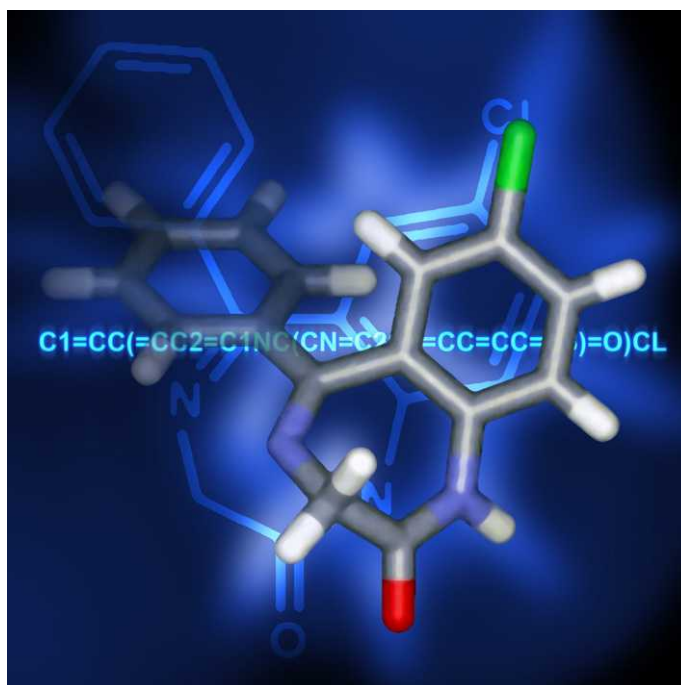


CORINA.*direct*

User Manual

Version 3.2



Molecular Networks GmbH Computerchemie
October 2008

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



Molecular Networks
Inspiring Chemical Discovery

Molecular Networks GmbH
Computerchemie
Henkestr. 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 © 2006 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.

(Author: Christof H. Schwab; Document version: 3.2-2008-10-01)

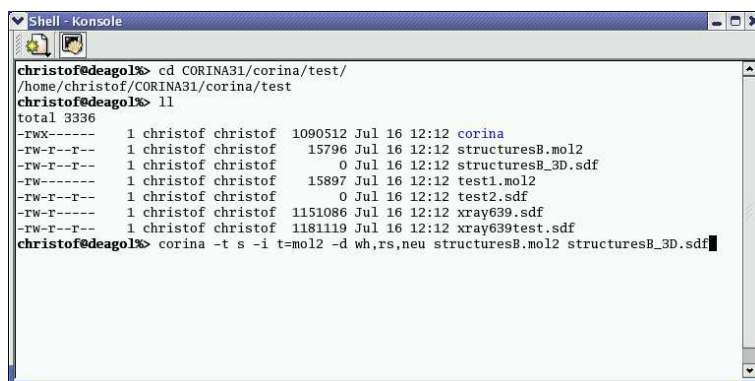
Contents

| | | |
|-------|--|----|
| 1 | Introducing CORINA and CORINA.direct | 1 |
| 2 | System Requirements | 3 |
| 3 | Program Installation | 4 |
| 3.1 | On-line Download from Molecular Networks' Web Server | 4 |
| 3.2 | New Installation | 4 |
| 3.2.1 | x86 Linux Systems | 4 |
| 3.2.2 | Microsoft Windows Platforms (2000/XP, win32) | 5 |
| 3.3 | Program Updates | 6 |
| 4 | Problems and Help! | 7 |
| 5 | Getting Started | 8 |
| 5.1 | Input Files | 8 |
| 5.2 | Output Files | 9 |
| 5.3 | General Options | 9 |
| 5.4 | Conformers & Isomers | 10 |
| 5.5 | 3D Structure Generation | 11 |
| 5.6 | 3D Structure Viewing | 12 |
| 5.7 | Editing Molecules | 13 |
| 5.8 | Project Files | 14 |
| 6 | Program Use | 15 |
| 6.1 | Options and Parameters | 15 |
| 6.1.1 | Main Menu | 15 |
| 6.1.2 | Button Bar | 16 |
| 6.1.3 | I/O Settings & Options Tab | 16 |
| 6.1.4 | General Options Tab | 20 |
| 6.1.5 | Conformers & Isomers Tab | 22 |

| | | |
|-------|--|----|
| 7 | Error Messages | 25 |
| 7.1 | General Errors | 25 |
| 7.2 | Input File Format Errors | 26 |
| 7.2.1 | MDL SDFFile | 26 |
| 7.2.2 | SMILES | 27 |
| 7.3 | Stereo Errors | 29 |
| 7.4 | Errors in the Generation of 3D Coordinates | 29 |
| 8 | Warning Messages | 31 |
| 8.1 | Warnings Related to Stereochemistry | 31 |
| 8.2 | Warnings in the Generation of 3D Coordinates | 31 |
| 9 | Acknowledgements | 33 |
| 10 | References | 34 |

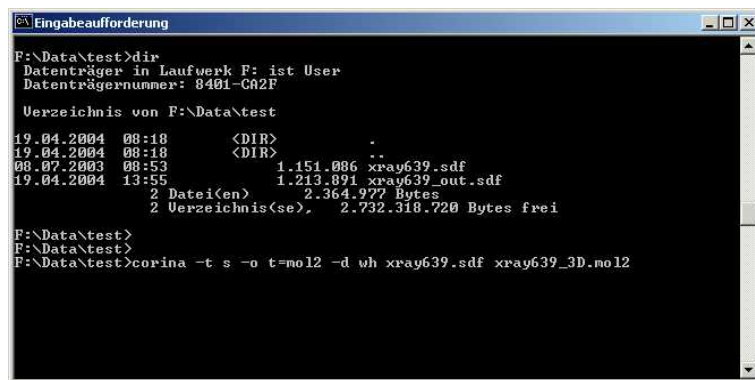
1 Introducing **CORINA** and **CORINA.direct**

The standard version of **CORINA** is a command line program for UNIX/Linux (executable file: *corina*) and Microsoft Windows platforms (executable file: *corina.exe*) [1]-[9]. Usually, the program has to be executed in a shell (e.g., *cs*h or *ba*sh on UNIX/Linux systems, see Figure 1) or at a Windows command prompt (see Figure 2).



```
Shell - Konsole
christof@deago1> cd CORINA31/corina/test/
/home/christof/CORINA31/corina/test
christof@deago1> 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@deago1> corina -t s -i t=mol2 -d wh,rs,neu structuresB.mol2 structuresB_3D.sdf
```

Figure 1 **CORINA** executed in a UNIX/Linux shell.



