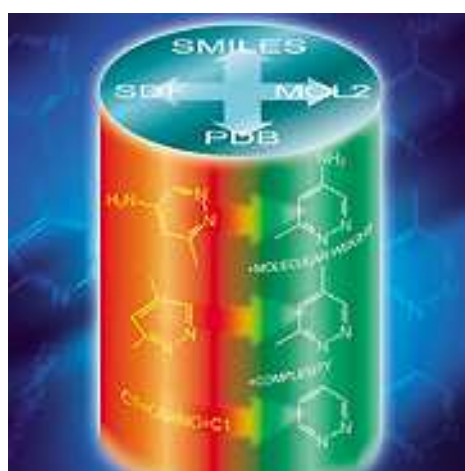




.CONVERT

Inter-conversion of Chemical Structure
and Reaction File Formats

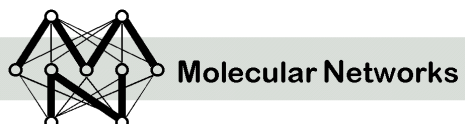


User Manual

Version 1.0

for program version 1.0 (or higher)

November 2003



Molecular Networks GmbH

Computerchemie

Nägelsbachstr. 25

91052 Erlangen

Germany

Phone: +49-(0)9131-815668

Fax: +49-(0)9131-815669

Email: info@mol-net.de

WWW: www.mol-net.de

This document is copyright © 2002 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 was originally designed and implemented for the CACTVS system by W. D. Ihlenfeldt. It is not part of the standard CACTVS toolkit distribution; it is furnished under a license and may be used and copied only in accordance with the terms of such license. For licensing this software please contact Molecular Networks GmbH, Nägelsbachstr. 25, 91052 Erlangen, Germany.

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.

Table of Contents

1.	General Information about MN.CONVERT	5
2.	Installation	6
2.1.	Requirements.....	6
2.2.	Installation Steps for UNIX Operating Systems (IRIX, Solaris, Linux)	6
2.3.	Installation Steps for Microsoft Windows Operating Systems (NT4/2000/XP).....	6
3.	Uninstallation.....	6
3.1.	Uninstallation Steps for UNIX Operating Systems (IRIX, Solaris, Linux).....	6
3.2.	Uninstallation Steps for Microsoft Windows Operating Systems (NT4/2000/XP) ..	6
4.	Problems and Help!.....	7
5.	Release Notes	8
5.1.	Version 1.0	8
5.2.	Version 1.1	8
6.	Getting Started.....	9
6.1.	UNIX operating systems.....	9
6.2.	Microsoft Windows operating systems.....	9
7.	Program Use	10
7.1.	Synopsis	10
7.2.	General Program Features.....	11
7.3.	Program Features for File Handling	11
	-format <abbreviation of the output format>	11
	-outfile <filename.extension>	13
	-directory <dirname>	13
	-count <n>	14
	-eoltype mac/pc/unix	14
	-miniheader	14
	-smileswithname.....	15
	-feedback 0/n	15
	-scope mol/ens/reaction	15
	-aromaticity 0/1	16
	-stat	16
	-version	16
	-h or -help.....	16
8.	Extended Features Only Available for the UNIX Operating Systems.....	17
9.	Frequently Asked Questions (FAQ)	18
10.	Error Messages	18
11.	Known Problems and Limitations	18
12.	Technical Support	19
	The MN.CONVERT Web Site.....	19
	Reporting Problems.....	19
	Updates.....	19
	Contact Information	19

13. Report Form..... 20
14. Index..... 21

1. General Information about MN.CONVERT

More than 40 different file formats have been designed by academic and commercial software developers to electronically store chemical structures and chemical reactions. Dealing with this diversity of file formats is a challenging issue for computational chemists that work in a heterogeneous software environment.

MN.CONVERT enables the seamless inter-conversion of 40 different structure and reaction file formats. This application automatically detects the format of the input file and converts it into the format specified by the user.

