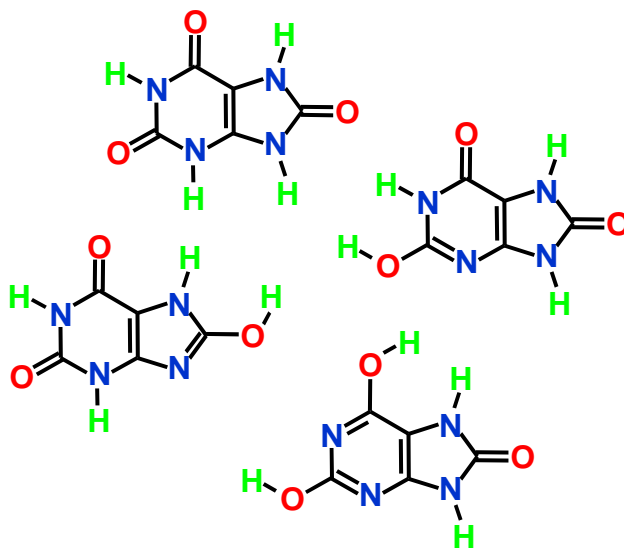


# MN.TAUTOMER

## Generation of Tautomeric Forms of Chemical Compounds

Version 1.12

### User Manual and Program Description



Molecular Networks GmbH Computerchemie  
March 2009

<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](mailto:info@molecular-networks.com)  
WWW: [www.molecular-networks.com](http://www.molecular-networks.com)

This document is copyright © 2007 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.

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.

(Document version: OS-CHS-1.12-2010-04-21)



## Contents

|     |  |    |
|-----|--|----|
| 1   | Introducing <b>MN.TAUTOMER</b>                       | 1  |
| 1.1 | Tautomerism  | 1  |
| 1.2 | Encoded Rules  | 1  |
| 1.3 | The Unique Tautomer                                  | 2  |
| 1.4 | Range of Application                                 | 2  |
| 2   | Release Notes  | 3  |
| 2.1 | Version 1.0  | 3  |
| 2.2 | Version 1.2  | 3  |
| 2.3 | Version 1.4  | 3  |
| 2.4 | Version 1.6  | 3  |
| 2.5 | Version 1.8  | 3  |
| 2.6 | Version 1.12   | 4  |
| 3   | Getting Started with <b>MN.TAUTOMER</b>              | 6  |
| 4   | Using <b>MN.TAUTOMER</b>                             | 8  |
| 4.1 | Synopsis   | 8  |
| 4.2 | General Features                                     | 8  |
| 4.3 | Options  | 9  |
| 5   | How to Use <b>MN.TAUTOMER</b>                        | 14 |
| 6   | Supported File Formats                               | 17 |
| 6.1 | MDL Molfile and Structure Data File (SDF)            | 17 |
| 6.2 | SMILES Linear Notation                               | 17 |
| 7   | Error Messages                                       | 18 |
| 7.1 | General Errors                                       | 18 |
| 7.2 | I/O Errors   | 18 |
| 7.3 | Errors in the Generation of Tautomers                | 19 |
| 8   | Technical Requirements and Known Limitations         | 21 |
| 8.1 | System Requirements                                  | 21 |
| 8.2 | Program Scope and Limitations                        | 21 |
| 9   | Program Installation                                 | 22 |
| 9.1 | On-line Download from Molecular Networks' Web Server | 22 |
| 9.2 | New Installation                                     | 23 |
| 9.3 | Program Updates                                      | 24 |



|    |                    |    |
|----|--------------------|----|
| 10 | Problems and Help! | 25 |
| 11 | Acknowledgements   | 26 |
| 12 | References         | 27 |
| 13 | Report Form        | 28 |



## 1 Introducing MN.TAUTOMER

### 1.1 Tautomerism

Tautomers have the same molecular graph (excluding bond types and bonds to hydrogen atoms) and the same molecular formula. They are interchangeable forms of one and the same molecule involving the displacement of hydrogen atoms and electrons. One of the most well-known and important tautomerism is the keto-enol tautomerism as depicted in Figure 1. Both forms exist to a certain extent in an equilibrium, however, exhibit huge differences in terms of stability, chemical reactivity and interactions with other chemical entities (e.g., as hydrogen bond acceptor or donor with a biological receptor).

