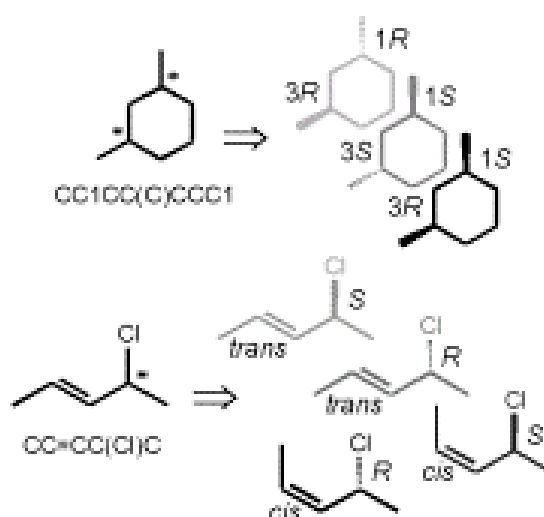


STERGEN

Stereoisomer Generator

Version 1.1

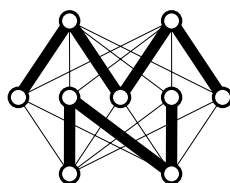
Program Description



Jens Sadowski, Christof H. Schwab and Johann Gasteiger

Molecular Networks GmbH Computerchemie
March 2003

<http://www.mol-net.de>



Molecular Networks GmbH
Computerchemie
Nägelsbachstr. 25
D-91052 Erlangen
Germany

Telephone: +49-(0)9131-815668
Telefax: +49-(0)9131-815669

Email: info@mol-net.de

This document is copyright © 2003 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.

Contents

1	Introducing STERGEN	1
2	Program Installation	2
2.1	SUN Sparc Stations	2
2.2	Silicon Graphics Workstations	3
2.4	IBM compatible systems – Linux	3
2.4	IBM compatible systems – Microsoft Windows NT 4.0 (win32)	3
2	Problems and Help!	4
3	Getting Started with STERGEN	5
4	Using STERGEN	7
5.1	Synopsis	7
5.2	Options	7
5	File Formats	10
5.1	MDL Structure Data File (SDF)	10
5.2	SMILES	10
5.3	Gasteiger Cleartext Format (CTX)	11
7	Genration of Stereoisomers	12
6	References	14
7	Report Form	15

1 Introducing STERGEN

Accessing the Configurational Space of Molecules

A substantial step towards the understanding of the physical, chemical, or biological properties of a molecule is to study and to analyze its spatial shape. Besides the constitution, a major shape determining features are the configuration of a molecule, i.e. the stereochemistry. **STERGEN** is a new software tool providing access to the configurational space of molecules.

Molecular chirality plays a major role in many areas of chemistry. Enantiomers often exhibit quite different physical, chemical, and biological properties. The exploration of the configurational space of a molecule and the analysis of the various isomers a molecule can adopt is therefore of great importance. Starting from a given connection table or linear string representation of a molecule, the stereoisomer generator **STERGEN** automatically identifies stereocenters and generates a set of isomeric compounds. It provides full or partial configurational searches, identifies and removes duplicate configurations (e.g., *meso*-compounds) and conformationally strained configurations and generates 2D coordinates, wedge symbols, and stereodescriptors.

2 Installing STERGEN

STERGEN verion 1.0 *alpha* is distributed on a CD-ROM, which contains the executable file(s) of STERGEN, this program description in PDF format, and some example files of structure information (see chapter 4 on page 5).

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

name of directory	description	name of file(s)
examples	example files for structure data (SMILES format)	<i>examples.sml</i>
manual	this program description in PDF format	<i>stergen10manual.pdf</i>

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

2.1 Sun SPARC Stations

The directory *sun* on the CD-ROM contains the Sun Solaris 2.6 executable file.

- 1) Create a subdirectory, e.g., *stergen* (for system administrators when installing software locally, e.g. */usr/local/bin/stergen*).
- 2) Copy the file *stergen.sun* from the CD-ROM to the subdirectory *rotate* and rename the file *stergen.sun* to *stergen*. **Please note:** *stergen.sun* is a binary file.
- 3) Add the *stergen* subdirectory name to the environment variable *PATH* in your *.login* or *.cshrc* files.

2.2 Silicon Graphics Workstations

The directory *sgi* on the CD-ROM contains the Silicon Graphics (IRIX) executable file.

- 1) Create a subdirectory, e.g., *stergen* (for system administrators when installing software locally, e.g. */usr/local/bin/ stergen*).
- 2) Copy the file *stergen.sgi* from the CD-ROM to the subdirectory *rotate* and rename the file *stergen.sgi* to *stergen*. **Please note:** *stergen.sgi* is a binary file.
- 3) Add the *stergen* subdirectory name to the environment variable *PATH* in your *.login* or *.cshrc* files.

