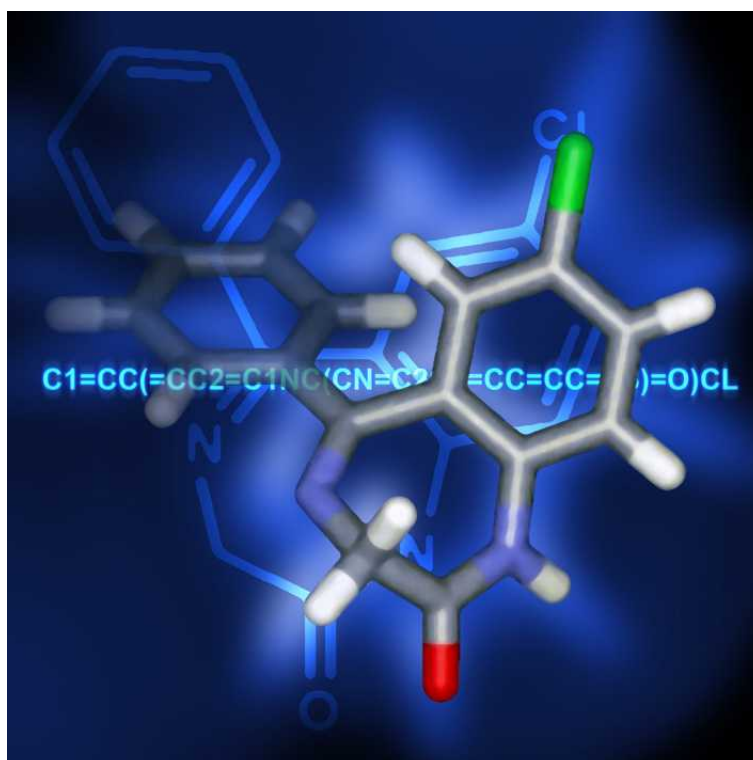


3D Structure Generator CORINA

Generation of High-Quality Three-Dimensional Molecular Models

Version 3.4

Program Description



Molecular Networks GmbH Computerchemie
December 2011

<http://www.molecular-networks.com>



Molecular Networks
Inspiring Chemical Discovery

Molecular Networks GmbH
Computerchemie
Henkestraße 91
91052 Erlangen
Germany

Phone: +49-9131-815668
Fax: +49-9131-815669

Email: info@molecular-networks.com
WWW: www.molecular-networks.com

This document is copyright © 1996-2011 by Molecular Networks GmbH Computerchemie. All rights reserved. Except as permitted under the terms of the Software Licensing Agreement of Molecular Networks GmbH Computerchemie, no part of this publication may be reproduced or distributed in any form or by any means or stored in a database retrieval system without the prior written permission of Molecular Networks GmbH Computerchemie.

The software described in this document is furnished under a license and may be used and copied only in accordance with the terms of such license.

CORINA is a registered trademark in the Federal Republic of Germany. Other product names and company names may be trademarks or registered trademarks of their respective owners, in the Federal Republic of Germany and other countries. All rights reserved.

(Authors: Jens Sadowski, Christof H. Schwab; Document version: 3.49-2011-12-01)

Contents

1	Introducing CORINA	1
1.1	Objective of CORINA	1
1.2	CORINA in Brief	1
2	Release Notes	3
2.1	CORINA (Full Version)	3
2.2	CORINA_F (Restricted FlexX Interface Version)	15
3	Getting Started with CORINA	16
4	Using CORINA	19
4.1	Synopsis	19
4.2	Options	19
5	How to ... Use CORINA	35
5.1	Converting a 2D MDL SDFFile into a 3D MDL SDFFile	35
5.2	Using 3D Input Information to Reproduce the Correct Stereochemistry	35
5.3	Generating Sets of Stereoisomeric Compounds	35
5.4	Using CORINA as a File Converter	35
5.5	Generating Sets of Multiple Ring Conformations	36
6	Supported File Formats and Interfaces	37
6.1	MDL Structure Data File (SDF) and Reaction Data File (RDF)	37
6.2	SMILES Linear Notation	39
6.3	SYBYL File Formats	39
6.4	Brookhaven Protein Data Bank Format (PDB)	41
6.5	The MacroModel Structure File Format (uncompressed)	42
6.6	The Maestro File Format	42
6.7	Gasteiger ClearText File Format (CTX)	42
6.8	Interface between CORINA and FlexX	43

7	Error Messages	44
	7.1 General Errors	44
	7.2 Input File Format Errors	45
	7.3 Stereo Errors	51
	7.4 Errors in the Generation of 3D Coordinates	52
	7.5 Output File Format Errors	54
8	Warning Messages	55
	8.1 General Warning Messages	55
	8.2 Warnings Related to Input File Formats	55
	8.3 Warnings Related to Stereochemistry	56
	8.4 Warnings in the Generation of 3D Coordinates	58
9	Technical Requirements	59
	9.1 System Requirements	59
	9.2 Program Scope and Known Limitations	59
10	Program Installation	60
	10.1 On-line Download from Molecular Networks' Web Server	60
	10.2 Distribution on CDROM	61
	10.3 New Installation	61
	10.4 Program Updates	62
11	Problems and Help!	63
12	Understanding CORINA	64
	12.1 Fast and Efficient Generation of High-Quality 3D Molecular Models	64
	12.2 Evaluation of 3D Structure Generators Using 639 X-Ray Structures	68
	12.3 Comparison of CONCORD and CORINA using 25,017 X-ray Structures	71
	12.4 Stereochemical Information	72
	12.5 Conformational Analysis of Ring Systems for Flexible Search Purposes	79
13	Acknowledgements	86
14	References	87
15	Report Form	91

1 Introducing **CORINA**

1.1 Objective of **CORINA**

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 and many other areas. Since the number of experimentally determined molecular geometries is limited—about 367,000 X-ray structures are presently contained in the Cambridge Structural Database (January 2006) [1] compared to more than 26 millions of known compounds—there is a need for methods to predicting 3D coordinates directly from the constitution of a molecule. As a consequence, in the last three decades a number of programs for automatic 2D-to-3D conversion have been reported (for reviews see reference [2]). Among them is the program **CORINA (COoRDINAtes)** [2]-[8] that automatically generates three-dimensional atomic coordinates from the constitution of a molecule (see Figure 1). The program scope, its reliability and speed as well as some special features for handling large rings and metal complexes make it extremely useful for any study or modeling purpose that requires 3D information of the molecules under investigation.

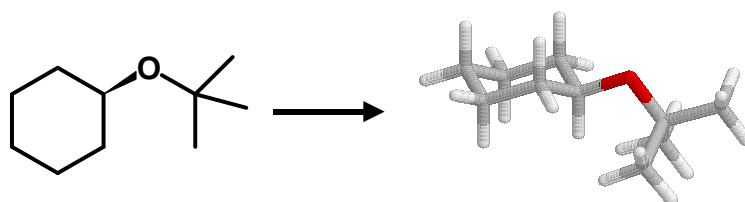


Figure 1 Generating a 3D model from the constitution of a molecule.

1.2 **CORINA** in Brief

CORINA is a rule and data based program system, that automatically generates three-dimensional atomic coordinates from the constitution of a molecule as expressed by a connection table or linear code, and which is powerful and reliable to convert large databases of several hundreds of thousand or even millions of compounds.

- **CORINA** is applicable to the entire range of organic chemistry. Structures which can be expressed in a valence bond notation can be processed.
- **CORINA** does not provide any upper limit to the size of the molecule or to the size of ring systems.
- **CORINA** fully considers stereochemical information and generates the defined stereoisomer (if the stereo information is defined properly in the input file).

- **CORINA** processes structures containing atoms with up to six neighbors. Thus, even metal complexes can be processed.
- **CORINA** generates by default one low energy conformation for each input structure. For ring systems consisting of up to nine atoms, multiple conformations can be generated—a useful feature for building flexible 3D databases.
- **CORINA** automatically detects stereo centers (tetrahedral centers and *cis/trans* double bonds) and is able to generate all possible isomers. Duplicate isomers, such as *meso* compounds are identified and removed as well as geometrically strained configurations.
- **CORINA** can process a variety of standard file formats for structure input and output (*e.g.*, MDL SD/RDFile [9], SMILES [10], SYBYL MOLFILE and MOL2 [11], PDB [12], MacroModel[13], Maestro [14] or CIF [15]).
- **CORINA** delivers structures of high quality. The RMS deviation of CORINA built models from published X-ray structures is among the best of all commercially available conversion programs.
- **CORINA** is fast (less than 0.1 sec for small and medium-sized organic molecules on a common x86 Linux workstation), robust and provides excellent conversion rates (99.5%) for the 250,251 structures of the National Cancer Institute (NCI) Open Database [16] without intervention or program crash.
- **CORINA** is general. A database with more than six million compounds has been converted with a conversion rate of more than 99%.
- **CORINA** offers many features to influence the 3D generation process, *e.g.*, addition of lacking or implicitly given hydrogen atoms, neutralization of formal charges, orientation of the 3D structures according to their moments of inertia, removing of counter ions in salts, ...
- **CORINA** provides an interface to the ligand-docking program FlexX [17]. During the docking process with FlexX, CORINA generates multiple conformations for the ring systems of the ligand in order to optimize the steric and electrostatic interactions between the small molecule ligand and the binding site of the protein.

A description of the program scope and known limitations is given in section 9 on page 59.

2 Release Notes

2.1 CORINA (Full Version)

2.1.1 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 [10].
- 2) The output file formats SYBYL MOL/MOL2 [11] and Brookhaven Protein Databank PDB [12] 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) [1], [18].
- 4) A substantial speed-up of almost a factor of 2 was achieved by optimizing 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 12.2 on page 68. 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.

2.1.2 Version 1.7

CORINA version 1.7 was tailored especially to the database business:

- 1) Two new driver options **-d flapn** and **sc** were added for generating multiple ring conformations.
- 2) Two additional PDB output options **-o pdbludi** and **pdbludilabel** allow the generation of fragments for databases interfacing to the *de novo*-design program Ludi [19].

In addition, an exhaustive study on the effect of multiple ring conformations on the performance of flexible 3D pharmacophore searches was performed (see section 12.5 on page 79).

2.1.3 Version 2.0

CORINA 2.0 is now able to interact with the ligand-docking program FlexX [17] as a conformer generator for ring systems (see section 6.8 on page 43). Thus, CORINA ring conformations can be used for flexible ligand docking into a receptor pocket. Changes were mainly made to the file format interfaces and to the ring conformation options.

- 1) Two new input file formats SYBYL MOL/MOL2 [11] (**-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.

2.1.4 Version 2.1

The following changes and improvements have been implemented.

- 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**, **-d amide**, and **-i sdfict** related to the handling of stereochemical information for MDL SDFiles [9] were added (see Section 4.2 on page 19).
- 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 no longer be generated.

2.1.5 Version 2.3

The following changes and improvements have been implemented.

- 1) A new option **-d no3d** allows using 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*.

2.1.6 Version 2.4

The following changes and improvements have been implemented.

- 1) The data files *stdval.ctx* and *rings.ctx* are now inline—easier installation, less mistakes with different versions.
- 2) A new driver option **-d 3dst** forces the use of a given 3D configuration instead of the atomic stereo descriptors. This might be useful if the stereo descriptors are not specified properly but the 3D structure is correct.
- 3) A new driver option **-d neu** neutralizes formal charges at acids, alcoholates, and basic nitrogen atoms 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) A new driver 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.

2.1.7 Version 2.6

The following improvements and changes have been implemented:

- 1) The file format MDL RDFFile [9] 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 [20] and X-PLOR [21] 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 SDFFile, SMILES linear code and Gasteiger ClearText format [22].
- 4) The SMILES reader and interpreter are now more general: SMILES strings containing hetero-aromatic 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 SDFfile, 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 [1], are written to the output file.
- 8) A 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) A new output option **-o mdlldb** 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) A new output option **-o noccat** switches off the automatic conversion of the carbon atom in amidinium-like structures ($[\text{NH}_2^+]=\text{CN}$) to the carbo-cation type SYBYL atom type C.cat ($\text{N}[\text{C}^+]\text{N}$). 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 avoid 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) [1], [18], [23].

- 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 1,216 sulfonamide nitrogen atoms as found in the Cambridge Structural Database (CSD) [1] has shown, that in the majority of cases (901 of 1216 sulfonamides – 74%) an "out-of-plane" distance of less than 0.3 Å is exhibited. Thus, the planar configuration is the preferred geometry compared to the pyramidal configuration.
- 14) A 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.

2.1.8 Version 3.0

The following improvements, changes, and new features have been implemented:

- 1) The functionality of the stereoisomer generator STERGEN [24] has been integrated in CORINA. The driver option **-d stergen** forces CORINA to determine all stereo centers 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 stereo centers 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 stereo centers that should be processed (**msc=<value>**) and to restrict the total number of generated stereoisomers (**msi=<value>**). Stereo centers which have a defined stereochemistry (stereo descriptor) 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 stereo descriptor is given in the input structure.
- 2) In order to provide interfaces to the molecular modeling package MacroModel [13], 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**) [14] as new input and output file formats.
- 3) In addition, the file format CIF (Crystallographic Information File, **-o cif**) [15] supported by a variety of crystallographic program packages, the file format ODB (O Database file format, **-o odb**) [25] to interface to the crystallographic modeling tool O, and the file format of the NMR structure calculation program DYANA (**-o dyana**) [26],[27] 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 that 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 writing 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) A 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 that 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 stereo descriptors following some implemented rules (see section 12.4 on page 72). 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 [28], [29] 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) By default, if the atom name and the atom type differ, CORINA tries to derive a reasonable 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) A new driver option **-d names** consecutively numbers the different conformations (**-d rc**) or stereoisomers (**-d stergen**) in ensembles that have been generated by CORINA. A counter is added to the compound name of each conformation (*compoundname_c00n*, $n=1,2,3,\dots,N$; where N is the total number of generated conformations) or stereoisomer (*compoundname_i00n*, $n=1,2,3,\dots,N$; 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.