```
Eingabeaufforderung
F:\Data\test>dir
Datenträger in Laufwerk F: ist User
Datenträgernummer: 8401-C02F

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 2 **CORINA** executed in a Microsoft Windows/DOS command prompt.

A detailed description of the basic algorithm, the scope and limitations and the applications of **CORINA** are given in the program manual "3D Structure Generator **CORINA** – Generation of High-Quality Three-Dimensional Molecular Models" (PDF copy available for download at Molecular Networks' web site, <http://www.molecular-networks.com>).

The Java-based graphical user interface (GUI) **CORINA.direct** can be regarded as a graphical front-end to the standard version of **CORINA**. It triggers the standard version

of **CORINA** and executes the 3D structure generation process in the background of the system. Furthermore, **CORINA.direct** has the following advantages.

- Comfortable selection of input and output files and setting of program parameters
- Integrated molecule editor (JME) to draw molecules
- Integrated 3D structure viewer to browse generated 3D models
- Access to all options and parameters provided by the standard version of **CORINA**
- Full functionality for batch mode execution or usage in a web-based environment
- Execution in the background of the operating system, i.e., GUI can be exited if large files are processed without interrupting the **CORINA** run
- Typical robust and reliable **CORINA** performance

CORINA.direct provides a main menu bar with the menus **Project**, **Options** and **Info** and three tabs **I/O Settings & Options**, **General Options** and **Conformers & Isomers** (see Figure 3).

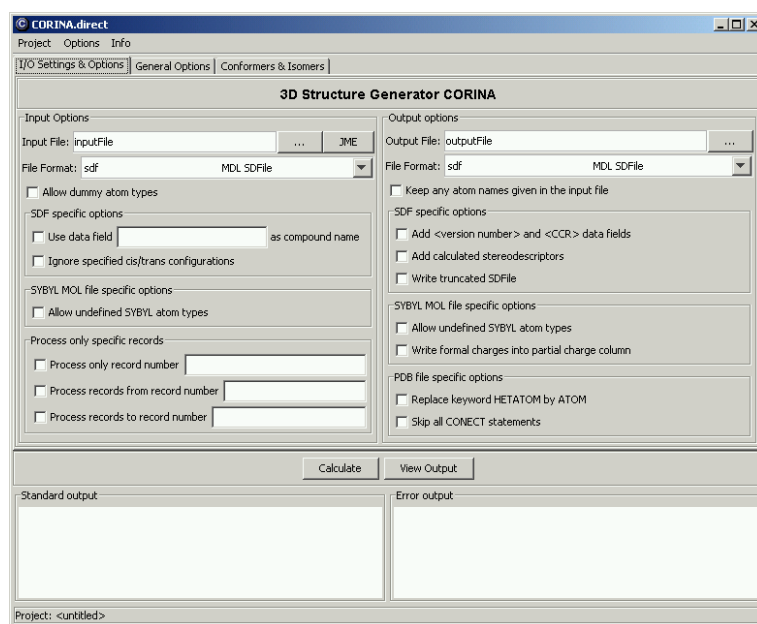


Figure 3 Graphical user interface of **CORINA.direct**.

A brief introduction to the usage of **CORINA.direct** is given in section 5 Getting Started. A detailed description of all options and parameters is given in section 6 Program Use.

2 System Requirements

The Java-based graphical user interface **CORINA.direct** runs on Microsoft® Windows® (2000, XP, it is recommended to have installed the latest service pack of Microsoft® Windows®) and on x86 Linux® workstations (kernel 2.4 or higher, distributions by SUSE and RedHat).

CORINA.direct requires access to the standard version of **CORINA** (versions 3.0, 3.1 or 3.2). The installation process of the standard version of **CORINA** is described in detail in the **CORINA** user manual "3D Structure Generator CORINA – Generation of High-Quality Three-Dimensional Molecular Models".

CORINA.direct requires a Java Virtual Machine installation and performs best with Java virtual machine 1.5.0 (1.5.0 Java 2 Runtime Environment, Standard Edition, <http://www.java.com>).

If you have any questions concerning the installation and usage of the **CORINA.direct** please contact us at info@molecular-networks.com (see also section 4 "Problems and Help!" on page 7).

3 Program Installation

3.1 On-line Download from Molecular Networks' Web Server

CORINA.direct is available for download via the Internet on the web server of Molecular Networks. At

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

an account can be created that provides access to evaluation copies, licensed or free software, program and user manuals, example files and tutorials of the 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 uncompressed easily with standard software tools for file compression and archiving, such as gzip (on Linux platforms) or WinZip and FileZip (on Windows or the system component of XP).

3.2 New Installation

The following chapters describe the procedures to install **CORINA.direct** on the various supported operating systems. Please ensure that the Java Virtual Machine version 1.5 or higher is installed before you proceed with the installation.

3.2.1 x86 Linux Systems

- 1) Install the standard version of **CORINA** following the installation instructions given in the program manual of **CORINA** "3D Structure Generator CORINA – Generation of High-Quality Three-Dimensional Molecular Models".
- 2) Uncompress the gzipped file *corina_direct_Linux2.4.gz*.
- 3) Copy the resulting Java archive file *CORINA.direct.jar* into the same sub directory in which the standard (command line) version of **CORINA** is located, e.g. */usr/local/packages/corina/corina.lnx*
- 4) Type the command *java -jar CORINA.direct.jar* at the prompt of a command shell. The **Settings** dialog box of **CORINA.direct** appears on your screen (see Figure 4).

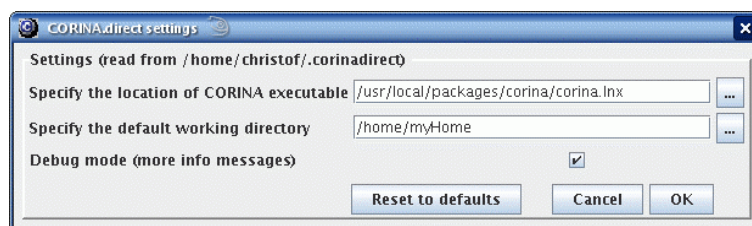


Figure 4 **Settings** dialog box of **CORINA.direct** on Linux systems.

- 5) Enter the following paths and commands in the fields of the dialog box.

| Field | Entry |
|---|--|
| Specify the location of CORINA executable | e.g., <i>/home/<name>/my_structures</i> |
| Specify the default working directory | e.g., <i>/usr/local/packages/corina/corina.lnx</i> |

- 6) Click on **OK**. **CORINA.direct** now starts displaying the **I/O Settings & Options** tab on your screen (see Figure 3).