2.3 IBM Compatible Systems - Linux

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

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

2.4 IBM Compatible Systems - Microsoft Windows NT 4.0 (win32) (not yet supported for version 1.1)

The directory *win32* on the CD-ROM contains the Win32 executable file.

- 1) Create a subdirectory, e.g., *stergen* (for system administrators when installing software locally, e.g. *X:\programs\ stergen*).
- 2) Copy the file *stergen.exe* from the CD-ROM to the subdirectory *stergen*. **Please note:** *stergen.exe* is a binary file.
- 3) Add the *stergen* executable file to the system variable *Path*.

3 Problems and Help!

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

Molecular Networks GmbH Computerchemie
Nägelsbachstr. 25
91052 Erlangen
Germany

or contact us by email support@mol-net.de,

or by Fax +49-(0)9131-815669.

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

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

4 Getting Started with STERGEN

The example file *example.sdf* submitted with the distribution contains the structure information of three molecules in MDL SDFfile format, [1] the default file format of STERGEN.

Copy this example file into your working directory and type the following command:

stergen example.sdf out.sdf

STERGEN now creates the output file *out.sdf* (in MDL SDFfile format) containing all possible stereoisomers starting from the given input constitution. Figure 1 shows one example structure.

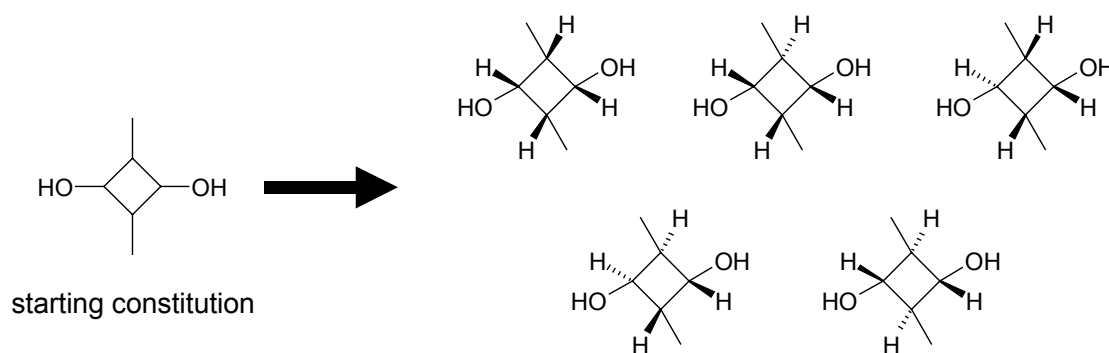


Figure 1 Generated configurations from one single initial constitution.

The additional option **-t s** redirects the trace information to standard output device. With the command

corina -t s example.sdf out.sdf

the following output appears on the screen:

```
stergen <serial number - compilation date - user - date - time - host>

INFORMATION stergen(): Input file type MDL SDFfile
INFORMATION stergen(): Output file type MDL SDFfile
INFORMATION stergen(): Parameters:
  Maximum number of stereo centers: 10
  Maximum number of stereo isomers: 1024
standard values, v2.6, December 2000

*** RECORD no.: 1 read *****
  Ident 1
  Name raz_01
  2 stereo isomer(s) written
  2 totally generated
```

```
0 rejected due to 3D incompatibilities
0 duplicate configuration(s) detected
Elapsed time: 0 ms

*** RECORD no.: 2 read *****
Ident 2
Name raz_02
36 stereo isomer(s) written
64 totally generated
0 rejected due to 3D incompatibilities
28 duplicate configuration(s) detected
Elapsed time: 40 ms

*** RECORD no.: 3 read *****
Ident 3
Name raz_03
16 stereo isomer(s) written
32 totally generated
0 rejected due to 3D incompatibilities
16 duplicate configuration(s) detected
Elapsed time: 10 ms

.
.
.

*** RECORD no.: 67 read *****
Ident 67
Name raz_67
2 stereo isomer(s) written
16 totally generated
14 rejected due to 3D incompatibilities
0 duplicate configuration(s) detected
Elapsed time: 0 ms

*** RECORD no.: 68 read *****
Ident 68
Name raz_68
1 stereo isomer(s) written
16 totally generated
14 rejected due to 3D incompatibilities
1 duplicate configuration(s) detected
Elapsed time: 10 ms

*** RECORD no.: 69 read *****
Ident 69
Name raz_69
9 stereo isomer(s) written
64 totally generated
0 rejected due to 3D incompatibilities
55 duplicate configuration(s) detected
Elapsed time: 10 ms

69 record(s) read, 69 converted
Totally elapsed time: 1 s
```

5 Using STERGEN

5.1 Synopsis

The general synopsis for using STERGEN is:

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

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

By default, STERGEN identifies all stereocenters in a given input structure and generates all possible, but unique stereoisomers. The output file contains the 2D coordinates of the structure and the stereochemical information coded as stereodescriptors and wedge symbols. **Please note:** Any 3D information is lost during the isomer generation process.