2.1.9 Version 3.1

The following improvements, changes and new features have been implemented:

- 1) CORINA can now add hydrogen atoms and their 3D coordinates to 3D input structures while retaining the input 3D geometry. This can be done by simply combining the driver options **wh** and **no3d** (**-d wh,no3d**). This option can be used to add 3D hydrogen atoms if they are missing in a 3D structure, but the 3D structure of all given atoms should be kept, e.g., for an experimentally determined geometry.
- 2) A new driver option **-d ringatom=<atom label>** can be used to denote a specific ring system for which multiple ring conformations should be generated, whereas **<atom label>** is the atom label of one atom contained in this specific ring system (requires driver option **-d rc**). This is helpful if multiple ring conformations should be generated only for one specific ring system in an input structure that contains more than one flexible ring systems (up to nine ring atoms).
- 3) The stereoisomer generation module of CORINA can now process double bonds in ring systems having more than 10 ring atoms and output *cis* and *trans*-isomers (driver option **-d stergen**). In addition, para-substituted ring systems are identified as pseudo stereo centers in order to generate di-equatorial and equatorial/axial substituted configurational isomers.

- 4) For UNIX and Linux systems, CORINA is now available as a library version (shared object). This enables software developers to easily include the CORINA functionalities into their chemoinformatics applications and to call CORINA for 3D structure generation purposes in own source code. The API for input is either a single line formatted MDL SDFfile or a SMILES linear string. The 3D structures are returned as an MDL SDFfile (again formatted as a single line with "\n" as new line character). In addition, all driver options can be triggered. Up to now, the library version is available for SGI workstations (IRIX 6.5), Sun SPARC stations (Solaris 8), and x86 Linux platforms (kernel 2.4). A version for MS Windows (win32) is currently under development and will be available soon.
- 5) The support of long compound names is secured, even for file formats which only allow a restricted number of characters for the name field (e.g., MDL SDFfile, 80 characters). However, CORINA prints a warning message to the trace file when encountering lines in the input file that are too long.
- 6) The algorithm for orienting 3D structures according to their moments of inertia has been made more reliable.
- 7) Several changes in the read and write interfaces to MDL SDF/RDF, SMILES, SYBYL MOL/MOL2, MacroModel, and Maestro file formats made the input and output more stable.
- 8) Improvements and bug fixing in the core algorithms for the handling of ring systems, generation of stereoisomers and usage of system routines increased the conversion rate and decreased CPU times.

2.1.10 Version 3.2

The following improvements, changes and new features have been implemented:

- 1) A new driver option **-d canon** prevents any atom numbering dependent artifacts in the 3D structure generation process. In some special cases, the 3D structure generation process may be sensitive to the atom numbering in an input structure, *i.e.*, for different atom numberings slightly different conformations may be obtained. This option canonicalizes the connection table of an input structure internally before the 3D model is built and then uses the reordered atoms for the generation of the 3D coordinates. After the 3D model is generated, the connectivity table is renumbered using the original atom numbering scheme.
- 2) Structures that cannot be converted by CORINA are now be written to a second output file (error file), if the new driver option **-d errorfile=<value>** (<value> = file name) is used. Currently, only SDF and SMILES format are supported. Collecting failed structures in a separate file is valuable for the batch processing of many input structures.

- 3) A new input option **-i force3d** forces CORINA to output molecules in SYBYL MOL/MOL2 or PDB file format (these file format by definition require 3D coordinates!), even if no 3D coordinates can be generated or are available in the input file. This option is especially useful if CORINA is used only as a file converter with the driver option **-d no3d**. The following two combinations are possible.

-i force3d	output of records (structures) for which no 3D coordinates can be generated in SYBYL MOL/MOL2 or PDB file format
-i force3d -d no3d	file conversion to SYBYL MOL/MOL2 or PDB file format if no 3D coordinates are present in the input file

The following two driver options address particular issues when CORINA is used in conjunction with FlexX and are only required in a small number of cases.

- 4) A new driver option **-d symoff** ignores the symmetry filtering during the ring conformation generation process. This option is especially helpful if CORINA is used in the interface mode for the docking program FlexX in order to generate multiple ring conformations of the ligands (see also sections 2.2 and 6.8). CORINA will only process the cyclic fragments originating from the larger ligand molecules. As acyclic atoms are removed by FlexX before sending fragments to CORINA, a higher symmetry might be perceived in the fragment compared to the symmetry of the entire molecule. As a consequence, fewer conformations may be generated than required.
- 5) A new driver option **-d ampax** amplifies axial substituent energy contributions during the ring conformation analysis. In some complex cases, the energy differences between axial and equatorial arrangements may become too small with the default parameters, *i.e.*, as for the amide ligand in the benzodiazepine compound c12ccccc1NC(=O)C(NC(=O)C)N=C2.
- 6) A new output file option **-o lname** allows the user to write out compound names to the MDL SDF name field and to exceed the 80 characters limit in the header block. **Note.** Per definition, the name field in the header block of an MDL SDF file may not be larger than 80 characters [9].
- 7) The default bond angles for oxygen and nitrogen atoms in ethers and amines (SYBYL atom types O.3 and N.3) have been changed to new values based on a recent analysis of drug-like molecules from the CSD. For sp^3 hybridized oxygen atoms the bond angles range from 114° to 134° depending on the two adjacent atoms. The bond angles of sp^3 hybridized nitrogen atoms in amines are set to 111° .

- 8) The SMILES interpreter now supports the OpenEye extensions for hybridization states. The following SMILES patterns for atom primitives are allowed:
^3 for sp³ hybridized atoms
^2 for sp² hybridized atoms
^1 or ^ for sp hybridized atoms
For example, the SMILES c1cccc1[NH2^2] will result in a planar amine group, whereas in c1cccc1[NH2^3] the amine group will exhibit a pyramidal geometry in the generated 3D structure.
- 9) The algorithm to detect duplicate stereoisomers that are generated using the driver option **-d stergen** has been revised. The detection is now based on an improved hashcoding technique. Originally, the 32bits algorithm could produce equivalent hashcode for different isomers. The new implementation, based on 64bits generates unique hashcodes.
- 10) For Microsoft® Windows® platforms (2000/XP) a static library of CORINA is now available. This enables software developers to easily include the CORINA functionalities into their cheminformatics applications and to call CORINA for 3D structure generation purposes in their own source code also on Windows platforms. The API guidelines require the input to be either a single line formatted MDL SDFfile (with the two text characters "\n" representing each new line character) or alternatively, a SMILES string. The 3D structures are returned as an MDL SDFfile (again formatted as a single line with "\n" as new line character). In addition, all driver options can be triggered.
- 11) Several improvements and bug fixes in the routines for the handling the MacroModel file format, the generation of stereoisomers, the orientation of 3D structures, the usage of system functions and the interface to the library versions (Windows and Linux/UNIX) were made to CORINA to address general stability and reliability issues.
- 12) Section 7 "Error Messages" and section 8 "Warning Messages" of this manual were revised. In addition, the new section 5 "How to ... Use CORINA" is introduced that lists some example uses for CORINA. We hope to revise this section of the manual every time the software is revised to add contributions from current and experienced users. Comments and suggestions can be sent to our support team at support@molecular-networks.com with a subject line header "user contributions: how to use corina".

2.1.11 Version 3.4

The following improvements, changes and new features have been implemented.

- 1) The new input option **-i sdfi2c=<value>** ("sdf item-to-comment") can be used to copy the contents of the SDF data field named <value> (data header) in an SDF input file to a comment line in the output file, e.g., to the 3rd line in the header block of an SDF file. Thus, a piece of information from a data field in an SDF file can be transported to any file format which supports a comment (such as SDF, SYBYL MOL2 or PDB).

For example, the option **-i sdfi2c=MySDFdataField** copies the data entry "My entry in my SDF data field" of the SDF data field

```
> <MySDFdataField>
My entry in my SDF data field
```

to the comment line of the SDF output file (3rd line in the header block):

```
ethane
CHStkserve06270620222D 0 0.00000 0.000004
My entry in my SDF data field
  8 7 0 0 0 0 0 0 0 0 1 V2000
```

Note. The original SDF data field of the input SDF file is preserved and also copied to the output SDF file. The size of the comment line may not exceed 80 characters.

- 2) The new output option **-o sdfc2i=<value>** ("sdf comment-to-item") can be used to copy the comment line of an input file, e.g., 3rd line in the header block of an SDF file, to a newly generated SDF data field named <value> (data header) in the SDF output file. Thus, a comment line from file formats which support comments (such as SDF or SYBYL MOL2) can be transported to a data field of an SDF output file.

For example, the option **-o sdfc2i=MyNewSDFdataField** copies the comment line of the SDF input file

```
ethane
CHStkserve06270620222D 0 0.00000 0.000004
My input SDF file comment line
  8 7 0 0 0 0 0 0 0 0 1 V2000
```

to the data field MyNewSDFdataField in the SDF output file:

```
> <MyNewSDFdataField>
My input SDF file comment line
```

Note. The original SDF file comment line is overwritten by the comment produced by CORINA. The size of the comment line may not exceed 80 characters.

- 3) The new output option **-o pascom** ("pass comment") passes comment lines between file formats which support comments (e.g., SDF, SYBYL MOL2 or PDB). In the case of SDFiles as input and output this option preserves the comment line in the header block (3rd line) of the input SDFile and writes it to the comment line of the SDF output file. By default, CORINA overwrites the comment line of the input SDFile with information about the program version. This option prevents overwriting and passes the comment line of the input to the output file.
- 4) The handling of large molecules has been changed. In general, CORINA does not have any limitations regarding the number of atoms or bonds of an input structure that should be converted into 3D. However, CORINA has been designed to process small to medium sized organic (typically "drug-like") molecules. The larger a molecule gets the more the intra-molecular interactions gain in importance influencing the secondary structure of a molecule. CORINA can model these interactions only to a limited extent and, therefore, is not able to correctly predict 3D structures of polymers and biopolymers such as proteins, enzymes or nucleic acids. For this reason, the following changes and improvements have been implemented.
 - By default, the maximum number of atoms and bonds of an input structure is now limited to 999. This default limitation can be extended by the new driver option **-d maxat=<value>**. For example, the option **-d maxat=1001** extends this limit to a maximum of 1001 atoms and bonds.
Note. Some file formats are limited by definition to a certain number of atom and bonds, e.g., SDFile is limited to 999. This driver option will not circumvent any of these limitations.
 - The functionality of file format conversion (driver option **-d no3d**) is not affected by this limitation. In addition, all file format interfaces that are commonly used for macromolecular structures (e.g., PDB, MacroModel, Maestro, CIF and SYBYL MOL2) have been carefully reviewed to guarantee a proper file format conversion.
- 5) The stereoisomer module (driver option **-d stergen**) has been improved.
 - Input 3D coordinates which may define the configuration at a stereo center are now ignored if no stereo descriptors (wedges, parity flags) are set. This is also the case if the additional driver option **-d preserve** is set in order to preserve defined stereo centers.
 - The handling of spiro compounds and the hashcoding algorithm have been further improved for avoiding the generation of duplicate stereoisomers.
- 6) New structural parameters (standard bond lengths and angles) have been added for the following atom types and substance classes according to some analysis of x-ray structures.
 - New standard bond angles for S(sp³) (103°), P(sp³) (102°) and Se(sp³) (101°) have been added.
 - The standard value for the CSC bond angle in thiophenes (s1cccc1) has been set to 91°.
- 7) Several improvements in the file format interfaces to MDL SDF/RDF (read/write), SMILES (read), SYBYL MOL/MOL2 (read/write), MacroModel (write) and Maestro (write) file formats made the read and write routines more stable and reliable.

- 8) The new driver option **-d ist** ("ignore stereo information") forces CORINA to ignore any stereo information given in the input file including parity flags, wedge bond symbols and definitions on *cis/trans* double bonds (e.g., via 2D coordinates in SDF files or "/" and "^" definitions in SMILES files). This option is helpful if the user is aware of stereo definitions in the input files that are geometrically forbidden (see e.g., *i,o*-norbornane in Figure 18 on page 76). Usually, CORINA will reject structures with geometrically forbidden stereo definitions. However, if the driver option **-d ist** is set, CORINA will generate a geometrically possible isomer. **Note.** Please use this option with care as all defined stereo information will be ignored.
- 9) The new output file option **-o hlabel** labels hydrogen atoms separately from the heavy atoms. By default, CORINA uses the same atom counter for heavy and hydrogen atoms. This option forces CORINA to start at the counter "0" for labeling the hydrogen atoms, independently of the counter for the heavy atoms and to allow for numbering of atoms in molecules with more than 99 atoms (including hydrogen atoms).
- 10) The new driver option **-d planil** forces a planar geometry at anilinic ring nitrogen atoms (see also section "Handling of Pyramidal Ring Nitrogen Atoms" at page 81).
- 11) The new output file option **-o flexrta** influences the REFMAC restraints of torsion angles of in aliphatic ring systems in CIF output files and sets all torsionen to "var" (variable) with a period of "3" and a standard deviation "esd" of "20" degrees in the loop "chem_comp_tor".

2.2 CORINA_F (Restricted FlexX Interface Version)

CORINA_F is a restricted version of CORINA that interfaces to the ligand docking program FlexX [17]. The interface functionality for FlexX is also contained in the full version of CORINA since version 2.0 and the driver option **-d flexx** switches on all command line options that are required to interface to FlexX (see section 4.2 on page 19). During the docking process, FlexX fragments the ligand into cyclic and acyclic parts. The ring systems including their first exocyclic neighbors are send to CORINA or CORINA_F, respectively, which then generates a set of low-energy conformations for these ring systems and sends them back to FlexX.