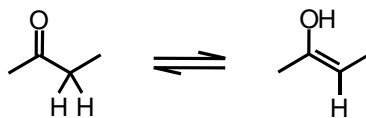


Figure 1 Simple keto-enol tautomerism.

Furthermore, databases of chemical compounds can contain different records of molecules that are, by having a closer look, tautomers. These records basically do not provide any new information and can be regarded as redundant when introducing the concept of tautomerism.

Due to these facts, a careful and reliable enumeration of possible tautomeric states of a molecule is an important, almost mandatory step when modeling the chemical, physicochemical or biological properties of a chemical compound.

### 1.2 Encoded Rules

**MN.TAUTOMER** is a fast and automatic rule-based enumerator of tautomeric forms of a chemical compound. The rules encoded automatically detect substructures in a given molecule that can undergo a tautomeric transformation. Currently, almost all of the most important types of tautomerism are supported:

- simple and long-range enol/thioenol exchange
- simple imine exchange
- nonsubstituted heteroaromatic exchange
- simple or long-range hetero atom hydrogen exchange
- keten/inol exchange

- nitro form/aci form of nitro compounds
- simple nitroso/oxim exchange or nitroso/oxim exchange of aromatic systems
- cyanuric acid, formamidinsulfon acid, hydrogen cyanide – hydrogen isocyanide, phosphonic acid

### 1.3 The Unique Tautomer

Internally, **MN.TAUTOMER** enumerates all possible tautomers by applying the encoded transformation rules to the input structure. Thereafter, an empirical scoring function is used in order to roughly estimate for each tautomer the probability of occurrence in solution relatively to the other tautomers. This empirical scoring function only takes into account some structural features and properties of a tautomer. Finally, all tautomers are ranked according to their scores. The highest ranked tautomer is called the **unique tautomer**, i.e., the proposed form being assumed to be one of the most prevalent in solution.

By default, the unique tautomer is written to the output file. However, **MN.TAUTOMER** can be forced to write out all possible tautomers that can be generated within the scope of the encoded rules or a ranked subset.

Furthermore, the total number of tautomers generated and the types of applied tautomeric transformations rules can be defined by the user.

### 1.4 Range of Application

**MN.TAUTOMER** has been designed to process a broad range of chemistry. It processes datasets of hundreds of thousands of chemical structures with a conversion rate of up to 99.9% and supports MDL Molfile [1], MDL SDF [1] and SMILES [2] file format for structure input and output.

**MN.TAUTOMER** is a valuable tool for scientists working in the areas of

- structure-based drug discovery, design and optimization
- ligand-based drug discovery, design and optimization
- structure and reaction database consolidation
- ADMETox studies
- prediction of chemical reactivity, synthetic accessibility and synthesis design
- structure elucidation and spectra prediction

**Note.** **MN.TAUTOMER** does not handle charged species and does not generate different protonation states of functional groups or mesomeric forms of the input molecules (see also section 8.2 "Program Scope and Limitations" on page 21).

## 2 Release Notes

### 2.1 Version 1.0

Version 1.0 is the first commercial release of **MN.TAUTOMER** that contains a basic set of 14 most important tautomeric transformations.

### 2.2 Version 1.2

The following new features and improvements have been implemented in version 1.2.

- (1) The supported list of transformations was extended by two additional rules.
- (2) The option `-number` was added to consecutively number the generated tautomers by adding a counter to the compound name.

### 2.3 Version 1.4

The following new features and improvements have been implemented in version 1.4.

- (1) If all transformation rules are excluded (option `-notypeN`,  $N$  of type integer) the original structure of the input file is written to the output file.
- (2) For the generated tautomers, the 2D coordinates of the heavy atoms given in the input file (MDL Molfile and SDFfile) are kept and the 2D coordinates of the hydrogen atoms are newly calculated.

### 2.4 Version 1.6

The following new features and improvements have been implemented in version 1.6.

- (1) The supported list of transformations was extended by one additional rule.
- (2) The handling of 2D and 3D atomic coordinates has been improved.

### 2.5 Version 1.8

The following new features and improvements have been implemented in version 1.8.

- (1) The new option `-nprefmax N` ( $N$  of type integer) internally generates all possible tautomers of an input structure and ranks them by applying the empirical scoring function (see section 1.3 on page 2). Finally, only the  $N$  top ranked (most probable) tautomers are written out, i.e., each input structure is represented by a maximum number of  $N$  tautomers.

