEL.CCRATIO>` to MDL SDF output file. If the output file type is set to MDL SDF to each record two additional fields are added in the output file. The data field `<MODEL.SOURCE>` gives information about the program version of CORINA, which was used to generate the 3D model. The data field `<MODEL.CCRATIO>` contains the smallest close contact ratio of the CORINA generated 3D molecular model.
mdlcompact	Write out a compact MDL SDFile. This options restricts the number of output fields in the atom lines of the atom block in MDL SDFiles (RDFiles) to the x-, y-, and z-coordinates, the atom type (symbol), the mass difference, the atom charge, and the stereochemical atom parity (columns 1 through 7 of the atom block). All other fields in the atom lines are omitted, since they contain no data which is mandatory for 3D structure information. The goal is to save disk space (up to 40%) in case of large files containing hundreds of thousand compounds.

mdl3dparity	Write out the atom stereo parity which have been calculated by CORINA for missing stereodescriptors to MDL SDF (RDF) output file. If stereo information is missing in the input file CORINA assumes reosonable stereodescriptors following some implemented rules (see section 0 on page 42). This option forces CORINA to output the stereo parity flags in MDL SDF and RDF files which were calculatetd and used by CORINA during the 3D structure generation process.
gold	Force the assignment of atom and bond types in SYBYL MOL/MOL2 output files according to the GOLD conventions for difficult functional groups. For functional groups which can be expressed by more than one canonical form (e.g., guanidinium groups), the ligand docking program GOLD requires a uniform and defined coding of the atom and bond types in SYBYL MOL/MOL2 files. This option forces CORINA to assign the atom and bond types for those groups following the GOLD conventions which are defined in the GOLD Version 2.0 documentation.
fcharges	Write formal atom charges into the partial charge column of SYBYL MOL/MOL2 output files. This option forces CORINA to write formal atom charges which are given in the input structure to the charge column (column 9) in the corresponding `@<TRIPOS>ATOM` data lines of the SYBYL MOL2 output file. In addition, the keyword `USER_CHARGES` is set in the field `@<TRIPOS>MOLECULE`.
keepnames	Keep any atom name given in the input file. Usually, if the atom name and the atom type differs, CORINA tries to derive a reosonable atom name from the atom type for PDB, MacroModel, and Maestro output files. This option forces CORINA to keep and to output atom names which are defined in the input file.
xelement	Allow extra elements in SYBYL MOL2 output files. If the input option **-i xelements** (see above) is set the automatically derived SYBYL atom types or interpreted element symbols ("artificial" SYBYL atom types) are written to the SYBL MOL2 output file.
resnam=<value>	Set the residue name to <value> in PDB, CCP4 topology, and X-PLOR parameter output files.

typchr=<value>	Set the atom type character(s) to <value> in CCP4 topology and X-PLOR parameter output files. The atom type names in top and par files are defined in the format <symbol><type character><index>. With this option the field <type character> can be assigned to <value>.
dicid=<value>	Set the group ID number to <value> in CCP4 dictionary output files.

-t Trace

s	Write trace output to standard error channel (default: *corina.trc*).
n	Suppress trace output. This option is useful for the conversion of large databases since the trace file *corina.trc* might become rather large.
tracefile=<value>	Set trace file name to <value> (default: *corina.trc*).

-n Record number

n=<value>	Process only record number <value>.
f=<value>	Process all records from record number <value>.
t=<value>	Process all records to record number <value>.

-d CORINA driver options

wh	Write hydrogen atoms. Write the added hydrogen atoms to the output file. Internally, CORINA adds missing or implicitly given hydrogen atoms before the generation of 3D coordinates in order to obtain structures with higher quality and better resolved close contacts. By default, the added hydrogen atoms are removed from the output file. This option forces CORINA to write out the added hydrogen atoms.
rs	Remove small fragments. Remove all but the largest fragments from multi-component records (e.g., counter-ions in salts, solvent molecules).

neu	Neutralize formal charges at $[C,S,P]-[O^-]$ and $[NH^+]$. This option can be used to achieve the same protonation state for acids, alcoholates and basic nitrogens by adding or removing protons. This option can be used together with the driver option **rs** (see above) in order to remove counter-ions from salts.
flexx	Tailor CORINA to the docking program FlexX [7]. CORINA can be used for generating ring conformations during the flexible docking process. This options sets the input and output file types and the conformations analysis options to suited values (see section 7.8 on page 39).
stergen	Generate stereoisomeric compounds. This option forces CORINA to automatically detect chiral centers and double bonds where *cis* and *trans* configuration may appear in an input structure and to generate all possible, but unique stereoisomeric compounds, regardless whether correct stereodescriptors are given in the input or not (see section 8 on page 40). By default, a maximum number of four stereocenters are processed and a maximum number of 16 stereoisomeric compounds are generated (see below).
msc=<value>	Set the maximum number of processed stereocenters per molecule to <value> (driver option **stergen** required). By default, the stereoisomer generating module of CORINA processes a maximum of four stereocenters of an input structure. This option forces CORINA to process the specified number <value> of stereocenters in order to restrict or to increase the number of output isomers.
msi=<value>	Set the maximum number of generated stereoisomers per molecule to <value> (driver option **stergen** required). By default, the stereoisomer generating module of CORINA generates a maximum of 16 possible but unique stereoisomers. This option forces CORINA to generate the specified number <value> of stereoisomers in order to restrict or to increase the number of output isomers.
preserve	Preserve defined stereocenters (driver option **stergen** required). If the configuration of an input structure is not specified for all chiral centers and *cis*/*trans* double bonds, this option forces the stereoisomer generating module of CORINA to process only those centers which do not have a defined stereochemistry or configuration.

rc	Generate multiple ring conformations. This options forces the conformational analysis module for small and medium sized ring systems of CORINA to output multiple ring conformations (see section 9 on page 47). The option does not support records which consist of more than one fragment (e.g., salts). Work-around: Combine with suboption **rs** (see above). The conformations are written in the order of increasing steric energy.
mc=<value>	Set the maximum number of generated conformations to <value> (driver option **rc** required). If the driver option **rc** is set (see above), CORINA generates by default a maximum number of 100 conformations per molecule. This option restricts the number of output conformations to <value>.
de=<value>	Set an energy window ΔE of <value> kJ/Mol for the ring conformations (driver option **rc** required). This option forces CORINA to output only those conformations which have an energy not higher than <value> (in kJ/mol) with respect to the lowest-energy conformation.
timeout=<value>	Restrict the computation time for the ring conformation analysis to <value> milliseconds (driver option **rc** required). For complicated fused and bridged ring systems the conformational analysis routine might be quite time consuming if multiple ring geometries should be generated. This option forces to stop the conformational analysis after the set timeout of <value> milliseconds and to output all conformations obtained so far.
flapn	Flap ring nitrogen atoms to generate multiple ring conformations (driver option **rc** required). This option allows pyramidal ring nitrogen atoms which have one exocyclic neighbor to invert their configuration in order to obtain all conformations (see section 9 on page 47).
sc	Generate ring conformations simultaneousely (driver option **rc** required). By default, when generating multiple ring conformations for compounds having more than one ring system CORINA generates all combinations of all conformations of these ring systems. This option reduces the number of conformations by simultaneously generating conformations for different ring systems (see section 9 on page 47).

names	Number the generated conformations or isomers consecutively by adding a counter to their names. If isomers or conformations are generated (driver options **stergen** and **rc**, see above) this option may be used to consecutivly number the different geometries by adding a counter to the compound name.
r2d	Remove 2D records from the output. If the input and the output file type are both set to MDL SDFile (default), CORINA by default writes the original 2D structure to the output file in cases where no 3D structure is or could be generated. This option is useful for database purposes in order to obtain consistent input and output files. This suboption prevents the output of 2D structures.
wb	Write bad models. 3D models having close contacts or other deficiencies are by default removed from the output file. This option enables the output of such models.
no3d	Skip the 3D coordinates generation. This option allows to use CORINA as a file format converter for the supported file formats without generating 3D coordinates. All appropriate options are valid—including the driver options **wh** and **rs**.
ori	Orient the 3D structure according to the moments of inertia. This option is useful when the structure is directly forwarded to a graphical viewer. The molecule then appears more often in an orientation that shows as much of it as possible on one sight.
ow	Override wedge symbols. Some input file formats (SDF, CTX) support both atom stereodescriptors and wedged bonds in the 2D drawings of the molecules. When these descriptors differ for one and the same stereocenter, CORINA by default overrides the atom descriptor and uses the wedged bond symbols for calculating the stereochemistry. This option allows to override the wedged bonds and to use the atomic descriptors instead. In any case, CORINA writes an error message when encountering different types of stereodescriptors for one and the same stereocenter (see section 10.3 on page 57).

amide	Use the configuration specified in the 2D input drawing for amide bonds, rather than the lowest energy configuration (which is the default). This option allows the specification of either *cis* or *trans* configurations for amide bonds by drawing them in the 2D diagram in the input file, accordingly.
3dst	Force stereodescriptors from the 3D structure. If there is a discrepancy between the stereodescriptors and the 3D structure in the input file, CORINA takes the configuration given in the 3D coordinates (default: usage of stereodescriptors).
newtypes	Generate new atom types and ignore all given atom types and also ignore aromaticity given in the input file. This option forces CORINA to newly generate atom types and aromaticity information instead of using the information given in the input file (atom and bond types). This can be useful, if the input file contains questionable atom or bond types.

-h CORINA on-line help options

i	Print help concerning CORINA input file options.
o	Print help concerning CORINA output file options.
t	Print help concerning CORINA trace file options.
n	Print help concerning CORINA record number options.
d	Print help concerning CORINA driver options.
all	Print help concerning all available CORINA options.

-v Print program version

-m Create a UNIX on-line reference manual page (man page)

This option can be used to generate a UNIX on-line reference manual page for the manual pager utility "man", e.g., by typing

```
corina -m > corina.1
```

6 CORINA: Automatic Generation of High-Quality 3D Molecular Models

6.1 Introduction

The three-dimensional structure of a molecule is closely related to a large variety of chemical, physical, and biological properties. The need for computer-generated 3D molecular structures has clearly been recognized in drug design. Searching in 3D databases is widely used for finding new lead compounds. Since the number of experimentally determined molecular geometries is limited—about 230,000 X-ray structures are presently contained in the Cambridge Structural Database [5] compared to 21 millions of known compounds—one needs a method for predicting 3D coordinates directly from the structural formula of a molecule. As a consequence, in recent years a number of programs for automatic 2D-to-3D conversion has been reported (for reviews see reference [22]). We have developed a program called CORINA (COoRdINAtes) [23]-[27] that automatically generates three-dimensional atomic coordinates from the constitution of a molecule as expressed by a connection table. The program scope and some special features for handling large rings and metal complexes make it extremely useful for any molecular modeling method requiring 3D atomic coordinates.

6.2 Program Scope

CORINA is applicable to the entire range of organic chemistry. This has been shown by converting large databases such as the National Cancer Institute file [28] (appr. 250,000 structures) or a database with appr. 7,000,000 structures with conversion rates of over 99%. All structures which can be expressed in a valence bond notation can be processed. Stereochemical information is also considered. Atoms with up to six neighbors are allowed. There are no upper limits to the molecule size or to the ring sizes. CORINA generates by default one low-energy conformation for each input structure. For rings consisting of less than nine atoms, multiple conformations can be generated—a useful feature for flexible 3D database search (see section 9 on page 47). The input structures can be processed in a variety of file formats (section 7 on page 34).