- 7) In order to create a desktop link (icon) to start **CORINA.direct** directly from your desktop, follow the instructions given in the help menu of your operating system.

3.2.2 Microsoft Windows Platforms (2000/XP, win32)

- 1) Install the standard version of **CORINA** following the installation instructions given in the program manual of **CORINA** "3D Structure Generator CORINA – Generation of High-Quality Three-Dimensional Molecular Models".
- 2) Uncompress the Zip file *corina_direct_win32.zip*.
- 3) Copy the resulting executable *CORINA.direct.exe* into the same sub directory in which the standard (command line) version of **CORINA** is located, e.g. *C:\programs\corina*.
- 4) Double click the file *CORINA.direct.exe*. The **Settings** dialog box of **CORINA.direct** appears on your screen (see Figure 5).

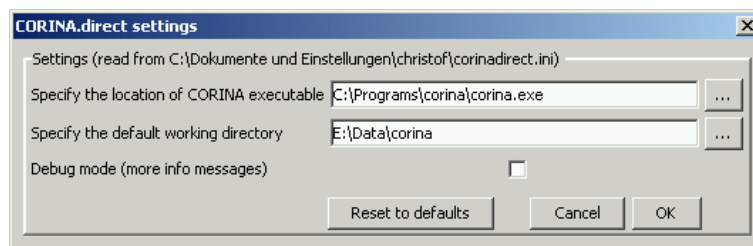


Figure 5 **Settings** dialog box of **CORINA.direct** on MS Windows platforms.

- 5) Enter the following paths and commands in the fields of the dialog box.

| Field | Entry |
|---|--|
| Specify the location of CORINA executable | e.g., <i>C:\Programs\corina\corina.exe</i> |
| Specify the default working directory | e.g., <i>E:\Data\corina</i> |

- 6) Click on **OK**. **CORINA.direct** now starts displaying the **I/O Settings & Options** tab (see Figure 3).
- 7) In order to create a desktop link (icon) to start **CORINA.direct** directly from your desktop, follow the instructions given in the help menu of your operating system.

3.3 Program Updates

- 1) Before installing the a version, please copy the old executable and configuration files to a new directory, e.g. *corinaVV* (*VV = old-version-number*, e.g., *corina30*).
- 2) According to the hardware platform install the new version following the instructions given in section 3.2 on page 4.

4 Problems and Help!

If you have any difficulties with the installation or usage of **CORINA** or **CORINA.direct** or if you encounter any problems when running **CORINA.direct**, 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 +49-9131-815669.

Please send the version and serial number of CORINA, the input and output file and the trace file *corina.trc* by email or fax. These files will help us to analyze the problem; if your system displays any error messages, please add them to your report. Thank you!

5 Getting Started

This section describes the usage and the functionalities of **CORINA.direct** using the sample MDL SDF file *example.sdf* that is provided together with this user manual. The sample file contains the structure information of three molecules in MDL SDF file format [10], which is the default file format for input and output of **CORINA**.

Please copy the file *example.sdf* into your default working directory, i.e., into the directory that you have specified as your working directory during the installation in the **Settings** dialog box (see section 3 "Program Installation" on page 4).

Start **CORINA.direct** (desktop icon or shell command) and the following window will appear on your screen.

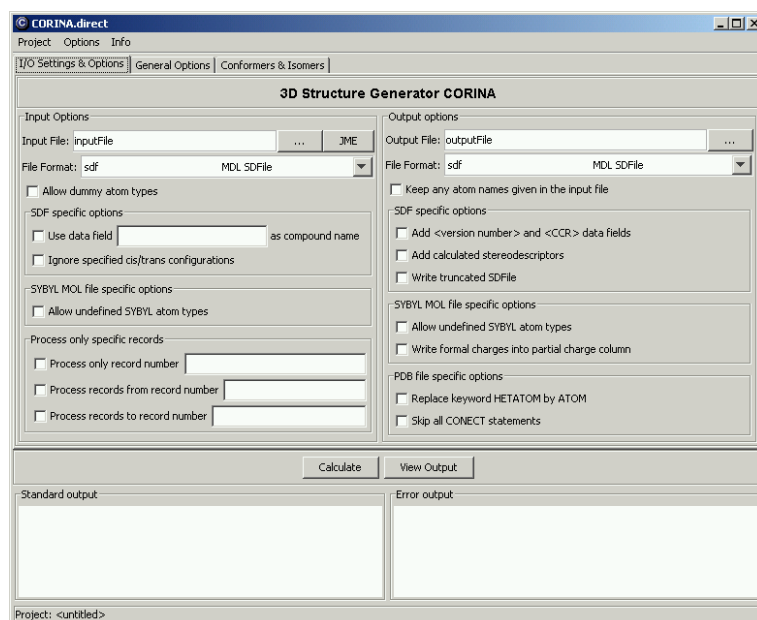


Figure 6 Starting **CORINA.direct**.

5.1 Input Files

In order to load the structure input file *example.sdf*, click the button ... right next to **Input File** field. The **Select Input File** dialog box opens in which the file *example.sdf* can be selected (see Figure 7).

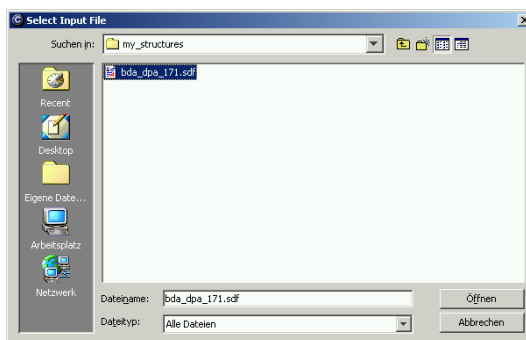


Figure 7 Selecting a structure input file.

The input file name is now displayed in the field **Input File**.

In the drop down menu **File Format** below the field **Input File**, a variety of different input file formats can be chosen. The file *example.sdf* is a structure file in MDL SDF file format, the default file format of **CORINA** and **CORINA.direct**.

Note. **CORINA** and **CORINA.direct** do not automatically recognize the file format of an input structure file. This has to be defined manually. The default file format is MDL SDF file.