- (2) The new option `-percentprefmax F` ( $F$  of type float and in percentage, %) internally generates all possible tautomers of an input structure and ranks them by applying the empirical scoring function (see section 1.3 on page 2). Finally, only the top ranked (most probable) tautomers are written out until the given threshold of  $F\%$ , whereas the unique tautomer is set to 100%. I.e., only those tautomers are written out that have a score of  $\geq F\%$  compared to the unique tautomer.

In contrast to the option `-nprefmax` (see above), the number of output tautomers may strongly vary for different input structures.

- (3) The new option `-typeN` ( $N$  of type integer) allows to apply only the specific tautomeric transformation rule  $N$  to the input structure and suppresses application of other transformation rules. Therefore, this option can be regarded as the opposite of the option `-notypeN` (see also section 4.3 on page 9). The application of more than one transformation rules can be forced by defining this option at the command line more than once.

A list of supported tautomeric transformation rules and their respective numbers  $N$  can be printed on the screen with the option `-types`.

- (4) The new option `-skiperrors` skips molecules (records) in a multi-record file which show some deficiencies or errors in the formatting or structure coding and continues processing the very next record without aborting the program run.

## 2.6 Version 1.12

The following new features and improvements have been implemented in version 1.10.

- (1) The behavior of the option `-percentprefmax F` ( $F$  of type float and in percentage, %) has been changed and now outputs the  $F\%$  top-ranked (most prevalent or probable) tautomers for a given input structure. This option is now consistent with the option `-nprefmax N`.
- (2) The new option `-preserveamides` forces MN.TAUTOMER to always generate (or to preserve) the most stable "carboxamide" form (SMILES code: "C(=O)N(H)", see also Figure 2) for amides and amide substructures and not to form a double bond at the carbonyl carbon atom.

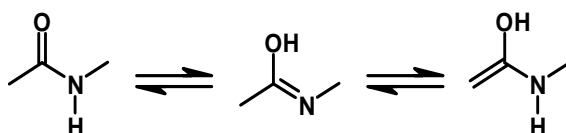


Figure 2 Tautomers of amides: most stable "carboxamide" form (left) and the less observed imidic acid (middle) and enol forms (right).

- (3) Since version 1.10, a component to run MN.TAUTOMER in the SciTegic® Pipeline Pilot environment, the data analysis and reporting platform from Accelrys® ([www.accelrys.com](http://www.accelrys.com)), is available. The components to run MN.TAUTOMER either on the SciTegic® Pipeline Pilot server or client computer can be downloaded free of charge from the web server of Molecular Networks at [www.molecular-networks.com](http://www.molecular-networks.com).

### 3 Getting Started with MN.TAUTOMER

**MN.TAUTOMER** is a command line oriented program system and has to be executed in a shell (e.g., *csch*, *tcsh*, or *bash* on UNIX/Linux systems, see Figure 3) or at a Microsoft® Windows® command prompt (see Figure 4). All command line options provided by **MN.TAUTOMER** are described in detail in Section 4 "Using **MN.TAUTOMER**" on page 8.

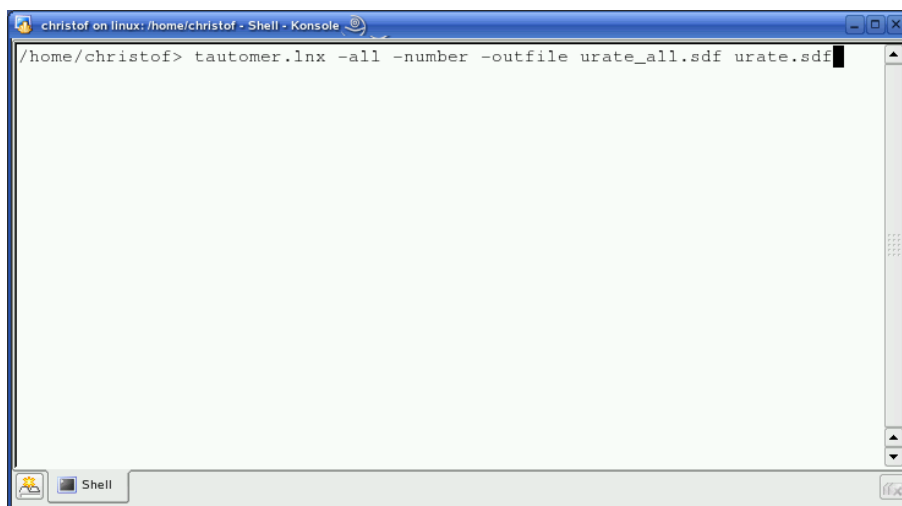


Figure 3 Executing **MN.TAUTOMER** in a UNIX/Linux shell.