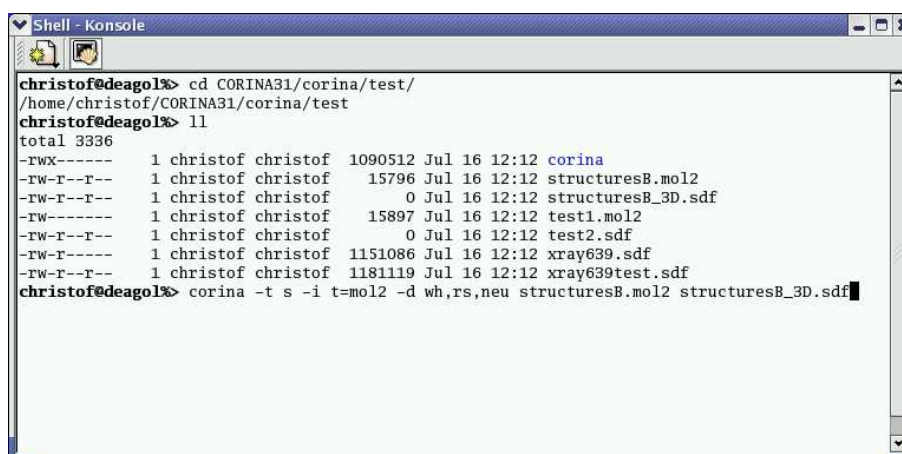
The only difference between CORINA and CORINA_F is that the latter runs only if a number of restrictions are fulfilled:

- 1) Only one input structure per program call is allowed.
- 2) The structure has to contain one but only one ring system (fused, bridged and spiro systems are regarded as single ring system!).
- 3) The only acceptable ring systems are those with no more than nine atoms in the ring.
- 4) Exocyclic parts that exceed two bonds are not allowed.

The FlexX-CORINA interface is described in more detail in section 6.8 on page 43. The method implemented in CORINA and CORINA_F, respectively, for generating multiple ring conformations is briefly described in section 12.5 on page 79.

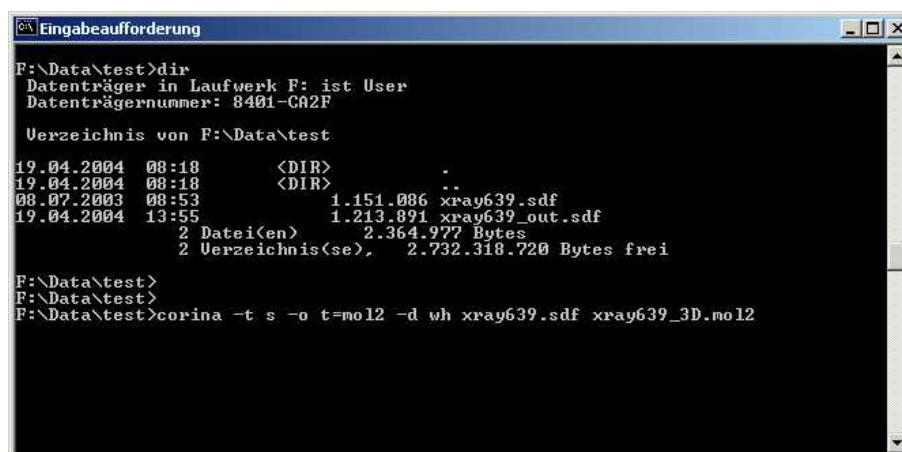
3 Getting Started with CORINA

CORINA is a command line oriented program system (executable file *corina* on UNIX/Linux systems and *corina.exe* on MS Windows platforms) and has to be executed in a shell (e.g., *csh*, *tcsh*, or *bash* on UNIX/Linux systems, see Figure 2) or at a Windows or DOS command prompt (see Figure 3). All command line options provided by CORINA are described in detail in section 4 "Using CORINA" of this manual.



```
christof@deagol> cd CORINA31/corina/test/  
/home/christof/CORINA31/corina/test  
christof@deagol> ll  
total 3336  
-rwx----- 1 christof christof 1090512 Jul 16 12:12 corina  
-rw-r--r-- 1 christof christof 15796 Jul 16 12:12 structuresB.mol2  
-rw-r--r-- 1 christof christof 0 Jul 16 12:12 structuresB_3D.sdf  
-rw----- 1 christof christof 15897 Jul 16 12:12 test1.mol2  
-rw-r--r-- 1 christof christof 0 Jul 16 12:12 test2.sdf  
-rw-r----- 1 christof christof 1151086 Jul 16 12:12 xray639.sdf  
-rw-r--r-- 1 christof christof 1181119 Jul 16 12:12 xray639test.sdf  
christof@deagol> corina -t s -i t=mol2 -d wh,rs,neu structuresB.mol2 structuresB_3D.sdf
```

Figure 2 Command line version of CORINA executed in a UNIX/Linux shell.



```
F:\Data\test>dir  
Datenträger in Laufwerk F: ist User  
Datenträgernummer: 8401-CA2F  
  
Verzeichnis von F:\Data\test  
  
19.04.2004 08:18 <DIR> .  
19.04.2004 08:18 <DIR> ..  
08.07.2003 08:53 1.151.086 xray639.sdf  
19.04.2004 13:55 1.213.891 xray639_out.sdf  
2 Datei(en) 2.364.977 Bytes  
2 Verzeichnis(se), 2.732.318.720 Bytes frei  
  
F:\Data\test>  
F:\Data\test>  
F:\Data\test>corina -t s -o t=mol2 -d wh xray639.sdf xray639_3D.mol2
```

Figure 3 Command line version of CORINA executed in a MS Windows/DOS command prompt.

In addition to the command line version of CORINA, the Java-based graphical user

interface **CORINA.direct** is available that triggers the command line version of CORINA and executes the 3D structure generation process in the background of the system. The installation and the usage of this graphical user interface are described in the document "CORINA.direct – Reference Manual" available from Molecular Networks.

The example file *example.sdf* provided with the distribution contains the structure information of three molecules in MDL SDFfile format [9], 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 4 shows the generated 3D structures.

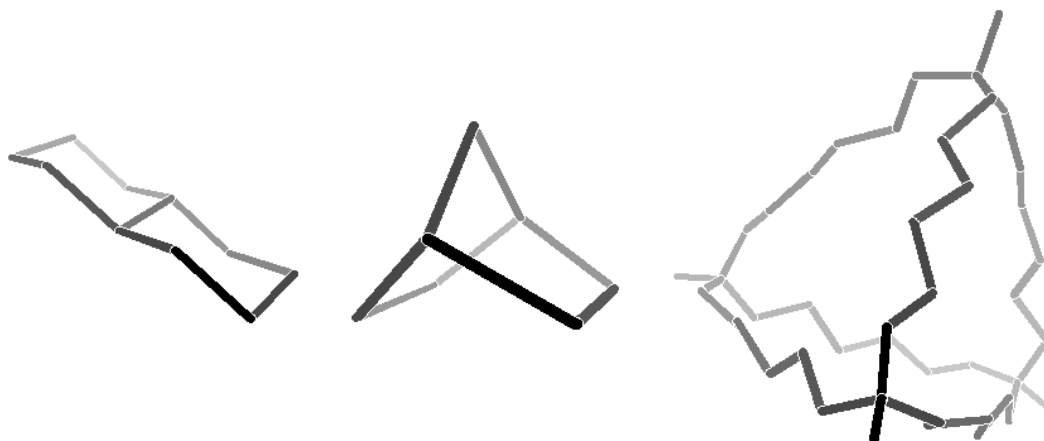


Figure 4 3D models of the structures of the example files.

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 error and warning messages is created in your working directory. The additional trace file option **-t s** redirects this trace information to the standard output device (usually the screen). With the command

corina -t s example.sdf out.sdf

the following output appears on the screen:

corina 3.40 <serial number - compilation date - user - date - time - host>

Input file type MDL SDFfile
Output file type MDL SDFfile

Options:

Standard values: version 3.4 January, 2006
Ring templates: version 3.0 March, 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: 1 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: 130 ms

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

4 Using CORINA

4.1 Synopsis

The general synopsis for using CORINA is:

corina [-option(s) [sub option[=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*.

4.2 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 SDFFile [9] (default),
rdf	MDL RDFFile [9],
smiles	SMILES linear code [1],
mol	SYBYL MOL file [11],
mol2	SYBYL MOL2 file [11],
ctx	Gasteiger ClearText file format [22],
mmod	MacroModel structure file format [13], and
mae	Maestro file format [14].

SDF related input file options

dummies Allow the interpretation of dummy atom types in SDFiles ("A,Q,*").

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>i.e.</i> , "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 that support atom names (<i>e.g.</i> , SYBYL MOL2).
force3d	Force that SDFiles having the 2D flag set are processed as if the 3D flag is set. This option is especially useful if CORINA is used only as a file converter with the driver option -d no3d (see below). This option allows to handle SDFiles that have the 2D flag set (2 nd header line) as if the 3D flag is set. Thus, 2D files can be interconverted into file formats that usually require 3D coordinates (<i>e.g.</i> , SYBYL MOL/MOL2, PDB).
sdfict	Ignore <i>cis/trans</i> configuration of double bonds in MDL SDF input files. In MDL SDFiles, configurations at double bonds are specified by the 2D coordinates of the substituents. This option suppresses the interpretation of the 2D coordinates and generates the most favorable configurations (E configuration in most cases). Note. The generated isomer might not be the required one.
sdfi2c=<value>	Copy the data field <value> (data header) to the comment line of the output file. This option copies the contents of the SDF data field named <value> in the SDF input file to the comment line field of the output file (<i>e.g.</i> , 3 rd line in the header block of an SDF output file).
sdfi2n=<value>	Copy the data field <value> to the name line of the output MDL SDFile (" <u>sdf</u> <u>item</u> <u>2</u> <u>name</u> "). This option copies the contents of the SDF data field named <value> in the SDF input file to the name line in the header block (1 st line) of the SDFile output file.

SYBYL MOL and MOL2 related input file options

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").
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).

SMILES related input file options

dummies	Allow the interpretation of dummy atom types in SMILES linear code ([*]) input files.
---------	---

-o Output file

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

sdf	MDL SDFFile [9] (default),
rdf	MDL RDFFile [9],
mol	SYBYL MOL file [11],
mol2	SYBYL MOL2 file [11],
pdb	Brookhaven Protein Data Bank format [12],
ctx	Gasteiger ClearText file format [22],
dic	CCP4 dictionary file [20],
top	X-PLOR topology file [21],
par	X-PLOR parameter file [21],
mmod	MacroModel structure file format [13],
mae	Maestro file format [14],
cif	Crystallographic Information File format [15],
odb	O Database file format [25], and
dyana	DYANA file format [27].

a Append the output to the input file instead of creating a new output file.

SDF related output file options

lname Allow compound names that are longer than 80 characters in the name field in the header block of an MDL SDF output file. **Note.** This option may violate the MDL SDFFile definitions (see also section 637).

mdl3db	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 SDF file. This option restricts the number of output fields in the atom lines of the atom block in MDL SDF files (RDF files) 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 that are 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 flags that have been calculated by CORINA for missing stereo descriptors to MDL SDF (RDF) output file. If stereo information is missing in the input file CORINA assumes reasonable stereo descriptors following some implemented rules (see section 12.4 on page 72). This option forces CORINA to output the stereo parity flags in MDL SDF and RDF files that were calculated and used by CORINA during the 3D structure generation process.
pascom	Pass the comment line from the input SDF file to the output file. This option preserves the comment line of the input SDF file and writes it to the comment line of the output file. This option works with any file formats that support comments (fields or lines). By default, CORINA overwrites the comment line of the input SDF file with information about the program version.
sdfc2i=<value>	Copy the comment line of the input file to a datafield <value> (data header) in the output SDF file. This option copies the comment line of the input file to a newly generated SDF data field named <value> in the SDF output file. This option works with any file formats that support comments (fields or lines).

SYBYL MOL and MOL2 related output file options

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 <code>USER_CHARGES</code> is set in the field @<TRIPOS>MOLECULE.
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 that 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.
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.
noccat	Suppress the automatic conversion of the carbon atom in amidinium-like structures ($[\text{NH}_2^+]=\text{CN}$) to the carbocation type SYBYL atom type C.cat ($\text{N}[\text{C}^+]\text{N}$) in SYBYL MOL/MOL2 output files. The conversion to this atom type, which is default, is still strongly recommended.
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.
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.

PDB related output file options

keepnames	Keep any atom name given in the input file. Usually, if the atom name and the atom type differ, CORINA tries to derive a reasonable atom name from the atom type for PDB output files. This option forces CORINA to keep and to output atom names which are defined in the input file.
pdbatom	Write the keyword <code>ATOM</code> instead of <code>HETATM</code> in PDB output files.
pdbludi	Create a PDB input file for a Ludi [19] fragment database.
pdbludilabel	Generate unique three-character labels for a Ludi [19] fragment database in PDB output files.
pdbnoconnect	Skip all <code>CONNECT</code> statements in PDB output files.
resnam=<value>	Set the residue name to <value> in PDB output files.
resno=<value>	Set the residue number to <value> in PDB output files.