The program MN.CONVERT

- recognizes about 40 formats
- converts the input file into one out of 40 file formats
- optionally generates canonical smiles
- processes datasets with 99.9% conversion rate
- handles datasets of hundreds of thousands of chemical structures

2. Installation

2.1. Requirements

MN.CONVERT is available for common UNIX platforms (x86 Linux, Sun Solaris, SGI IRIX, DEC AlphaStation). It is also available for Microsoft Windows NT4/2000/XP. The program runs in a batch mode.

2.2. Installation Steps for UNIX Operating Systems (IRIX, Solaris, Linux)

- 1.) Create a subdirectory, e.g., `mn_convert`
(for system administrators when installing software locally, e.g. `/usr/local/bin/mn_convert`).
- 2.) Copy the file `mn_convert_<version>.<os>.gz` to the subdirectory `mn_convert`
- 3.) Unpack the distribution by executing the `gunzip` command:
`gunzip mn_convert_<version>.<os>.gz`
- 4.) Rename the file `mn_convert_<version>.<os>` to `mn_convert`.
Please note: `mn_convert_<version>.<os>` is a binary file.
- 5.) Add the `mn_convert` subdirectory name to the environment variable `PATH` in your `.login` or `.cshrc` files (`.profile` or `.bashrc`).

Launch MN.CONVERT with the command

```
mn_convert -version or /usr/local/bin/mn_convert/mn_convert -  
version
```

2.3. Installation Steps for Microsoft Windows Operating Systems (NT4/2000/XP)

Although administrator privileges are not necessary, we recommend logging in as administrator. Double-click on the executable setup program and follow the instructions on the screen.

After successful installation there is no need to reboot your PC.

3. Uninstallation

3.1. Uninstallation Steps for UNIX Operating Systems (IRIX, Solaris, Linux)

Log in as root and delete the file `mn_convert` in your installation directory carefully (default path during installation was `/usr/local/bin/mn_convert/`).

3.2. Uninstallation Steps for Microsoft Windows Operating Systems (NT4/2000/XP)

Log in as administrator, launch the uninstaller and follow the on-screen instructions.

4. Problems and Help!

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

Molecular Networks GmbH Computerchemie
Nägelsbachstr. 25
91052 Erlangen
Germany,

or contact us by email support@mol-net.de,
or by fax +49-(0)9131 - 81 56 69.

Please mention the program version of MN.CONVERT (`mn_convert -version`), include your input file and the output file on an MS/DOS diskette (3½") or send it to us by email. These files will help us to analyze the problem; if your system displays any error messages, please add them to your report.
You can also use the report form at the end of this manual.

5. Release Notes

5.1. Version 1.0

First release of MN.CONVERT

5.2. Version 1.1

Support of MDL's file format SDF V 3000 (read and write)

6. Getting Started

6.1. UNIX operating systems

The example file `ccc.smi` submitted with the distribution contains the structure information of one molecule in SMILES format. Copy this example file into your working directory and type the following command:

```
mn_convert -format sdf ccc.smi
```

MN.CONVERT now creates the output file named `ccc.mdl` written to the same directory where the file `ccc.smi` is located. Figure 1 shows the content of this file.

6.2. Microsoft Windows operating systems

The example file `ccc.smi` submitted with the distribution contains the structure information of one molecule in SMILES format. Copy this example file into your working directory and open a DOS shell. Change the working directory to the directory where you installed `mn_convert` by using the `cd` command then type the following command:

```
mn_convert -format sdf ccc.smi
```

MN.CONVERT now creates the output file named `ccc.mdl` written to the same directory where the file `ccc.smi` is located. Figure 1 shows the content of this file.