5.2 Output Files

CORINA.direct automatically suggests an output file name by adding "_3D" to the base name of the input file, e.g., *example_3D.sdf*. This suggested file name is displayed in the field **Output File**. Furthermore, the output file is written in the default working directory (as specified in the **Settings**). However, the file name (as well as the directory in which the output file should be written) can be changed by clicking the ... button right to the field **Output File**. The **Select Output File** dialog box opens in which a user-defined output file can be selected or newly created.

By default, an MDL SDF file is written out. Other file formats can be chosen in the drop down menu **File Format**.

Note. The file extension of the output file (e.g., .pdb for PDB output files) has to be changed manually in the field **Output File**.

All further options that are available on this tab are described in detail in section 6 "Program Use" on page 15.

5.3 General Options

Change to the tab **General Options** (see Figure 8). In this tab, all general options to control the 3D structure generation process are handled. By default, the following three

options (check boxes) are activated:

- **Write added hydrogen atoms** in order to add implicitly given hydrogen atoms to the 3D output file.
- **Remove small fragments** in order to remove small fragments in a record, e.g., counter ions in salts.
- **Write not-converted molecules "as is"** in order to pass structures into the output file that couldn't be converted into 3D.

All further options that are available in this tab are described in detail in section 6 "Program Use" on page 15.

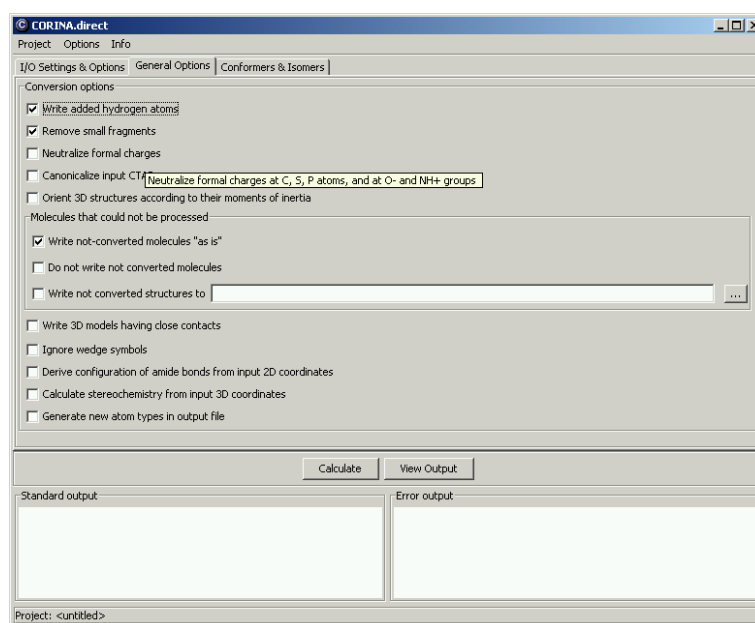


Figure 8 The tab **General Options** (figure shows tool tip help).

5.4 Conformers & Isomers

CORINA can generate multiple ring conformation (for ring systems with up to nine ring atoms) as well as stereoisomeric compounds using tetrahedral chiral centers and *cis/trans* double bonds. By default, this generation process is switched off and can be activated and controlled in this tab (see Figure 9).

Note. The generation of conformational ensemble and stereoisomers might result in a large number of output 3D models for one input structure.

All further options that are available in this tab are described in detail in section 6 "Program Use" on page 15.

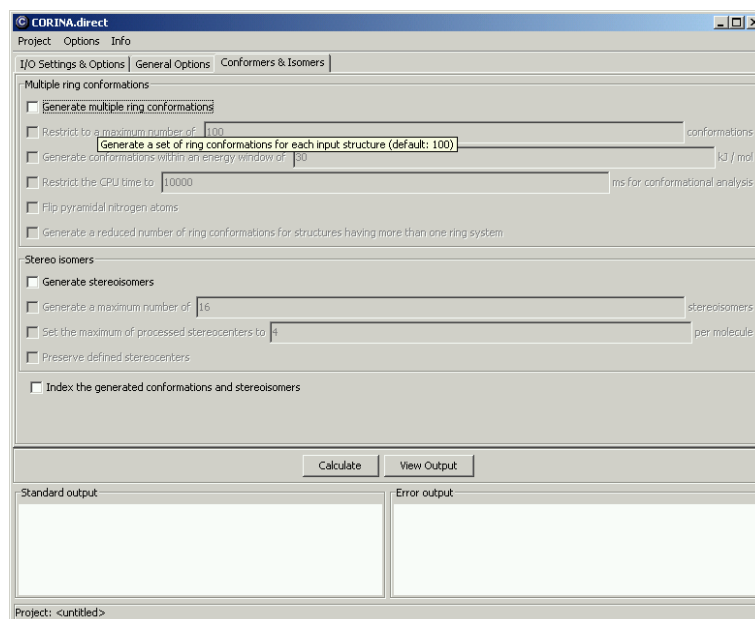


Figure 9 The tab **Conformers & Isomers** (figure shows tool tip help).

5.5 3D Structure Generation

In order to start the 3D structure generation process press the **Calculate** button. In the **Standard Out** console below the **Calculate** button, information about the CORINA run, such as used parameters, defined options, computation times, details on the 3D structure generation process as well as warning and error messages (see Figure 10 and section 7 "Error Messages" on page 25 and section 8 "Warning Messages" on page 31) is printed (the standard version of CORINA—command line version—writes this information into the file *corina.trc*).

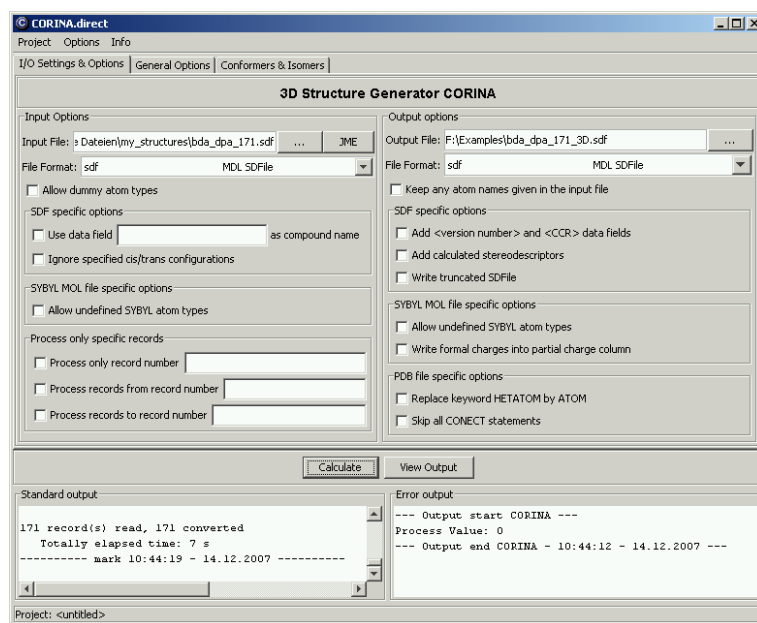


Figure 10 Running CORINA with CORINA.direct.