CCP4 dictionary, X-PLOR topology/parameter, MacroModel, Maestro, Crystallographic Information File (CIF), O Database, and DYANA related output file options

digid=<value>	Set the group ID number to <value> in CCP4 dictionary output files.
flexrta	Set all torsion angles in aliphatic ring systems (with more than 4 atoms) to "var" (variable) with a period of "3" and a standard deviation "esd" of "20" degrees in the loop "chem_comp_tor".
hlabel	Label hydrogen atoms separately. By default, CORINA uses the same atom counter for heavy and hydrogen atoms. This option forces CORINA to start at the counter "0" for hydrogen atoms, independently of the counter for the heavy atoms and to allow for numbering of atoms in molecules with more than 99 atoms (including hydrogen atoms).
keepnames	Keep any atom name given in the input file. Usually, if the atom name and the atom type differ, CORINA tries to derive a reasonable atom name from the atom type for DYANA output files. This option forces CORINA to keep and to output atom names which are defined in the input file.

multor	Write out all possible torsion angle patterns (CIF only). By default, CORINA writes out a single line ("loop_chem_comp_tor" statement) for each torsion angle. This option forces CORINA to write out all possible torsion angle patterns using the loop "chem_comp_tor" statement.
novar	Suppress the writing of torsion angle patterns (CIF only). By default, CORINA writes out a single line ("loop_chem_comp_tor" statement) for each torsion angle. This option forces CORINA not to write out any torsion angle information.
resnam=<value>	Set the residue name to <value> in CCP4 dictionary, X-PLOR topology, MacroModel, Maestro, Crystallographic Information File, O Database, and DYANA output files.
resno=<value>	Set the residue number to <value> in MacroModel, Maestro, and DYANA output files.
typchr=<value>	Set the atom type character(s) to <value> in X-PLOR topology and 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>.
-t	Trace
s	Write trace output to standard error channel (default: <i>corina.trc</i>).
n	Suppress trace output. This option is useful for the conversion of large databases since the trace file <i>corina.trc</i> might become rather large.
tracefile=<value>	Set trace file name to <value> (default: <i>corina.trc</i>).

-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**General driver options**

- 3dst Force stereo descriptors from the 3D structure. If this option is switched on and there is a discrepancy between the stereo descriptors and the 3D structure in the input file, CORINA takes the configuration derived from the 3D coordinates (default: usage of stereo descriptors).
- 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.
- ampax Amplify energy penalties for axial over equatorial substituents. This option tries to direct exocyclic substituents always into the equatorial position unless specified differently or close contacts are detected. **Note.** Using this option may generate geometries that are not the lowest energy conformation CORINA can generate (e.g., 1,2-dimethyl-cyclohexane with two equatorial methyl substituents instead of the lower energy conformation with one in equatorial and the second in axial position).

canon	Canonicalize the input structure before the calculation of 3D coordinates. This option canonicalizes the connection table of an input structure internally before the 3D model is built and uses the canonical atom numbering for the calculation of the 3D coordinates. After the 3D model is generated the original numbering is regained. In some special cases, some algorithms of the 3D structure generation process may be sensitive to the atom numbering of the input structure, <i>i.e.</i> , for different atom numberings slightly different conformations may be obtained. This option prevents any atom numbering dependency.
errorfile=<value>	Write failed structures to the output file <value>. Structures that couldn't be converted by CORINA can be written to a second output file/error file with the file name <value>. The error file <value> will have the same file format as the input file and currently only SDF and SMILES format are supported.
ist	Ignore all stereo information that is given in the input file. By default, CORINA reads in and interprets stereo information that is given in the input file. This option forces CORINA to ignore any stereo information. Note. Please use this option with care as ALL defined stereo information will be ignored.
maxat=<value>	Set the maximum allowed number of atoms per molecule to <value>. By default, the number of atoms is restricted to 999 (due to limitations of file formats, <i>e.g.</i> , SDFfile and to prevent the conversion of macromolecular structures). This option resets this limitation to a user-defined value. Note. This option does NOT circumvent any limitations of file formats.
no3d	Skip the 3D coordinates generation. This option allows using 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 .
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 nitrogen atoms by adding or removing protons. This option can be used together with the driver option rs (see below) in order to remove counter-ions from salts.

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.
ori	Orient the 3D structure according to its principal 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 stereo descriptors and wedged bonds in the 2D drawings of the molecules. When these descriptors differ for one and the same stereo center, 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 stereo descriptors for one and the same stereo center (see section 12.4 on page 72).
planil	Force anilinic ring nitrogen atoms to a planar geometry. This option forces a planar geometry at anilinic ring nitrogen atoms.
r2d	Remove 2D records from the output. If the input and the output file type are both set to MDL SDFfile (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 sub option prevents the output of 2D structures.
rs	Remove small fragments. Remove all but the largest fragments from multi-component records (e.g., counter-ions in salts, solvent molecules).
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.

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.

Generation of stereoisomers

stergen	Generate stereoisomeric compounds. This option forces CORINA to automatically detect chiral centers and double bonds where <i>cis</i> and <i>trans</i> configuration may appear in an input structure and to generate all possible, but unique stereoisomeric compounds, regardless whether correct stereo descriptors are given in the input or not (see section 12.4 on page 72). By default, a maximum number of four stereo centers are processed and a maximum number of 16 stereoisomeric compounds are generated (see below).
msc=<value>	Set the maximum number of processed stereo centers per molecule to <value> (driver option stergen required). By default, the stereoisomer-generating module of CORINA processes a maximum of four stereo centers of an input structure. This option forces CORINA to process the specified number <value> of stereo centers 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.
names	Number the generated conformations or isomers consecutively by adding a counter to their names. If isomers or conformations are generated (driver option stergen required, see above) this option may be used to consecutively number the different geometries by adding a counter to the compound name.
preserve	Preserve defined stereo centers (driver option stergen required). If the configuration of an input structure is not specified for all chiral centers and <i>cis/trans</i> 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.

Generation of multiple ring conformations (for ring systems up to 9 ring atoms)

- 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 12.5 on page 79). The option does not support records that consist of more than one fragment (*e.g.*, salts). Work-around: Combine with sub option **rs** (see above). The conformations are written in the order of increasing steric energy.
- 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.
- flapn** Flap ring nitrogen atoms to generate multiple ring conformations (driver option **rc** required). This option allows pyramidal ring nitrogen atoms that have one exocyclic neighbor to invert their configuration in order to obtain all conformations (see Section 12.5 on page 79).
- mc=<value>** Set the maximum number of generated conformations to <value> (driver option **rc** required). By default, CORINA generates a maximum number of 10 conformations per molecule if the driver option **rc** is set (see above). This option sets the maximum number of output conformations to <value>.
- names** Number the generated conformations or isomers consecutively by adding a counter to their names. If isomers or conformations are generated (driver options **rc** required, see above) this option may be used to consecutively number the different geometries by adding a counter to the compound name.

- ringatom=<value>** Denote a ring system by the atom with label <value> that is part of the ring system to generate multiple ring conformations (requires driver option **rc**). This sub option can be used to denote a specific ring system for which multiple ring conformations should be generated, whereas <value> is the atom label of one atom contained in this specific ring system. It allows generating multiple ring conformations only for one specific ring system in an input structure that contains more than one flexible ring systems.
- sc** Generate ring conformations simultaneously (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 12.5 on page 79).
- symoff** Switch off the symmetry check for multiple ring conformations. In order to generate different ring conformations that are unique, CORINA checks for symmetries but only in the flexible ring skeletons (including the first exocyclic substituents). This option switches off this check and thus, allows to output also conformations that can be interconverted by symmetry operations (e.g., the two conformations of CCC1=C(CC)CCC1).
- 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.

Interface to FlexX

- flexx** Tailor CORINA to the docking program FlexX [17]. CORINA can be used for generating ring conformations during the flexible docking process. This option sets the input and output file types and the conformations analysis options to suited values (see section 6.8 on page 43).

-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
```

5 How to ... Use CORINA

The following section lists some typical use cases of CORINA and shows the corresponding command lines and command line options that have to be switched on for the specific tasks.

5.1 Converting a 2D MDL SDFFile into a 3D MDL SDFFile

A MDL SDFFile should be converted into 3D. Implicit hydrogen atoms should be added, small fragments (e.g., counter ions in salts) should be removed and all molecules should be neutralized. In addition, structures that couldn't be converted should be excluded from the 3D output file but written to a separate error file. The output file should also be formatted in MDL SDFFile format.

Command line:

```
corina -d wh,rs,neu,r2d,errorfile=errors.sdf in.sdf out3D.sdf
```

5.2 Using 3D Input Information to Reproduce the Correct Stereochemistry

A MDL SDFFile containing crude 3D structures with missing stereo information (either no parity flags nor wedge symbols) should be converted into 3D. The stereo information should be derived from the crude input geometries, implicit hydrogen atoms should be added and the generated 3D structures should be oriented by their moments of inertia. Since the 3D structures should be docked with the ligand docking program GOLD, the output file has to be in SYBYL MOL2 file format and all atom and bond types should be assigned according to the GOLD conventions for functional groups

Command line:

```
corina -o t=mol2,gold -d 3dst,wh,ori in.sdf out3D.mol2
```

5.3 Generating Sets of Stereoisomeric Compounds

A file containing several hundreds of SMILES strings should be converted into 3D. For chiral compounds a set of a maximum of 20 stereoisomers should be generated but any defined stereochemistry in the input structures should be preserved. The stereoisomers should be numbered and implicit hydrogen atoms added. The output file format should be MDL SDFFile.

Command line:

```
corina -i t=smiles -d stergen,msi=20,preserve,names,wh in.smi out3D.sdf
```

5.4 Using CORINA as a File Converter

A MDL SDFFile containing structures from X-ray experiments should be converted in a SYBYL MOL2 file. The 3D coordinates of the missing hydrogen atoms should be

added, however, the coordinates of any atom in the input structures should not be changed at all.

Command line:

corina -o t=mol2 -d no3d,wh in3D.sdf out3D.mol2

5.5 Generating Sets of Multiple Ring Conformations

For structures in a SYBYL MOL2 file, new 3D coordinates should be generated. Missing hydrogen atoms should be added, small fragments (e.g., counter ions in salts) and failed structures should be removed from the output file. For each input structure a set of a maximum of ten ring conformations should be generated, nitrogen atoms in flexible rings are allowed to be inverted and the search for different ring geometries should generate a reduced set of conformations (see also Section 12.5 on page 79).

Command line:

corina -i t=mol2 -o t=mol2 -d wh,rs,r2d,rc,mc=10,flapn,sc in3D.mol2 out3D.mol2

6 Supported File Formats and Interfaces

This section shows the file formats that are currently supported by CORINA and gives additional information on their use in CORINA. Table A gives an overview of the input/output formats and a reference to the file format description.

Table A Supported input and output file formats.

Format	Input	Output	Reference
MDL SDF (RDF)	x	x	[9]
SMILES	x		[1]
SYBYL MOLFILE	x	x	[11]
SYBYL MOL2	x	x	[11]
PDB		x	[12]
CTX	x	x	[22]
CCP4 dictionary file		x	[20]
X-PLOR topology/parameter file		x	[21]
MacroModel structure file	x	x	[13]
Maestro file	x	x	[14]
CIF		x	[15]
ODB		x	[25]
DYANA		x	[26],[27]

6.1 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 [9]. 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 implicitly 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.

6.1.1 Options to manipulate MDL SDF and RDF files

The input option **-i sdfi2n=value** copies a one-line data item named <value> to the compound name line (1st line of the header block) in the SDF output file. For example, **-i sdfi2n=MY_FIELD** forces CORINA to copy the data line immediately following the data item header <MY_FIELD> into the compound name line. This can be used to export a single special data line into any other output file format which supports a compound name.

The input option **-i sdfi2c=<value>** copies the contents of the SDF data field named <value> (data header) in an SDF input file to a comment line in the output file, e.g., to the 3rd line in the header block of an SDF file. Thus, a piece of information from a data field in an SDF file can be transported to any file format which supports a comment (e.g., SDF, SYBYL MOL2, PDB).

By default, CORINA considers atom stereo descriptors and wedge symbols for chiral centers (see also driver option **-d ow** in section 4.2 on page 19), as well as bond descriptors indicating *cis* or *trans* double bonds that are given in the input structure for generating a 3D structure. Unfortunately, stereo descriptors are quite often 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*-configured 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 stereo descriptors that were calculated by CORINA for centers with undefined stereo information.

The output options **-o mdlldb** and **-o mdlcompact** are useful for the conversion of large datasets or databases. If **-o mdlldb** 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).

The output option **-o pascom** passes comment lines between file formats which support comments (*e.g.*, SDF, SYBYL MOL2, PDB). If the SDF file format is set as input and output file format the comment line in the header block (3rd line) of the input SDF file is preserved and passed to the output file. By default, CORINA replaces the comment line given in the input SDF file with information about the program version and writes it to the output SDF file.

The output option **-o sdfc2i=<value>** copies the comment line of an input file, *e.g.*, 3rd line in the headerblock of an SDF file, to a newly generated SDF data field named <value> (data header) in the SDF output file. Thus, a comment line from file formats which support comments (*e.g.*, SDF, SYBYL MOL2, PDB) can be transported to a data field of an SDF output file.

6.2 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**). **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 that 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.

6.2.1 Options to manipulate SMILES

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

6.3 SYBYL File Formats

Both SYBYL MOL and MOL2 file formats were implemented following the SYBYL program manual [11]. 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.

6.3.1 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 parameterized 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 that 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 that 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 that 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 ($[H_2N^+]=CN: N.2^+=C.2-N.pl3$) to the SYBYL atom type "C.cat" ($N[C^+]N: N.pl3-C.cat-N.pl3$). **Note.** The conversion to this atom type that 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 [29]). 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.

6.4 Brookhaven Protein Data Bank Format (PDB)

The PDB format was implemented following the literature [12]. The following keywords are used: `HEADER`, `COMPND`, `REMARK`, `HETAM`, `CONNECT`, 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 `CONNECT` statements.

6.4.1 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 that need to read in the `ATOM` statement or cannot handle `HETATOM` statements.

The output option **-o pdbnoconnect** forces CORINA to skip the `CONNECT` statements in PDB output files. **Note.** The `CONNECT` 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`, `CONNECT`, 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.

6.5 The MacroModel Structure File Format (uncompressed)

The MacroModel structure file format was implemented following the literature [13]. All 58 different atom types that are defined in MacroModel are supported. In addition, three different bond types (single, double, and triple bonds) that 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.

6.6 The Maestro File Format

The Maestro file format was implemented following the literature [14]. 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 `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.

6.7 Gasteiger ClearText File Format (CTX)

CTX is a keyword oriented ASCII format developed in the research group of Prof. Dr. Johann Gasteiger [22]. 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.

6.8 Interface between CORINA and FlexX

The flexible ligand docking program FlexX [17] 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 that 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**.

7 Error Messages

7.1 General Errors

ERROR pfopen(): Can't open file *filename (path)*.

The program cannot open the specified file (path in parentheses).

ERROR iopen(): Can't open trace file.

The trace file *corina.trc* cannot be opened.

ERROR iopen(): Identical input/output files.

Identical file names for the input and output files are not allowed.

ERROR iopen(): Can't open input file *filename*.

An error occurred while opening the specified input file.

ERROR iopen(): Can't open output file *filename*.

An error occurred while 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 4 on page 19).

ERROR corina(): File type not allowed.

A file type that is not allowed for input and/or output files is specified (see section 4 on page 19).

ERROR restrict(): FlexX restrictions violated.

The restrictions of the FlexX interface are violated or not fulfilled (see section 6.8 on page 43).

7.2 Input File Format Errors

7.2.1 MDL SDFFile

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

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

The counts line cannot be read in. The current record is discarded.

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

The atom block cannot be read in. The current record is discarded.

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

The atom bond cannot be read in. The current record is discarded.

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

The 2nd header line cannot be read in. The current record is discarded.

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

The 2nd 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 have to be assumed.

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

The 3rd header line is misformatted and cannot be read in. The current record is discarded.

ERROR rmdl(): Unknown element symbol.

An unknown element symbol is encountered. The current record is discarded.

ERROR rmdl(): Bond atoms out of range.

A bond between atoms out of the range 1.. N_{Atom} was encountered. The current record is discarded.

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. The current record is discarded.

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. The current record is discarded.

ERROR rmdl(): Non-standard element.

A non-standard element symbol or atom type is encountered and assumed to be a dummy atom type unless dummy atom types are explicitly forbidden.

ERROR rmdl(): Extra character after element symbol.

The atom symbol consists of three character and only two are allowed. The current record is discarded.

ERROR rmdl(): Dimensions code is 2D in an obvious 3D record.

The dimension flag in the 2nd header line is set to 2D, although the x, y and z coordinates are available for all atoms of the input structure.

Errors with missing columns:

ERROR rmdl(): No mass difference column specified.

The 5th column in the atom block to specify isotopic mass differences is missing.

ERROR rmdl(): No atomic charge column specified.

The 6th column in the atom block to specify charges is missing.

ERROR rmdl(): No atom stereo column specified.

The 7th column in the atom block to specify atom parity stereo flags is missing.

ERROR rmdl(): Too many additional lines.

Only a limited number of additional lines (max. 50,000) is allowed in the data section of an MDL SDFFile.

ERROR rmdl(): Line too long.

Only 80 characters per line are allowed in MDL SDFFiles.

Errors in charge, radical, isotope, and attachment point lines:

ERROR rmdl(): CHG atom out of range.

A charge (M CHG) is specified for an atom with an atom label that does not exist.

ERROR rmdl(): RAD atom out of range.

A radical (M RAD) is specified for an atom with an atom label that does not exist.

ERROR rmdl(): APO atom out of range.

An attachment point (M APO) is specified for an atom with an atom label that does not exist.

ERROR rmdl(): ISO atom out of range.

An isotope (M ISO) is specified for an atom with an atom label that does not exist.

7.2.2 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 "(" parenthesis. The current record is discarded.

ERROR smilesct(): General error while interpreting this character.

A not interpretable character was encountered. The current record is discarded.

ERROR smilesct(): Unknown element.

An unknown element symbol was found. The current record is discarded.

ERROR smilesct(): Too many ('s.

A closing ")" parenthesis is missing. The current record is discarded.

ERROR smilesct(): Closing ring.

No suited ring bond label. The current record is discarded.

ERROR smilesct(): Valency problem making a new bond.

The standard valency of an atom is exceeded. The current record is discarded.

ERROR smilesct(): Too many ['s.

A closing "]" bracket is missing. The record is discarded.

ERROR smilesct(): Reading bracketed atom.

The atom type in square brackets is incorrect. The current record is discarded.

- 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. The current record is discarded.
- ERROR smilesct(): Duplicate slash at double bond.
Two bonds at one atom of a stereo double bond have a slash character as stereo descriptor. Only one slash is allowed.
- ERROR smilesct(): Not all rings closed.
Not for all ring bonds a second label was defined.
- ERROR smilesct(): Reading atomic charge qualifier.
A charge of an atom is defined incorrectly. The current record is discarded.
- ERROR smilesct(): Too many { 's.
A closing "}" brace is missing. The current record is discarded.
- ERROR smilesct(): Unknown stereo class.
An unknown stereo class is specified. The current record is discarded.
- ERROR smilesct(): No organic element. Use square brackets.
Inorganic elements must be written in square brackets. The current record is discarded.
- ERROR smilesct(): Ring closure label not following immediately the atomic symbol.
Ring bond labels must follow immediately after the atomic symbol. The current record is discarded.
- ERROR smilesct(): Conflicting ring closure bond types.
The type of the ring closure bond was defined twice with differing values. The current record is discarded.
- ERROR smilesct(): Chiral center has wrong connectivity.
A tetrahedral center with less than four neighbors was encountered. A common error is to forget to specify a hydrogen atom neighboring the stereo center within the square brackets.
- ERROR smilesct(): Expecting atomic symbol.
The first character inside square brackets must be an atomic symbol. The current record is discarded.
- ERROR smilesct(): Illegal valence state.
An atom in an illegal valence state was encountered. The current record is discarded.
- ERROR smilesct(): Inconsistent EZ specification.
The EZ configuration of a double bond was defined redundantly with different values.
- ERROR smilesct(): Isotopic mass error.
The given isotopic mass is out of range. The current record is

discarded.

ERROR smilesct(): Stereo permutation not implemented.
The specifications for square-planar, trigonal-bipyramidal and octahedral chirality is not implemented and is therefore ignored.

ERROR smilesct(): Illegal hybridisation.
An illegal hybridization state of an atom according to the SMARTS extensions is encountered. The current record is discarded.

7.2.3 SYBYL MOL2 File

ERROR rsyb2(): Missing name in line #.
The name of the molecule is missing (1st line after the MOLECULE keyword). The current record is discarded.

ERROR rsyb2(): Missing counts in line #.
The atom and bond counts (number of atoms and bonds within the molecule) are missing (2nd line after the MOLECULE keyword).
The current record is discarded.

ERROR rsyb2(): Error reading counts in line #.