5.2 Options

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

-i Input file

t=<value>	Set the input file type. Allowed values are: sdf—MDL SDFfile [1] (default), smiles—SMILES linear code [2], ctx—Gasteiger ClearText file format [3].
sdfi2n=<value>	SDF item to name: For MDL SDFfiles—copy the data item <value> to the compound name.
dummies	Allow dummy type atoms in SMILES and MDL SDFfile input files.
sdfict	Ignore <i>cis/trans</i> configuration of double bonds in MDL SDF input files.

-o Output file

- t=<value> Set the output file type. Allowed values are:
sdf—MDL SDFFile [1] (default),
ctx—Gasteiger ClearText file format [3].
- a Append the output to the output file instead of creating a new output file.

-t Trace

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

-n Record number

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

-d STERGEN driver options

- mc=<value> Set the maximum number of stereocenters, which should be processed to <value> in order to restrict the number of generated isomers. By default, STERGEN processes a maximum number of four stereocenters of an input structure. This option forces STERGEN to process the given number <value> of stereocenters.
- mi=<value> Set the maximum number of stereoisomers, which should be generated to <value>. By default, STERGEN generates a maximum of 16 stereoisomers (maximum of four stereocenters), which are unique. This option forces STERGEN to generate only the given number <value> of stereoisomers.

- preserve Preserve defined stereocenters. If the configuration of an input structure is not specified for all chiral centers and *cis/trans* double bonds, this option forces STERGEN to process only those centers which do not have a defined stereochemistry or configuration.
- duplicates Retain all duplicate configurations. By default, STERGEN identifies identical configuration, e.g. *meso*-compounds and only unique isomeric structures are written to the output file. This option forces STERGEN to output all configurations which are generated.
- nowedges Suppress the generation of wedge symbols. By default, STERGEN generates wedge symbols as well as stereodescriptors and 2D coordinates and writes them to the output file. This option forces STERGEN to suppress the generation of wedge symbols and only stereodescriptors are written to the output file.
- v Print program version**
- h Print on-line help**
- i Print online help for input file specifications.
- o Print online help for output file specifications.
- t Print online help for trace output specifications.
- n Print online help for record number specifications.
- d Print online help for driver specifications.
- all Print the complete online.

6 File Formats

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

Table A Input/output file formats.

Format	Input	Output	Reference
MDL SDF	x	x	[1]
SMILES	x		[2]
CTX	x	x	[3]

6.1 MDL Structural Data File (SDF)

The MDL SDF file format was implemented following the description in the literature [1]. STERGEN reads only the name (line 1), the status line (line 2), the atom and bond counts (line 4), the atom block, and the bond block. In addition, the RAD and CHG atom properties from the properties block are read in. From the atom block, columns 1 through 7 are read in. From the bond block, columns 1 through 4 are read in. All other information remains unread but is piped unchanged to the output when the output file type is also set to MDL SDF file. This is a great advantage for database purposes since all information except the stereodescriptors remains unchanged. The individual records are assumed to be closed by a \$\$\$\$ delimiter.

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

6.2 SMILES

The SMILES linear notation was implemented following the literature [1]. In addition, non-standard formal charge qualifiers immediately following the atomic symbol in curly brackets have been implemented. As lower case aromatic atoms only C, N, O, and S are allowed. Implicit hydrogen atoms given inside square brackets are expanded and written to the output file whereas all other missing implicit hydrogen atoms are not written to the output file. The first 80 characters of the SMILES string are copied into the compound name. Alternatively, the option **-i smilesname** allows to interpret the first string in the line as SMILES linear code and all following strings in this line as the compound name. **Please note:** Every line has to begin with a SMILES linear code and only one SMILES string per line is allowed.

6.3 Gasteiger ClearText File Format (CTX)

CTX is a keyword oriented ASCII format developed in the research group of Prof. Dr. Johann Gasteiger [3]. The following keywords are read in: IDENT, NAME, MOLECULS, ATOMS, BONDS, BLABEL, 2DCOORD, HIGEOM, INTCOORD, and END. These keywords and all additional information are directly piped to the output when the output file type also was set to CTX. The additional keywords STEREO and BLABEL is written the stereodescriptors of the isomer under consideration.

7 Generation of Stereoisomers

STERGEN automatically identifies stereocenters and is able to generate all possible, but unique and chemically reasonable, isomeric compounds starting from a given connection table or linear string representation of a molecule. Tetrahedral chiral centers as well as *cis/trans* isomerism is taken into account. Duplicate configurations (e.g., *meso*-compounds) and conformationally strained configurations (e.g., the *i,o*-isomer of norbornane) are rejected. Wedge symbols and/or stereodescriptors are generated and written to the output file. Figure 2 shows some examples.