6.3 The Core System

CORINA can be viewed as an automatic 3D model building kit. Combining monocentric fragments with standard bond lengths and angles and using appropriate dihedral angles a 3D model of a molecule is built. Bond lengths and angles possess only one rigid minimum and can be taken from a table. Since multiple solutions exist for torsional angles, two major problems arise: First, in ring systems only sets of torsional angles are allowed which ensure proper ring closure. Secondly, non-bonded interactions between flexible chain portions have to be minimized. Thus, CORINA handles rings and chains separately.

Rings of up to a size of nine atoms are processed using a table of single ring conformations which implicitly ensure ring closure. In the case of fused or bridged systems, a backtracking search procedure finds a contradiction-free set of conformations for each single ring following some geometric and energetic restrictions. Since this strategy works on the torsion angle representations of the ring conformations and uses only logical operations and integer arithmetic it is extremely fast. The ring conformations are then translated into 3D coordinates and further refined using a simplified pseudo force field which contains only special geometrical terms for the optimization of ring systems.

For acyclic fragments and molecules, the principle of longest pathways has been implemented in CORINA (see Figure 2). The main chains are extended as much as possible by setting the torsion angles to *anti* or *trans* configurations, unless a *cis* double bond is specified. This method effectively minimizes non-bonding interactions.

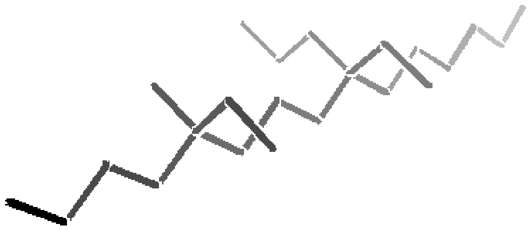


Figure 2 The principle of longest pathways for acyclic fragments and molecules.

After combination of the three-dimensional fragments of the ring systems and of the acyclic parts, the complete 3D model is checked for overlap of atoms and for close contacts. If such situations are detected, CORINA performs a reduced conformational analysis in order to avoid these interactions. Firstly, a strategic rotatable bond within the pathway connecting the two interacting atoms is determined, depending on topological features and double bond character. Secondly, the torsional angle of this bond is changed until the non-bonded interactions are eliminated (see Figure 3). For appropriate torsion angles, CORINA uses a set of rules and data obtained from a statistical analysis of the conformational preferences of open-chain portions in small molecule crystal structures. This knowledge was derived from the Cambridge Structural Database (CSD) and is stored in the Torsion Angle Library [4],[5],[12].

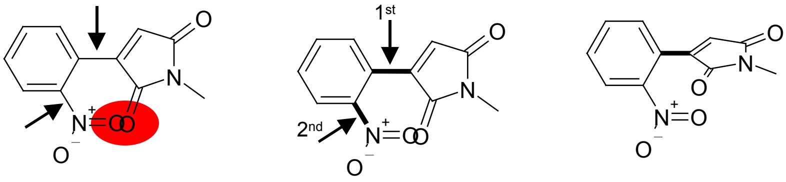


Figure 3 Reduced conformational analysis to avoid non-bonded interactions.

6.4 The Challenge: Large Rings

Large rings represent a special challenge and most of the other published 3D structure generators fail to process such systems. The conventional approach of taking small ring conformations from a table will not work for large, flexible rings. The ring table used by CORINA contains conformations only for rings with up to eight atoms. Thus, for large rings another method is necessary. However, polymacrocyclic structures often show a general outline, a superstructure [26]. The porphyrine bridged cyclophane molecule in Figure 4 shows a cage-like superstructure that retains the approximate shape and symmetry of the entire system.

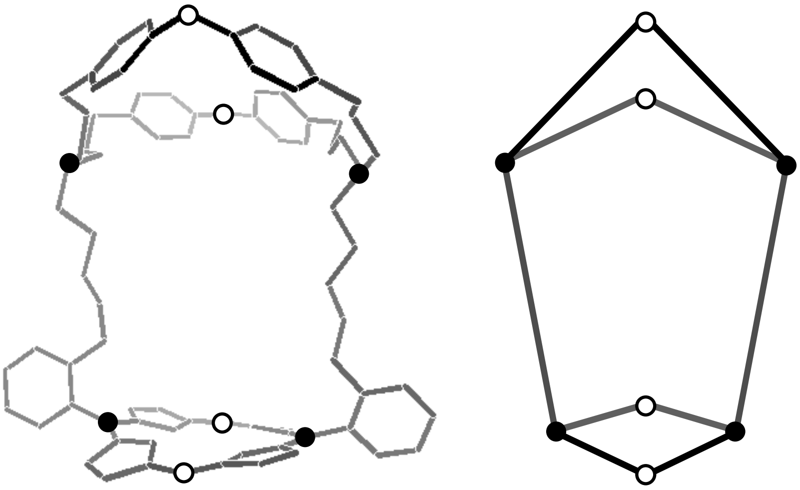
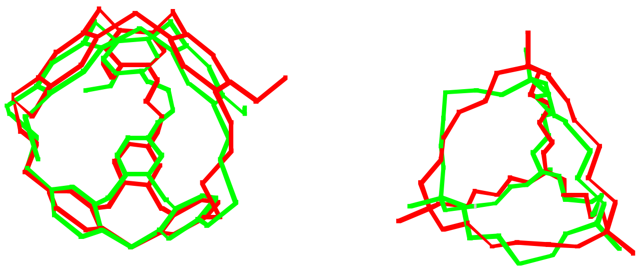


Figure 4 A macrocyclic molecule and the corresponding superstructure. The anchor atoms in both structures are marked by circles.

The procedure for generating a 3D structure for polymacrocycles follows the so-called "principle of superstructure". First, the ring system is reduced to its superstructure as shown in Figure 4. Then, a 3D model for the superstructure which contains only small rings can be generated by using the methods for small rings. Finally, the removed atoms are restored and a complete 3D model of the entire ring system is obtained. Figure 5 compares the X-ray structures of three polymacrocycles with the corresponding CORINA models and the RMS_{XYZ} deviations between them. Although rather large RMS_{XYZ} deviations of 0.14 to 0.95 Å are measured, it can be seen that CORINA succeeded to predict correctly the overall shape and symmetry.

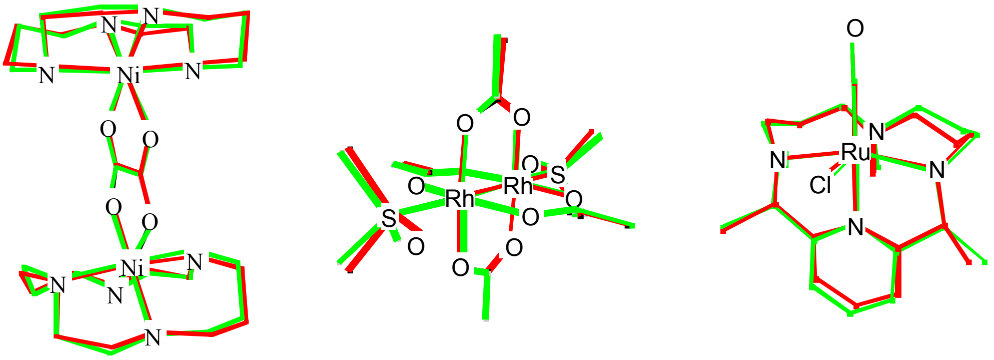


DIJJUB: RMS_{XYZ} = 0.69 Å CISZUZ: RMS_{XYZ} = 0.95 Å FEFZIZ: RMS_{XYZ} = 0.14 Å

Figure 5 Comparison of the X-ray structures of three polymacrocyclic systems with the corresponding CORINA models and their RMS_{XYZ} deviation. (The experimental structure of CISZUZ contains an I_3^- anion inside the ring system.)

6.5 Another Challenge: Metal Complexes

Another type of structures commonly avoided by conventional structure generators are organometallic compounds. CORINA can process compounds containing atoms with up to six neighbors. Thus, metal complexes with up to octahedral centers can be handled. The extensions made are quite simple: First, the input structures must fulfill the restrictions of the valence bond concept. Secondly, appropriate monocentric geometries are predicted for the metal centers. Third, the lengths of metal-ligand bonds are corrected by specific factors taking into consideration their non-covalent character. The resulting structures correspond quite good to the experimentally determined geometries. Figure 6 shows three examples: a nickel, a ruthenium, and a rhodium complex and the RMS_{XYZ} deviations from the X-ray structures.



VADTOJ: RMS_{XYZ} = 0.15 Å PVARHC: RMS_{XYZ} = 0.24 Å DIJKAI: RMS_{XYZ} = 0.20 Å

Figure 6 Comparison of the X-ray structures of three metal complexes with the corresponding CORINA models and their RMS deviation.

6.6 Evaluation of 3D Structure Generators Using 639 X-Ray Structures

A paper comparing six automatic 3D structure generators (CONCORD, ALCOGEN, Chem-X, MOLGEO, COBRA, and CORINA) using 639 X-ray structures and demonstrating the power of CORINA has been published [29]. We will summarize the most important results and present some additional investigations in order

- to include the Biosym CONVERTER [30] (version 950 alpha) into the comparison, and
- to demonstrate the major improvements in the CORINA algorithm from version 1.5 to 1.6.

We are not aware of substantially new versions of the other programs that were involved in this test.

Evaluation Procedure

A dataset of 639 X-ray structures was taken from the Cambridge Crystallographic Database. For all programs a set of quality criteria was determined: the conversion rate, the number of program crashes, the number of stereo errors, the average computation time per molecule, the percentage of reproduced X-ray geometries, the percentage of reproduced ring geometries, the percentage of reproduced chain geometries, and the percentage of structures without crowded atoms.

An X-ray geometry is considered to be reproduced reasonably well if the RMS_{XYZ} deviation of the atomic positions is less than 0.3 Å. For acyclic geometries an RMS_{TA} deviation of the torsion angles at rotatable bonds of less than 15° is taken to consider the model compared to the X-ray geometry as well reproduced. A 3D model is regarded to be free of non-bonded interactions if the close contact ratio—the ratio of the smallest non-bonded distance to the smallest acceptable value for this distance—is greater than 0.8.

Results and Discussion

Table A shows the values for the quality criteria determined for the different model builders. For CORINA, both the results for version 1.5 and 1.6 are shown.

Conversion Rate. CORINA and CONVERTER come up with the largest conversion rate (98-100%). This indicates that these programs have the most wide-spread scope.

Robustness. CONCORD and ALCOGEN encountered one and two program crashes, respectively, a rather high rate considering the rather limited size of the dataset.

Correctness of Stereochemistry. All programs except Chem-X (23 failures) retained the stereochemistry of almost all stereocenters.

Computation Time. CONCORD requires extremely short computation times (0.14 s/mol), whereas MOLGEO, CONVERTER, and COBRA need substantially larger times (3.49-8.98 s/mol). All other programs need times of less than 1 s/mol. The computation times refer to the number of structures converted by the different programs.

Reproduction of X-Ray Geometries. CORINA reproduced the largest portion of X-ray structures (46%). Considering structural details such as ring systems as rigid, this rate becomes 87-90% for all programs except MOLGEO (69%). This is a hint that MOLGEO produces random conformations whereas the other programs try to find low-energy conformations. The highest rate of reproduced chain geometries were generated by ALCOGEN, CONVERTER, and CORINA (53-58%). Please note that the criterion for reproduced chain geometries has been redefined. In the original paper [29] all torsion angles at a rotatable bond have been taken into account. Since this over-estimates some types of bonds, we now count only one torsion angle per rotatable bond. Thus, the percentages in the RMS_{TA} row of Table 1 have slightly changed.