If you have no permission writing to the directory in which the program was installed, set the **-directory** option for specifying another directory:

```
UNIX: mn_convert -format sdf -directory /tmp ccc.smi
```

```
Windows: mn_convert -format sdf -directory C:/temp ccc.smi
```

```
C3H8
OL   mn_convert05070209282D

11 10 0 0 0 0 0 0 0 0999 V2000
0.0000 0.0000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.0000 0.0000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.0000 0.0000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.0000 0.0000 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.0000 0.0000 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.0000 0.0000 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.0000 0.0000 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.0000 0.0000 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.0000 0.0000 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.0000 0.0000 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.0000 0.0000 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 2 1 0 0 0 0
2 3 1 0 0 0 0
1 4 1 0 0 0 0
1 5 1 0 0 0 0
1 6 1 0 0 0 0
2 7 1 0 0 0 0
2 8 1 0 0 0 0
3 9 1 0 0 0 0
3 10 1 0 0 0 0
3 11 1 0 0 0 0
M  END
$$$$
```

Figure 1: Content of the output file `ccc.mdl`

7. Program Use

7.1. Synopsis

The general synopsis for using MN.CONVERT is:

```
mn_convert [ -option(s) ] [ infile ]
```

An overview of the various options is given in Table 1 and in a more detailed one in the following chapter. `Infile` is the input file name. If no file name is given, the program reads from standard input.

[-aromaticity 0/1]	resolves aromatic bonds into a Kekulé system
[-count number]	limits the amount of converted structures
[-directory dirname]	specifies the output directory
[-eoltype mac/pc/unix]	selects between different end-of-line characters
[-feedback 0/n]	prints a control message after processing a block of n items
[-format fmt]	specifies the output format name
[-h] or [-help]	shows a brief help message about the usage of the program
[-miniheader]	writes some file formats with only minimal header information
[-scope mol/ens/reaction]	controls the basic units read from the input file (molecules or reactions)
[-smileswithname]	writes the compound name (if absent, the calculated formula) to the right of the actual SMILES data in the output file
[-stat]	writes statistical information about the number of successfully processed records and processing failures
[-outfile filename]	defines the name of the output file
[-version]	prints version and licensing information

Table 1: Overview of all options

Executing the program without any option will write structures in SD file format to the output file.

7.2. General Program Features

The file type of the input file is automatically recognized. If no input file is specified, or the file name „-“ is used, the program reads from standard input.

If you are running MN.CONVERT under a UNIX operating system, there are some more features reading input files (see chapter 8 “Extended Features Only Available for the UNIX Operating Systems” for more details).

The file name of the output file is either explicitly set with the **-outfile** option or automatically derived from the input file and the given output file format (**-format**). The special filename `stdout` can be used to direct output to the standard output channel.

7.3. Program Features for File Handling

-format <abbreviation of the output format>

The parameter **format** is specified for selecting the output format. As MN.CONVERT automatically recognizes the format of the input file, there is no need for an analog parameter specifying the input format.

The software can process both structure files and reaction files. The supported file formats are listed in the table below.

Full Name	Format	Abbrev.	Output-Extens.	Read	Write	Single/Multiple	Comment
441		441	441	Yes	Yes	M	
Alchemy		alc	alc	Yes	Yes	M	
Cactvs/Ascii		ascii	cac	Yes	Yes	M	structures and reactions
Cactvs/Scan		base	cbr	Yes	Yes	M	structures and reactions
Cactvs/Binary		bin	bin	Yes	Yes	M	structures and reactions
CAR		car	car	Yes	Yes	M	
Cerius II		cerius	msi	Yes	Yes	M	
Charmm		charmm	crd	Yes	Yes	M	
ChemDraw		cdx	cdx	Yes	Yes	S	
ChemDraw XML		cdxml	cdxml	Yes	Yes	S	
CIF (CHIRON)		cif	cif	Yes	Yes	S	this is not the crystallographic CIF
CML		cml	xml	No	Yes	M	CML 2.0
Compass		compass	cpa	Yes	Yes	M	
Cosmo		-	cosmo	Yes	No		
CTX		ctx	ctx	Yes	Yes	M	structures and reactions
Gaussian Archives		garc	garc	Yes	No	M	
Gaussian Cube		cube	cube	Yes	No	S	
Gaussian Input		gaussian	gin	Yes	Yes	S	
Hitlist		hitlist	hit	Yes	Yes	M	
Hyperchem		hyperchem	hin	Yes	Yes	S	

