



Splitting Chemical Files

User Manual

Version 1.0

for program version 1.0 (or higher)

November 2003

---

Molecular Networks GmbH – Computerchemie  
Nägelsbachstraße 25, 91052 Erlangen, Germany



Molecular Networks GmbH

Computerchemie

Nägelsbachstr. 25

91052 Erlangen

Germany

Phone: +49-(0)9131-815668

Fax: +49-(0)9131-815669

Email: [info@mol-net.de](mailto:info@mol-net.de)

WWW: [www.mol-net.de](http://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.SPLIT .....	4
2.	Installation .....	5
2.1.	Requirements.....	5
2.2.	Installation Steps for UNIX Operating Systems (IRIX, Solaris, Linux) .....	5
2.3.	Installation Steps for Microsoft Windows Operating Systems (NT4/2000/XP).....	5
3.	Uninstallation.....	5
3.1.	Uninstallation Steps for UNIX Operating Systems (IRIX, Solaris, Linux).....	5
3.2.	Uninstallation Steps for Microsoft Windows Operating Systems (NT4/2000/XP) ..	5
4.	Problems and Help!.....	6
5.	Release Notes .....	7
5.1.	Version 1.0 .....	7
6.	Getting Started.....	8
6.1.	UNIX operating systems.....	8
6.2.	Microsoft Windows operating systems.....	8
7.	Program Use .....	9
7.1.	Synopsis .....	9
7.2.	General Program Features.....	10
7.3.	Supported file formats for input files.....	10
7.4.	Program Features in More Detail .....	10
	-format <abbreviation of the output format> .....	10
	-outfile <filename.extension> .....	11
	-directory <dirname> .....	11
	-feedback 0/n .....	11
	-stat .....	12
	-version .....	12
	-h or -help.....	12
8.	Extended Features Only Available for the UNIX Operating Systems.....	12
9.	Frequently Asked Questions (FAQ) .....	13
10.	Error Messages .....	13
11.	Known Problems and Limitations .....	13
12.	Technical Support .....	14
	The MN.SPLIT Web Site.....	14
	Reporting Problems.....	14
	Updates.....	14
	Contact Information .....	14
13.	Report Form.....	15
14.	Index.....	16

## 1. General Information about MN.SPLIT

MN.SPLIT splits a file including n structures into n files.

The program MN.SPLIT

- reads structure files saved in SD file format
- processes datasets with 99.9% conversion rate
- handles datasets of hundreds of thousands of chemical structures
- supports the SD file format for saving the output files

## 2. Installation

### 2.1. Requirements

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

Launch MN.SPLIT with the command

```
mn_split -version or /usr/local/bin/mn_split/mn_split -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_split` in the installation directory carefully (default path during installation was `/usr/local/bin/mn_split/`).

### 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.SPLIT or if any problems occur while running MN.SPLIT, please send all your inquiries to the following address:

Molecular Networks GmbH Computerchemie  
Nägelsbachstr. 25  
91052 Erlangen  
Germany,

or contact us by email  
or by fax

support@mol-net.de,  
+49-(0)9131 - 81 56 69.

Please mention the program version of MN.SPLIT (`mn_split -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.SPLIT

## 6. Getting Started

### 6.1. UNIX operating systems

The example file `alkanes1_12.sdf` submitted with the distribution contains the structure information of twelve molecules in SD format. Copy this example file into your working directory and type the following command:

```
mn_split alkanes1_12.sdf
```

MN.SPLIT now creates the output files named `alkanes_<n>.mdl` written to the same directory where the file `alkanes1_12.sdf` is located.

### 6.2. Microsoft Windows operating systems

The example file `alkanes1_12.sdf` submitted with the distribution contains the structure information of twelve molecules in SD 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_split` by using the `cd` command then type the following command:

```
mn_split alkanes1_12.sdf
```

MN.SPLIT now creates the output file named `alkanes_<n>.mdl` written to the same directory where the file `alkanes1_12.sdf` is located.

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_split -directory /tmp alkanes1_12.sdf
```

```
Windows: mn_split -directory C:/temp alkanes1_12.sdf
```

## 7. Program Use

### 7.1. Synopsis

The general synopsis for using MN.SPLIT is:

```
mn_split [ -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.

<b>[-directory dirname]</b>	specifies the output directory
<b>[-feedback 0/n]</b>	prints a control message after processing a block of n items
<b>[-format fmt]</b>	specifies the output format name
<b>[-h] or [-help]</b>	shows a brief help message about the usage of the program
<b>[-stat]</b>	writes statistical information about the number of successfully processed records and processing failures
<b>[-outfile filename]</b>	defines the name of the output file
<b>[-version]</b>	prints version and licensing information

Table 1: Overview of all options

Executing the program without any option will write each structure of the input file into a separate output file. Please note that a record count string in the form `_<number>` will be inserted before the suffix to indicate the original record number. Such record numbers always begin with number 1.

## 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.SPLIT 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. Supported file formats for input files

The program will automatically detect the supported file format of the input files. Thus, there is no need for a parameter specifying the input format.

The supported file format is listed in the table below.

Full Format Name	Default Input-Extension	Read	Comment
MDL SDF	sdf	Yes	

Table 2: Overview of the supported input file formats

## 7.4. Program Features in More Detail

### **-format <abbreviation of the output format>**

The parameter **format** is specified for selecting the output format.

The supported file formats are listed in the table below.

Full Format Name	Default Output-Extension	Write	Comment
MDL Molfile	mol	Yes	
MDL SDF	mdl	Yes	

Table 3: Overview of the supported output 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_split -format sdf ./examples/alkanes1_12.sdf
```

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.SPLIT automatically recognizes the desired output format, thus in most cases it is not necessary to specify the output format.

If you are using MN.SPLIT 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_split -outfile c.sdf ./examples/alkanes1_12.sdf
```

Remark:

Please note that the new file names are c\_1.sdf, c\_2.sdf, c\_3.sdf, ...

#### **-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_split -outfile c.sdf -directory /tmp/alkanes  
./examples/alkanes1_12.sdf
```

```
Windows: mn_split -outfile c.sdf -directory C:/Temp/alkanes  
./examples/alkanes1_12.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 fifth records:

```
mn_split -outfile c.sdf -feedback 5 ./examples/alkanes1_12.sdf
```

### **-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_split -outfile c.sdf -stat ./examples/alkanes1_12.sdf
```

Output:

```
Convert file alkanes1_12.sdf
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_split -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_split -h or mn_split -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**

## **11. Known Problems and Limitations**

## **12. Technical Support**

### **The MN.SPLIT Web Site**

If you have problems while running MN.SPLIT please have a look at the Support- and FAQ web site of MN.SPLIT. 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.SPLIT 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.SPLIT with maintenance you will automatically receive updates every time a new release is launched.

### **Contact Information**

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

Molecular Networks GmbH  
Computerchemie  
Nägelsbachstraße 25  
91052 Erlangen  
Germany

e-Mail: [support@mol-net.de](mailto: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.SPLIT, 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.SPLIT program and version number (mn\_split -version):

---

Command line to run MN.SPLIT:

---

Error and warning messages by MN.SPLIT:

---

System messages:

---

Short description:

---

Please include the input file and output file generated by MN.SPLIT 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.sdf 8, 11, 12  
option  
  directory 11  
  feedback 11  
  format 10  
  h 12  
  help 12  
  outfile 11  
  stat 12  
  version 12  
outputfile  
  alkanes1\_12.sdf 11  
  c.sdf 11, 12