Close Contacts. The CONVERTER structures are completely free of close contacts. CONCORD, ALCOGEN, and CORINA generated between 91% and 97% structures without close contacts. The Chem-X builder produced only 71% of such overlap-free structures—an indication that the program does not perform any check for atom crowding.

Table A Summary of results. The percentages refer to the total number of structures converted by each of the different programs and not to the total number of 639 structures in the original dataset.

	Concord	**Alcogen**	**Chem-X**	**Molgeo**	**Cobra**	**Corina 1.5**	**Corina 1.6**	**Converter**
conversion rate [%]	84	79	74	79	75	100	100	98
program crashes	1	2	0	0	0	0	0	0
stereo errors	0	1	23	1	0	0	0	0
RMS_{XYZ} < 0.3 Å [%]	38	40	33	19	38	42	46	37
RMS_{XYZ}^{rings} < 0.3 Å [%]	89	88	89	69	89	89	90	87
RMS_{TA}^{chains} < 15° [%]	49	55	45	41	49	55	58	53
CCR > 0.8 [%]	91	94	71	86	87	93	97	100

Quantity-Quality Characteristics. The impression by the numbers in Table A is somewhat biased by the different conversion rates. As stated above, the percentages refer to the number of structures converted by the individual programs and not to the total number of 639 X-ray structures in the study. Thus, there is a sensitive relation between conversion rate and quality. Figure 7 characterizes the relationship between quantity (conversion rate) and quality (the degree of reproduction of the X-ray structures), i.e., the efficiency of the different programs [31]. For each program the ordered RMS_{XYZ} values of the non-hydrogen atoms are plotted vs. the number of converted structures. Thus, the ends of the curves mark the number of totally converted structures and the ascents of the curves characterize the quality of the structures in terms of similarity to the X-ray structures. These quantity-quality characteristics show again the different suitability of the seven programs for automatic 2D-to-3D conversion.

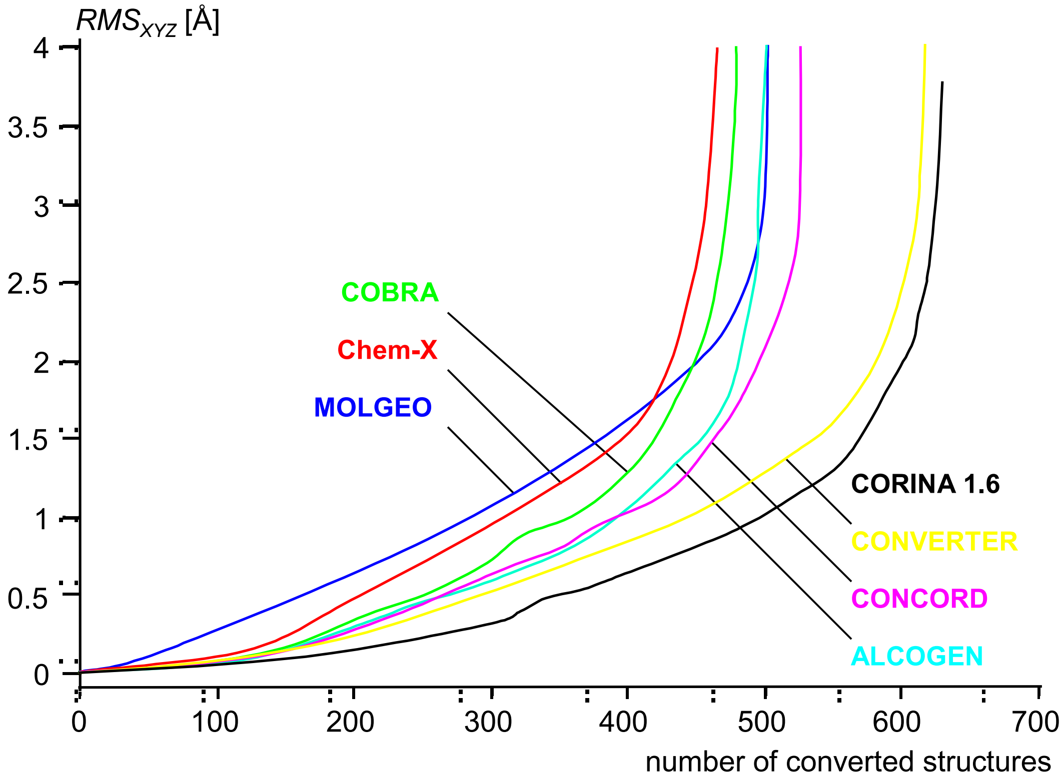


Figure 7 Quantity-quality characteristics of the seven 3D structure generators: Conversion rate vs. RMS_{XYZ} value of the non-hydrogen atoms [31].

Improvements from CORINA Version 1.5 to 1.6. Comparing the results of the two different CORINA versions the effect of three major improvements can be seen:

1) The inclusion of different rules for finding preferable torsion angles for rotatable bonds based on X-ray statistics (see RMS_{XYZ} and RMS_{TA}^{chains}) [4].

2) Improvements in the algorithm which refines atom overlap (see CCR).

3) A substantial speed-up of almost a factor of 2 (see CPU time).

6.7 Comparison of CONCORD and CORINA using 25,017 X-ray Structures

To address both the higher computational throughput of nowadays computers and the larger number of experimental 3D structures available now, the above evaluation study was repeated in year 2001 using 25,017 X-ray structures. This evaluation was applied to the two now mostly used converters, CONCORD and CORINA. The new dataset should provide less bias and a more realistic impression of the performance of the programs under real-world conditions: both are designed to convert millions of structures as fast as possible while maintaining a good quality.

Dataset

The new dataset was obtained from the Cambridge Structural Database using the retrieval program QUEST in batch mode. The query was simply a combination of screens which selected error-free organic compounds which had been fully resolved, for which the connection table had been completely assigned, and which had an R-factor of less than or equal to 5%. The compounds were exported in SYBYL MOL2 format. This initially gave 36,085 compounds. They were then converted into the MDL SDFile format and compounds with obvious errors in the connection tables were removed. This resulted in 35,556 compounds. From these, all purely inorganic compounds not containing any carbon atom, all compounds outside a molecular weight range between 100 and 750, compounds having more than six rotatable bonds, and compounds with rings larger than nine atoms were removed. These criteria should reduce the dataset to reasonably small and moderately flexible compounds resulting in a total of 27,688 compounds. Finally, in cases with multiple species in the unit cell, all fragments but the largest one were removed (i.e., counter-ions, solvents, etc.). In a last filtering step, all duplicate compounds were removed from the dataset. This finally gave 25,017 compounds. After calculating stereo parity values for stereocenters, this dataset was used for the new evaluation study.

Criteria

The same criteria were used as in the smaller evaluation study above with one minor change: The percentage of reproduced ring geometries (RMS < 0.3 Å) was restricted to flexible rings and calculated relative to the number of compounds having flexible rings instead of the number of all compounds. This should provide a more realistic figure since it would exclude, e.g., easy cases like phenyl.

Programs

The program versions used for this study were CONCORD 4.0.4 and CORINA 3.0.

Results and Discussion

Table B summarizes the results. The results are shown for both the complete dataset of 25,017 X-ray structures and for the subset of 22,768 compounds converted by both programs. None of the programs crashed or produced any stereo errors. Again, CORINA had a conversion rate near 100% whereas CONCORD converted only 91%. However, CONCORD was faster than CORINA with an average conversion time of

0.014 seconds per compound compared to 0.049 seconds per compound for CORINA. This relation changes if the smaller subset of 22,768 compounds converted by both programs is considered. Then, the timings for CONCORD and CORINA are 0.013 and 0.033 seconds per molecule. Thus, the subset seems to include less time-consuming cases on average. Looking at the structure related quality criteria, it becomes obvious that by using this dataset the percentages of compounds fulfilling them are a bit lower for both programs compared with the smaller set of 639 compounds discussed above. This might have to do with a higher flexibility of the compounds in the larger set on average. Again, the relative differences of the percentages are much in favour of CORINA. Both programs seem to perform a robust and reasonably good 3D conversion. Whereas CONCORD performs 2.5-3.5 times faster, CORINA converts a significantly higher rate of structures with a better reproduction of the experimental geometries on average.

Table B Comparison of CONCORD and CORINA using 25,017 X-ray structures.

	25,017 compounds		22,768 compounds[a]	
	CONCORD	**CORINA**	**CONCORD**	**CORINA**
conversion rate [%]	91.2	99.7	100	100
program crashes	0	0	0	0
stereo errors	0	0	0	0
CPU time [s/mol][b]	0.014	0.049	0.013	0.033
RMS_{XYZ} < 0.3 Å [%][c]	20	28	20	28
RMS_{XYZ}^{rings} < 0.3 Å [%][d]	71	78	71	78
RMS_{TA}^{chains} < 15° [%][e]	32	43	32	42
CCR > 0.8 [%][f]	95	98	95	98

---

[a] Subset converted by both programs.

[b] On a SG R12000 workstation.

[c] Percentage of structures with an RMS deviation of the non-hydrogen atoms of less than 0.3 Å.

[d] Percentage of structures with an RMS deviation of the ring atoms of less than 0.3 Å (flexible rings only).

[e] Percentage of structures with an RMS deviation of the torsion angles in acyclic portions of less than 15°.

[f] Percentage of structures with a close contact ratio of greater than 0.8.

7 File Formats and Interfaces

This section shows the implemented file formats and gives additional information on their use in CORINA. Table C gives an overview of the input/output formats and a reference to the file format description.

Table C Supported input and output file formats.

Format	Input	Output	Reference
MDL SDF (RDF)	x	x	[8]
SMILES	x		[1]
SYBYL MOLFILE	x	x	[2]
SYBYL MOL2	x	x	[2]
PDB		x	[3]
CTX	x	x	[11]
CCP4 dictionary file		x	[9]
X-PLOR topology/parameter file		x	[10]
MacroModel structure file	x	x	[14]
Maestro file	x	x	[15]
CIF		x	[16]
ODB		x	[17]
DYANA		x	[18],[19]

7.1 MDL Structure Data File (SDF) and Reaction Data File (RDF)

The MDL SDF and RDF file formats were implemented following the description in the literature [8]. CORINA reads only the name (line 1), the status line (line 2), the atom and bond counts (line 4), the atom block, and the bond block. In addition, the RAD and CHG atom properties from the properties block are read in. From the atom block, columns 1 through 7 are read in. From the bond block, columns 1 through 4 are read in. All other information remains unread but is piped unchanged to the output when the output file type is also set to MDL SDF/RDF. This is a great advantage for database purposes since all information except the 3D coordinates remains unchanged. However, CORINA behaves quite differently if the connection table changed during the 3D structure generation process because of, e.g., adding implicitely given hydrogen atoms, removing small fragments (counter ions), or neutralizing formal charges. In these case the program generates the counts line and the atom and bond blocks newly from the changed molecule information and discards all additional, not read-in columns of these blocks. The individual records are assumed to be closed by a $$$$ delimiter.