Index	index	idx	Yes	Yes	M	to select subsets of larger files
JCAMP-CS	jcamp	jcp	Yes	Yes	M	
JCAMP-DX	jcamp	jcp	Yes	Yes	M	
JME	jme	jme	Yes	Yes	S	
M3D	m3d	m3d	Yes	Yes	S	
Molconn-Z	molconnz	b	Yes	Yes	M	
Molfile (MDL)	mol	mdl	Yes	Yes	S	
Molgen	molgen	gen	Yes	Yes	M	
Mopac Input	mopacin	mop	Yes	Yes	S	
Mopac Output	mopac-output	out	Yes	No	S	
NETCDF	netCDF	netcdf	Yes	No	S	(Network Common Data Form), does not support all conventions for spectral data exchange
PDB	pdb	pdb	Yes	Yes	M	
RDF	rdf	rdf	Yes	Yes	M	structures and reactions
RXN	rxn	rxn	Yes	Yes	S	
SCF	scf	scf	Yes	Yes	M	
SDDATA	sddata	sdd	Yes	Yes	M	SD-style data file without structure block
SDF	mdl/sdf	mdl	Yes	Yes	M	most ISIS query attributes, including 3D search and R-groups, are supported
SDF V3000	sdf3	sdf3	Yes	Yes	M	
Sharc	sharc	sharc	Yes	Yes	M	
Shel-X	shelx	-	Yes	No	S	
SKC (MDL Isis Sketch)	skc	-	Yes	No	S	minimal implementation
SMARTS	smarts	sma	Yes	Yes	M	SMARTS uses explicit hydrogen enumeration: [CH3][CH3]
SMD 4	smd/smd4	smd	Yes	Yes	M	
SMD 5	smd5	smd5	Yes	Yes	M	not a complete implementation
SMILES	smi/smiles	smi	Yes	Yes	M	without explicit hydrogen enumeration: CC
STF	stf	stf	Yes	Yes	S	
Sybyl	sybyl	syb	Yes	Yes	M	
Sybyl II	sybyl2	syb2	Yes	Yes	M	
Sybyl Line Notation	sln	sln	Yes	Yes	M	
TGF	tgf	tgf	No	Yes	S	supports molecules and structure queries, but not yet reactions.
Thor Data Tree/ Stigmata	stigmata/tdt	stg	Yes	Yes	M	

Vamp	vamp	-	Yes	No	S	
XBSA	xbsa	bs	Yes	Yes	M	
Xtelplot	xtelplot/ xtel	xtel	Yes	Yes	M	
XYZ	xyz	xyz	Yes	Yes	S	trajectories are supported
XYZR	xyzr	xyzr	Yes	Yes	S	file format only contains atomic radii and no element information

Table 2: Overview of the supported file formats

Please use the abbreviation of the format names for specifying your desired file format using the **-format** option. If no output file is specified, the output has the same name (but with an updated suffix) and is written in the same directory as the input file. The extension of the resulting output file is sometimes different to the given abbreviation (see the previous table). If the output file is specified explicitly with the **-outfile** parameter, this file name including the chosen suffix, will be used.

Default value:

Parameter without a default value

Example:

Generating a MDL SD-file:

```
mn_convert -format sdf ./examples/ccc.smi
```

Remarks:

If this option is not used, an attempt is made to guess the output file format from its suffix by the given **-outfile** parameter.

-outfile <filename.extension>

The parameter **outfile** defines the name of the output file. MN.CONVERT automatically recognizes the desired output format, thus in most cases it is not necessary to specify the output format.

If you are using MN.CONVERT on a UNIX operating system the output file name can also be an anonymous ftp URL.