5.6 3D Structure Viewing

The output file *example_3D.sdf* contains the generated 3D coordinates for each molecule. The 2D coordinates of the atoms provided in the input file are replaced by their 3D coordinates. However, any other original structure and data information given in the input file *example.sdf* is preserved in the output file.

The generated 3D structures can be displayed in a 3D structure viewer by clicking the **View Output** button (see Figure 11).

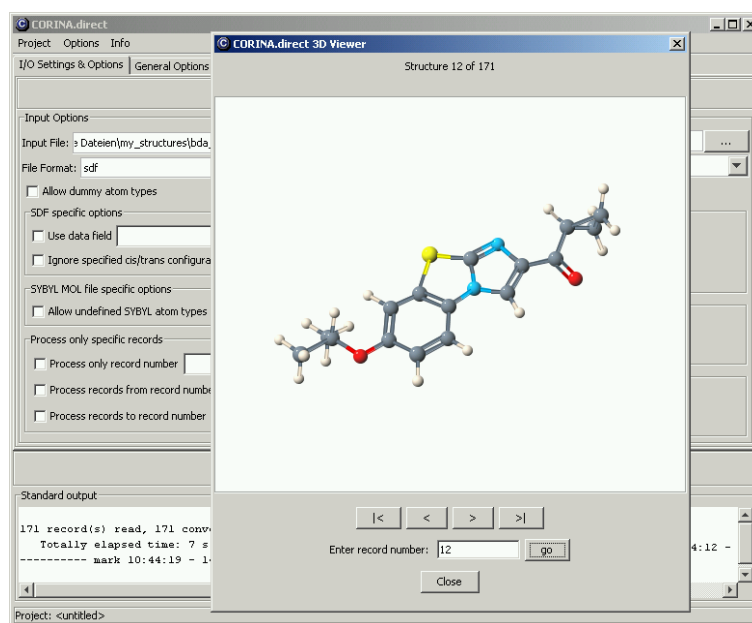


Figure 11 3D structure browser to display output files.

5.7 Editing Molecules

By default, **CORINA** expects structure information in standard file formats, such as MDL SDF files, SMILES or SYBYL MOL/MOL2 files. However, **CORINA.direct** has an integrated version of the molecule editor JME Classic (Java Molecule Editor) that can be used to edit (to draw) a molecule, to submit the sketch directly to 3D structure generation and to save the 3D molecular model in a standard file format.

In order to start JME Classic press the button **JME** right to the field **Input File**. Draw a molecule and press the button **OK** of the JME Classic window (see Figure 12).

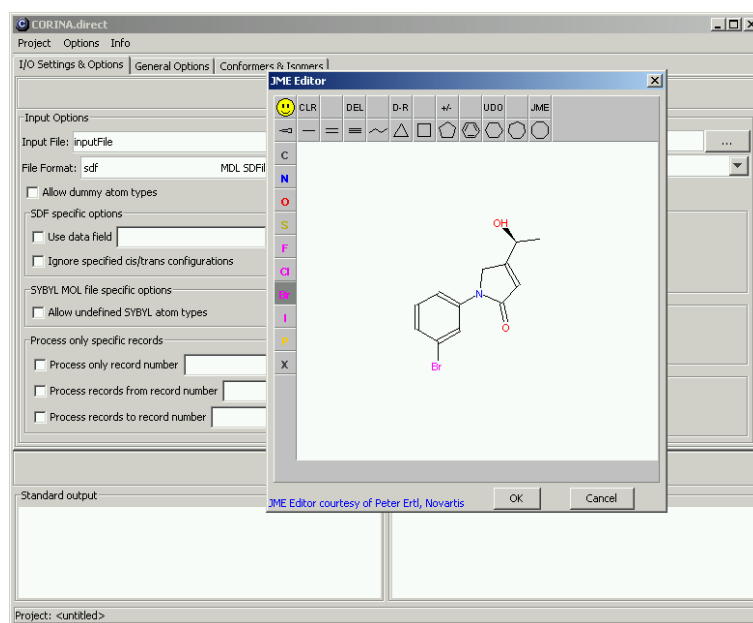


Figure 12 Editing a molecule with JME Classic.

Select an output file and press the button **Calculate**. The generated 3D model can be visualized using the 3D structure browser of **CORINA.direct** (see Figure 11).

5.8 Project Files

Project files store all parameters that are used during a **CORINA.direct** session, e.g., input and output files and all used options and control parameters. Project files can be reloaded in order to restore all parameters in a later session of **CORINA.direct**.

A project file is saved using the menu **Project** of the main menu bar by the menu item **Save Project**. A dialog box appears where the path and the file name of the project file can be specified (default file extension *wco*).

Project files can be reloaded using the menu item **Load Project** in the menu **Project** of the main menu bar.

In order to reset all parameters to the default settings create a new project using the menu item **New Project** in the menu **Project** of the main menu bar.

6 Program Use

6.1 Options and Parameters

In the following section, all program options available in the tabs of **CORINA.direct** (check boxes, drop down menus, fields) are described. More details about CORINA can be found in the program manual "3D Structure Generator CORINA – Generation of High-Quality Three-Dimensional Molecular Models".

6.1.1 Main Menu

In the main menu of **CORINA.direct** three different menus are available, i.e., **Project**, **Options** and **Info** (see Figure 6).