7.1.1 Options to manipulate MDL SDF and RDF files

The input option **-i sdfi2n=value** (SDF item to name) allows to copy a one-line data item <value> to the compound name. For example, **-i sdfi2n=CASRN** forces CORINA to copy the data line immediately following the data item header <CASRN> into the compound name. This can be used to export one special data line into any other output file format which supports a compound name.

By default, CORINA considers atom stereodescriptors and wedge symbols for chiral centers (see also driver option **-d ow** in section 5.2 on page 14), as well as bond descriptors indicating *cis* or *trans* double bonds which are given in the input structure for generating a 3D structure. Unfortunately, quite oftently stereodescriptors are even not specified or specified not correctly. Two options specifically designed for MDL SDF (RDF) input files influence the handling of stereochemistry during the 3D structure generation process. If the input option **-i sdfict** (SDF ignore *cis*/*trans*) is set CORINA ignores all bond descriptors which define *cis*- or *trans*-configurated double bonds in order to convert also those structures with unreasonably defined descriptors, e.g., if a *trans* double bond is specified in a small ring system, or with ambiguous definitions, e.g., contradictory definitions in conjugated systems. In addition, the output option **-o mdl3dparity** forces CORINA to write out all stereodescriptors which were calculated by CORINA for centers with undefined stereo information.

The output options **-o mdldb** and **-o mdlcompact** are usefull for the conversion of large datasets or databases. If **-o mdldb** is set the additional data fields `<MODEL.SOURCE>` giving information about the program version of CORINA, which was used to generate the 3D models, and `<MODEL.CCRATIO>` containing the smallest close contact ratio encountered in generated 3D molecular model are added to each record in the output file. The option **-o mdlcompact** forces CORINA to output only the fields containing the x-, y-, and z-coordinates, the atom type (symbol), the mass difference, the atom charge, and the atom stereo parity (columns 1 through 7 of the atom block) in the atom lines of the atom block. The columns 8 through 16 (in most cases assigned with values of 0) mainly contain information related to chemical reactions and, thus, are not mandatory for 3D structure generation and representation. This option may save disk space of up to 40%. Please always ensure that the information given in the omitted columns is really not needed for any other purposes before using this option.

The input option **-i expandapo** only has an impact if attachment points `M  APO` are defined in input structures. If this option is set all attachment points are expanded into 3D space. The attachment points are added as "artificial" atoms to the connection table (both to the atom and bond list) and 3D coordinates are calculated. Dummy atom types are assigned to the "artificial" atoms, i.e. "Du" in SYBYL MOL/MOL2 files, "*" (first attachment point) and "**" (second attachment point), respectively, and "X" in PDB files. In addition, the names of the attachment point atoms are set to "R1" (first attachment point) and "R2" (second attachment point), respectively, in file formats which support atom names (e.g., SYBYL MOL2).

7.2 SMILES Linear Notation

The SMILES linear notation was implemented following the literature [1]. In addition, non-standard formal charge qualifiers immediately following the atomic symbol in curly brackets have been implemented. As lower case aromatic atoms only C, N, O, and S are allowed. Implicit hydrogen atoms given inside square brackets are expanded and written to the output file whereas all other missing hydrogen atoms are only written on user requirement (driver option **-d wh**). **Please note:** Each line has to begin with a SMILES linear code and only one SMILES string per line is allowed. Any additional information in the same line which is separated by a white space (or tab) from the SMILES string is interpreted as the compound name and copied into the corresponding data field in the output file. If no compound name is given, the first 80 characters of the SMILES string are copied into the compound name field of the output file.

7.2.1 Options to manipulate SMILES

The input option **-i dummies** allows the interpretation of unknown or dummy atom types in SMILES linear code. **Please note:** The correct definition of a dummy atom according to the SMILES language syntax is "[*]".

7.3 SYBYL File Formats

Both SYBYL MOL and MOL2 file formats were implemented following the SYBYL program manual [2]. Since both formats are based on rather special atom types, their applicability as a standard database format is limited and many cases can occur where no meaningful atom type can be assigned. Dummy atom types are assigned to atoms with lacking atom types. MOL2 files are written by using the keywords `@<TRIPOS>MOLECULE`, `@<TRIPOS>ATOM`, and `@<TRIPOS>BOND`.

7.3.1 Options to manipulate SYBYL File Formats

SYBYL file formats are restricted to a limited number of different atom types, so-called SYBYL atom types, according to the parametrized atom types in the SYBYL force field package. Therefore, SYBYL file formats are rather restricted for general structure representation purposes. However, many program systems and software packages support SYBYL MOL/MOL2 file formats with various extensions in order to overcome the lacking atom type definitions. By default, CORINA only accepts and interprets atom types in SYBYL files which are properly defined as SYBYL atom types (a straightforward philosophy since the SYBYL interface was implemented accurately following the SYBYL program manual). Furthermore, several input and output options allow CORINA to handle also atom types which cannot be regarded as generic SYBL atom types. The input option **-i csdmol2** and **-i xelement**, as well as **-i dummies** force

CORINA to interpret CSD specific extensions (e.g., transition metal atom types such as Ni, Zn, or Cu), to internally use atom types which are estimated when encountering element symbols or ambiguous defined SYBYL atom types, or to allow dummy atom types ("Du") in SYBYL input files.

Furthermore, the output option **-o nodummies** suppresses the output of dummy atom types ("Du") in SYBYL files and records which contain dummy atom types or unknown SYBYL atom types are discarded. This option is useful if post-processing software requires or can handle only atom types which are "true" SYBYL atom types. In contrast to this, the output option **-o xelement** allows to write out atom types which are not listed as SYBYL atom types ("artificial" SYBYL atom types, e.g., a SYBYL atom type "Zn" for a zinc atom) or to output atom types which were derived from their element symbol and their chemical environment given in the input file.

The output option **-o nocat** suppresses the automatic conversion of the carbon atom in amidinium-like structures and substructures ($[H_2N^+]=$**C**N: N.2$^+$=**C.2**-N.pl3) to the SYBYL atom type "C.cat" (N[**C**$^+$]N: N.pl3-**C.cat**-N.pl3). Please note that the conversion to this atom type which is done by default is highly recommended. This option should only be used if the amidinium-like group is actually required with a charged nitrogen atom, e.g, by any post-processing software.

The output option **-o fcharges** creates a charge column (column 9) in a SYBYL MOL/MOL2 output file containing the atom charges (e.g., formal charges) given in the input file. In addition, the charge type contained under the `@<TRIPOS>MOLECULE` is set to `USER_CHARGES`.

If the CORINA generated models are used as starting geometries of ligands for docking experiments with the flexible docking program GOLD, the output option **-o gold** forces the automatic assignment of atom and bond types according to the GOLD conventions for difficult groups (see [21]). The SYBYL MOL2 output file of CORINA can then be directly used as input file for GOLD and ensures a proper and correct atom type assignment in GOLD.

7.4 Brookhaven Protein Data Bank Format (PDB)

The PDB format was implemented following the literature [3]. The following keywords are used: `HEADER`, `COMPND`, `REMARK`, `HETAM`, `CONECT`, and `END`. The compound name is written to the `COMPND` statement. The atomic symbols and the 3D coordinates are written to `HETATM` statements. The bond graph (connectivity information) is reflected by `CONECT` statements.

7.4.1 Options to manipulate PDB files

The output option -o **pdbatom** replaces all `HETATOM` statements which are set by default for the 3D coordinates of non-standard residues (groups) in biological macromolecules in PDB output files by the `ATOM` statements. This is useful if the structures will be post-processed by program systems which need to read in the ATOM statement or cannot handle `HETATOM` statements.

The output option **-o pdbnoconect** forces CORINA to skip the `CONECT` statements in PDB output files. **Please note:** The `CONECT` statements are mandatory for non-standard residues (`HETATOM`), but can be neglected for standard groups (`ATOM`).

Furthermore, the two output options **-o pdbludi** and **-o pdbludilabel** have been especially designed to generate PDB output files which can be used as input for fragment databases in the *de novo* design program system LUDI, i.e., the `HEADER`, `COMPND`, `REMARK`, `CONECT`, and `END` statements are skipped and the `HETATOM` statement is replaced by `ATOM`, and, if **-o pdbludilabel** is set, unique fragment labels consisting of a three letter code are generated for each input structure.

7.5 The MacroModel Structure File Format (uncompressed)

The MacroModel structure file format was implemented following the literature [14]. All 58 different atom type which are defined in MacroModel are supported. In addition, three different bond types (single, double, and triple bonds) which can be expressed in a valence bond notation (VB method) are supported. The first line of the file or entry contains the number of atoms in the entry and the name of the compound. The atom entries start at line 2, whereas each atom in the entry is described by one single line. The generated Cartesian coordinates of each atom are added by CORINA in the columns 55 through 87 if the output file type is set to the MacroModel file format.

7.6 The Maestro File Format

The Maestro file format was implemented following the literature [15]. The following blocks and keywords are supported: `s_m_m2io_version`, `f_m_ct`, `s_m_title`, and `m_atom` (containing: `i_m_mmod_type`, `r_m_x_coord`, `r_m_y_coord`, `r_m_z_coord`, `i_m_residue_number`, `s_m_insertion_code`, `s_m_mmod_res`, `s_m_chain_name`, `i_m_color`, `r_m_charge1`, `r_m_charge2`, `s_m_pdb_residue_name`, `s_m_pdb_atom_name`, `s_m_grow_name`, `i_m_atomic_number`, `i_m_formal_charge`, `s_m_atom_name`), as well as the `m_bond` block (containing `i_m_from`, `i_m_to`, `i_m_order`). Similar to the MacroModel file format, each atom is described by one single line. The generated Cartesian coordinates of each atom are added by CORINA in the columns 13 through 45 if the output file type is set to the Maestro file format.

7.7 Gasteiger ClearText File Format (CTX)

CTX is a keyword oriented ASCII format developed in the research group of Prof. Dr. Johann Gasteiger [11]. The following keywords are read in and interpreted: `IDENT`, `NAME`, `MOLECULS`, `ATOMS`, `BONDS`, `BLABEL`, `2DCOORD`, `STEREO`, `HIGEOM`, `INTCOORD`, and `END`. These keywords and all additional information are directly piped to the output if the output file type is also set to CTX. An additional keyword `3DCOORD` is written containing the generated 3D coordinates.

7.8 Interface between CORINA and FlexX

The flexible ligand docking program FlexX [7] can use CORINA for the generation of low-energy conformations of ring systems with up to nine atoms per ring. During the docking process FlexX is able to send the cyclic parts of the ligand to CORINA module which then generates an ensemble of ring conformations. The exchange file format is SYBYL MOL2. In order to restrict CORINA to the ring systems of a molecule and to provide as much additional information as necessary the molecule is fragmented by FlexX according to the following rules:

1) Every ring system forms a new fragment. Two ring systems are in the same ring system if they have at least one atom in common.

2) Exocyclic substituents of a ring system and their first neighbors are included in order to provide the information necessary for the correct discrimination between equatorial and axial substituents.

3) All SYBYL atom and bond types of the fragment are retained as in the source molecule.

The option **-d flexx** sets all necessary program parameters to the required values. It is identical to the sequence **-i t=mol2 -o t=mol2 -d rc,mc=25,de=30,timeout=30000,nh**.

8 Stereochemical Information

8.1 2D Coding of Stereochemical Information

The different file formats for chemical structure information support a number of possibilities for coding the stereochemistry of chemical structures. With the ubiquitous availability of interactive graphical structure editors the 2D coding of stereochemistry became the most widely used and most convenient method. By using up and down bond symbols (wedges) the local configuration at atom centers is defined as shown for the bridgehead atoms of *cis*- and *trans*-decalin (see Figure 8).