Default value:

Parameter without a default value

Example:

Generating a MDL SD-file:

```
mn_convert -outfile ccc.sdf ./examples/ccc.smi
```

-directory <dirname>

This parameter sets the target directory. If the directory does not yet exist, it will be created.

Default value:

The directory of the output files is the same as of the corresponding input files, or the current directory, if the input file names do not contain directory information.

Example:

Generating a MDL SD-file saved in a given directory:

```
UNIX: mn_convert -outfile ccc.sdf -directory /tmp/output
./examples/ccc.smi
Windows: mn_convert -outfile ccc.sdf -directory C:/Temp/alkane
./examples/ccc.smi
```

-count <n>

The parameter **count** is able to limit the amount of converted structures. A maximum of <n> records from the input file(s) is processed. Especially while handling large input files with hundreds of thousands of structures this parameter is useful to convert only the first hundred structures, if you would like to test other parameters and would not like to wait until the entire file was processed.

Default value:

No limitation (value= ∞)

Example:

Generating a MDL SD-file with four alkanes (from methan to butane):

```
mn_convert -outfile alkanes.sdf -count 4
./examples/alkanes1_12.smi
```

-eoltype mac/pc/unix

This parameter is used for choosing between different end-of-line characters.

Default value:

NL (Unix)

CR (Macs)

CR/NL (PC)

Example:

Generating a MDL SD-file with Windows end-of-line characters:

```
mn_convert -outfile ccc.sdf -eoltype pc ./examples/ccc.smi
```

Remarks:

This option is ignored if the output file format is binary.

-miniheader

If the parameter miniheader is set, some file formats are written with only minimal header information.

Default value:

This option is not active by default.

Example:

Generating a MDL SD-file with only a minimal header information:

```
mn_convert -outfile ccc_miniheader.mol -miniheader
./examples/ccc.smi
```

Remarks:

This option is intended to be used for software which cannot process the full header specifications of formats such as MDL molfiles.

-smileswithname

If this parameter is set, the compound name (if present, the formula otherwise) is written to the right of the actual SMILES data in the output. Most SMILES readers will be able to read this format extension.

Default value:

This option is not active by default.

Example:

Generating a MDL SD-file with the compound name or formula:

```
mn_convert -outfile ccc_withname.smi -smileswithname
./examples/ccc.sdf
```

-feedback 0/n

If the parameter **feedback** is set to a value larger than zero, a control message is printed after processing a block of *n* structures. The current record number and the object name are printed on the standard error channel. Only structures which are actually written out are counted.

Default value:

It is not active by default.

Example:

Generating a MDL SD-file printing dots for every one hundred records:

```
mn_convert -outfile maybridge.sdf -feedback 100
./examples/maybridge.smi
```

-scope mol/ens/reaction

This parameter controls the basic units read from the input file (normally, molecules or molecular ensembles from structure files and reactions from reaction files). However, if the scope is set to *ens*, it is for example possible to read reagent and product ensembles as two separate records from a reaction data file. The precise meaning of *mol* vs. *ens* data input depends on the file format. Only a few formats make a distinction at all. It is not possible to use a mode higher in the hierarchy than the file contents actually present (i.e. it is not possible to read reactions from a simple structure file).

Default value:

Automatically set according to the input format.

Example:

Reading molecules from the input file and generating a MDL SD-file:

```
mn_convert -outfile da_molecules.sdf -scope mol
./examples/diels_alder.rdf
```

Remarks:

This parameter is intended for expert use.

-aromaticity 0/1

If this parameter is set to 1, aromaticity information is used to format the output. For SMILES, this means that the π -center variant of atom encoding is used. For example, benzene in the Kekulé-style default encoding is C1=CC=CC=C1, but c1ccccc1 in aromatic encoding. For MDL formats, aromatic structure bonds can be output as aromatic query bonds, if necessary. The latter is not a correct use outside an ISIS structure query specification according to the original format definition, but there are a number of programs which expect their input encoded this way.