A general error occurred when reading the atom lines after the ATOM keyword. The current record is discarded.

ERROR rsyb2(): Missing charge type in line #.
The definition of the type of charges is missing (4th line after the MOLECULE keyword). The current record is discarded.

- ERROR rsyb2(): Can't read atoms.
A general error occurred when reading the atom lines after the ATOM keyword. The current record is discarded.
- ERROR rsyb2(): Can't read bonds before atoms.
The atom block (after the ATOM keyword) has to be defined before the bond block (after the BOND keyword). The current record is discarded.
- ERROR rsyb2(): Can't read bonds.
A general error occurred when reading the bond lines after the BOND keyword. The current record is discarded.
- ERROR rsyb2(): Can't read sets before atoms and bonds.
Sets (SET keyword) have to be defined after the atom block (ATOM keyword) and the bond block (BOND keyword). The current record is discarded.
- ERROR rsyb2(): Can't read all sets.
A general error occurred when reading the set lines after the SET keyword. The current record is discarded.
- ERROR rsyb2(): Can't read rotatable bonds before atoms and bonds.
Rotatable bonds (ROTATABLE_BOND keyword) have to be defined after the atom block (ATOM keyword) and the bond block (BOND keyword). The current record is discarded.
- ERROR rsyb2(): Can't convert aromatic system.
The assignment of electrons to an aromatic system failed. Probably, the VB structure is corrupted. The current record is discarded.
- ERROR rsyb2(): Can't read atom in line #.
The atom in line # cannot be read. The current record is discarded.
- ERROR rsyb2(): Error parsing atom in line #.
The information given for the atom in line # is not sufficient. The current record is discarded. The current record is discarded.
- ERROR rsyb2(): Unknown atom type in line #.
The atom type in line # does not comply with definition of allowed SYBYL atom types. The current record is discarded.
- ERROR rsyb2(): Can't read bond in line #.
The bond in line # cannot be read. The current record is discarded.

- ERROR rsyb2(): Error parsing bond in line #.
The information given for the bond in line # is not sufficient. The current record is discarded.
- ERROR rsyb2(): Unknown bond type.
The bond type in line # does not comply with definition of allowed SYBYL bond types. The current record is discarded.
- ERROR rsyb2(): Valence problem reading bond #-#.
The assignment of the electrons for the bond between the atoms # and # failed. Probably, the VB structure is corrupted. The current record is discarded.
- ERROR rsyb2(): Reading set header.
A general error occurred while reading the header for a set (SET keyword). The current record is discarded.
- ERROR rsyb2(): Too few tokens in set header.
The information given in the set header (SET keyword) is not sufficient. The current record is discarded.

7.3 Stereo Errors

Some further information on the handling of stereochemistry by CORINA and for the interpretation of the following error messages is given in section 12.4 on Page 72.

- 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 CORINA.
- ERROR initster(): More axes than expected at atom #.
The maximum number of six neighboring atoms is exceeded. The current record is discarded.
- ERROR initster(): No suitable configuration for atom #.
No suitable configuration for atom # can be derived. The number of axes does not correspond to the number of neighboring atoms. The current record is discarded.
- ERROR clcster2(): Ambiguous configuration of ligands.
The specification of a chiral center is ambiguous in the input 2D structure diagram and has to be ignored.

- ERROR clcster2(): The ligands at atom # don't span up a volume.
The specification of a chiral center is ambiguous in the input 2D structure diagram. CORINA tries to correct it.
- ERROR clcster2(): Insufficient number of axes at atom #.
The number of axes for a higher coordinated center # (5 or 6 ligand atoms) that are specified in the input 2D structure diagram is not enough. CORINA tries to correct it.
- ERROR clcster3(): Insufficient number of axes at atom #.
The number of axes for atom # that are specified in the input 3D structure is too small. CORINA tries to correct it.
- ERROR allcis(): No suited stereo descriptors for bridge #- #-#.
The coding of the stereochemistry of the bridgehead atoms #-# is incorrect. No 3D coordinates can be generated.
- ERROR stergen(): Maximum number of # stereo centers exceeded.
By default, the stereoisomer generator in CORINA only accepts a maximum of four (4) stereo centers per molecule unless specified differently with the driver option -d msc=<value>. If the maximum number of stereo centers is exceeded, only one stereoisomer is generated.

7.4 Errors in the Generation of 3D Coordinates

- ERROR corina(): Input structure incorrect. No 3D generation.
The input structure is corrupted and CORINA cannot generate a 3D structure.
- ERROR gen3d(): Unable to build a 3D structure/single conformation.
No 3D structure can be generated. A fatal error occurred during the generation process.
- ERROR gen3d(): 3D structure didn't pass the quality check.
The quality of generated 3D structure is insufficient and not written to 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 gen3d(): To many atoms (max. 999).
The number of atoms in the current molecule exceeded the internal limitation of 999. This limitation can be re-defined by the user with the driver option -d maxat=<value> (e.g., -d maxat=2000).
- ERROR quality(): Some internal coordinates are heavily distorted.
A fatal error occurred during the 3D generation. No 3D structure can be generated.
- ERROR quality(): Violated cis/trans bond.
A stereo double bond changed its configuration during the geometry optimization. No 3D structure is output.
- ERROR quality(): Violated stereo atom.
A chiral atom changed its configuration during the geometry optimization. No 3D structure is output.
- ERROR quality(): Bad contacts detected.
Unfavourable non-bonded interactions (crowded atoms) are detected in the generated 3D structure and therefore not written to the output file.
- ERROR alltempl(): Missing ring template.
For one ore more rings no suitable ring template can be found in the list of predefined ring templates. No 3D structure can be generated for the current record.
- ERROR ringfrag(): Fragment contains ring(s) > 9.
CORINA handles only rings up to a size of nine ring atoms members by predefined ring templates. Larger rings are reduced to a secondary structure that have less than ten anchor atoms (see section 12.1.2 on page 65). This reduction failed. No 3D structure can be generated for the current record.
- ERROR ringfrag(): Can't combine the templates.
The available ring templates cannot be combined to a one single ring conformation. No 3D structure can be generated for the current record.
- 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. No 3D structure can be generated for the current record.

ERROR procfrag(): Can't process small ring system.

A small ring system cannot be translated into a 3D structure. No 3D structure can be generated for the current record.

ERROR bigsys(): Unable to process all fragments.

Some of the fragments of a large ring system cannot be translated into a 3D structure. No 3D structure can be generated for the current record.

ERROR getconf(): Can't build ring fragment.

For a small ring system no suitable ring conformation could be generated. No 3D structure can be generated for the current record.

7.5 Output File Format Errors

7.5.1 MDL SDFfile

ERROR wmdl(): Cannot write more than 999 atoms.

The current molecule has more than 999 atoms which exceed the limitations of an MDL SDFfile.

ERROR wmdl(): Cannot write more than 999 bonds.

The current molecule has more than 999 bonds which exceed the limitations of an MDL SDFfile.

8 Warning Messages

8.1 General Warning Messages

WARNING mnLicenseCheck(): License expired on <date>! Exiting ...
The license time has on <date> expired. The usage of the program is not allowed anymore.

8.2 Warnings Related to Input File Formats

8.2.1 SMILES

WARNING smilesct(): Non-standard formal charge qualifier in curly brackets.
A charge value is defined in "{}" braces (curly brackets) instead of in "[]" square brackets. This does not comply with the SMILES language definition. CORINA tries to correct it.

WARNING smilesct(): Unnormal valence state.
An unnormal valence state is encountered. The number of bonds exceeds the free valences of an atom. CORINA tries to correct it.

WARNING smilesct(): Un-paired label--inserting APO
An unclosed ring system is encountered since the second label for ring closure is missing. An attachment point (APO) is inserted in the internal CT representation to take into account the open valence.

WARNING smilesct(): Duplicate slash at double bond--second one ignored.
A *cis* or *trans* double bond is marked with two consecutive slash characters ("/" or "\\"). The second slash character is ignored in order to derive a proper definition of the double bond configuration.

WARNING smilesct(): Ignoring chirality at atom with more than one H.
An atom that is marked as chiral has more than one bonded hydrogen atom. Therefore, the chirality is ignored.

WARNING smilesct(): Incomplete EZ specification--ignored
A *cis* or *trans* double bond is incompletely specified, *i.e.*, one of the two mandatory slash characters is missing. Therefore, the descriptors are ignored.

WARNING smilesct(): Label following branch.

A label indicating a ring closure that is directly placed behind a branch is encountered. This combination might cause problems but CORINA tries to solve this problem.

8.2.2 SYBYL MOL2 File

WARNING rsyb2(): Discarding record due to problems with aromatic system.

The assignment of electrons to an aromatic system failed. The current record is discarded.

WARNING rsyb2(): Discarding record due to dummy atoms/bonds.

The record has to be discarded due to dummy atom and/or bond types that cannot be interpreted. The current record is discarded.

WARNING rsyb2(): Improper atom and bond types.

Some atom and/or bond types do not comply with the definition of allowed SYBYL atom and bond types. CORINA tries to derive correct types.

WARNING rsyb2(): Dummy atom in line <#> interpreted as <type> from atom name.

The dummy atom type in line # is interpreted as atom symbol <type> derived from the atom name. This message only appears if the input option "xelement" is set.

WARNING rsyb2(): Unknown atom type in line <#> (<type>) interpreted as element symbol

The unknown SYBYL atom type <type> in line # is interpreted as an element symbol. This message only appears if the input option "xelement" is set.

WARNING sybchkn(): Setting atom <#> from <type> to <type> based on 3D.

Based on the input 3D structure, the SYBYL atom type of atom # does not match to the geometry of the atom. Therefore, the atom type is changed internally (e.g., N.pl3 to N.3).

WARNING sybplaus(): Probably wrong SYBYL type <type> at atom #.

The SYBYL atom type <type> at atom # may be wrong due to geometric reasons. It seems that the atom has more neighbors than the geometry allows. CORINA tries to derive a proper atom type.

8.3 Warnings Related to Stereochemistry

Some further information on the handling of stereochemistry by CORINA and for the interpretation of the following warning messages is given in section 12.4 on Page 72.

WARNING initster(): Stereo atom # without stereo descriptor.

No stereo descriptor (parity flag, wedge symbol) is given for the

chiral atom #. The output 3D structure might have an unexpected configuration as CORINA has to use default rules or assume an arbitrary stereochemistry.

WARNING clcster2(): Possibly stereo problem at atom #.

A general problem while calculating the configuration of a stereo center from the 2D structure diagram was encountered. CORINA tries to correct it (see below).

WARNING clcster2(): Trying to ignore H-atom at stereo center #.

An ambiguous 2D configuration was encountered. CORINA 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. CORINA tries to assign a wedge descriptor to an additional bond in order to solve the problem.

WARNING clcster2(): Trying to correct by moving the central atom #.

An ambiguous 2D configuration was encountered. CORINA tries to correct it by moving the central atom #.

WARNING clcster2(): Collision of wedge symbol and stereo descriptor 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 (parity flag) 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.

WARNING clcster2(): Wedge symbol(s) pointing with the basis to the stereo center.

Wedge symbols that are pointing with the basis (broad end) to the stereo center lead to an ambiguous stereo definition. CORINA tries to correct this by finding an alternative coding (wedge symbol).

WARNING clcster3(): Collision of implicit stereo descriptor and 3D coords at atom #.

If 3D coordinates are given in the input structure CORINA checks whether they match to the stereo descriptors. By default, the stereo descriptors are used, however, if the driver option -d 3dst is set the stereo information is calculated from the 3D coordinates.

WARNING clcster3(): Number of possible axes at atom # insufficient.

The number of axes for atom # that are specified in the input 3D structure is too small. CORINA tries to correct this.

8.4 Warnings in the Generation of 3D Coordinates

- WARNING bondlen(): No bond length #-#.
CORINA 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 atoms, e.g., cubene.
- WARNING getta(): No TA #-#-#-#.
In case of linear systems (e.g., acetylene), torsion angle values (TA's) cannot be specified for some bonds.
- WARNING hmoboord(): No HMO constant for atom #.
No Coulomb integral parameter is found for a hetero atom.
CORINA uses the value for carbon.
- WARNING hmoboord(): No HMO constant for bond #-#.
No parameter for a bond resonance integral is found. The value for a C-C bond is used instead.
- WARNING genconf(): Time out.
The time limit for a complete ring conformation analysis was exceeded. The conformer with lowest energy found so far is not necessarily the global CORINA minimum.
- WARNING rrefine(): # pair(s) of crowded atoms
Some pairs of atoms have a distance less than 75% of the sum of their atomic radii.

9 Technical Requirements

9.1 System Requirements

CORINA is a command line tool and has to be executed in a shell (e.g., csh or tcsh on UNIX/Linux systems) or at a MS Windows or DOS command prompt (see also section 3 on page 16). The following hardware platforms and operating systems are supported:

- x86 platforms, Microsoft® Windows® 2000/XP (win32, latest service pack required)
- x86 platforms (32bit, 64bit processors), SUSE Linux®, version 10, kernel 2.4/2.6
- x86 platforms (32bit processors), RedHat® Linux®, version 9, Kernel 2.4
- Silicon Graphics SGI® workstations, IRIX® 6.5
- Sun® Sparc stations, Solaris® 9 (SunOS® 5.9)

9.2 Program Scope and Known Limitations

CORINA has been designed to process a broad range of chemistry. The scope is on organic chemistry and typically drug-like, small to medium sized organic molecules.

In general, there are no limitations concerning the number of atoms or bonds of a molecule to be processed. **Note.** Some supported structure file formats as well as hardware and operating system related issues may cause such limitations.

Structure input file formats must fully comply with their definitions according to the literature references given in this document.

The periodic table is parameterized up to atomic number 103 (Lawrencium). Bond types are limited to single, double, triple and aromatic bonds. Atoms having up to six neighbors can be processed. Bond geometries are limited to linear, planar, tetrahedral, trigonal bipyramidal and octahedral types.

Organic chemical compounds that can be correctly expressed in a valid valence bond (**VB**) notation can be processed. Multi-center electron-deficient bonds, π -complexes, metal/transition metal multi-core compounds and clusters are not in the scope.

Stereochemical information is fully considered as far as the supported file formats allow for the definition of stereochemistry.

Multi-fragment structures (e.g., salts) can be processed.

By default, a single low-energy conformation is generated for an input structure. For ring systems having more than three and less than ten ring atoms, a limited number of reasonable multiple conformations can be generated.

10 Program Installation

10.1 On-line Download from Molecular Networks' Web Server

Since version 3.1, CORINA is available for electronic download via the Internet on the web server of Molecular Networks (Download Area). At

<http://www.molecular-networks.com/php/profile.php>

an account can be created that provides access to licensed software, evaluation copies, program manuals, example files, and tutorials of CORINA as well as to test copies of a variety of chemoinformatics applications offered by Molecular Networks.

The software packages are submitted electronically to the user as compressed files in order to increase the download speed. The downloaded files can be easily uncompressed with standard software tools for file compressing and archiving, such as WinZip, FileZip (<http://www.filezip.com>), or gzip (<http://www.gzip.org>)

CORINA is currently available for the following hardware platforms and operating systems.

hardware platform operating system	downloaded file name
SGI workstations IRIX 6.5	<i>corina_<annual eval>_IRIX6.5_<YYYY_MM_DD>.sgi.gz</i>
x86 Linux workstations kernel 2.4 (gcc 3.2)	<i>corina_<annual eval>_Linux2.4_<YYYY_MM_DD>.lnx.gz</i>
x86 Linux workstations RedHat 9 distribution, kernel 2.4 (gcc 3.3)	<i>corina_<annual eval>_Linux2.4_RedHat9_<YYYY_MM_DD>.lnx.gz</i>
Sun SPARC stations Solaris9 (SunOS 5.9)	<i>corina_<annual eval>_SunOS5.8_<YYYY_MM_DD>.sun.gz</i>
MS Windows platforms win32: NT4/2000/XP	<i>corina_<annual eval>_Win32_<YYYY_MM_DD>.exe.gz</i>

Please uncompress the downloaded file of CORINA

corina_<annual|eval>_<OS>_<YYYY_MM_DD>.sgi|sun|lnx|exe.gz

(*annual* = licensed version with annual run time; *eval* = evaluation version; OS = operating system; YYYY_MM_DD = expiration date in the format year_month_day)

and rename the resulting file to *corina.sgi|.sun|.lnx|.exe* with respect to the operating system (see also table above).

Furthermore, additional information such as this program manual in PDF format or

examples of structure files (see section 3 on page 16) can be downloaded from the web server of Molecular Networks (<http://www.molecular-networks.com>).

10.2 Distribution on CDROM

On request, a CDROM can be ordered from Molecular Networks for backup purposes. To order a CDROM, please contact Molecular Networks GmbH via email at support@molecular-networks.com or at any other address given in section 11 "Problems and Help!" on page 63 of this manual.

10.3 New Installation

10.3.1 UNIX Systems (SGI, Sun SPARC, x86 Linux)

CORINA is a command line oriented program system (executable file *corina*) that has to be executed in UNIX/Linux shell (e.g., *csh*, *tcsh*, or *bash*). The usage of CORINA as well as all available command line options are described in detail in the sections 3 "Getting Started" and 4 "Using CORINA" of this manual.

To install the command line version of CORINA (*corina.sgi.sun.lnx*) please follow the instructions below.

- 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* to the subdirectory *corina* and rename the file *corina.sgi.sun.lnx* to *corina*.
Note. *corina.sgi.sun.lnx* is a binary file.
- 3) Add the *corina* subdirectory name to the environment variable *PATH* in your *.login* or *.cshrc* files (*.profile* or *.bashrc*).

In addition to the command line version of CORINA, a Java-based graphical user interface (CORINA.*direct*) is available that triggers the command line version of CORINA and executes the 3D structure generation process in the background of the system. The installation and the usage of this graphical user interface are described in the document "CORINA.*direct* – Reference Manual" available in the Download Area of Molecular Networks' web server (<http://www.molecular-networks.com>).

10.3.2 Microsoft Windows Platforms (win32, 2000/XP)

CORINA is a command line oriented program system (executable file *corina.exe*) that has to be executed at a Windows or DOS command prompt. The usage of CORINA as well as all available command line options are described in detail in the sections 3 "Getting Started" and 4 "Using CORINA" of this manual.

To install the command line version *corina.exe* please follow the instructions below.

- 1) Create a subdirectory *corina*, e.g., *C:\programs\corina*.
- 2) Copy the file *corina.exe* to the subdirectory *corina*.
- 3) In order to execute CORINA from any other folder or directory, add the *corina* executable file (*corina.exe*) and the path where the program resides (e.g., *C:\programs\corina*) to your environment variables of your system settings (variable: *corina*; value: *C:\programs\corina*).

In addition to the command line version of CORINA, a Java-based graphical user interface (*CORINA.direct*) is available that triggers the command line version of CORINA and executes the 3D structure generation process in the background of the system. The installation and the usage of this graphical user interface are described in the document "*CORINA.direct* – Reference Manual" available in the Download Area of Molecular Networks' web server (<http://www.molecular-networks.com>).

10.4 Program Updates

- 1) Before installing the new version, please copy the old executable and configuration files to a new directory, e.g., *corinaVV* (*VV = old-version-number*, e.g., *corina24*).
- 2) According to the hardware platform install the new version following the installation instructions given in section 10.3 on page 61.
- 3) **Note.** Since CORINA version 2.4, the data files *stdval.ctx* and *rings.ctx* are no longer part of the distribution. All data have been included in the binary file of CORINA (see section 2.1.6 on page 5).

11 Problems and Help!

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

Molecular Networks GmbH Computerchemie
Henkestr. 91
91052 Erlangen
Germany

or contact us by email support@molecular-networks.com,

or by Fax +49 9131 815 669.

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

You can also use the report form in section 15 on page 91 of this manual.

12 Understanding CORINA

12.1 Fast and Efficient Generation of High-Quality 3D Molecular Models

12.1.1 The Core System

CORINA can be regarded as an automatic 3D model building kit. By combining monocentric fragments with standard bond lengths and angles and by 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 torsion angles, two major problems arise: First, in ring systems only restricted sets of torsion angles are allowed that ensure proper ring closure. Secondly, non-bonded interactions due to flexible chain portions have to be minimized. Therefore, CORINA handles rings and chains separately.

Rings of up to a size of nine atoms are processed by using a table of single ring conformations that 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 energy 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 that contains only special geometric terms for the optimization of ring systems.

For acyclic fragments and molecules, the principle of longest pathways has been implemented in CORINA (see Figure 5). 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.

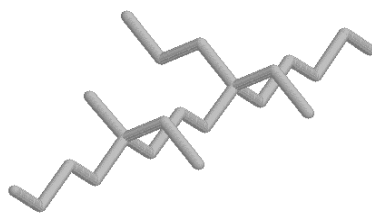


Figure 5 The principle of longest pathways for acyclic fragments and molecules.

After the 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. First, a strategic rotatable bond within the pathway connecting the two interacting atoms is determined, depending on topological features and double bond character. Secondly, the torsion angle of this bond is changed until the non-bonded interactions are eliminated (see Figure 6). 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 [1], [18], [23].

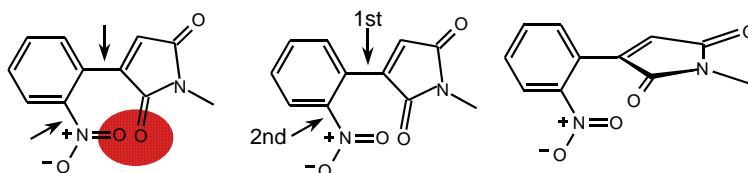


Figure 6 Reduced conformational analysis to avoid non-bonded interactions.

12.1.2 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 nine atoms. Therefore, for large rings another method is necessary. However, polymacrocyclic structures often show a general outline, a superstructure [6]. The porphyrine bridged cyclophane molecule in Figure 7 shows a cage-like superstructure that retains the approximate shape and symmetry of the entire system.

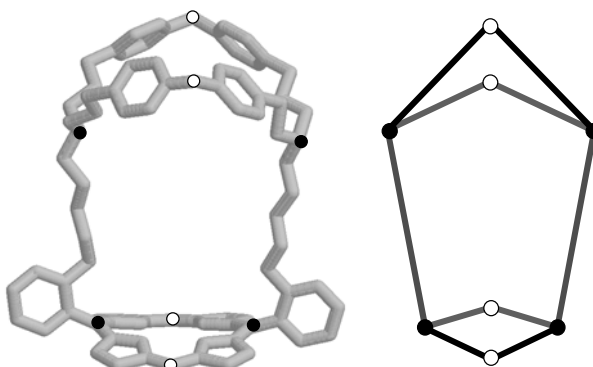


Figure 7 A macrocyclic molecule and the corresponding superstructure. The anchor atoms in both structures are marked by circles.

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 7. Then, a 3D model for the superstructure that contains only small rings can be generated applying the methods for small rings. Finally, the removed atoms are restored and a complete 3D model of the entire ring system is obtained. Figure 8 compares the X-ray structures of three polymacrocycles with the corresponding CORINA models and shows 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.

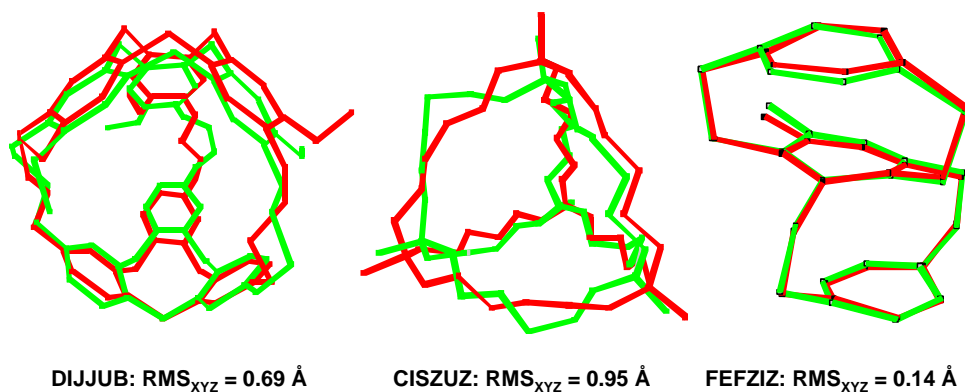


Figure 8 Comparison of the X-ray structures of three polymacrocyclic systems with the corresponding CORINA models and their RMS_{XYZ} deviation. (The experimental structure of CISZUZ contains an I_3^- anion inside the ring system that is not modelled by CORINA.)

12.1.3 Another Challenge: Metal Complexes

Another type of structures commonly neglected 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 well to the experimentally determined

geometries. Figure 9 shows three examples: a nickel, a ruthenium, and a rhodium complex and the RMS_{XYZ} deviations from the X-ray structures.

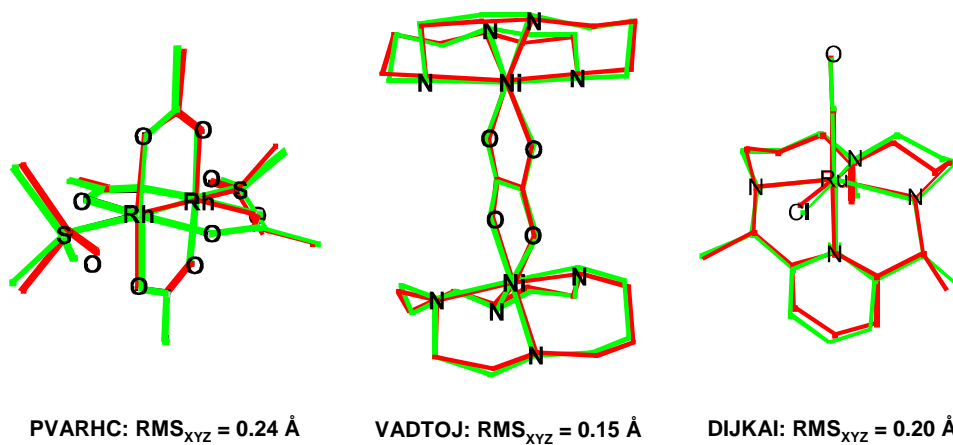


Figure 9 Comparison of the X-ray structures of three metal complexes with the corresponding CORINA models and their RMS deviation.

12.2 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 [7]. The following section will summarize the most important results and present some additional investigations in order

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

The authors 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 B 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 broadest 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 stereo centers.

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%). **Note.** The criterion for reproduced chain geometries has been redefined. In the original paper [7] all torsion angles at a rotatable bond have been taken into account. Since this over-estimates some types of bonds, only one torsion angle per rotatable bond is counted. 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 B 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 B 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 10 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.

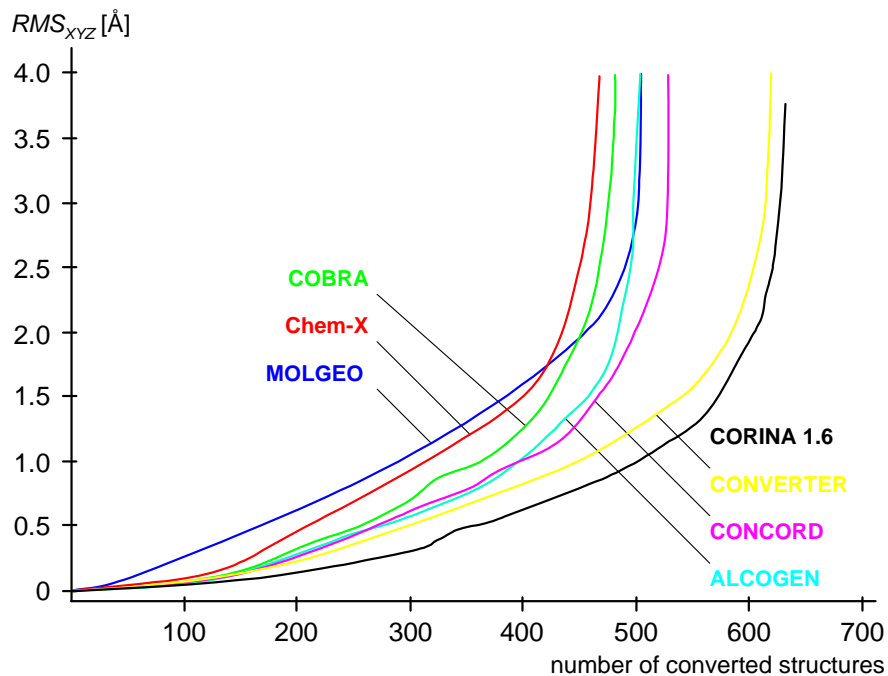


Figure 10 Quantity-quality characteristics of the seven 3D structure generators: Conversion rate vs. RMS_{XYZ} value of the non-hydrogen atoms [31].

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}) [18].
- 2) Improvements in the algorithm that refines atom overlap (see CCR).
- 3) A substantial speed-up of almost a factor of 2 (see CPU time).

12.3 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 SDF file 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 stereo centers, 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 ($\text{RMS} < 0.3 \text{ \AA}$) 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 C 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 favor 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 C 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

12.4 Stereochemical Information

12.4.1 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

^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.

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 11).

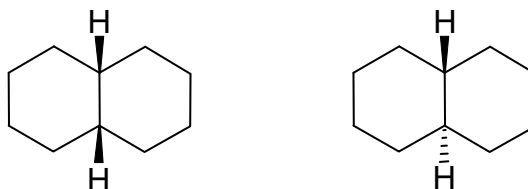


Figure 11 Coding of the stereo information of *cis*- and *trans*-decalin by up and down bond symbols.

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 12 shows two recommended 2D drawings of a chiral center. Other correct variations are shown in Figure 13.



Figure 12 Recommended input of stereochemistry.

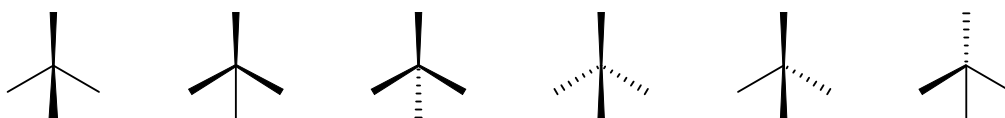


Figure 13 Examples of other correct 2D drawings of a chiral center.

Pitfalls. A number of ambiguous specifications of stereo centers are possible which are often not easily recognized. Figure 14 shows some examples. They all have in common that the ligands of the stereo center 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 stereo descriptor can be calculated. An error message (see Section 7.3 on page 51) is written and the calculation is continued with an arbitrary descriptor.

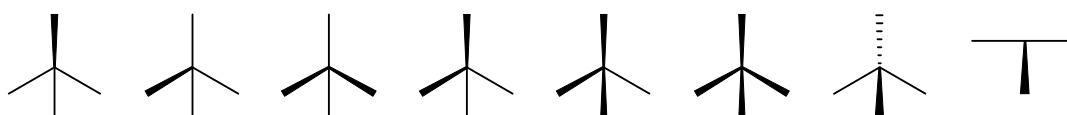


Figure 14 Examples of incorrect 2D drawings of a chiral center.

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 stereo center 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 stereo descriptor (see Figure 15). This is automatically tried and a warning is written (see section 8.3 on page 56) since there is no guaranty that the correction was the intended one.