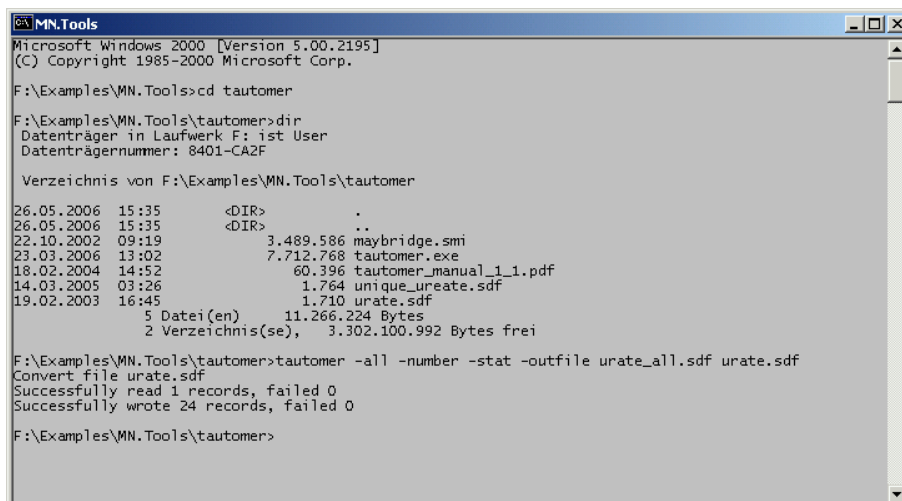


Figure 4 Executing **MN.TAUTOMER** at a MS Windows® command prompt.

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

```
mn_tautomer -all -outfile urate_all.sdf urate.sdf
```

**MN.TAUTOMER** now creates the output file named `urate_all.sdf` written to the same directory where the input file `urate.sdf` is located. Figure 5 shows a part of the content of the output file.

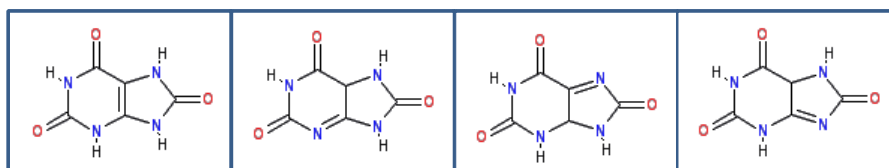


Figure 5 The first four of a total of 60 generated tautomeric forms for urate in the output file `urate_all.sdf`.

If you don't have a permission to write into the directory in which the program was installed, use the `-directory` option for specifying another directory.

For UNIX/Linux systems:

```
tautomer -all -directory /tmp -outfile urate_all.sdf  
urate.sdf
```

For Microsoft® Windows® systems:

```
tautomer -all -directory C:/temp -outfile urate_all.sdf  
urate.sdf
```

## 4 Using **MN.TAUTOMER**

### 4.1 Synopsis

The general synopsis for using **MN.TAUTOMER** is:

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

### 4.2 General Features

**MN.TAUTOMER** automatically recognizes the file format of the input file. If no input file is specified, or the file name "-" is used, the program reads from standard input (`stdin`).

The file format of the output file is set accordingly to the input file format (unless specified differently by the user, see option `-format`). The file name of the output file can be explicitly set with the option `-outfile`. If no output file name is specified a suitable file name is derived. The special filename `stdout` can be used to direct output to the standard output channel.

For UNIX/Linux systems, **MN.TAUTOMER** offers the following additional features for reading input files.

- Input files can be processed in compressed or gzip-ed form without prior unpacking.
- The input file name arguments may be a local file, an URL (`http`, `ftp`, `gopher` or `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.