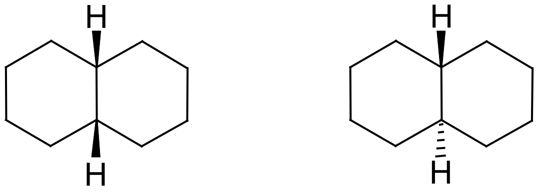


Figure 8 Coding of the stereo information of *cis*- and *trans*-decalin by up and down bond symbols.

Recipe. A number of common errors in specifying up and down bond descriptors often occur. In order to avoid problems the following procedure is recommended. The atom center in question should be drawn in a quasi-tetrahedral configuration with all four ligand atoms (including hydrogen atoms where appropriate). First, draw three of the four ligand atoms with angles of approximately 120° between the bonds. Then, place the fourth ligand between two of the other ligands and assign the up/down bond symbol to this fourth bond. Figure 9 shows two recommended 2D drawings of a chiral center. Other correct variations are shown in Figure 10.

Figure 9 Recommended input of stereochemistry.

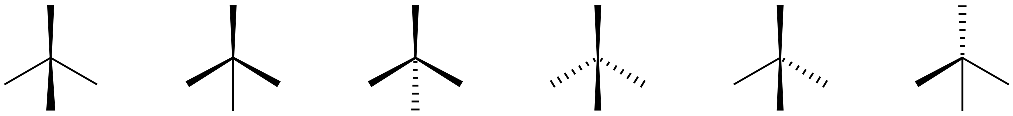


Figure 10 Examples of other correct 2D drawings of a chiral center.

Pitfalls. A number of ambiguous specifications of stereocenters are possible which are often not easily recognized. Figure 11 shows some examples. They all have in common that the ligands of the stereocenter do not span up a suited volume when translating the up and down bond symbols into three dimensions. Thus, the result is ambiguous and no stereodescriptor can be calculated. An error message (see section 10.3 on page 57) is written and the calculation is continued with an arbitrary descriptor.

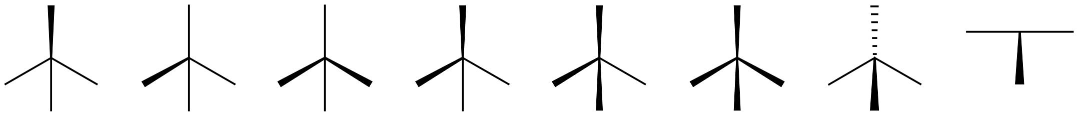


Figure 11 Examples of incorrect 2D drawings of a chiral center.

Automatic Correction. Often, these errors occur after the automatic addition of hydrogen atoms by the structure editor program. Thus, the specification of all ligand atoms of a stereocenter by hand is strongly recommended. However, in some cases the problem of incorrect coding of stereochemistry can be solved by ignoring one hydrogen atom for the calculation of the stereodescriptor (see Figure 12). This is automatically tried and a warning is written (see section 11.1 on page 59) since there is no guaranty that the correction was the intended one.

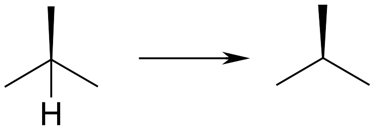


Figure 12 Correction by ignoring one hydrogen atom.

Another possibility for correcting this type of error is to assume an additional up or down bond descriptor for the bond describing the smallest angles with the other bonds at the center in question (see Figure 13). This is automatically tried and a warning is written (see section 11.1 on page 59) since there is no guaranty that the correction was the intended one.

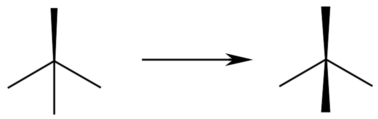


Figure 13 Correction by assuming an additional bond descriptor.

The last type of ambiguous coding of stereochemistry is corrected by moving the central atom. Figure 14 gives an example. The three neighbors at the central atom of the fragment on the left hand side do not span a volume due to the linear position of two of the atoms. This can be corrected by moving the central atom into a direction opposite to the third atom (assuming the hypothetical fourth neighbor in the opposite position). This is automatically tried and a warning is written (see section 11.1 on page 59) since there is no guaranty that the correction was the intended one.

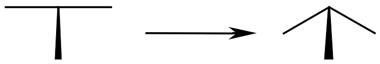


Figure 14 Correction by moving the central atom.

8.2 Addition of Missing Stereodescriptors

The specification of the configuration of chiral centers is essential in generating 3D structures. Thus, the complete definition of stereochemistry is strongly recommended. CORINA does not seek the energetically most favorable configuration. However, in most cases a reasonable 3D structure can be generated with arbitrary chosen stereodescriptors. An exception are ring systems containing chiral atoms. For these systems other than arbitrary values must be found for the stereodescriptors of unspecified stereocenters. Geometric and energetic constraints reduce the number of possible stereoisomers and require a careful choice of suited stereodescriptors. Some examples shall illustrate this idea.

Bridged Systems. Bridged ring systems as, e.g., norbornane (see Figure 15) require a *o,o*-configuration of the bridgehead atoms. The *i,o*-isomer is geometrically forbidden since the bridge cannot be closed as illustrated in Figure 15. CORINA defines the stereochemistry at unspecified bridgehead atoms according to this rule. When the input file contains stereodescriptors violating this rule, the processing of the molecule is abandoned with an error message (see section 10.3 on page 57).

Figure 15 Bridged systems: *o,o*- and *i,o*-norbornane.

Fused Systems. Fused ring systems as, e.g., decalin (see Figure 16) can occur as different stereoisomers that differ in energy. In the case of decalin, the *trans*-isomer has a lower energy than the *cis*-isomer. When the stereo information of the input structure is incomplete, an energy criterion is used for setting a default configuration. For the decision whether two fused rings shall prefer *cis*- or *trans*-configuration a set of rules is used depending on the sizes of the two rings.

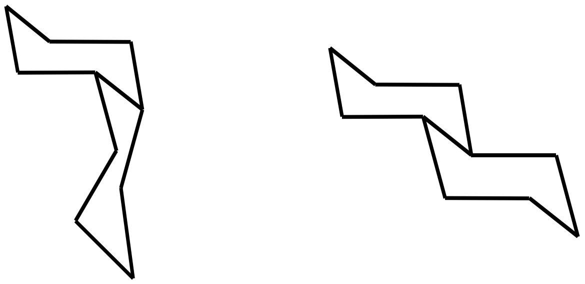


Figure 16 *Cis*- and *trans*-decalin.

Exocyclic Substituents. Exocyclic substituents of ring systems can occupy different spatial positions depending on the stereochemistry and on the conformation. The transition from the equatorial to the axial form of methyl-cyclohexane (see Figure 17) depends on the conformation whereas the difference between the diequatorial and the equatorial/axial forms of 1,4-dimethylcyclohexane (see Figure 18) is a problem of stereochemistry. Thus, only in the second case there is a connection between stereochemistry and steric energy. In case of unspecified stereocenters CORINA tries to maximize the number of equatorial substituent positions.

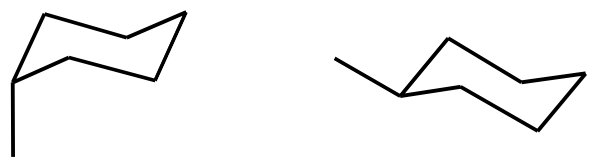


Figure 17 Equatorial and axial forms of methylcyclohexane.

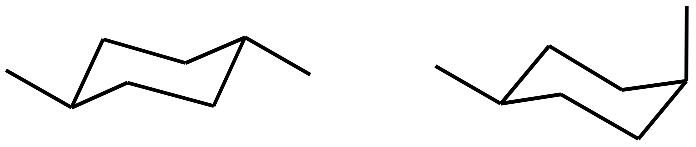


Figure 18 Diequatorial and equatorial/axial forms of 1,4-dimethylcyclohexane.

8.3 Generation of Stereoisomers

A substantial step towards the understanding of the physical, chemical, or biological properties of a molecule is to study and to analyze its spatial shape. Besides the constitution, a major shape determining feature is the configuration of a molecule, i.e. the stereochemistry.

Furthermore, molecular chirality plays a major role in many areas of chemistry. Enantiomers often exhibit quite different physical, chemical, and biological properties. The exploration of the configurational space of a molecule and the analysis of the various isomers a molecule can adopt is therefore of great importance. CORINA now provides access to the configurational space of molecules.

As mentioned above, CORINA generates by default one single stereoisomer by taking into consideration the stereo information given in the input connection table and by making resonable assumptions for missing stereo information. The driver option **-d stergen** forces CORINA to automatically identify stereocenters and to generate all possible, but unique and chemically reasonable, isomeric compounds starting from a given connection table or linear string representation of a molecule before it is converted into 3D space. Tetrahedral chiral centers as well as *cis*/*trans* isomerism is taken into account. Duplicate configurations (e.g., *meso*-compounds) and conformationally strained configurations (e.g., the *i*,*o*-isomer of norbornane, see Figure 15) are rejected. Wedge symbols and/or stereodescriptors are generated and written to the output file. Figure 19 shows som examples.

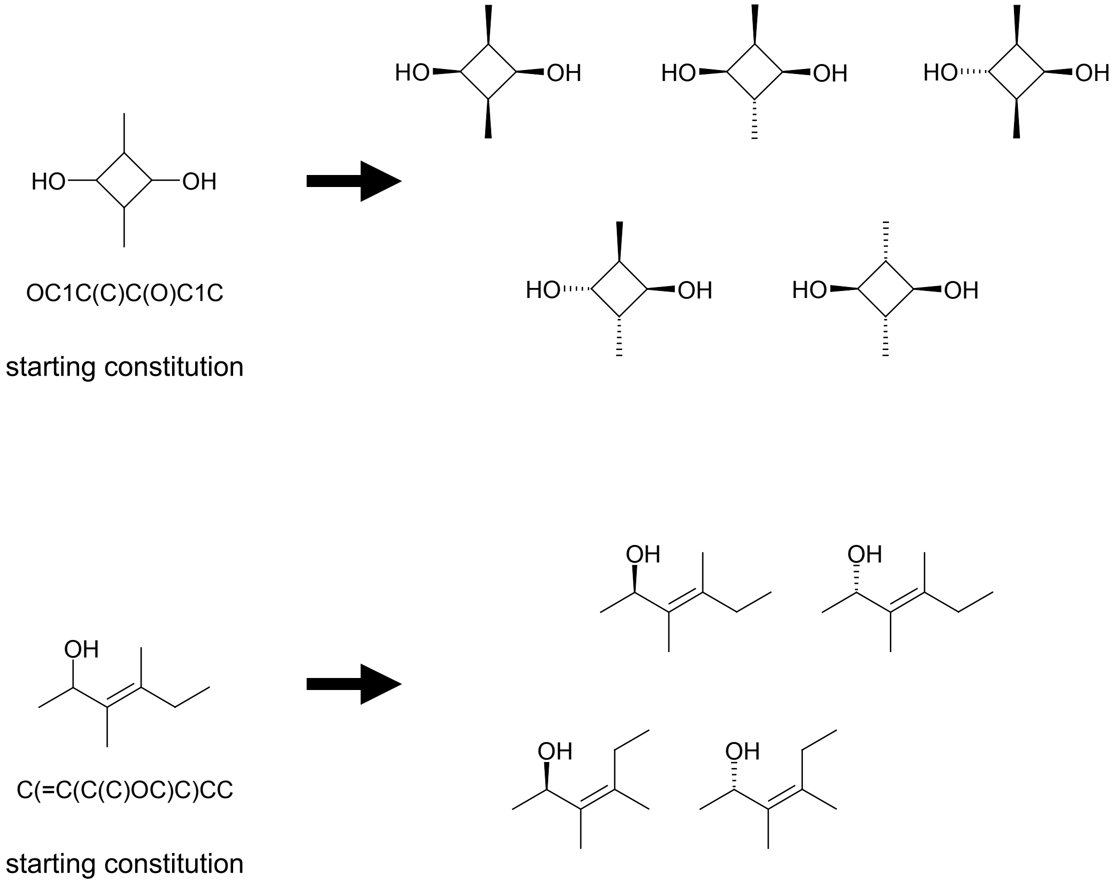


Figure 19 Generated configurations from one single initial constitution.