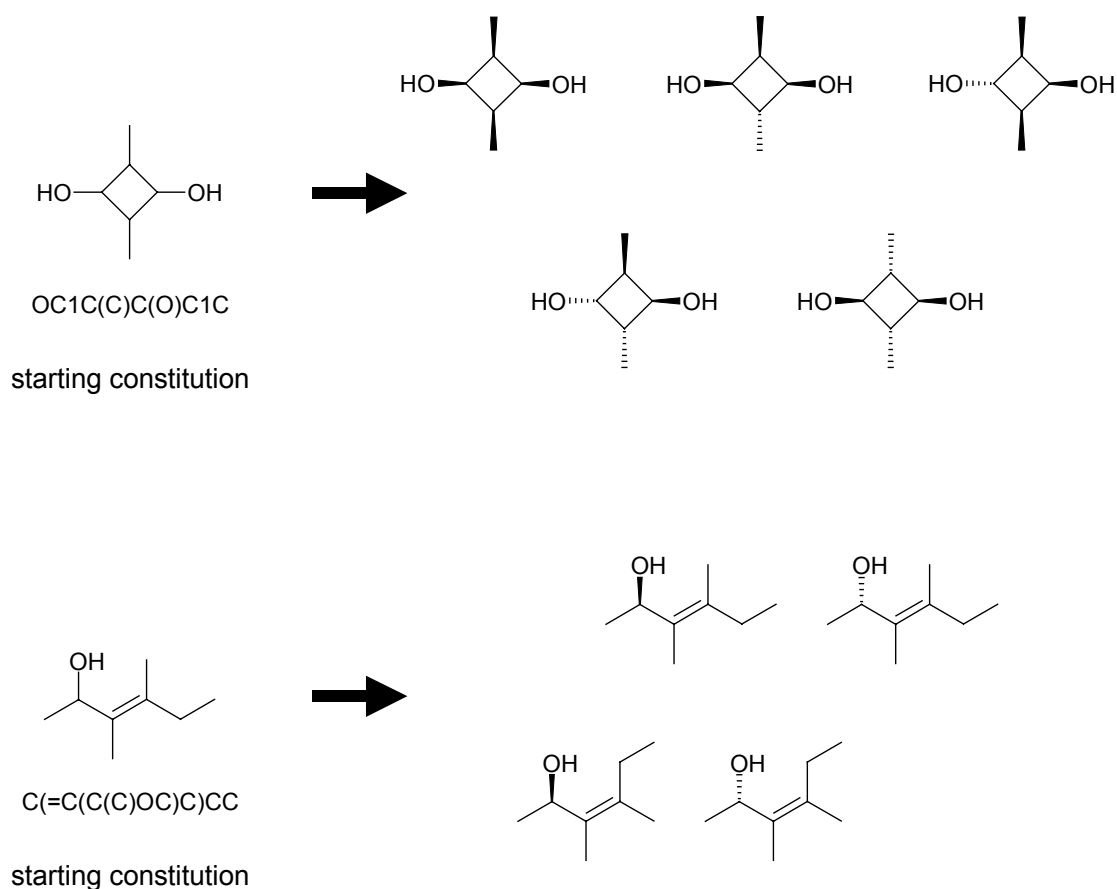


Figure 2 Generated configurations from one single initial constitution.

By default, a maximum number of four stereocenters are processed and a maximum number of 16 stereoisomers per molecule are generated. However, as the number of possible isomers exponentially increases with the number of stereocenters (2^n , where n is the number of stereocenters in a molecule), additional options allow to set the

maximum number of output isomers (**-d ms=<value>**, maximum number of stereoisomers) or to define a maximum number of stereocenters which should be processed (**-d mc=<value>**, maximum number of stereocenters). Furthermore, the option **-d preserve** allows to retain the configuration at atoms which have a defined stereochemistry (i.e., a stereodescriptor is given in the input file).

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

Command line example. The following command line

stergen -d ms=20,mc=6,preserve <in> <out>

generates a maximum of twenty stereoisomers per molecule (**ms=20**) by processing a maximum of six stereocenters (**mc=6**) and preserves stereocenters which have a defined stereochemistry (i.e., a stereodescriptor is given in the input file, **preserve**).

8 References

- [1] 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.
- [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.
- [3] Gasteiger, J. et al. CTX Keyword Reference Manual. University of Erlangen-Nürnberg: 1995, unpublished results.

9 Report Form

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

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

User:

STERGEN program and version number ("stergen -v"):

Command line to run STERGEN:

Error and warning messages by STERGEN:

System messages:

Short description:

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