Project

| | |
|--------------|--|
| New Project | Close the current project and start with a new project. CORINA.direct can save project files that contain file names and all options and parameters used in a specific project (file extension: <i>wco</i>). Project files can be reloaded in a following session in order to restore all parameters used in the previous one. Please note: Selecting New Project resets all options and parameters to the default (factory) settings. |
| Load Project | Load an existing project file. All options, parameters, and file names are set according to the information contained in the selected project file. A file selection dialog box is opened when choosing this menu item. |
| Save Project | Save current project into project file. A file selection dialog box is opened when choosing this menu item and all options, parameters, and file names can be saved in a project file. The file name can be chosen by the user. |
| Exit | Exit the program. If the current settings were not saved in a project file the user is asked whether they should be saved in a project file before CORINA.direct exits. |

Options

| | |
|----------|--|
| Settings | Edit the system settings of CORINA.direct Work-Dir, Shell Command, and Path CORINA (see also section 3 on page 4 of this manual). |
|----------|--|

Info

| | |
|------|---|
| Info | Some additional information on CORINA.direct . |
|------|---|

6.1.2 Button Bar (see Figure 6)

| | |
|-------------|--|
| Calculate | Start the generation of 3D structures with the current settings and parameters. Before the generation of 3D structures is started CORINA.direct checks whether all input options and parameters are set to appropriate values. If any problem is encountered a message is printed to the console (see below). |
| View Output | Display the generated 3D structures in an external 3D structure viewer. The generated 3D structures can be displayed in an external 3D structure viewer by pressing the View Output button (see Figure 11 on page 13 of this manual). |

6.1.3 I/O Settings & Options Tab

The **I/O Settings & Option** tab is the first tab of **CORINA.direct** (see Figure 6).

Input file options (middle right, see Figure 6)

| | |
|------------|---|
| Input File | Select an input file. A file selection box will be opened started in the specified working directory |
| JME | Start the molecule editor JME Classic (see Figure 12) |
| File Type | Select an input file type. Allowed file formats are: sdf MDL SDFfile [10] (default) rdf MDL RDFfile [10] smiles SMILES linear code [11] mol SYBYL MOL file [12] mol2 SYBYL MOL2 file [12] |

ctx Gasteiger ClearText [13]

Allow dummy atom types Allow the interpretation of dummy atom types in SYBYL MOL/MOL2 ("Du") and SMILES linear code ([*]) input files.

SDF specific options

Use data field <field> as compound name Copy the data item <field> in the input MDL SDF file to the compound name field (1st line) in MDL SDF input files.

Ignore specified cis/trans configurations Ignore *cis/trans* configuration of double bonds in MDL SDF input files.

SYBYL MOL file specific options

Allow undefined SYBYL atom types 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 **Allow undefined SYBYL atom types**).

Record number (lower right, see Figure 6)

Process only record number <field> Process only record number <field> in a multi-record file. The value of <field> must be of type integer.

Process all records from record number <field> Process all records from record number <field> in a multi-record file. The value of <field> must be of type integer.

Process all records to record number <field> Process all records to record number <field> in a multi-record file. The value of <field> must be of type integer.

Output file (middle right, see Figure 6)

| | |
|--|--|
| Output File | Select an output file. A file selection box will be opened started in the specified working directory |
| File Type | Set the output file type. Allowed formats are: sdf MDL SDFfile [10] (default) rdf MDL RDFfile [10] mol SYBYL MOL file [12] mol2 SYBYL MOL2 file [12] pdb Brookhaven Protein Data Bank [14] cif Crystallographic Information File format [15] ctx Gasteiger ClearText [13] |
| Keep any atom name given in the input file | 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. |

SDF specific options

| | |
|---|---|
| Add version number and CCR to output file | 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. |
| Add calculated stereodescriptors | Write out the atom stereo parity flags that have been calculated by CORINA for missing stereodescriptors to MDL SDF (RDF) output file. If stereo information is missing in the input file CORINA assumes reasonable stereodescriptors following some implemented rules. 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. |

| | |
|-------------------------|---|
| Write truncated SDFFile | Write out a compact, truncated MDL SDFFile. This options restricts the number of output fields in the atom lines of the atom block in MDL SDFiles (RDFiles) to the x-, y-, and z-coordinates, the atom type (symbol), the mass difference, the atom charge, and the stereochemical atom parity (columns 1 through 7 of the atom block). All other fields in the atom lines are omitted, since they contain no data 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. |
|-------------------------|---|

SYBYL MOL file specific options

| | |
|----------------------------------|--|
| Allow undefined SYBYL atom types | Allow extra/undefined elements in SYBYL MOL2 output files. If the input option Allow undefined SYBYL atom types (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. |
|----------------------------------|--|

| | |
|--|--|
| Write formal atom charges into partial charge column | Write formal atom charges into the partial charge column of SYBYL MOL/MOL2 output files. This option forces CORINA to write formal atom charges which are given in the input structure to the charge column (column 9) in the corresponding @<TRIPOS>ATOM data lines of the SYBYL MOL2 output file. In addition, the keyword USER_CHARGES is set in the field @<TRIPOS>MOLECULE. |
|--|--|

PDB file specific options

| | |
|---------------------------------|---|
| Replace keyword HETATOM by ATOM | Write the keyword ATOM instead of HETATM in PDB output files. |
|---------------------------------|---|

| | |
|----------------------------|---|
| Skip all CONECT statements | Discard all CONECT statements and keywords in PDB output files. |
|----------------------------|---|

6.1.4 General Options Tab

The **General Option** tab is the second tab of **CORINA.direct** (see Figure 8).

| | |
|---|---|
| Write added hydrogen atoms | 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. |
| Remove small fragments | Remove small fragments. Remove all but the largest fragments from multi-component records (e.g., counter-ions in salts, solvent molecules). |
| Neutralize formal charges | 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 option Remove small fragments (see above) in order to remove counter ions from salts. |
| Canonicalize input CTAB | 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.e., for different atom numberings slightly different conformations may be obtained. This option prevents any atom numbering dependency. |
| Orient 3D structures according to the moments of inertia. | Orient the 3D structures according to the moments of inertia. This option is useful when the structure is directly forwarded to a graphical viewer. The molecule then appears more often in an orientation that shows as much of it as possible on one sight. |

| | |
|---|---|
| Write not converted molecules "as is" | Write out the original records also for failure cases (where CORINA fails to generate a 3D model) If the input and the output file type are both set to MDL SDFfile (default), CORINA writes the original structure record to the output file in cases where no 3D structure could be generated (see also option below: Do not write not converted molecules). |
| Do not write not converted molecules | Remove 2D records from the output file. 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 suboption prevents the output of 2D structures (see also option above: Write not converted molecules "as is"). |
| Write not-converted molecules to <field> | Write failed structures to the output file <field>. Structures that couldn't be converted by CORINA can be written to a second output file/errorfile with the file name specified in <field>. The errorfile specified in <field> will have the same file format as the input file and currently only SDF and SMILES format are supported. |
| Write models having close contacts | 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. |
| Ignore wedge symbols | Override wedge symbols. Some input file formats (SDF, CTX) support both atom stereodescriptors and wedged bonds in the 2D drawings of the molecules. When these descriptors differ for one and the same stereocenter, CORINA by default overrides the atom descriptor and uses the wedged bond symbols for calculating the stereochemistry. This option allows to override the wedged bonds and to use the atomic descriptors instead. In any case, CORINA writes an error message when encountering different types of stereodescriptors for one and the same stereocenter (see section 7.3 on page 29). |
| Derive configuration of amide bonds from input 2D coordinates | 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 <i>cis</i> or <i>trans</i> configurations for amide bonds by drawing them in the 2D diagram in the input file, accordingly. |

| | |
|---|---|
| Calculate stereochemistry from input 2D coordinates | Force stereodescriptors from the 3D structure. If there is a discrepancy between the stereodescriptors and the 3D structure in the input file, CORINA takes the configuration given in the 3D coordinates (default: usage of stereodescriptors). |
| Generate new atom types in output file | 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. |

6.1.5 Conformers & Isomers Tab

The **Conformers & Isomers** tab is the third tab of **CORINA.direct** (see Figure 9).