By default (if the driver option **-d stergen** is set), a maximum number of four stereocenters are processed and a maximum number of 16 stereoisomers per molecule are generated. However, as the number of possible isomers exponentially increases with the number of stereocenters (2^n, where n is the number of stereocenters in a molecule), additional options allow to restrict the maximum number of output isomers (**-d msi=<value>**, maximum number of stereoisomers) or to define a maximum number of stereocenters which should be processed (**-d msc=<value>**, maximum number of stereocenters). Furthermore, the option **-d preserve** allows to retain the configuration at atoms which have a defined stereochemistry (i.e., a stereodescriptor is given in the input file).

Please note: CORINA does not identify the lowest-energy configuration of a molecule or suggests any ranking of the generated isomers.

Command line example. The following command line

```
corina -d stergen,msi=20,msc=6,preserve,wh,rs,r2d <in> <out>
```

generates a maximum of twelve stereoisomers per molecule (**msi=12**) by processing a maximum of six stereocenters (**msc=6**) and preserves stereocenters which have a defined stereochemistry (i.e., a stereodescriptor is given in the input file, **preserve**). The stereoisomers are converted to 3D, implicitely given hydrogen atoms in the input structures are added and written to the output file (**wh**), small fragments (e.g., counterions in salts) are removed from the output file (**rs**) as well as structures which couldn't be converted by CORINA (**r2d**).

9 Conformational Analysis of Ring Systems for Flexible Search Purposes

The method of choice for flexible 3D database search is the use of 3D databases containing single low-energy conformations for each molecule and to solve the flexibility problem on the fly instead of storing multiple conformations in the database—a rather time and disk space consuming approach. A number of methods such as the directed tweak method exist for flexible searching [32]. These methods are rather efficient for chain portions of the molecules but run into problems when performed on ring systems (*vide infra*). A solution can be provided by the storing of 3D models with multiple ring conformations and applying the flexible search only to the chain portions. CORINA supports this technique by providing the possibility of generating multiple ring conformations.

9.1 Generation of Multiple Ring Conformations

For ring systems consisting of rings having up to eight atoms CORINA performs internally a systematic conformational analysis in order to find the ring conformation with lowest energy. Thus, a partial conformational analysis can easily be performed by outputting all conformations found in this way. The conformations of side chains remain unchanged unless there are problems with non-bonded interactions caused by the different ring conformations. In any case, only one conformation for each side chain is generated. Thus, the output consists of a series of 3D models having different ring conformations and almost the same chain conformations. The philosophy behind this approach is that a conformational analysis for ring systems requires much more program intelligence and that an analysis of the side chains can be performed by a much simpler postprocessor (e.g., by permuting systematically all rotatable bonds) using the structures with different ring conformations as input. In addition, the number of reasonable conformations of ring systems is often orders of magnitude smaller than the number of chain conformations.

CORINA is forced by the option **-d rc** to write multiple ring conformations to the output file. The command **corina -n n=1 -d rc example.sdf out.sdf** produces multiple conformations for *trans*-decalin—the first record of the example file example.sdf. Figure 20 shows the five conformations obtained. The conformations are written in the order of increasing steric energy. The maximal number of conformations per molecule can be set by the **-d mc=<value>** option where <value> is the desired number of conformations.

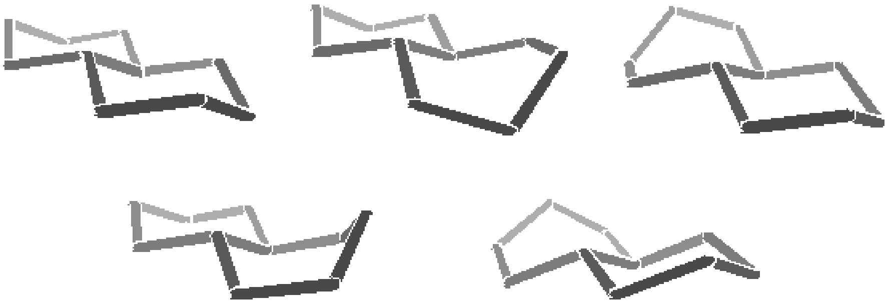


Figure 20 Conformations of *trans*-decalin.

9.2 Handling of Pyramidal Ring Nitrogen Atoms

For pyramidal ring nitrogen atoms having one exocyclic substituent, CORINA can produce, in conjunction with the multiple conformer generation option, conformations having both possible configurations at the nitrogen atoms (**option -d rc,flapn**). CORINA would generate for 1,4-dimethyl-piperidine (SMILES: CN1CCC©CC1) (when no stereochemistry is specified), four chair conformations with all combinations of the two substituents in equatorial and axial positions (see Figure 21).

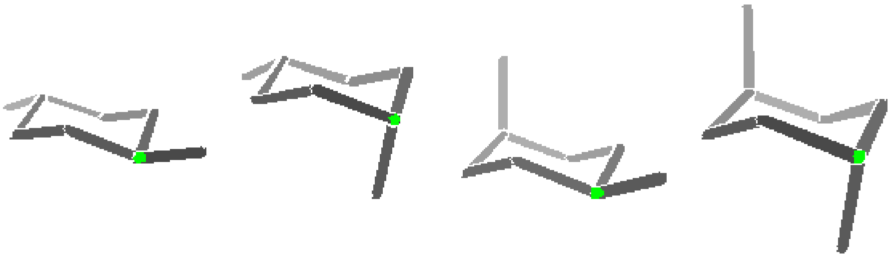


Figure 21 Chair conformations of 1,4-dimethyl-piperidin.

9.3 Handling of Molecules Having More Than One Ring System

For molecules having more than one ring system connected by chains, CORINA offers two techniques:

1) The exhaustive method (default). All combinations of all conformations of the individual ring systems are generated. A possible combinatorial explosion is inherent to this method.

2) The compact method (option **-d sc**). All ring systems change simultaneously their conformations from the low- to the high-energy levels. All conformations of a particular ring system will be seen, but not all combinations of them. A possible loss of bio-active conformations is the price for a significant smaller number of conformations generated.

These methods are illustrated by using 1-cyclohexyl-2-cyclohex-3-enyl-ethane (SMILES: C1CCCCC1CCC1CC=CCC1) as an example (see Figure 22).

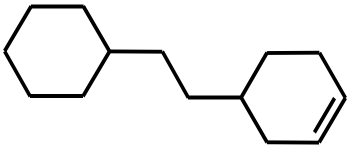


Figure 22 1-Cyclohexyl-2-cyclohex-3-enyl-ethane.

For the cyclohexane and cyclohexene rings each having one exocyclic substituent, CORINA generates 14 and 4 conformations, respectively. Thus, the exhaustive method (1) results in 14 x 4 = 56 conformations. The compact method (2) yields MAX(14, 4) = 14 conformations.

9.4 Multiple Ring Conformations in 3D Database Searches

An exhaustive study of the effect of using multiple ring conformations for 3D database searches was performed [33]. In this section, a short summary of the results will be given.

Directed tweak and ring flexibility. One approach to 3D database searching is to address torsional flexibility by a fitting technique called directed tweak [32]. Directed tweak optimizes the torsion angles at rotatable bonds with respect to a penalty function describing the distance of the actual conformation of a molecule to a given pharmacophor. In principle, this technique is also applicable to ring systems. However, the application of the tweak technique to rings runs into problems: First, ring systems are orders of magnitude less flexible than chain fragments and have a limited number of significantly different conformations. Thus, a continuous fit technique will not be as good as for chain fragments. Secondly, the handling of ring closure by opening one bond per ring and introducing additional constraints into the penalty function makes the hypersurface of the optimization function much more complicated and is a potential source for numerical problems. Convergence problems and questionable hit geometries are the result. In general, the geometries of the hit structures are often rather distorted. Frequently found problems are those of violated sp^2 atoms or stereocenters and atom clashes at the ring closure bonds even when searching with explicit van der Waals check.

Hybrid-approach. The following ideas led to a hybrid approach that overcomes the problems sketched above of the directed tweak method. First, directed tweak performs very efficient for chain portions of molecules but runs into problems when applied to rings. Secondly, the explicit storage of multiple conformations for addressing flexibility suffers both from the immense requirements of computer resources and from possible losses of bio-active conformations due to a too coarse search grid. On the other hand, ring systems consisting of small rings (e.g., with up to eight atoms) show only a small number of conformations that represent a coarse grid. Thus, a hybrid approach combining the two techniques can increase the search efficiency. The proposed technique splits the handling of flexibility into two routes, one for handling rings and another one for chain portions of the molecules:

1) For chains, the directed tweak method is used.

2) For rings, multiple conformations are stored in a 3D database.

This method was implemented by combining two programs: UNITY [34]. for the flexible search (directed tweak) and for the database management and CORINA for the 2D-to-3D conversion and the generation of multiple ring conformations.

Evaluation method. In order to compare the performance of the hybrid technique with alternative approaches, five reference queries were searched in a public domain database. The public part of the NCI database [28] containing 126,705 molecular structures was used as an evaluation database. Two 3D databases were constructed by using CORINA: One database (NCI) containing only one conformation per molecule and the second one (NCI_FLEX) containing up to 25 ring conformations per molecule. CORINA was forced to add missing hydrogen atoms, to remove small fragments, to generate a maximum of 25 conformers per molecule, to invert pyramidal ring nitrogen atoms, and to use the compact method for molecules having more than one ring system (*vide supra*) by using the command line options **-d wh,rs,rc,mc=25,flapn,sc**. Test queries for dopamine agonists (derived from the dopamine structure), kinase C agonists [35] histamine agonists 1 and 2 agonists [36], and for antiarrythmic agents [37] where taken from the literature. These queries were searched in three different ways:

1) Search the NCI database only with chain flexibility (reference run).

2) Search the NCI database with both flexible rings and chains (original approach).

3) Search the NCI_FLEX database with flexible chains (hybrid-approach).

In order to exclude artifacts and unreasonable geometries as mentioned above, all initially found hits were relaxed by a robust force field [38] and then searched again without ring flexibility in order to figure out the genuine hits with a new ring conformation.

Results and discussion. Table D shows the results of the three search runs. Clearly, the hybrid approach (NCI_FLEX, tweak chains) results in a significant additional portion of hits (6-23%) compared to the reference run without ring flexibility. The application of the directed tweak method to rings (original approach) yielded a significant smaller additional amount of hits. Moreover, in one case (histamine 2) the tweaking of rings decreased the hit number by 2%. This indicates that a number of "stable" hit structures found without ring flexibility in the reference run may be hidden by the flexible ring search due to an "instable" hit structure falling into a local minimum during relaxation which cannot fulfill the query in the subsequent search run without ring flexibility.