By default, only the unique tautomer, i.e., the proposed form being assumed to be one of the most prevalent in solution, is written out, unless specified differently by the user with the options `-all`, `-max`, `-nprefmax` or `-percentprefmax`, respectively.

### 4.3 Options

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

#### Input file format

The input file format is recognized automatically. The supported file formats are:

|        |                        |
|--------|------------------------|
| mol    | MDL Molfile [1]        |
| sdf    | MDL SDFfile [1]        |
| smiles | SMILES linear code [2] |

#### Output file format and location

**-format S** Set the output file format to *S* (*S* of type character string).

Allowed file formats (character strings) are:

|        |                        |
|--------|------------------------|
| mol    | MDL Molfile [1]        |
| sdf    | MDL SDFfile [1]        |
| smiles | SMILES linear code [2] |

If this option is not set the output file format is set to the input file format accordingly.

**-outfile S** Specify the name *S* of the output file (*S* of type character string).

This option defines the name of the output file. The parameter *S* **must** be given in the format *filename.extension*. Recommended file name extensions are:

|     |                    |
|-----|--------------------|
| mol | MDL Molfile        |
| sdf | MDL SDFfile        |
| smi | SMILES linear code |

**-directory *S*** Specify the directory *S* to which the output file should be written (*S* of type character string).

This option specifies the target directory. If the directory does not yet exist, it will be created. By default, 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.

The parameter *S* should be given as full or relative path name (e.g., C:\myFiles\tautomers or ../tautomers).

### **Information on Tautomer Generation Process**

**-feedback *N*** Print after each *N* of processed records some information of the program run (*N* of type integer).

If *N* is set to greater 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 output channel. Only structures which are actually written out are counted.

**-skiperrors** Skip erroneous records and continue with the next record.

This option ignores records that have errors (e.g., erroneous connection table or SMILES string) and continues to process the very next record.

**-stat** Print some statistical information after all records have been processed.

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

## Tautomer Generation Options

- all** Generate all possible tautomers for each molecule.
- This option controls the enumeration of tautomeric forms. If the parameter is set, the output file includes all possible tautomeric forms. If this option is **not** set, only the unique tautomer, i.e., the proposed form being assumed to be one of the most prevalent (most stable) in solution, is written out.
- max  $N$**  Restrict the number of output tautomers to  $N$  for each molecule ( $N$  of type integer).
- This option controls the maximum number of tautomeric forms which are generated and output for each molecule, whereas the parameter  $N$  can be 1, 2, 3, ...,  $n$ .
- Note.** This option may not be used in combination with the options **-all** (see above), **-nprefmax** or **-percentprefmax** (see below). In contrast to the options **-nprefmax** and **-percentprefmax** (see below) the tautomers are not output in a ranked order.
- nprefmax  $N$**  Write out only the  $N$  top-ranked tautomers ( $N$  of type integer).
- This option generates all possible tautomers of the input molecule(s) and ranks them according to their probability of occurrence in solution (by an empirical scoring function, see also section 1.3 on page 2). Finally, only the  $N$  top-ranked tautomers are written out.
- Note.** If less than  $N$  tautomers can be generated for a given input structure, all tautomers are written out.

- notype***N* Exclude tautomeric transformation rule *N* (*N* of type integer).
- This option excludes a specific type of a tautomeric transformation from the generation process. An entire list of all supported tautomeric transformation can be printed to the screen by the command `tautomer -types` (see below).
- Note.** This option can be specified more than once in order to exclude more than one transformation rule.
- number** Number the generated tautomers consecutively by adding a counter to the compound name.
- This option consecutively adds a counter to the compound name for each different tautomeric forms (e.g., `myMolecule_TAUTOMER_1`).
- percentprefmax** *F* Write out only the *F*% top-ranked tautomers (*F* of type float and in percentage, %).
- This option generates all possible tautomers of the input molecule(s) and ranks them according to their probability of occurrence in solution (by an empirical scoring function, see also section 1.3 on page 2). Finally, only the *F*% top-ranked are written out.
- Note.** In contrast to the option `-nprefmax` (see above), the number of output tautomers may vary for different input structures.
- type***N* Only apply tautomeric transformation rule *N* (*N* of type integer).
- This option applies only the tautomeric transformation rule *N* to the input structure(s). All other transformation rules are excluded (opposite of option `-notypeN`, see above). An entire list of all supported tautomeric transformation can be printed to the screen by the command `tautomer -types` (see below).
- Note.** This option can be specified more than once in order to apply more than one transformation rule.