Generation of multiple ring conformations

| | |
|---|---|
| Generate multiple ring conformations | Generate multiple ring conformations. This option forces the conformational analysis module for small and medium sized ring systems of CORINA to output multiple ring conformations. The option does not support records which consist of more than one fragment (e.g., salts). Work-around: Combine with the option Remove small fragments (see above). The conformations are written in the order of increasing steric energy. |
| Restrict to a maximum number of <field> conformations | Set the maximum number of generated conformations to <value>. If the option Generate multiple ring conformations is set (see above), CORINA generates by default a maximum number of 100 conformations per molecule. This option restricts the number of output conformations to the value input in <field>. The value of <field> must be of type integer. |

| | |
|--|--|
| Generate conformations within an energy window of <field> kJ/mol | Set an energy window ΔE of <field> kJ/Mol for the ring conformations. This option forces CORINA to output only those conformations which have an energy not higher than the value (in kJ/mol) that is input in <field> with respect to the lowest-energy conformation. The value of <field> must be of type integer or float. The default value is 30 kJ/mol. |
| Restrict the CPU time to <field> ms for conformational analysis | Restrict the computation time for the ring conformation analysis to <field> milliseconds. 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 timeout (in milliseconds) that is input in <field> and to output all conformations obtained so far. |
| Flip pyramidal nitrogen atoms | Flip ring nitrogen atoms to generate multiple ring conformations. This option allows pyramidal ring nitrogen atoms that have one exocyclic neighbor to invert their configuration in order to obtain all conformations. |
| Generate a reduced number of conformations for structures having more than one ring system | Generate ring conformations simultaneously. 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 changing the conformations for the different ring systems. |

Generation of stereoisomeric compounds

| | |
|------------------------|--|
| Generate stereoisomers | 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 stereodescriptors are given in the input or not. By default, a maximum number of four stereocenters are processed and a maximum number of 16 stereoisomeric compounds are generated (see below). |
|------------------------|--|

| | |
|--|--|
| <p>Generate a maximum number of <field> stereoisomers</p> | <p>Set the maximum number of processed stereocenters per molecule to the value input in <field>. By default, the stereoisomer-generating module of CORINA processes a maximum of four stereocenters of an input structure. This option forces CORINA to process the specified number (in <field>) of stereocenters in order to restrict or to increase the number of output isomers.</p> |
| <p>Set the maximum number of processed stereocenters to <field> per molecule</p> | <p>Set the maximum number of generated stereoisomers per molecule to the value input in <field>. 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 (in <field>) of stereoisomers in order to restrict or to increase the number of output isomers.</p> |
| <p>Preserve defined stereocenters</p> | <p>Preserve defined stereocenters. 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.</p> |
| <p>Index the generated conformations and stereoisomers</p> | <p>Number the generated conformations or isomers consecutively by adding a counter to their names. If isomers or conformations are generated this option may be used to consecutively number the different geometries by adding a counter to the compound name (<i>compound_name_its001, ...</i>).</p> |

7 Error Messages

7.1 General Errors

ERROR license(): License time out.

The license time has expired. You are not allowed to use the program any more.

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

The program is unable to open the specified file (path in parentheses).

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

The trace file *corina.trc* can't 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*.

Error opening the specified input file.

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

Error opening the specified output file.

ERROR corina(): Too many non-option parameters.

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

ERROR corina(): Option invalid in version 1.6 and later.

An old-fashioned version 1.5 option was encountered (see section 6 on page 15).

ERROR corina(): File type not allowed.

A file type not allowed for input and/or output files was specified encountered (see section 6 on page 15).

ERROR restrict(): FlexX restrictions violated.

The restrictions of the FlexX interface are violated or not fulfilled.

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.

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

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

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

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

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

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

ERROR rmdl(): Unknown element symbol.

ERROR rmdl(): Bond atoms out of range.

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

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

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

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

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

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

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

Errors with missing columns:

ERROR rmdl(): No atomic charge column specified.

ERROR rmdl(): No atom stereo column specified.

ERROR rmdl(): No mass difference column specified.

ERROR rmdl(): Too many additional lines.

Only a limited number of additional lines in MDL SDFFile is allowed (5,000).

ERROR rmdl(): Line too long.

Only 80 characters per line are allowed in MDL SDFiles.

Errors in charge, radical, and attachment point lines:

ERROR rmdl(): CHG atom out of range.

ERROR rmdl(): RAD atom out of range.

ERROR rmdl(): APO atom out of range.

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 () brackets.

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

A not interpretable character was encountered.

ERROR smilesct(): Unknown element.

An unknown element symbol was found.

ERROR smilesct(): Too many ('s.

Missing a closing) bracket.

ERROR smilesct(): Closing ring.

No suited ring bond label.

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

The standard valency of an atom is exceeded.

ERROR smilesct(): Too many ['s.

Missing a closing] bracket.

ERROR smilesct(): Reading bracketed atom.

The atom type in square brackets is incorrect.

- ERROR smilesct(): Generating Kekule structure for an aromatic system.
A Kekulé structure with alternating single and double bonds cannot be found for an aromatic system.
- ERROR smilesct(): Duplicate slash at double bond.
Two bonds at one atom of a stereo double bond are assigned a slash stereodescriptor. Only one slash is allowed.
- ERROR smilesct(): Not all rings closed.
Not for all ring bonds was a second label defined.