Table D Hit numbers of the three search runs.

query	NCI tweak chains (reference run)	NCI tweak chains tweak rings (original method)	NCI_FLEX tweak chains (hybrid approach)
dopamine	117	122 (+4%)	144 (+23%)
kinase C	490	553 (+13%)	601 (+23%)
histamine 1	3736	3999 (+7%)	4247 (+14%)
histamine 2	1932	1885 (-2%)	2050 (+6%)
antiarrythmic	1180	1197 (+1%)	1441 (+22%)

In addition, the dependence of the hit rate on the maximum number of conformations per molecule was investigated (see Figure 23). In most cases, the hit rates converged to a value of 10 conformations and this value is recommended. At that point, the size of the NCI_FLEX database was increased by a factor of 1.7 compared to the one conformer database (NCI), a rather moderate requirement of additional resources.

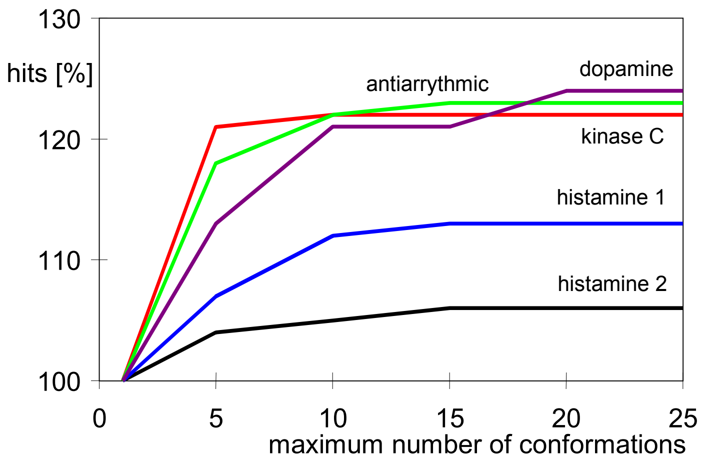


Figure 23 Hit rate vs. maximum number of conformations per molecule.

Recommendations. As a consequence of the study presented above, the following CORINA command line is recommended for the generation of 3D databases for pharmacophor search purposes:

```
corina -d wh,rs,r2d,rc,mc=10,flapn,sc <in> <out>
```

10 Error Messages

10.1 General Errors

ERROR licence(): Licence time out.
The license time has expired. You are not allowed to use the program any more.

ERROR pfopen(): Can’t open file *filename (path)*.
The program is unable to open the specified file (path in parentheses).

ERROR ioopen(): Can’t open trace file.
The trace file *corina.trc* can’t be opened.

ERROR ioopen(): Identical input/output files.
Identical file names for the input and output files are not allowed.

ERROR ioopen(): Can’t open input file *filename*.
Error opening the specified input file.

ERROR ioopen(): Can’t open output file *filename*.
Error opening the specified output file.

ERROR corina(): Too many non-option parameters.
The command line must not contain more than two non-option parameters (the input and output file names). Options start with a "-". After the first non-option parameter no more options are allowed.

ERROR corina(): Option invalid in version 1.6 and later.
An old-fashioned version 1.5 option was encountered (see section 5 on page 14).

ERROR corina(): File type not allowed.
A file type not allowed for input and/or output files was specified encountered (see section 5 on page 14).

ERROR restrict(): FlexX restrictions violated.
The restrictions of the FlexX interface are violated or not fulfilled.

10.2 Input File Format Errors

10.2.1 MDL SDFile

The error messages given below are completed by the line number in the MDL SDFile and the record number where the error occurred.

ERROR rmdl(): Can't read counts line.

ERROR rmdl(): Can't read atom block.

ERROR rmdl(): Can't read bond block.

ERROR rmdl(): Can't read 2nd header line.

ERROR rmdl(): Dimensional code (2D/3D) not specified.

The second header line did not specify whether the given atomic coordinates are 2D or 3D. The program automatically checks whether z-coordinates or up/down bond descriptors are given and thus, which type of coordinates must be assumed.

ERROR rmdl(): Can't read 3rd header line.

ERROR rmdl(): Unknown element symbol.

ERROR rmdl(): Bond atoms out of range.

A bond between atoms out of the range 1..N_{Atom} was encountered.

ERROR rmdl(): Unknown bond type (set 1).

A bond type not equal to 1, 2, 3, or aromatic is specified. The bond is assigned a bond order of 1.

ERROR rmdl(): Valence error reading a new bond.

The maximum valence state of an atom forming the bond was exceeded or a bond of an atom to itself was encountered.

ERROR rmdl(): Can't distribute double bonds over an aromatic system.

The distribution of alternating single and double bonds over an aromatic system failed. This may result from the fact that this distribution requires the introduction of charges. However, the best way to avoid such problems is to specify the correct valence bond notation in the input file.

Errors with missing columns:

ERROR rmdl(): No atomic charge column specified.

ERROR rmdl(): No atom stereo column specified.

ERROR rmdl(): No mass difference column specified.

ERROR rmdl(): Too many additional lines.
Only a limited number of additional lines in MDL SDFile is allowed (5,000).

ERROR rmdl(): Line too long.
Only 80 characters per line are allowed in MDL SDFiles.

Errors in charge, radical, and attachment point lines:

ERROR rmdl(): CHG atom out of range.

ERROR rmdl(): RAD atom out of range.

ERROR rmdl(): APO atom out of range.

10.2.2 SMILES

The error messages below are completed by indicating the position in the SMILES string where the error occurred.

ERROR smilesct(): Reading branch.
Error while reading a branch in () brackets.

ERROR smilesct(): General error while interpreting this character.
A not interpretable character was encountered.

ERROR smilesct(): Unknown element.
An unknown element symbol was found.

ERROR smilesct(): Too many (‘s.
Missing a closing) bracket.

ERROR smilesct(): Closing ring.
No suited ring bond label.

ERROR smilesct(): Valency problem making a new bond.
The standard valency of an atom is exceeded.

ERROR smilesct(): Too many [‘s.
Missing a closing] bracket.

ERROR smilesct(): Reading bracketed atom.
The atom type in square brackets is incorrect.

ERROR smilesct(): Generating Kekule structure for an aromatic system.
A Kekulé structure with alternating single and double bonds cannot be found for an aromatic system.

ERROR smilesct(): Duplicate slash at double bond.
Two bonds at one atom of a stereo double bond are assigned a slash stereodescriptor. Only one slash is allowed.

ERROR smilesct(): Not all rings closed.
Not for all ring bonds was a second label defined.

ERROR smilesct(): Reading atomic charge qualifier.
Incorrect charge specification.

ERROR smilesct(): Too many {‘s.
Missing a closing } bracket.

ERROR smilesct(): Unknown stereo class.
An unknown stereo class was specified.

ERROR smilesct(): No organic element. Use square brackets.
Inorganic elements must be written in square brackets.

ERROR smilesct(): Ring closure label not following immediately the atomic symbol.
Ring bond labels must follow immediately after the atomic symbol.

ERROR smilesct(): Conflicting ring closure bond types.
The type of the ring closure bond was defined twice with differing values.

ERROR smilesct(): Chiral center has wrong connectivity.
A tetrahedral center with less then four neighbors was encountered. A common error is to forget to specify an eventual implicit hydrogen neighbor of the stereocenter within the square brackets.

ERROR smilesct(): Expecting atomic symbol.
The first character inside square brackets must belong to an atomic symbol.

ERROR smilesct(): Illegal valence state.
An atom in an illegal valence state was encountered.

ERROR smilesct(): Inconsistent EZ specification.
The EZ configuration of a double bond was defined redundantly with different values.

10.3 Stereo Errors (See Section 8 on Page 40)

ERROR initster(): Resetting a trans double bond in a small ring: #-#.
A trans double bond in a ring with less than eight atoms was encountered. Since this is geometrically impossible it can be corrected by the program.

ERROR clcster2(): Ambiguous configuration of ligands.
The specification of a chiral center is ambiguous in the 2D structure diagram.

ERROR clcster2(): The ligands at atom # don‘t span up a volume.
The specification of a chiral center is ambiguous in the 2D structure diagram.

ERROR clcster2(): Insufficient number of axes.
For a higher coordinated center (5 or 6 ligand atoms) are not enough axes specified in the 2D structure diagram.

ERROR allcis(): No suited stereodescriptors for bridge #- -#.
Incorrect coding of the stereochemistry of some bridgehead atoms.

10.4 Errors in the Generation of 3D Coordinates

ERROR corina(): Input structure incorrect. No 3D generation.

ERROR gen3d(): Unable to build a 3D structure.
No 3D structure can be generated and no structure is written to the output. A fatal error occurred during the generation process.

ERROR quality(): Some internal coordinates are heavily distorted.
A fatal error occurred during the 3D generation.

ERROR quality(): Violated cis/trans bond.
A stereo double bond changed its configuration during the geometry optimization.

ERROR quality(): Violated stereo atom.
A chiral atom changed its configuration during the geometry optimization.

ERROR quality(): Bad contacts detected.
Unfavourable non-bonded interactions (crowded atoms) detected.

ERROR gen3d(): 3D structure didn‘t pass the quality check.
The generated 3D structure is insufficient and removed from the output file.

ERROR gen3d(): No generation of multiple conformations for multiple fragments.
For records containing multiple fragments the output of multiple ring conformations (-d rc) is not supported. Work-around: remove all but the largest fragments (-d rc,rs).

ERROR alltempl(): Missing ring template.
For one ore more rings no template is found in the file *rings.ctx*.

ERROR ringfrag(): Fragment contains ring(s) > 8.
CORINA handles only rings up to a size of eight members by templates. (Rings with more then eight atoms are reduced to a secondary structure. Only when this reduction does not result in ring sizes < 9 the above message is written.)

ERROR ringfrag(): Can’t combine the templates.
The available ring templates cannot be combined to a ring conformation.

ERROR bigring(): Unable to build the secondary structure.
The secondary structure of a large ring system with ring sizes > 9 is too complex or contains rings > 9.

ERROR procfrag(): Can’t process small ring system.
A small ring system cannot be translated into a 3D structure.

ERROR bigsys(): Unable to process all fragments.
Some of the fragments of a large ring system cannot be translated into a 3D structure.

11 Warning Messages

11.1 Warnings Related to Stereochemistry (See Section 8 on Page 40)

WARNING initster(): Stereo atom # without stereodescriptor.

WARNING clcster2(): Possibly stereo problem at atom #.
A general problem while calculating the configuration of a stereocenter from the 2D structure diagram was encountered.

WARNING clcster2(): Trying to ignore H-atom at stereocenter #.
An ambiguous 2D configuration was encountered. The program tries to ignore the 2D coordinates of one hydrogen atom in order to solve the problem since these hydrogen atoms are often automatically added to the 2D structure diagram without regarding the stereochemistry of the central atom.

WARNING clcster2(): Trying to give a direction to bond #-#.
An ambiguous 2D configuration was encountered. The program tries to assign a wedge descriptor to an additional bond in order to solve the problem.

WARNING clcster2(): Moving the central atom #.
An ambiguous 2D configuration was encountered. The program tries to correct it by moving the central atom.