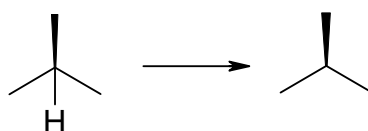


Figure 15 Correction by ignoring one hydrogen atom.

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 16). This is automatically tried and a warning is written (see section 8.3 on page 56) since there is no guaranty that the correction was the intended one.

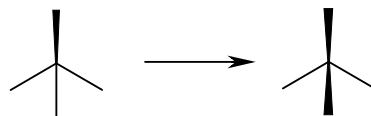


Figure 16 Correction by assuming an additional bond descriptor.

The last type of ambiguous coding of stereochemistry is corrected by moving the central atom. Figure 17 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 8.3 on page 56) since there is no guaranty that the correction was the intended one.

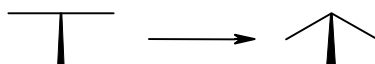


Figure 17 Correction by moving the central atom.

12.4.2 Addition of Missing Stereo Descriptors

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 stereo descriptors. An exception are ring systems containing chiral atoms. For these systems other than arbitrary values must be found for the stereo descriptors of unspecified stereo centers. Geometric and energetic constraints reduce the number of possible stereoisomers and require a careful choice of suited stereo descriptors. Some examples shall illustrate this idea.

Bridged Systems. Bridged ring systems as, *e.g.*, norbornane (see Figure 18) require an *o,o*-configuration of the bridgehead atoms. The *i,o*-isomer is geometrically forbidden since the bridge cannot be closed as illustrated in Figure 18. CORINA defines the stereochemistry at unspecified bridgehead atoms according to this rule. When the input file contains stereo descriptors violating this rule, the processing of the molecule is abandoned with an error message (see section 7.3 on page 51).



Figure 18 Bridged systems: *o,o*- and *i,o*-norbornane.

Fused Systems. Fused ring systems as, *e.g.*, decalin (see Figure 19) 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.

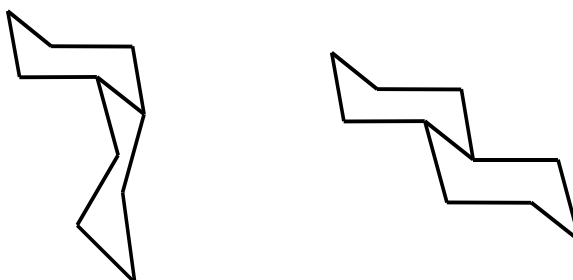


Figure 19 *Cis*- and *trans*-decalin.

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 20) depends on the conformation whereas the difference between the diequatorial and the equatorial/axial forms of 1,4-dimethylcyclohexane (see Figure 21) is a problem of stereochemistry. Thus, only in the second case there is a connection between stereochemistry and steric energy. In case of unspecified stereo centers CORINA tries to maximize the number of equatorial substituent positions.

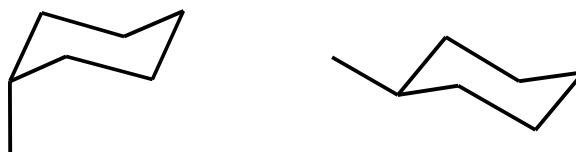


Figure 20 Equatorial and axial forms of methylcyclohexane.

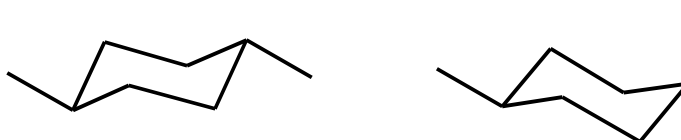


Figure 21 Diequatorial and equatorial/axial forms of 1,4-dimethylcyclohexane.

12.4.3 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 reasonable assumptions for missing stereo information. The driver option **-d sterger** forces CORINA to automatically identify stereo centers 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 18) are rejected. Stereo descriptors (parity markers) are generated and written to the output file. Figure 22 shows some examples.

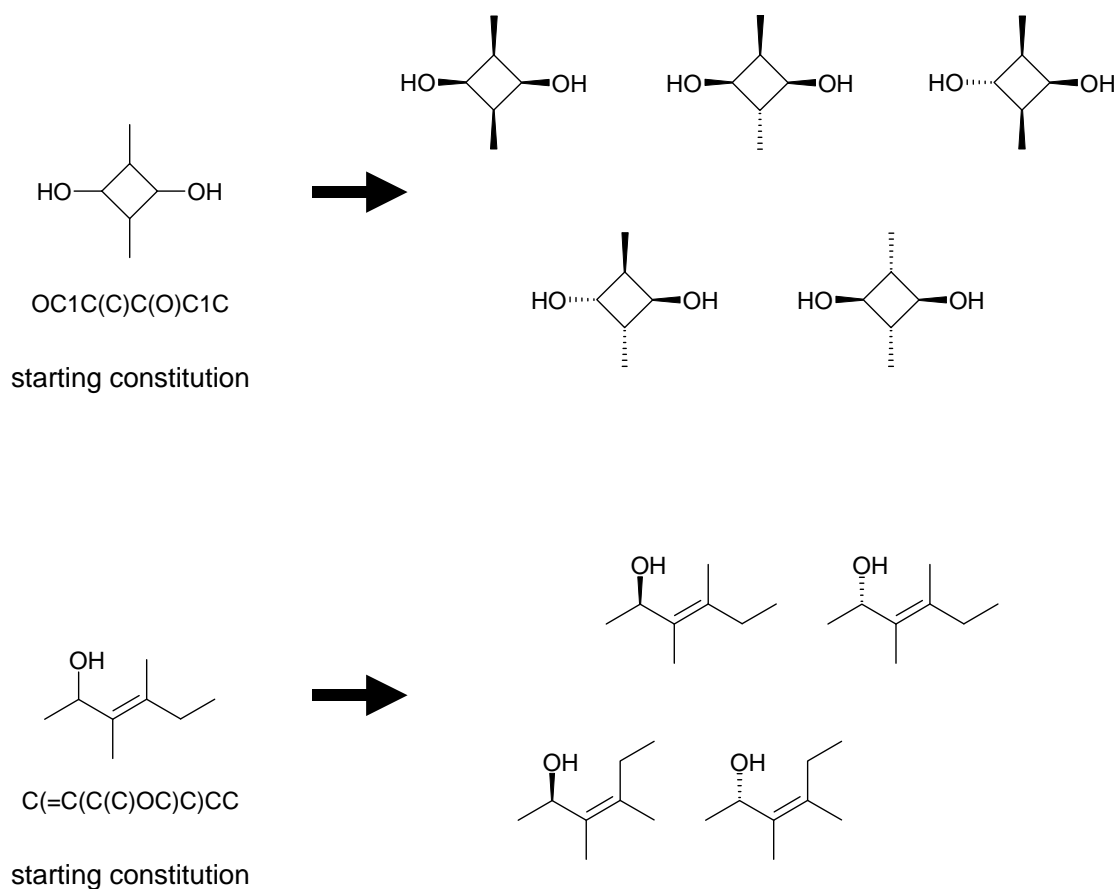


Figure 22 Generated configurations starting from one single initial constitution.

Double bonds in ring systems are only processed in ring systems having more than 10 ring atoms. In addition, *para*-substituted ring systems are identified as *pseudo* stereo centers in order to generate diequatorial and equatorial/axial substituted configurational isomers (see Figure 23).

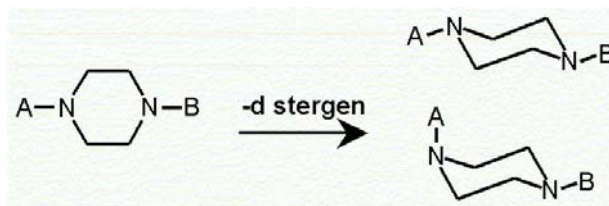


Figure 23 Generation of isomers of para-substituted rings.

By default (if the driver option **-d stergen** is set), a maximum number of four stereo centers are processed and a maximum number of 16 stereoisomers per molecule are generated.

As the number of possible isomers exponentially increases with the number of stereo centers (2^n , where n is the number of stereo centers 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 stereo centers which should be processed (**-d msc=<value>**, maximum number of stereo centers). Furthermore, the option **-d preserve** allows retaining the configuration at atoms that have a defined stereochemistry (*i.e.*, a stereo descriptor is given in the input file).

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

Note. If an input structure contains more than the specified number of stereo centers that should be permuted (or four stereo centers if the default values are used) only **one** isomer is generated. CORINA does not assess or rank the stereo centers in order to select only those for permutation that are the most reasonable. To warn the user the message "Maximum number of # stereo centers exceeded" is printed to the trace file.

Command line example. The following command line

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

generates a maximum of twenty stereoisomers per molecule (**msi=20**) by processing a maximum of six stereo centers (**msc=6**) and preserves stereo centers that have a defined stereochemistry (*i.e.*, a stereo descriptor is given in the input file, **preserve**). The stereoisomers are converted to 3D, implicitly given hydrogen atoms in the input structures are added and written to the output file (**wh**), small fragments (*e.g.*, counter ions in salts) are removed from the output file (**rs**) as well as structures which couldn't be converted by CORINA (**r2d**).

12.5 Conformational Analysis of Ring Systems for Flexible Search Purposes

The method of choice for flexible 3D database searches is to use compound databases that contain single low-energy conformations for each molecule and to solve the flexibility problem on the fly instead of storing conformational ensembles in the

database, a rather disk space consuming approach. A variety of methods such as the directed tweak algorithm exist for flexible searching [32]. These methods are efficient for chain portions of the molecules but run into problems when applied to ring systems (*vide infra*). One approach is to store 3D models with multiple ring conformations and to apply the flexible search only to the chain portions. CORINA supports this approach by providing methods for generating multiple ring conformations.

12.5.1 Generation of Multiple Ring Conformations

For ring systems consisting of up to nine ring atoms, CORINA performs internally a systematic conformational analysis in order to find the ring conformation with the lowest energy. Thus, a partial conformational analysis can easily be performed by outputting all conformations found in this way. For each side chains only one conformation is generated that remains unchanged unless there are no problems with non-bonded interactions caused by the different ring conformations. Therefore, the output consists of a set of 3D models having different ring and almost the same chain conformations. The philosophy behind this approach is that a conformational analysis for ring systems requires more program intelligence and that an analysis of the side chains can be performed by a much simpler postprocessor (*e.g.*, by systematically permutating all rotatable bonds) using 3D structures with different ring conformations as input. Furthermore, the number of reasonable ring geometries is often orders of magnitude smaller than the number of chain conformations.

The driver option **-d rc** forces CORINA to generate multiple ring conformations. The command **corina -n n=1 -d rc example.sdf out.sdf** produces a conformational ensemble for *trans*-decalin, the first record of the example file *example.sdf*. Figure 24 shows the five conformations obtained. The conformations are output in the order of increasing steric energy. The maximum number of conformations per molecule can be restricted by the driver option **-d mc=<value>** where <value> is the required number of conformations.



Figure 24 Conformations of *trans*-decalin.

12.5.2 Handling of Pyramidal Ring Nitrogen Atoms

For pyramidal ring nitrogen atoms having one exocyclic substituent, CORINA can generate conformations having both possible configurations at the nitrogen atoms (driver option **-d rc,flapn**). For 1,4-dimethyl-piperidine (SMILES: CN1CCCC1 without specified stereochemistry), four chair conformations with all combinations of the two substituents in equatorial and axial positions (see Figure 25) are generated by CORINA.

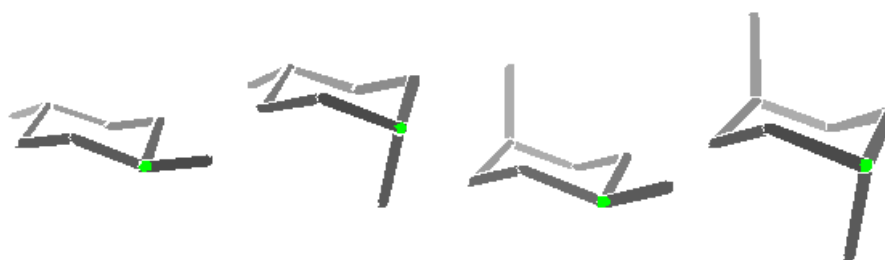


Figure 25 Chair conformations of 1,4-dimethyl-piperidin.

The driver option **-d noflapn** suppresses the flapping of any pyramidal nitrogen atoms only when stereoisomers should be generated (*i.e.*, in combination with the driver option **-d stergen**).

The driver option **-d planil** forces anilinic nitrogen atoms which are part of a ring system into a planar geometry.

12.5.3 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 bioactive 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=CC1) as an example (see Figure 26).

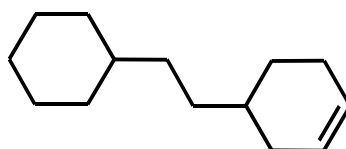


Figure 26 1-Cyclohexyl-2-cyclohex-3-enyl-ethane.

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 $14 \times 4 = 56$ conformations. The compact method (2) yields $\text{MAX}(14, 4) = 14$ conformations.

12.5.4 Multiple Ring Conformations in 3D Database Searches

An exhaustive study on the impact of using multiple ring conformations for 3D database searches was performed [33]. A short summary of the results are given in this section.

Directed tweak and ring flexibility. One approach to 3D database searching is to address torsion 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 pharmacophore. In principle, this technique is also applicable to ring systems but leads to several 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 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 stereo centers 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 bioactive 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 [16] containing 126,705 compounds 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 (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 antiarrhythmic agents [37] were 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, 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 relaxed 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%)
antiarrhythmic	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 27). In most cases, the hit rates converged to a value of 10 conformations. At that point, the size of the NCI_FLEX database was increased by a factor of 1.7 compared to the database with one conformer per record (NCI), a rather moderate requirement of additional resources.

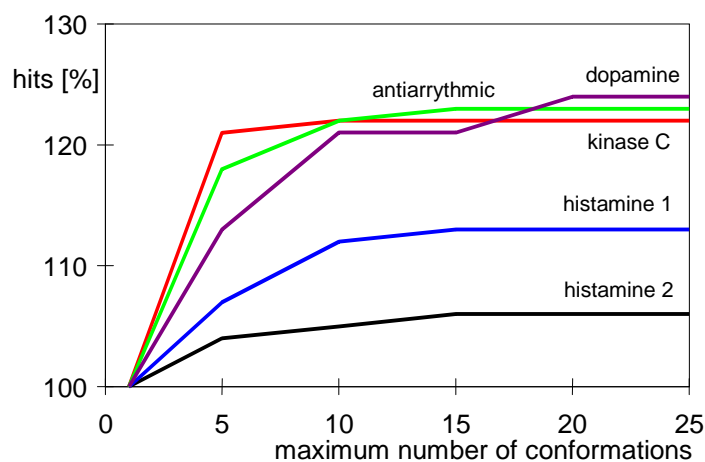


Figure 27 Hit rate vs. maximum number of conformations per molecule.

Recommendations. As a consequence of the study presented above, the following CORINA command line is recommended for the generation of 3D databases for pharmacophore search purposes: **corina -d wh,rs,r2d,rc,mc=10,flapn,sc <in> <out>**

13 Acknowledgements

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, reimplementations of CORINA in C programming language).

The authors 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). Furthermore, the authors 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, Erlangen (by Dr. Christof H. Schwab).

14 References

- [1] 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.
- [2] 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, 2004; pp. 151-212. d) Sadowski, J. 3D Structure Generation. In *Handbook of Chemoinformatics - From Data to Knowledge*. J. Gasteiger, J.; Engel, T., Eds., Wiley-VCH: Weinheim, 2003, pp. 231-261.
- [3] Hiller, C.; Gasteiger, J. Ein automatisierter Molekülbaukasten. In *Software-Entwicklung in der Chemie*, Vol 1; Gasteiger, J., Ed.; Springer: Berlin, 1987; pp 53-66.
- [4] Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic Generation of 3D Atomic Coordinates for Organic Molecules. *Tetrahedron Comp. Method.* **1990**, 3, 537-547.
- [5] Sadowski, J.; Rudolph, C.; Gasteiger, J. The Generation of 3D Models of Host-guest Complexes. *Anal. Chim. Acta* **1992**, 265, 233-241.
- [6] 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.
- [7] 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.
- [8] Schönberger, H.; Schwab, C.H.; Hirsch, A; J. Gasteiger, J. Molecular Modelling of Fullerene Dendrimers. *J. Mol. Model.* **2000**, 6, 379-395.

- [9] 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>.
- [10] 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>.
- [11] 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>.
- [12] 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.
- [13] 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>.
- [14] 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>.
- [15] 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>.
- [16] 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://cactus.nci.nih.gov/ncidb2/download.html>.
- [17] 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.
- [18] Klebe, G.; Mietzner, T. A Fast and Efficient Method to Generate Biologically Relevant Conformations. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 583-606.

- [19] 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.
- [20] Collaborative Computational Project, Number 4, 1994: The CCP4 Program Suite: Programs for Protein Crystallography, *Acta Cryst.* **1994**, *D50*, 760-763.
- [21] 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.
- [22] Gasteiger, J. et al. CTX Keyword Reference Manual. University of Erlangen-Nürnberg: 1995, unpublished results.
- [23] 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.
- [24] STERGEN Version 1.1 is developed and distributed by Molecular Networks GmbH, Erlangen, Germany (<http://www.molecular-networks.com>).
- [25] 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>.
- [26] 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.
- [27] 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>.
- [28] 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.
- [29] 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.
- [30] Biosym Technologies, Inc., San Diego, USA. The authors 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.

15 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
Henkestraße 91
91052 Erlangen, Germany
FAX: +49 9131 815 669

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 by email to support@molecular-networks.com These files will help us to analyze your problems. All data will be treated confidentially.