- ERROR smilesct(): Reading atomic charge qualifier.
Incorrect charge specification.
- ERROR smilesct(): Too many {'s.
Missing a closing } bracket.
- ERROR smilesct(): Unknown stereo class.
An unknown stereo class was specified.
- ERROR smilesct(): No organic element. Use square brackets.
Inorganic elements must be written in square brackets.
- ERROR smilesct(): Ring closure label not following immediately the atomic symbol.
Ring bond labels must follow immediately after the atomic symbol.
- ERROR smilesct(): Conflicting ring closure bond types.
The type of the ring closure bond was defined twice with differing values.
- ERROR smilesct(): Chiral center has wrong connectivity.
A tetrahedral center with less than four neighbors was encountered. A common error is to forget to specify an eventual implicit hydrogen neighbor of the stereocenter within the square brackets.
- ERROR smilesct(): Expecting atomic symbol.
The first character inside square brackets must belong to an atomic symbol.
- ERROR smilesct(): Illegal valence state.
An atom in an illegal valence state was encountered.
- ERROR smilesct(): Inconsistent EZ specification.
The EZ configuration of a double bond was defined redundantly with different values.

7.3 Stereo Errors

- ERROR initster(): Resetting a trans double bond in a small ring: #-#.
A trans double bond in a ring with less than eight atoms was encountered. Since this is geometrically impossible it can be corrected by the program.
- ERROR clcster2(): Ambiguous configuration of ligands.
The specification of a chiral center is ambiguous in the 2D structure diagram.
- ERROR clcster2(): The ligands at atom # don't span up a volume.
The specification of a chiral center is ambiguous in the 2D structure diagram.
- ERROR clcster2(): Insufficient number of axes.
For a higher coordinated center (5 or 6 ligand atoms) are not enough axes specified in the 2D structure diagram.
- ERROR allcis(): No suited stereodescriptors for bridge #- #.
Incorrect coding of the stereochemistry of some bridgehead atoms.

7.4 Errors in the Generation of 3D Coordinates

- ERROR corina(): Input structure incorrect. No 3D generation.
- ERROR gen3d(): Unable to build a 3D structure.
No 3D structure can be generated and no structure is written to the output. A fatal error occurred during the generation process.
- ERROR quality(): Some internal coordinates are heavily distorted.
A fatal error occurred during the 3D generation.
- ERROR quality(): Violated cis/trans bond.
A stereo double bond changed its configuration during the geometry optimization.
- ERROR quality(): Violated stereo atom.
A chiral atom changed its configuration during the geometry optimization.
- ERROR quality(): Bad contacts detected.
Unfavourable non-bonded interactions (crowded atoms) detected.

- ERROR gen3d(): 3D structure didn't pass the quality check.
The generated 3D structure is insufficient and removed from the output file.
- ERROR gen3d(): No generation of multiple conformations for multiple fragments.
For records containing multiple fragments the output of multiple ring conformations (-d rc) is not supported. Work-around: remove all but the largest fragments (-d rc,rs).
- ERROR alltempl(): Missing ring template.
For one ore more rings no template is found in the file *rings.ctx*.
- ERROR ringfrag(): Fragment contains ring(s) > 8.
CORINA handles only rings up to a size of eight members by templates. (Rings with more then eight atoms are reduced to a secondary structure. Only when this reduction does not result in ring sizes < 9 the above message is written.)
- ERROR ringfrag(): Can't combine the templates.
The available ring templates cannot be combined to a ring conformation.
- ERROR bigring(): Unable to build the secondary structure.
The secondary structure of a large ring system with ring sizes > 9 is too complex or contains rings > 9.
- ERROR procfraq(): Can't process small ring system.
A small ring system cannot be translated into a 3D structure.
- ERROR bigsys(): Unable to process all fragments.
Some of the fragments of a large ring system cannot be translated into a 3D structure.

8 Warning Messages

8.1 Warnings Related to Stereochemistry

WARNING initster(): Stereo atom # without stereodescriptor.

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

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

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

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

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

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

8.2 Warnings in the Generation of 3D Coordinates

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

WARNING initba(): Geometry type for atom # PLANAR --> TETRAEDER changed
The bridgehead atoms in strained ring systems cannot be planar.
This warning occurs in the case of unsaturated ring systems
containing rings smaller than 5, e.g., cubene.

WARNING getta(): No TA #-#-#-#.
In case of linear systems (e.g., acetylene) torsional angles (TA's)
for some bonds cannot be specified.

WARNING hmoboord(): No HMO constant for atom #.
No Coulomb integral parameter is found for a hetero atom.
CORINA takes the value for carbon.

WARNING hmoboord(): No HMO constant for bond #-#.
No parameter for a bond resonance integral is found. The C-C
value is taken.

WARNING genconf(): Time out.
The time limit for a complete ring conformation analysis was
exceeded. The conformer with lowest energy so far found is not
necessarily the global minimum.

WARNING rrefine(): # pair(s) of crowded atoms
Some pairs of atoms became closer to each other than 75% of
their atomic radii.

9 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).

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

CORINA.direct was developed at Molecular Networks GmbH, Erlangen, Germany.

CORINA and **CORINA.direct** is now further developed and maintained for general usage by Molecular Networks GmbH Erlangen, Germany (by Dr. C.H. Schwab).

10 References

- [1] Hiller, C.; Gasteiger, J. Ein automatisierter Molekülbaukasten. In *Software-Entwicklung in der Chemie*, Vol 1; Gasteiger, J., Ed.; Springer: Berlin, 1987; pp 53-66.
- [2] Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic Generation of 3D Atomic Coordinates for Organic Molecules. *Tetrahedron Comp. Method.* **1990**, *3*, 537-547.
- [3] Sadowski, J.; Rudolph, C.; Gasteiger, J. The Generation of 3D Models of Host-guest Complexes. *Anal. Chim. Acta* **1992**, *265*, 233-241.
- [4] 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.
- [5] Sadowski, J., Gasteiger, J. From Atoms and Bonds to Three-dimensional Atomic Coordinates: Automatic Model Builders. *Chemical Reviews* **1993**, *93*, 2567-2581.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] CORINA Version 3.1 is available from Molecular Networks GmbH, Germany (<http://www.mol-net.de>, inf@mol-net.de).
- [10] 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>.
- [11] 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>.

-
- [12] 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>.
- [13] Gasteiger, J. et al. CTX Keyword Reference Manual. University of Erlangen-Nürnberg: 1995, unpublished results.
- [14] 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.
- [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>.