WARNING clcster2(): Collision of wedge symbol and stereodescriptor at atom #.
The stereochemistry of an atom was defined by an up/down (wedge) bond descriptor in the 2D structure diagram and by an atom parity descriptor with the two specifications giving opposite configurations. The atom descriptor has the higher priority and overrides the bond descriptor.

WARNING clcster2(): Number of axes at atom # not sufficient.
The number of axes at a coordination center higher than 4 (5 or 6 ligand atoms) in the 2D structure diagram is too small.

11.2 Warnings in the Generation of 3D Coordinates

WARNING bondlen(): No bond length #-#.
The system is unable to calculate a bond length from standard atomic parameters.

WARNING initba(): Geometry type for atom # PLANAR --> TETRAEDER changed

The bridgehead atoms in strained ring systems cannot be planar. This warning occurs in the case of unsaturated ring systems containing rings smaller than 5, e.g., cubene.

WARNING getta(): No TA #-#-#-#.

In case of linear systems (e.g., acetylene) torsional angles (TA‘s) for some bonds cannot be specified.

WARNING hmoboord(): No HMO constant for atom #.

No Coulomb integral parameter is found for a hetero atom. CORINA takes the value for carbon.

WARNING hmoboord(): No HMO constant for bond #-#.

No parameter for a bond resonance integral is found. The C-C value is taken.

WARNING genconf(): Time out.

The time limit for a complete ring conformation analysis was exceeded. The conformer with lowest energy so far found is not necessarily the global minimum.

WARNING rrefine(): # pair(s) of crowded atoms

Some pairs of atoms became closer to each other than 75% of their atomic radii.

12 Acknowledgments

CORINA was developed in the research group of Prof. Johann Gasteiger at the Technical University of Munich and at the University of Erlangen-Nürnberg since 1985. The program development was initiated by Dr. Christian Hiller (initial implementation in Fortran programming language) and continued by Dr. Christine Rudolph (treatment of large rings, macrocyclic, and polymacrocyclic structures) and Dr. Jens Sadowski (reduced force field, large rings, flexible macrocyclic structures, metal complexes, conformational analysis and refinement of close contacts, reimplementation of CORINA in C programming language).

We gratefully acknowledge the permission obtained from the Cambridge Crystallographic Data Centre (Cambridge, UK) to implement in CORINA the more detailed and advanced version of the Torsion Angle Library (915 torsion angle patterns), which was derived from the Cambridge Structural Database (CSD). We thank Prof. Gerhard Klebe (University of Marburg, Germany), Dipl.-Math. Thomas Mietzner and Dr. Frank Weber (BASF-AG, Ludwigshafen, Germany) for the close cooperation in implementing this program part.

CORINA is now maintained for general usage by Molecular Networks GmbH Computerchemie (by Dr. Christof H. Schwab).

13 References

[1] a) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31-36. b) Daylight Software Manual. Daylight Chemical Information Systems: Santa Fe, NM, USA, 1993, http://www.daylight.com.

[2] a) SYBYL is developed and distributed by Tripos Associates Inc.: St. Louis, MO, USA (http://www.tripos.com). b) A detailed description of SYBYL MOL/MOL2 file formats is available on the Internet for download as a PDF document at http://www.tripos.com.

[3] a) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535-542. b) Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Research* **2000**, *28*, 235-242.

[4] Klebe, G.; Mietzner, T. A Fast and Efficient Method to Generate Biologically Relevant Conformations. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 583-606.

[5] Cambridge Structural Database: a) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, D. S.; Rodgers, J. R.; Watson, D. G. *Acta Cryst. Sect. B: Struct. Crystallogr. Cryst. Chem.* 1979, B35, 2331. b) Allen, F. H.; Kennard, O. *Acc. Chem. Res.*, 1983, 16, 146. c) Allen, F.H.; Hoy, V.J. Cambridge Structural Database. In *Encyclopedia of Computational Chemistry*, Schleyer, P.v.R.; Allinger, N.L.; Clark, T.; Gasteiger, J.; Kollman, P.A.; Schaefer, III, H.F.; Schreiner, P.R. (Eds.), John Wiley & Sons, Inc., Chichester, UK, 1998; pp. 155-167.

[6] Böhm, H.-J. The Computer Program Ludi: A New Method for the De Novo Design of Enzyme Inhibitors. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61-78.

[7] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470-489.

[8] a) Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 244-255. b) A detailed description of MDL file formats (Mol, SDF, and RDF) is available on the Internet for download as a PDF document at http://www.mdli.com.

[9] Collaborative Computational Project, Number 4, 1994: The CCP4 Program Suite: Programs for Protein Crystallography, *Acta Cryst.* **1994**, *D50*, 760-763.

[10] X-PLOR, A System for X-ray Crystallography and NMR: Brünger, A.T., The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA.

[11] Gasteiger, J. et al. CTX Keyword Reference Manual. University of Erlangen-Nürnberg: 1995, unpublished results.

[12] Klebe, G.; Mietzner, T.; Weber, F. Methodological Developments and Strategies for a Fast Flexible Superposition of Drug-Size Molecules. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 35-49.

[13] STERGEN Version 1.1 is developed and distributed by Molecular Networks GmbH, Erlangen, Germany (http://www.mol-net.de).

[14] a) MacroModel is developed and distributed by Schrödinger, LLC, Portland, OR, USA (http://www.schroedinger.com). b) MacroModel User Manual Version 7.0 http://www.schrodinger.com/Support/pdf.html.

[15] a) Maestro is developed and distributed by Schrödinger, LLC, Portland, OR, USA (http://www.schroedinger.com) b) Maestro User Manual Version 5.0 http://www.schrodinger.com/Support/pdf.html.

[16] a) Hall, S.R.; Allen, F.H.; Brown, I.D. The Crystallographic Information File (CIF): a New Standard Archive File. *Acta Cryst.* **1991**, *A47*, 655-685. b) See also: http://www.iucr.org.

[17] Jones, T.A.; Zou, J.Y.; Cowan, S.W.; Kjeldgaard, M. Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in These Models. *Acta Cryst.* **1991**, *A47*, 110-119. b) See also: http://xray.bmc.uu.se/usf.

[18] Güntert, P.; Mumenthaler, C.; Wüthrich, K. Torsion Angle Dynamics for NMR Structure Calculation with the New Program DYANA. *J. Mol. Biol.* **1997**, *273*, 283-298.

[19] a) DYANA (CYANA) is developed and distributed by P. Güntert Scientific Software, Zürich, Switzerland. b) DYANA Version 1.5 User's Manual: http://www.guentert.com.

[20] a) Jones, G.; Willett, P.; Glen, R.C. Molecular Recognition of Receptor Sites Using a Genetic Algorithm with a Description of Desolvation. *J. Mol. Biol.* **1995**, *254*, 43-53. b) Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727-748.

[21] a) GOLD is distributed by Cambridge Crystallographic Data Centre, Cambridge, UK (http://www.ccdc.cam.uk. b) GOLD 2.0 Documentation (User Manual): http://www.ccdc.cam.uk/support/csd_doc/zdoc_main.html.

[22] a) Sadowski, J.; Gasteiger, J. From Atoms and Bonds to Three-dimensional Atomic Coordinates: Automatic Model Builders. *Chemical Reviews* **1993**, *93*, 2567-2581. b) Sadowski, J., Three-Dimensional Structure Generation: Automation. In *Encyclopedia of Computational Chemistry*, Schleyer, P.v.R.; Allinger, N.L.; Clark, T.; Gasteiger, J.; Kollman, P.A.; Schaefer, III, H.F.; Schreiner, P.R. (Eds.), John Wiley & Sons, Inc., Chichester, UK, 1998; pp.2976-2988. c) Sadowski, J.; Schwab, C.H.; 3D Structure Generation and Conformational Searching. In *Computational Medicinal Chemistry and Drug Discovery*, Bultinck, P.; De Winter, H.; Langenaeker, W.; Tollenaere J.P., Eds., Dekker Inc., New York, in Press (2003).

[23] Hiller, C.; Gasteiger, J. Ein automatisierter Molekülbaukasten. In *Software-Entwicklung in der Chemie*, Vol 1; Gasteiger, J., Ed.; Springer: Berlin, 1987; pp 53-66.

[24] Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic Generation of 3D Atomic Coordinates for Organic Molecules. *Tetrahedron Comp. Method.* **1990**, *3*, 537-547.

[25] Sadowski, J.; Rudolph, C.; Gasteiger, J. The Generation of 3D Models of Host-guest Complexes. *Anal. Chim. Acta* **1992**, *265*, 233-241.

[26] Sadowski, J.; Gasteiger, J. Polygon Patterns for the Generation of Conformations of Large Rings. In *Software Development in Chemistry*, Vol 7; Ziessow, D., Ed.; Gesellschaft Deutscher Chemiker: Frankfurt am Main, 1993; pp 65-76.

[27] Schönberger, H.; Schwab, C.H.; Hirsch, A; J. Gasteiger, J. Molecular Modelling of Fullerene Dendrimers. *J. Mol. Model*. **2000**, *6*, 379-395.

[28] a) Milne, G. W. A.; Nicklaus, M. C.; Driscoll, J. S.; Wang, S.; Zaharevitz, D. W. The NCI Drug Information System 3D Database. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1219-1224. b) See also: http://cactuscit.nih.gov/ncidb/download.html.

[29] Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-Ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000-1008.

[30] Biosym Technologies, Inc., San Diego, USA. We thank Dr. M. Waldman (Biosym) for providing us with the CONVERTER results.

[31] This representation was proposed by Dr. V. van Geerestein (Akzo Organon, Oss, The Netherlands; unpublished results).

[32] Hurst, T. Flexible 3D Searching: The Directed Tweak Technique. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 190-196.

[33] Sadowski, J. A Hybrid-Approach for Addressing Ring Flexibility in 3D Database Searching. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 53-60.

[34] UNITY is developed and distributed by Tripos Ass., St. Louis, MO, USA (http://www.tripos.com).

[35] Wang, S.; Zaharevitz, D. W.; Sharma, R.; Marquez, V. E.; Lewin, N. E.; Du, L.; Blumberg, P. M.; Milne, G. W. A. The Discovery of Novel, Structurally Diverse Protein Kinase C Agonists through Computer 3D Database Pharmacophore Search. Molecular Modelling Studies. *J. Med. Chem.* **1994**, *37*, 4479-4489.

[36] Good, A. C.; Kuntz, I. D. Investigating the Extension of Pairwise Distance Pharmacophore Measures to Triplet-Based Descriptors. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 373-379.

[37] Moock, T. E., Henry, D. R., Ozkabak, A. G., and Alamgir, M., Conformational Searching in ISIS/3D Databases. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 184-189.

[38] Davies, E. K.; Murrall, N. W. How Accurate Does a Force Field Need To Be? *Computers Chem.* **1989**, *13*, 149-156.

14 Report Form

In the case of problems occurring during installation or running CORINA, please complete the following form and send it or fax it to

Molecular Networks GmbH Computerchemie
Nägelsbachstr.25
91052 Erlangen, Germany
FAX: +49-(0)9131-815669

User:

CORINA program and version number ("corina -v"):

Command line to run CORINA:

Error and warning messages by CORINA:

System messages:

Short description:

Please include the input file, output file and trace file (*corina.trc*) generated by CORINA on a 3½" diskette written in MS/DOS format or forward it via email to support@mol-net.de These files will help us to analyze your problems. All data will be treated confidentially.