Default value:

This flag is deactivated by default.

Example:

Generating a MDL SD-file using the aromaticity information:

```
mn_convert -outfile aro_res.smi -aromaticity 1
./examples/kek_aro.smi
```

Remarks:

A number of well-known chemistry software packages do not implement the MDL structure exchange formats correctly. According to the original specifications, an aromatic bond in these files can only be used as a query attribute, and it is read as such by MN.CONVERT and therefore does not have a bond order, electron count, etc.

-stat

If this flag is set, statistical information about the number of successfully processed records and conversion failures is written to the standard error channel.

Default value:

This flag is deactivated by default.

Example:

Generating a MDL SD-file showing the statistical information of the conversion:

```
mn_convert -outfile alkanes.sdf -stat ./examples/alkanes1_12.smi
```

Output:

```
Successfully read 12 records, failed 0
Successfully wrote 12 records, failed 0.
```

-version

If this flag is set, the version and licensing information is printed.

Default value:

This flag is deactivated by default.

Example:

Showing the program version:

```
mn_convert -version
```

-h or -help

If this flag is set, a brief help message about the usage of the program is shown.

Default value:

This flag is deactivated by default.

Example:

Show the help message:

```
mn_convert -h or mn_convert -help
```

8. Extended Features Only Available for the UNIX Operating Systems

Input files can be processed in compressed or gzip-ed form without prior unpacking. The input file name arguments may each be a local file, an URL (http, ftp, gopher, file) or an email message file containing the structure data in the main body or as one or more attachments. URL retrieval and compression can be combined.

9. Frequently Asked Questions (FAQ)

10. Error Messages

MN.CONVERT does not process the single input structure although a valid SMILES string is given

Running MN.CONVERT under Windows the very last line of the input structure must be empty.

11. Known Problems and Limitations

12. Technical Support

The MN.CONVERT Web Site

If you have problems while running MN.CONVERT please have a look at the Support- and FAQ web site of MN.CONVERT. The pages are available at <http://www.mol-net.de>

Reporting Problems

If your problem is not listed in these web pages please report it to the MN.CONVERT team at Molecular Networks. Please make sure to provide us with all important data for replicating your problem on our machines. Therefore please use the report form on the next page.

Updates

If you have licensed the program MN.CONVERT with maintenance you will automatically receive updates every time a new release is launched.

Contact Information

Distribution and Maintenance for MN.CONVERT is handled by Molecular Networks Computerchemie, Erlangen, Germany.

Molecular Networks GmbH
Computerchemie
Nägelsbachstraße 25
91052 Erlangen
Germany

e-Mail: support@mol-net.de

Tel. +49 9131/815668

Fax +49 9131/815669

13.Report Form

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

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

User:

MN.CONVERT program and version number (mn_convert -version):

Command line to run MN.CONVERT:

Error and warning messages by MN.CONVERT:

System messages:

Short description:

Please include the input file and output file generated by MN.CONVERT on a 3½" diskette written in MS/DOS format or send an e-mail to support@mol-net.de attaching these files. These files will help us to analyze your problems. All data will be treated confidentially.

14.Index

inputfile

- alkanes1_12.smi 14, 16
- ccc.sdf 15
- ccc.smi 9, 13, 14
- diels_alder.rdf 15
- kek_aro.smi 16
- maybridge.smi 15

option

- aromaticity 16
- count 14
- directory 13
- eoctype 14
- feedback 15
- format 11
- h 16
- help 16

- miniheader 14

- outfile 13

- scope 15

- smileswithname 15

- stat 16

- version 16

outputfile

- alkanes.sdf 14, 16

- aro_res.smi 16

- ccc.sdf 14

- ccc.smi 14

- ccc_miniheader.mol 14

- ccc_withoutname.sdf 15

- da_molecules.sdf 15

- maybridge.sdf 15