**Supported Tautomeric Transformations, Help and Version**

|          |  |
|----------|--|
| -types   | <p>List all types of supported tautomeric transformation rules.</p> <p>This option shows the entire list of tautomeric transformations that are supported by the current version, e.g.,</p> <p>type1: simple enol/thioenol<br/>type2: long-range enol/thioenol<br/>...</p> |
| -help    | Print the online help to the screen  |
| -version | Print the version number, license type and timeout to the screen.  |

## 5 How to Use **MN.TAUTOMER**

The following section lists some typical use cases of **MN.TAUTOMER** and shows the corresponding command lines and command line options that have to be switched on for the specific tasks. Some more examples are also given in section 3 "Getting Started with **MN.TAUTOMER**" on page 6.

### List all types of supported tautomeric transformation

The list of all tautomeric transformations that are supported should be printed to the screen.

Command line:

```
tautomer -types
```

### Generate the unique tautomeric form

For each molecule of a dataset of  $n$  molecules given in SMILES linear notation, the unique tautomer (i.e., the tautomeric state being assumed to be one of the most prevalent in solution) should be generated and written out in MDL SDFfile format.

Command line:

```
tautomer -format sdf -outfile uniqueTautomers.sdf  
infile.smi
```

**Note.** The option `-format sdf` may be omitted since the extension of the output file name is named `sdf`.

### Generate all possible tautomeric forms

For each molecule of a dataset of  $n$  molecules given in MDL SDFfile format, all possible tautomeric states should be generated and written out in MDL SDFfile format.

Command line:

```
tautomer -all -outfile allTautomers.sdf infile.sdf
```

### Generate a limited number of tautomers

For each molecule of a dataset of  $n$  molecules given in MDL SDF file format, a limited but user-defined number, i.e., 12, of tautomeric states should be generated and written out in MDL SDF file format. **Note.** The option `-max N` may not be used in combination with the option `-all`. The option `-max N` invokes internally the generation of multiple tautomeric states but limited to the given number  $N$ .

Command line:

```
tautomer -max 12 -outfile 12tautomers.sdf infile.sdf
```

### Number the generated tautomers

For each molecule of a dataset of  $n$  molecules given in MDL SDF file format, a limited but user-defined number, i.e., 8, of tautomeric states should be generated, numbered and written out in SMILES format.

Command line:

```
tautomer -max 8 -number -outfile 8tautomers.smi  
-format smiles infile.sdf
```

### Suppress specific types of tautomeric transformations

For the generation of tautomeric states, the keten/inol and the cyanuric acid tautomerism **should not** be taken into account.

Command line:

```
tautomer -all -notype9 -notype14 -outfile  
specificTautomers.sdf infile.sdf
```

### Apply only specific tautomeric transformations

For the generation of tautomeric states, only the simple enol/thioenol and the simple imine tautomerism **should** be taken into account.

```
tautomer -all -type1 -type3 -outfile specificTautomers.sdf  
infile.sdf
```

### Process all files within a directory

Generate the unique tautomer of all molecules stored in the files with the extension `.mol` in the working directory and write the output files into a newly created directory.

Command line:

```
tautomer -directory newDirectory ./inputdir/*.mol
```

### **Output the 4 top-ranked tautomers**

For each input molecule a maximum of 4 but the top-ranked tautomers (most prevalent or probable in solution) should be output.

Command line:

```
tautomer -nprefmax 4 -outfile topRanked_4.sdf infile.sdf
```

### **Output the 35% top-ranked tautomers**

For each input molecule the 35% top-ranked tautomers (most prevalent or probable in solution) should be output.

Command line:

```
tautomer -percentprefmax 35 -outfile topRanked_35p.sdf  
infile.sdf
```

## 6 Supported File Formats

This section briefly describes the file formats that are supported by **MN.TAUTOMER**. Table A lists the input/output formats with a reference to the file format description.

Table A Supported input and output file formats.

| Format      | Input | Output | Default Extension for Input File | Reference |
|-------------|-------|--------|----------------------------------|-----------|
| MDL Molfile | x     | x      | mol                              | [1]       |
| MDL SDFfile | x     | x      | sdf                              | [1]       |
| SMILES      | x     | x      | smi                              | [2]       |

**Note.** The format of the input file is detected automatically, even if the file name does not contain an appropriate file name extension.

### 6.1 MDL Molfile and Structure Data File (SDF)

The MDL SDF and RDF file formats were implemented following the description in the literature [1]. **MN.TAUTOMER** 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. Additional information (e.g., in SDF data fields) remains unread but is piped unchanged to the output when the output file type is also set to MDL SDF (see note below). The individual records are assumed to be terminated by a \$\$\$\$ delimiter.

**Note.** The generation of tautomers mandatorily leads to changes in the connection table (atom/bond block) or atomic positions. **MN.TAUTOMER** may not be able to update this information for newly generated tautomers in the output file (e.g., 3D atomic coordinates).

### 6.2 SMILES Linear Notation

The SMILES linear notation was implemented following the literature [2]. Lower case C, N, O and S are interpreted as aromatic. Implicit hydrogen atoms given inside square brackets are expanded and written to the output file.

**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 of the output file. If no compound name is given, the gross formula is used as compound name in the output file. The very last line **must** be an empty line if it contains only a single record (molecule).

## 7 Error Messages

### 7.1 General Errors

option <name> not supported

The specified option <name> at the command line is not supported.

Timeout: <time>

The license has expired. Contact Molecular Networks for renewal.

### 7.2 I/O Errors

file <inputfile> is not readable

Make sure that you have permissions to read the inputfile.

file or pattern <inputfile> not found

The specified input file <inputfile> at the command line cannot be read. Check if the input file exists.

molfile open failed: File format of file <inputfile> not identified  
inputfile is not of type smiles, mol or sdf

read error: <message>

An error occurred during processing a file. One record in the file could not be read successfully.

failed to open output file <outputfile>

The specified output file <outputfile> cannot be opened. Check the permissions in the specified or current directory.

format <format> not supported

The specified file format <format> (option `-format`) for the output file is not supported. Supported specifications for the output file format are: smiles, smi, sdf, mdl, mol.

<outputdirectory> exists, but is no directory

The specified output directory <outputdirectory> (option `-directory`) is not a directory. Check if the specified output directory is a file.

directory <outputdirectory> is not writable

The specified output directory <outputdirectory> (option `-directory`) cannot be opened. Check the permissions in the specified directory.

### 7.3 Errors in the Generation of Tautomers

no tautomer rule given

This error occurs when all tautomeric rules are excluded using the option `-notype`. Check the command line.

failed to set tautomeric number

An error occurred while numbering the generated tautomers that are written to the output file. Check if compound names exist or if they exceed the allowed length.



## 8 Technical Requirements and Known Limitations

### 8.1 System Requirements

**MN.TAUTOMER** is a command line tool and has to be executed in a shell (e.g., csh, tcsh, or bash on UNIX/Linux systems) or at a Windows or DOS command prompt (see also Section 3 on page 6). The following hardware platforms and operating systems are supported:

- x86 platforms running Microsoft® Windows® 2000/XP (win32)
- x86 platforms running Linux®, Kernel 2.4
- Silicon Graphics SGI® workstations running IRIX® 6.5
- Sun® Sparc stations running Solaris® 9 (SunOS® 5.9)

For all hardware platforms and operating systems, disk space of 10 MB and 256 MB of RAM are required.

### 8.2 Program Scope and Limitations

**MN.TAUTOMER** has been designed to process a broad range of chemistry.

There are no limitations concerning the number of atoms or bonds of a molecule.

**Note.** Some structure file formats that are supported might have such limitations.

The entire periodic table is supported (up to atomic number 103, Lawrencium). Any chemical compound that can be expressed correctly in a valence bond (**VB**) notation can be processed.

Stereochemical information is fully considered.

Atoms with up to six neighbors are allowed.

Only **uncharged molecules** can be processed. Furthermore, **MN.TAUTOMER** does not generate different protonation states of functional groups or mesomeric forms of the input molecules.

The number of records in an input file is limited to **4.29 x 10<sup>9</sup>**.

2D or 3D atomic coordinates are not generated (e.g., when reading a SMILES input file and writing a MDL SDF file). To add 2D or 3D atomic coordinates in a post-processing step the software systems **MN.2DCOOR** [3] or **CORINA** [4] are recommended.

## 9 Program Installation

### 9.1 On-line Download from Molecular Networks' Web Server

**MN.TAUTOMER** 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 and example files of **MN.TAUTOMER** 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.

**MN.TAUTOMER** is currently available for the following hardware platforms and operating systems.

| hardware platform<br>operating system          | downloaded file name  |
|--|---|
| SGI workstations<br>IRIX 6.5                   | <i>tautomer_&lt;eval&gt;_IRIX6.5_&lt;YYYY_MM_DD&gt;.sgi.gz</i>  |
| x86 Linux workstations<br>kernel 2.4 (gcc 3.2) | <i>tautomer_&lt;eval&gt;_Linux2.4_&lt;YYYY_MM_DD&gt;.lnx.gz</i> |
| Sun SPARC stations<br>Solaris9 (SunOS 5.9)     | <i>tautomer_&lt;eval&gt;_SunOS5.8_&lt;YYYY_MM_DD&gt;.sun.gz</i> |
| MS Windows platforms<br>win32: NT4/2000/XP     | <i>tautomer_&lt;eval&gt;_Win32_&lt;YYYY_MM_DD&gt;.exe.gz</i>    |

Please uncompress the downloaded file of **MN.TAUTOMER**

*tautomer\_<eval>\_<OS>\_<YYYY\_MM\_DD>.sgi|sun|lnx|exe.gz*

(*eval* = evaluation version; *OS* = operating system; *YYYY\_MM\_DD* = expiration date in the format year\_month\_day)

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

## 9.2 New Installation

### 9.2.1 UNIX Systems (SGI, Sun SPARC, x86 Linux)

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

To install the command line version of **MN.TAUTOMER** (*tautomer.sgi.sun.lnx*) please follow the instructions below.

- 1) Create a subdirectory, e.g., *tautomer*, (for system administrators when installing software locally, e.g., */usr/local/bin/tautomer*).
- 2) Copy the executable file of **MN.TAUTOMER** *tautomer.sgi.sun.lnx* to the subdirectory *tautomer* and rename the file *tautomer.sgi.sun.lnx* to *tautomer*.  
**Please note:** *tautomer.sgi.sun.lnx* is a binary file.
- 3) Add the *tautomer* subdirectory name to the environment variable *PATH* in your *.login* or *.cshrc* files (*.profile* or *.bashrc*).

### 9.2.2 Microsoft Windows Platforms (win32: 2000/XP)

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

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

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

### 9.3 Program Updates

- 1) Before installing the new version, please copy the old executable to a new directory, e.g. *tautomerVV* (*VV = old-version-number*, e.g., *tautomer10*).
- 2) According to the hardware platform install the new version following the installation instructions given in Section 9.2 on page 23.

## 10 Problems and Help!

If you have any difficulties with the installation of **MN.TAUTOMER** or if you encounter any problems when running **MN.TAUTOMER**, 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](mailto:support@molecular-networks.com),

or by Fax +49-9131-815669.

Please send the input file, the output file, error and warning messages generated by **MN.TAUTOMER** and your system specifications to us by email. These files will help us to analyze your problem. Thank you!

You can also use the report form in Section 13 on page 28 of this manual.

## 11 Acknowledgements

**MN.TAUTOMER** was developed in the research group of Prof. Johann Gasteiger at the University of Erlangen-Nürnberg since 2001. The program is based on the CACTVS system initiated and developed by Dr. Wolf-Dietrich Ihlenfeldt. It is not part of the standard CACTVS toolkit distribution.

## 12 References

- [1] 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>.
- [2] 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>).
- [3] 2D Coordinates Generator **MN.2DCOOR**, developed and distributed by Molecular Networks GmbH, Erlangen, Germany (<http://www.molecular-networks.com>).
- [4] 3D Structure Generator **CORINA**, developed and distributed by Molecular Networks GmbH, Erlangen, Germany (<http://www.molecular-networks.com>).

### 13 Report Form

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

Molecular Networks GmbH Computerchemie  
Henkestr. 91  
91052 Erlangen, Germany  
FAX: +49-9131-815669

---

User:

---

**MN.TAUTOMER** program and version number ("tautomer -version"):

Command line to run **MN.TAUTOMER**:

Error and warning messages by **MN.TAUTOMER**:

---

System messages:

---

Short description: