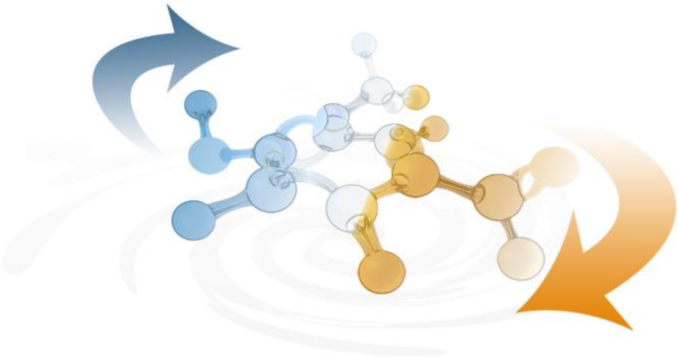
ROTATE Classic

Conformer Generator
For Acyclic Molecules And Fragments

Version 1.6

Program Description



Molecular Networks GmbH Computerchemie
April 2016
www.mn-am.com

Molecular Networks GmbH
Neumeyerstr. 28
90411 Nuremberg
Germany

Altamira LLC
1455 Candlewood Drive
Columbus, OH 43235-1623
USA

mn-am.com

This document is copyright © 2007-2016 by Molecular Networks GmbH Computerchemie and Altamira LLC. All rights reserved. Except as permitted under the terms of the Software Licensing Agreement of Molecular Networks GmbH Computerchemie or Altamira LLC, 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 or Altamira LLC.

The software described in this document is furnished under a license and this document may be used and copied only in accordance with the terms of such license. (Doc version: 1.0-2016-04-25)

Content

1 Introducing ROTATE Classic 1

2 Installing ROTATE Classic 2
  2.1 Download from the Web Server of Molecular Networks 2
  2.2 Linux 2
  2.3 Microsoft Windows 3

3 Reporting Problems with ROTATE Classic 4

4 Getting Started with ROTATE Classic 5

5 Using ROTATE Classic 7
  5.1 Synopsis 7
  5.2 Options 7

6 Generating Multiple Conformations by CORINA Classic and ROTATE Classic 11
  6.1 Generation of Multiple Ring Conformations with CORINA Classic 11
  6.2 Generation of Multiple Conformation for Acyclic Parts with ROTATE Classic 12
  6.3 Recommendations 12

7 Understanding ROTATE Classic 13
  7.1 Program Scope 13
  7.2 The Algorithm: A Short Overview 13
  7.3 Assignment of Rotatable Bonds 14
  7.4 The Torsion Angle Library 15
  7.5 Generation of Conformations 17
  7.6 Classification of Conformations 20
    7.6.1 Classification in Cartesian Space 20
    7.6.2 Classification in Torsion Angle Space 21
    7.6.3 Comparison of the Classification Methods 21
  7.7 The Automatic Run Mode "-d auto" 22
  7.8 A Case Study: 2-L-Benzylsuccinate 22

8 File Formats 31
  8.1 MDL Structural Data File (SDF) 31
  8.2 SYBYL File Formats 31
  8.3 Brookhaven Protein Data Bank Format (PDB) 32
  8.4 Gasteiger ClearText File Format (CTX) 32

9 Error Messages 33
- 9.1 General Errors 33
- 9.2 Input File Format Errors 34
  - 9.2.1 MDL SDFile 34
- 9.3 Errors in the Generation and Classification of Conformations 35

10 Warning Messages 37
- 10.1 General Warnings 37
- 10.2 Warnings during the Generation and Classification of Conformations 37

11 Acknowledgments 38

12 References 39

13 Report Form 41

1 Introducing ROTATE Classic

Addressing Conformational Flexibility

Most molecules of organic, biochemical, or pharmacological interest can adopt more than one conformation of nearly equal energy by rotation around single bonds. The interaction of the molecule with its environment, e.g. with the binding site of a biological receptor, may determine which of these conformations is the preferred one. Therefore, conformational flexibilty has to be taken into account in the development of new compounds with certain biological or chemical properties.

Small ring systems are orders of magnitude more rigid than open-chain portions and therefore, the number of low-energy conformations is much smaller. Thus, in our rule and data based approach to addressing conformational flexibility, cyclic and acyclic molecules and fragments are treated separately.

Cyclic Structures. The 3D structure generator CORINA Classic is able to perform a systematic conformational analysis for ring systems of up to a size of nine ring atoms. CORINA is using a list of different ring templates derived from statistical and empirical data to generate a set of different ring geometries. These ring templates are stored as lists of torsion angles, for each ring size and number of unsaturations in the ring, ordered by their conformational energy.

Acyclic Structures. The rule and data based program system ROTATE Classic, which is described in this program manual, was developed to explore the conformational space of acyclic parts in medium-sized (drug-like) molecules.

Starting with a given 3D structure, ROTATE Classic generates a set of multiple conformations by rotation around bonds, which are regarded to be rotatable. The program uses a set of rules and data derived from a statistical analysis of the conformational preferences of open-chain fragments in small molecule crystal structures. For this purpose the **Cambridge Structural Database** (CSD, developed and distributed by the Cambridge Crystallographic Data Centre, Cambridge, UK, http://www.ccdc.cam.ac.uk) was used. This analysis led to sets of preferred dihedral angles for fragments consisting of four atoms. These torsion angle patterns are stored in a library, the so-called Torsion Angle Library. Thus, ROTATE Classic generates reasonable geometries which have been observed, rather than sampling local energy minima or generating conformations randomly.

The number of conformations drastically increases with the number of rotatable bonds. In order to restrict the number of generated conformers, while retaining a maximum of structural diversity, ROTATE Classic offers two different methods for classifying the rotamers. Each class thus obtained is represented by a single conformation.

2 Installing ROTATE Classic

2.1 Download from the Web Server of Molecular Networks

Since version 1.6, ROTATE Classic is available for electronic download via the Internet on the web server of Molecular Networks (Download Area). An account can be requested at

www.mn-am.com/php/profile.php

that provides access to licensed software, evaluation copies, program manuals, example files, and tutorials of ROTATE Classic 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, such as WinZip, FileZip (http://www.filezip.com), or gzip (http://www.gzip.org)

ROTATE Classic is currently available for the following hardware platforms and operating systems.

Hardware platform & operating system	Download file name
x86 Cent OS Linux 6 (gcc 4.4.7, libc 2.12, 32 bit)	*rotate_<annual|eval>_x86-32_rhel6_<date>.lnx.gz*
Microsoft Windows win32, 7/8/10	*rotate_<annual|eval>_Win32_<date>.exe.gz*

corina_<annual|eval>_<OS>_<Date>.lnx|exe.gz

(*annual* = licensed version with annual run time; *eval* = evaluation version)

2.2 Linux

1) Create a subdirectory, e.g., *rotate* (for system administrators when installing software locally, e.g. */usr/local/bin/rotate*).
2) Uncompress the file *rotate_<annual|eval>_x86-32_rhel6_<date>.lnx.gz* and copy the resulting file *rotate_<annual|eval>_x86-32_rhel6_<date>.lnx* into the subdirectory *rotate*.
3) Rename the file *rotate_<annual|eval>_x86-32_rhel6_<date>.lnx* to *rotate*. **Please note:** *rotate.lnx* is a 32-bit binary file.
4) Add the *rotate* subdirectory name to the environment variable *PATH* in your *.login* or *.cshrc* files (*.profile* or *.bashrc*).

2.3 Microsoft Windows

1) Create a subdirectory, e.g., *rotate* (for system administrators when installing software locally, e.g. *C:\Program Files (x86)\rotate*).
2) Uncompress the file *rotate_<annual|eval>_win32_<date>.exe.gz* and copy the resulting file *rotate_<annual|eval>_win32_<date>.exe* into the subdirectory *rotate*.
3) Rename the file *rotate_<annual|eval>_win32_<date>.exe* to *rotate.exe* **Please note:** *rotate.exe* is a 32-bit binary file.
4) Add the *rotate* executable file (*rotate.exe*) and the path where the program resides (e.g. *C:\Program Files (x86)\rotate\rotate.exe*) to your environment variable PATH of your system settings.

3 Reporting Problems with ROTATE Classic

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

Molecular Networks GmbH Computerchemie
Neumeyerstr. 28
90411 Nuremberg, Germany

or contact us by email	support@mn-am.com,
or by FAX	+49-911-597 424 09.

Please include your input file, the output file, and the ROTATE trace file *rotate.trc* generated by ROTATE and send it to us by email to support@mn-am.com. 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 ROTATE Classic

The file *example.sdf* submitted with the distribution contains the connection tables and the 3D structure information on the amino acids valine, leucine, and methionine in MDL SDFile format [1], the default file format of ROTATE.

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

rotate example.sdf out.sdf

ROTATE now creates the output file *out.sdf* containing a set of conformations for each amino acid. Figure 1 shows the superposition of all generated rotamers separately for each amino acid.

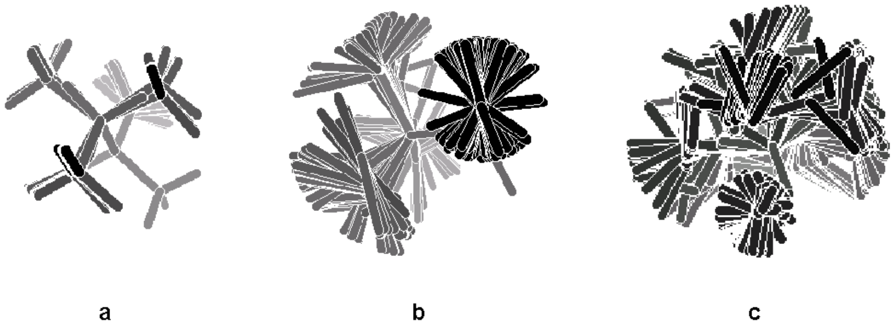


Figure 1 Superimposition of the generated conformers of the amino acids valine (a, with hydrogen atoms), leucine (b), and methionine (c) (b and c without hydrogen atoms).

Retype the command with the additional option **-t s**, which redirects the trace information to standard output device.

rotate -t s example.sdf out.sdf

The following output appears on the screen:

```
INFORMATION rotate(): Input file type MDL SDFile
INFORMATION rotate(): Output file type MDL SDFile
INFORMATION rotate(): Standard values, version 3.4, January 2006
INFORMATION rotate(): Ring templates, version 3.0, March 2001
INFORMATION rotate(): TA library, version 1.0, December 1999 (1088 TA
                      patterns)
                      All rights CCDC, UK (www. ccdc.cam.ac.uk)

INFORMATION rotate(): User defined driver options:
        processing all rotatable bond(s)
        removing duplicate conformations
        no classification method specified


*** Record No.:  1 read **********
    Ident: 2
    Name:  valine

INFORMATION generate(): Found 2 rotatable bond(s).
INFORMATION generate(): Processing all rotatable bonds.
INFORMATION confanalys(): Using a maximum of 6 preferred TA values per
                          rotor.
INFORMATION duplicate(): Removing duplicate conformations.
INFORMATION confanalys(): 36 conformations generated,
                          9 classes found.

   Elapsed time: 40 msec

*** Record No.:  2 read **********
    Ident: 3
    Name:  leucine

INFORMATION generate(): Found 3 rotatable bond(s).
INFORMATION generate(): Processing all rotatable bonds.
INFORMATION confanalys(): Using a maximum of 6 preferred TA values per
                          rotor.
INFORMATION duplicate(): Removing duplicate conformations.
INFORMATION confanalys(): 208 conformations generated,
                          71 classes found.

   Elapsed time: 330 msec

*** Record No.:  3 read **********
    Ident: 03
    Name:  methionine

INFORMATION generate(): Found 4 rotatable bond(s).
INFORMATION generate(): Processing all rotatable bonds.
INFORMATION confanalys(): Using a maximum of 6 preferred TA values per
                          rotor.
INFORMATION duplicate(): Removing duplicate conformations.
INFORMATION confanalys(): 1226 conformations generated,
                          247 classes found.

   Elapsed time: 1880 msec

   Statistics on Conformation Generation

   3 record(s) read, 3 processed.

   Total number of generated conformations:      327
   Average values per record:
     Number of conformations:                  109.0
     Number of rotatable bonds:                  3.0
     Number of processed rotatable bonds:        3.0

   Totally elapsed time: 2 sec (0.67 sec per molecule)
```

5 Using ROTATE Classic

5.1 Synopsis

The general synopsis for using ROTATE is:

```
rotate [ -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 interpreted as input file and the output will be written to standard output. Minimum trace output is by default written to the file *rotate.trc*.

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 SDFile [1] (default), mol—SYBYL MOL [2] mol2—SYBYL MOL2 [2], and ctx—Gasteiger ClearText file format [3].

-o **Output file**

t=<value>	Set the output file type. Allowed values are sdf—MDL SDFile [1] (default), mol—SYBYL MOL [2], mol2—SYBYL MOL2 [2], pdb—Brookhaven Protein Data Bank format [4], and ctx—Gasteiger ClearText file format [3].
a	Append the output to the output file instead of creating a new output file.
pdbatom	PDB ATOM: For PDB files—write ATOM keywords instead of HETATM keywords.

-t **Trace**

s	Write trace output to the standard error channel (default: *rotate.trc*).

n	Suppress trace output. This option is useful for the conversion of large datasets or databases since the trace file *rotate.trc* might become rather large.
tracefile=<value>	Set the trace file name (default: *rotate.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 ROTATE driver options

auto	Set all driver options automatically. This option enables the fully automatic mode of ROTATE. Each input structure is analyzed for its rotatable bond and all important driver options (**rot**, **pta**, **dE**, **rms_xyz/rms_ta**, see below) are set automatically. **The usage of this option is highly recommended. Please note:** This driver option must **not** be used in combination with any other driver option (see also section 7.7 on page 22).
all	Process all rotatable bonds (default). All open-chain portions within the molecule are analyzed for rotatable bonds (see section 7.3 at page 14). This option forces ROTATE to process all bonds which are regarded as rotatable in order to generate multiple conformations. Thus, this is the most exhaustive method to explore the conformational space. **Please note:** Processing all rotors is the default unless no other suboption is specified, which restricts the number of rotatable bonds to be processed (**rot=<value>** or **bnd=<value>**, *vide infra*).
rot=<value>	Process only the <value> most central (inner most) rotatable bonds. Each bond which is considered as rotatable is examined and ordered according to its topological position within the molecular graph. If not all rotors are processed, the largest effect on the changes of the overall shape of the molecule can be obtained by rotation around the most central rotatable bonds (see section 7.3 at page 14).
bnd=<value>	Process only the bond with bond label <value>. This option forces ROTATE to process only a bond with a specific bond label given by the input structure file. This option is usefull if only a single molecule or a dataset of molecules of a homologeous series, where a specific bond in all structures is marked by the same bond label, is processed (see section 7.3 at page 14).

pta=<value>	Use the most <value> preferred torsion angles for each rotatable bond. By default, ROTATE uses the six most preferred torsion angle values per rotor found in the Torsion Angle Library (see section 7.5 on page 17). This option restricts the number of torsion angle values that are used per rotor to generate the initial geometries. Allowed values are 2, 3, 4, 5, and 6 (default). This option is useful for structures that have a large number of rotatable bonds. Instead of restricting the number of bonds to be rotated (see above for driver option rot=<value>) the user can lower the number of torsion angle values per rotor but process all rotors (or at least a higher number of rotors).
dE=<value>	Generate only conformations having an energy not higher than <value> kJ/mol with respect to the lowest energy conformation. This option reduces the number of output conformations: The lower the preset energy value, the lower the number of conformations which are written to the output file.
rms_xyz=<value>	Classify the conformations in Cartesian space by applying an RMS_{XYZ} threshold of <value> Å. After the generation process each conformer is compared with all previously generated conformations by superimposition in 3D space (see section 7.6.1 at page 20). Two conformers are regarded to exhibit the same geometry and are united into one family (cluster) if the RMS deviation of the Cartesian coordinates (RMS_{XYZ} in [Å]) of all non-hydrogen atoms does not exceed a threshold of <value> Å. **Please note:** There is no preset default value for <value>. The usage of this suboption requires the specification of <value> of type floating point number (see section 7.6.1 at page 20).
rms_ta=<value>	Classify the conformations in torsion angle space by applying an RMS_{TA} threshold of <value>°. After the generation process each conformer is compared with all previously generated conformations in torsion angle space (see section 7.6.2 at page 21). Two conformers are regarded to exhibit the same geometry and are united into one family (cluster) if the RMS deviation of the torsion angles (RMS_{TA} in [°]) of the rotated bonds does not exceed a threshold of <value>°. **Please note:** There is no preset default value for <value>. The usage of this suboption requires the specification of <value> of type floating point number (see section 7.6.2 at page 21).

-h ROTATE on-line help options

i	Print help concerning ROTATE input file options.
o	Print help concerning ROTATE output file options.

t	Print help concerning ROTATE trace file options.
n	Print help concerning ROTATE record number options.
d	Print help concerning ROTATE driver options.
all	Print help concerning all available ROTATE options.

-v **Print program version**

6 Generating Multiple Conformations by CORINA Classic and ROTATE Classic

Most of the molecules of interests are consisting of ring systems and open-chain portions. In order to generate a set of conformations for a molecule under consideration and to take into account the conformational flexibility of both cyclic and acyclic fragments, the Conformer Generator ROTATE and the 3D Structure Generator CORINA have to be used in combination [5]-[10]: CORINA for the generation of multiple ring conformations and ROTATE to process the acyclic parts of the molecule to generate a set of rotamers.

6.1 Generation of Multiple Ring Conformations with CORINA Classic

For small ring systems consisting of less than nine atoms, the number of reasonable conformations is rather limited. Thus, CORINA processes these systems by using a table of allowed single ring conformations, so-called ring templates. These ring templates are stored as lists of torsion angles, for each ring size and the number of unsaturations in the ring, ordered by their conformational energy. Figure 2 gives an example for the conformations of cyclohexane and cyclohexene as they are stored in the ring conformation table.

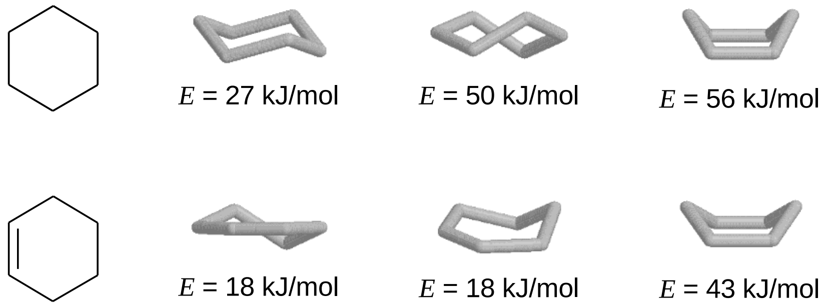


Figure 2 Ring templates for cyclohexane and cyclohexene used by CORINA for the generation of multiple ring conformations.

The different ring geometries can be used for a conformational analysis and for outputting a set of ring conformations. In the case of fused or bridged ring systems all possible, but structurally reasonable combinations of the different ring templates are generated. Only one conformations for each side chain is generated, which remains unchanged unless there are problems with non-bonded interactions caused by the different ring geometries. Thus, the output consists of a series of 3D models having different ring conformations and almost the same conformations of the acyclic part of the molecule. CORINA is forced by the driver option **-d rc** to write multiple ring conformations into the output file[11].

6.2 Generation of Multiple Conformation for Acyclic Parts with ROTATE Classic

The output of the multiple ring conformations generated by CORINA can then be used as input information for ROTATE. Each ring conformation is treated separately and remains unchanged. Now, only the side chains are processed in order to generate multiple conformations. The general procedures of ROTATE are described in chapter 7 on page 13.

6.3 Recommendations

To generate a set of conformations taking into account the conformational flexibility of both cyclic and acyclic fragments of a molecule, CORINA and ROTATE should be used in the following order:

1) Generation of multiple conformations for ring systems with the 3D Structure Generator CORINA, e.g. (see [11]):

   ```
   corina -d wh,rs,r2d,rc,mc=10,flapn,sc <infile2d> <corina3d_rc>
   ```

2) Generation of multiple conformations for the acyclic parts with the Conformer Generator ROTATE, e.g.:

   ```
   rotate -d auto <corina3d_rc> <rotate3d>
   ```

   or

   ```
   rotate -d rot=7,pta=4,rms_ta=45.0 <corina3d_rc> <rotate3d>
   ```

7 Understanding ROTATE Classic

7.1 Program Scope

ROTATE is applicable to the entire range of organic chemistry. Atoms with up to six neighbors can be handled according to the Valence Shell Electron Pair Repulsion (VSEPR) [12] model and all structures which can be expressed correctly in a *valence bond* notation can be processed [13]. There is no upper limit to the size of the molecules. Most of the methods for file format handling and the internal data structure for the handling of connection tables and 3D atom coordinates are based on the algorithms implemented in the 3D Structure Generator CORINA, one of the most general and robust 3D model builder which is presently available. Thus, ROTATE shows the same performance regarding the broad range of application and robustness.

7.2 The Algorithm: A Short Overview

Figure 3 shows the general principles of ROTATE.

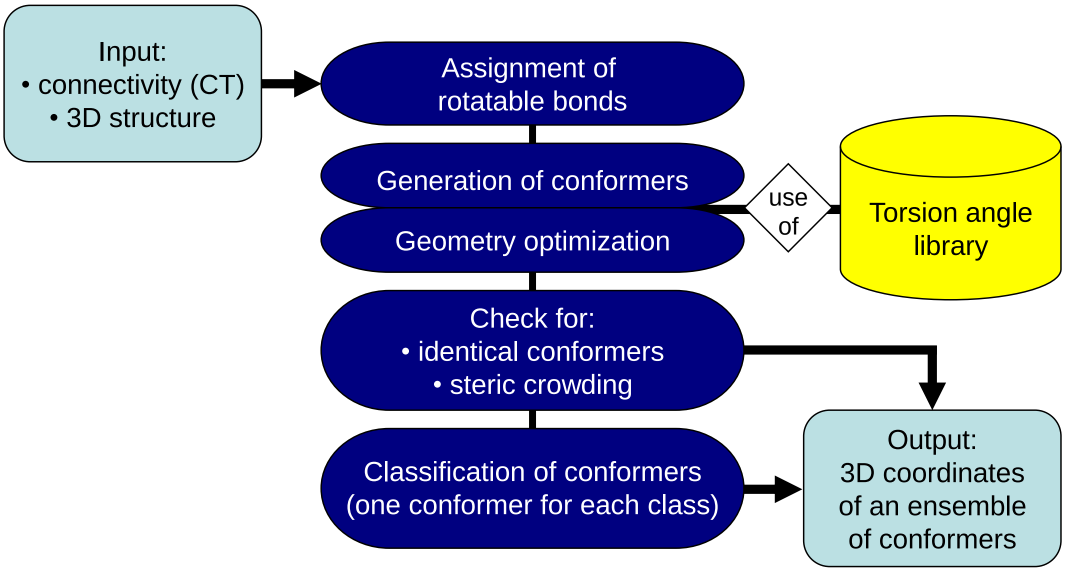

Input:
• connectivity (CT)
• 3D structure
Assignment of rotatable bonds
Generation of conformers
Geometry optimization
use of
Torsion angle library
Check for:
• identical conformers
• steric crowding
Classification of conformers (one conformer for each class)
Output:
3D coordinates of an ensemble of conformers

Figure 3 General principles of ROTATE.

ROTATE requires a connection table, containing information on the number of atoms, atom types, and bonds, and a 3D model of the molecule to process as input information. After the assignment of the rotatable bonds (see section 7.3 on page 14) the rotamers are generated by using the rules and data contained in the Torsion Angle Library (see section 7.4 on page 15). The geometry of each newly generated rotamer is optimized and checked for overlap of atoms and for close contacts. If such situations are detected, the conformation is rejected (see section 7.5 on page 17). Finally, ROTATE offers the possibility to print out all possible conformation or to classify the conformers during the generation process. The current rotamer is then compared to all previously generated geometries and similar or identical conformation are combined to classes, whereby the similarity criterion can be set by the user. Each class is represented by one single conformer, which is the output information of ROTATE.

7.3 Assignment of Rotatable Bonds

After reading in the structure, the molecule is examined for its rotatable bonds. The following restrictions in the assignment of the rotors are made:

1) Bonds with bond orders higher than two are not rotatable. Only single bonds allow a conversion between two conformations without breaking the bond.
2) Single bonds within ring systems are not rotatable. The degrees of freedom of rotation are rather restricted for ring bonds compared to those in open-chain portions. Only sets of torsion angles are allowed that fulfill the ring closure condition properly. Thus, cyclic and acyclic parts of the molecule are treated separately, the major philosophy of our approach to address conformational flexibility. Multiple conformations for ring systems can be generated by the 3D Structure Generator CORINA. Open-chain molecules or fragments are handled by ROTATE.
3) Single bonds to terminating groups are not rotatable. Rotation around a single bond to e.g., a methyl group does not produce significantly new structural information compared to a rotation in the center of the molecule. This assumption is also valid for terminating functional groups containing hetero atoms, such as alcohol, thiole, amine and phosphane groups. Also, the C-N bond in carbon acid amides is not regarded as rotatable, because of its partial double bond character.

ROTATE generates a list of rotatable bonds for the molecule under consideration. All rotors contained in this list are then rated and sorted according to their topological position in the molecule. Rotation around bonds which lie in the center of a molecule have the largest effect on the overall shape of the structure compared to those which are obtained by processing rotors which lie at the "outskirts" of the molecule. To evaluate the topological position of a bond, the Morgan algorithm [14] is implemented. Figure 4 illustrates this procedure with the example of 3-ethyl-hexane.

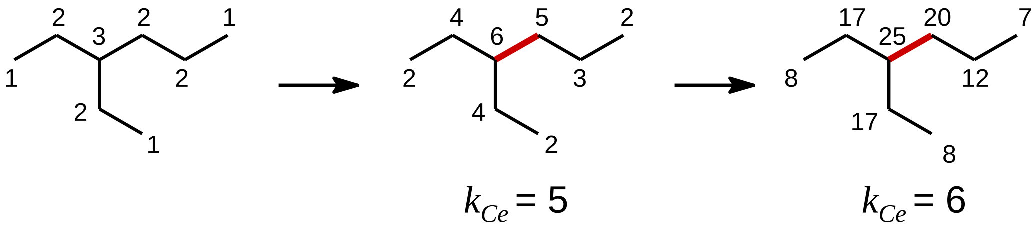


Figure 4 Morgan's algorithm to determine the topological position of a rotatable bond within the molecule (C_e: extended connectivity shown next to the atoms, k_{Ce}: number of extended connectivities).

First, each atom is assigned with the number of all non-hydrogen atom neighbors (connectivity). In the next step, the extended connectivity of each atom C_e is calculated, which is the sum of the connectivities of all non-hydrogen atom neighbors of the atom under consideration. This summation is continued until the number k_{Ce} of different extended connectivities C_e is not increasing anymore. In the case of 3-ethyl-hexane this is fulfilled after the third summation. The bond between the atoms with the highest extended connectivities has the most central (inner most) topological position in the molecule. This bond is then rated with the highest priority in the list of bonds to rotate. Finally, all rotatable bonds are sorted in decreasing manner according to the values of the extended connectivities of their bond partners.

Program usage notes:

By default, ROTATE processes all rotatable bonds. The driver suboption **rot=<value>** (<value> is of the type integer number) offers the possibility to process only the <value> most topologically central rotatable bonds of the molecule. The topological positions of the rotors have been calculated with the method described above.

The driver suboption **bnd=<value>** (<value> is of the type integer number) allows the specification of a certain bond with the bond label <value>, which should be processed only.

7.4 The Torsion Angle Library

For the generation of the conformations ROTATE uses a set of rules and data obtained from a statistical analysis of the conformational preferences of open-chain portions in small molecule crystal structures. This knowledge is stored in the Torsion Angle Library and has been derived from the Cambridge Structural Database (CSD) [15] (Please note: The current version may use a Torsion Angle Library that has been derived from the Macromolecular Structure Database – MSD, 15,000 co-crystallized ligand structures, http://www.ebi.ac.uk/msd/ – instead of the Torsion Angle Library that has been derived from the Cambridge Structural Database.)

The Torsion Angle Library contains over 900 (MSD: 288) entries of torsion angle fragments (torsion patterns) consisting of four atoms and their adjacent neighbors and a single bond in the center. For each pattern the distributions of torsion angles τ as they have been observed in the crystal structures, are stored as histograms [16], [17]. The sampling of the individual frequencies of the torsion angle values was done from 5° to 355° in steps of 10° for each fragment according to its symmetry and periodicity. Figure 5 shows four histograms of different torsion angle patterns.

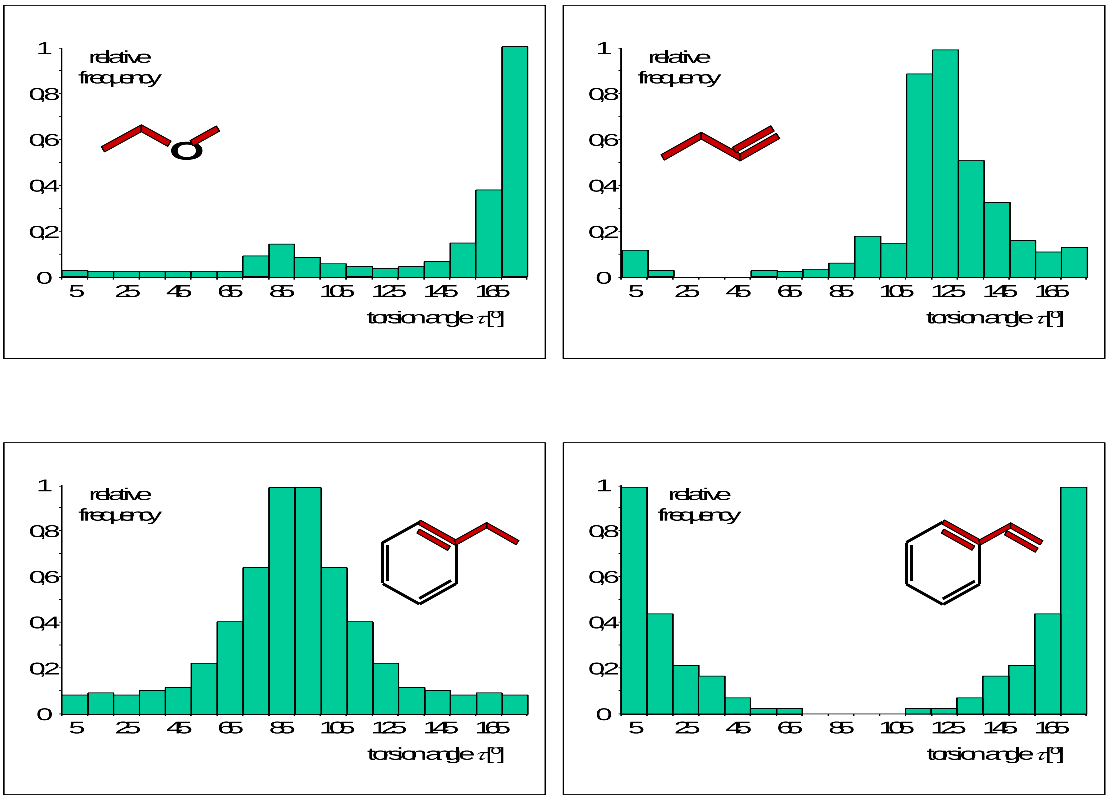


Figure 5 Histograms of torsion angle distributions in different torsion angle patterns as observed in crystal structures contained in the CSD.

These histograms contain the implicit information about the conformational behavior of molecules in a structured molecular environment with varying intermolecular directional forces and dielectric conditions in the different crystal packings. It has been shown that the geometries exhibited by small molecules in crystal structures are correlated to those of ligand conformations at the binding site of a biological receptor [18]. Thus, ROTATE generates only observed conformations likely to be of biological relevance.

7.5 Generation of Conformations

The distributions of the torsion angles derived from the CSD, which are contained in histograms can be transformed into an empirical potential energy function for each torsion angle fragment. A low frequency in the distribution of a certain angle value is correlated with a high energy content for the structural fragment exhibiting this geometry and vice versa. This behavior can be described by applying Equation 1. This approach was originally introduced by Murray-Rust [19] to derive a symbolic energy function from the relative distributions.

$$E(\tau) = -A \cdot \ln f(\tau)$$

$E(\tau)$	symbolic energy value at torsion angle τ
A	adjustable parameter
$f(\tau)$	relative frequency of torsion angle τ

Equation 1: Derivation of a symbolic energy function.

The potential values are calculated in steps of 10°. A spline function is interpolated between these steps to obtain a smooth and derivable functional form. If more than one entry of the Torsion Angle Library can be assigned to the probe fragment under consideration (maximum of 9), the individual distributions of each matching pattern are subsequently superimposed and averaged. Thus, for a particular torsion fragment under consideration a specific energy function can be derived.

This empirical potential energy function can be used to select a set of preferred torsion angles for a certain fragment. These values are taken as initial angles at the beginning of the build-up procedure. Another application of the symbolic energy function is its use in the optimization of the initial geometries during the generation process.

Selection of the Initial Torsion Angles. By applying a grid of 30° the symbolic energy values of 12 torsion angles are calculated and ordered according to their conformational energy. To obtain a set of the most preferred torsion angles, a maximum of six torsion angle values which lie under a certain energy threshold are selected and are used as initial torsions at the beginning of the build-up procedure. Figure 6 shows the energy function for a C_{sp3}-C_{sp3}-O_{sp3}-C_{sp3} torsion and the preferred angles as selected as described above. The selection procedure led to six most preferred torsion angles for this pattern (in order of increasing energy): 180°, 150°, 210°, 90°, 270°, and 120°. These values are used as initial torsions for this fragment.

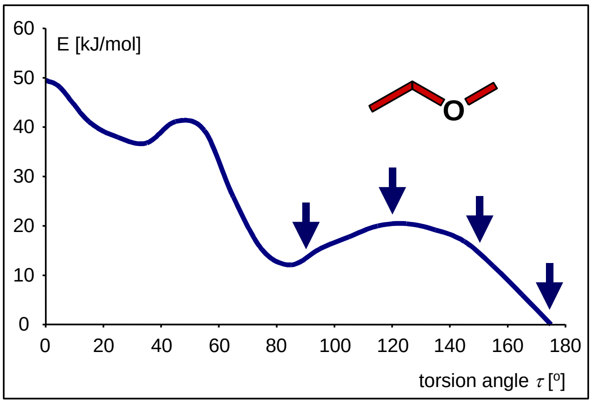


Figure 6 Symbolic energy function for a C_{sp3}-C_{sp3}-O_{sp3}-C_{sp3}-torsion and its preferred torsion angle values as selected by ROTATE. (Please note: Due to the symmetry of this torsion angle pattern the following angle values are also considered to be preferred: 210°, 270°.)

Generation of Conformations. A set of preferred torsion angles has now been determined for each torsion angle pattern of the molecule under consideration. A *depth-first search* algorithm is implemented to generate all possible combinations of torsion angles to build the conformers. After a new conformation has been generated its geometry is optimized by using the empirically derived energy function.

Geometry Optimization. A simple gradient optimizer is implemented to search the symbolic energy function of the torsion angle fragment under consideration for local minimum which lies next to the initial torsion angle. For the example of an C_{sp3}-C_{sp3}-O_{sp3}-C_{sp3} torsion (see Figure 6) starting with an initial angle of 90° the optimization process is finished, if an angle of 85.9° is reached (see Figure 7).

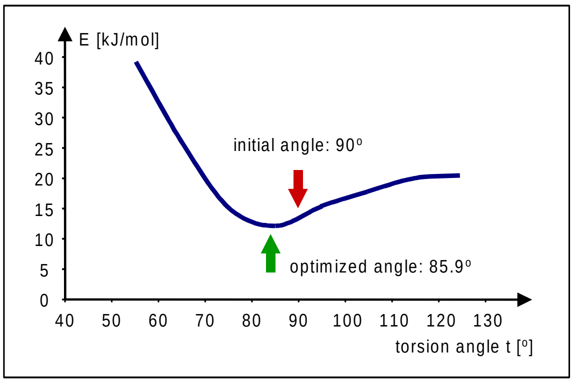


Figure 7 Gradient optimizer for searching the symbolic energy function for local minima.

Program usage notes:

By default, ROTATE uses the six most preferred torsion angle values for each rotatable bond that are found in the Torsion Angle Library in order to generate initial geometries. The driver option **pta=<value>** (<value> is of the type integer number and may range from 2 to 6) offers the possibility to use only the <value> most preferred torsion angle values for each rotatable bond (e.g., the 3 most preferred torsion angle values). Thus, the final number of conformations is, of course, lower. However, for molecules with a large number of rotatable bonds more rotors can be processed (see driver suboption **-d rot=<value>**). For example, the command line

```
rotate -d rot=5 infile.sdf outfile.sdf
```

processes the 5 most central rotatable bonds of a molecule and uses the 6 most preferred torsion angle values for each rotatable bond (default setting). The maximum number of conformations that can be generated is 6^5 = 7,776. The command line

```
rotate -d rot=8,pta=3 infile.sdf outfile.sdf
```

processes the 8 most central rotatable bonds of a molecule but uses only the 3 most preferred torsion angle values for each rotatable bond. The maximum number of conformations that can be generated is 3^8 = 6,561, however, more degrees of freedom (i.e., 8 instead of 6 rotatable bonds) have been used.

7.6 Classification of Conformations

The number of conformations increases exponentially with the number of rotatable bonds. For most of the molecules it is therefore not feasable to take all possible conformations into account. On the other side, however, a balanced sampling of the conformational space should be ensured if only subsets are considered. Thus, several constraints have to be applied in order to reduce the number of conformations. ROTATE realizes these restrictions in two different approaches: First of all, only conformation which can be observed experimentally are generated by using the rules and the data contained in the Torsion Angle Library. The second possibility to reduce the number of conformations implemented in ROTATE are classification methods. Each newly generated rotamer is compared to all previously generated geometries. Similar or identical conformations are combined to classes. Finally, each class is represented by the conformer which established the class. Two different and adjustable similarity criteria can be chosen in order to influence the number of conformations of the subsets obtained. These constraints provide a balanced sampling of the conformational space because all other class members do not contribute significantly new structural information within the defined similarity criterion.

7.6.1 Classification in Cartesian Space

The classification of conformations in Cartesian space is based on the RMS_{XYZ} (root mean square) deviation of the Cartesian coordinates of all non-hydrogen atoms of the two conformers which are compared (see equation 2).

$$RMS_{XYZ} = \sqrt{\frac{\sum_i^N d^2}{N}}$$

N number of non-hydrogen atoms

Δd distance of the i- th corresponding atom pair in [Å]

Equation 2 Calculation of the RMS_{XYZ} deviation in Cartesian space.

Two conformations are considered to be identical if their RMS_{XYZ} deviation is less than 0.3 Å (empirically derived value). However, this is an adjustable parameter which can be set by the user before running ROTATE. The higher this threshold value is set, the more conformations are regarded to be similar or identical and are combined into classes. This results in a decreasing number of outputted rotamers with increasing structural diversity.

Program usage notes:

By default, ROTATE generates and writes out all possible conformations which can be obtained, unless no classification method is specified. The driver suboption **rms_xyz=<value>** (<value> is of type floating point number) specifies the classification method in Cartesian space and <value> defines the RMS_{XYZ} deviation (threshold value) in Ångström [Å] within which two conformations are considered to be identical and are combined into classes. To remove similar conformations, a RMS_{XYZ} threshold value of 0.3 Å is recommended.

7.6.2 Classification in Torsion Angle Space

The classification of rotamers in torsion space is based on the RMS_{TA} (root mean square) deviation of the torsion angles along the rotated bonds of the two conformers which are compared (see equation 3).

$$RMS_{TA} = \sqrt{\frac{\sum_i^N \Delta\tau^2}{N}}$$

N number of changed torsion angles

$\Delta\tau$ deviation of the i- th corresponding torsion angle in [°]

Equation 3 Calculation of the RMS_{TA} deviation in torsion angle space.

Two conformations are considered to be identical if their RMS_{TA} deviation is less than 15° (empirically derived value). However, this is an adjustable parameter which can be set before running ROTATE. The higher this threshold value is set, the more conformations are regarded to be similar or identical and are combined into classes. This results in a decreasing number of outputted rotamers with increasing structural diversity.

Program usage notes:

By default, ROTATE generates and writes out all possible conformations which can be obtained, unless no classification method is specified. The driver suboption **rms_ta=<value>** (<value> is of type floating point number) specifies the classification method in torsion angle space and <value> defines the RMS_{TA} deviation (threshold value) in degree [°] within which two conformations are considered to be identical and are combined into classes. To remove similar conformations, a RMS_{TA} threshold value of 15° is recommended.

7.6.3 Comparison of the Classification Methods

The classification in Cartesian space is the more accurate one but is computationally the more time consuming method of comparing two conformations. Both structures have to be superimposed in 3D space. This requires the translation of all coordinates into the same space unit. A minimization of the RMS_{XYZ} deviation of the Cartesian coordinates of all non-hydrogen atoms has to be performed by varying the coordinates of the two conformers against each other, but keeping the structures rigid.

The classification in torsion angle space is much faster. Hereby, only internal coordinates of the conformations have to be compared to calculate the RMS_{TA} deviation. These coordinates are independent of the position of the structures in 3D space. No superimposition and no minimization of the RMS_{TA} deviation has to be performed.

The results, i.e., the structures, which are obtained by both classification methods are different. In Cartesian space the conformers are combined into classes if they have a similar overall shape according to the preset RMS_{XYZ} threshold value. In torsion angle space a relatively small change of a torsion angle in the center of the molecule can cause significantly large spatial changes in the outer regions of the structure ("leverage" effect).

Both methods are compared in the following (see section 7.8 on page 22).

7.7 The Automatic Run Mode "-d auto"

ROTATE offers a fully automatic mode by setting the driver option **-d auto**. Each input structure is then analyzed for its number of rotatable bonds and their topological position within the molecule (see section 7.3) and the program automatically determines default values for

- how many and which of all rotatable bonds that are processed,
- the number of preferred torsion angle values that are taken from the TA library (**pta=<value>**) and
- the classification method including an appropriate RMS threshold value (**rms_xyz=<value>** or **rms_ta=<value>**)

The usage of this option is highly recommended since ROTATE then tries to generate a diverse and balanced set of conformations as fast as possible and the user does not need to care about all the different driver options.

Please note that the driver option **-d auto** must not be used in combination with any other driver option. If so, an error message is printed to the trace file (and to standard-out) and the program stops.

7.8 A Case Study: 2-L-Benzylsuccinate

The following study addresses the question if ROTATE is able to reproduce molecular geometries of biological relevance and how well the conformational space is sampled.

2-L-Benzylsuccinate (BZS) is one of the most effective inhibitors of the enzyme carboxypeptidase A (CPA) [20]. This zinc containing enzyme (exopeptidase) specifically catalyses the hydrolysis of peptides and esters bearing large hydrophobic side chains in mammals. The receptor/ligand complex of CPA/BZS has been determined experimentally (by X-ray studies) and is available from the Protein Data Bank (PDB ID: 3cbs) [21]. Figure 8 shows the receptor-bound geometry (X-ray) and two calculated models, the CORINA built model and the structure obtained by using the semiempirical method AM1. In addition, the RMS_{XYZ} deviations of the models to the X-ray structure (receptor-bound geometry) are given.

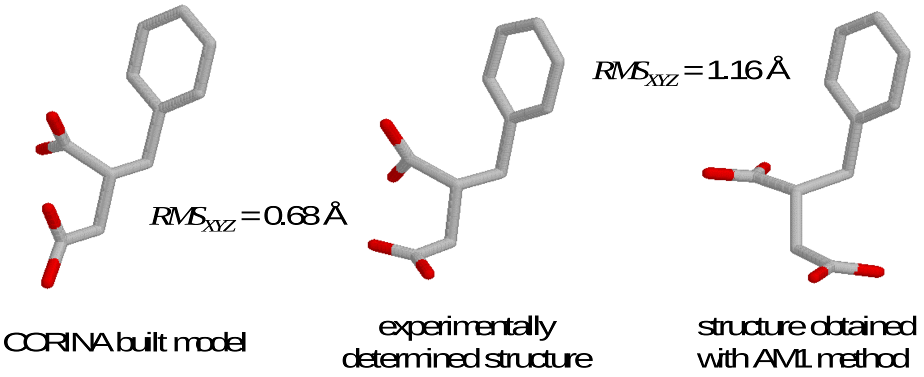

$RMS_{XYZ} = 1.16$ Å
$RMS_{XYZ} = 0.68$ Å
CORINA built model
experimentally determined structure
structure obtained with AM1 method

Figure 8 Three different geometries of 2-L-benzylsuccinate (BZS): Experimentally determined structure (center), the CORINA built model (left) and the structure obtained by using the quantum mechanical AM1 method (right).

The CORINA built model was taken as input structure for ROTATE. 2-L-Benzylsuccinate has five rotatable bonds which are all processed. Figure 9 shows the rotatable bonds and the torsion angle histograms taken from the Torsion angle Library which are assigned to the rotors.

In order to compare the different classification methods with regard to the computation time and the number of generated conformations (classes) several program runs were performed with different RMS_{XYZ} and RMS_{TA} threshold values, respectively, leading to various subsets of conformations. The generation of all possible combinations of preferred initial torsion angles gives a total number of 7776 conformations. Due to the geometry optimization routine (see section 7.6 at page 20), this set contains identical conformations, which are rejected during the generation process. This yields a total number of 3701 conformations. Table A summarizes the statistics on the ROTATE runs.

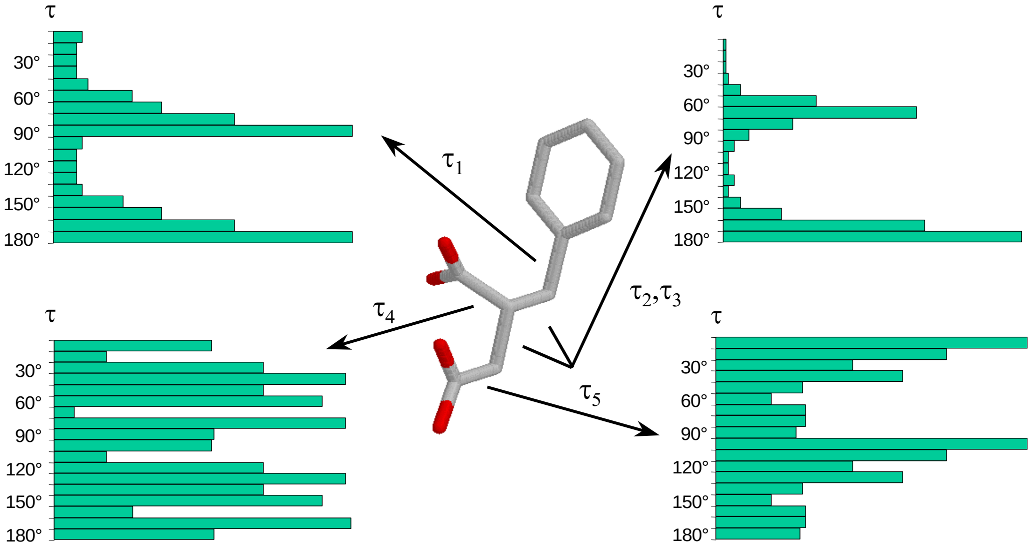


Figure 9 Rotatable bonds of 2-L-benzylsuccinate (BZS) and their corresponding torsion angle histograms taken from the Torsion Angle Library.

The RMS_{XYZ} deviation to the experimentally determined structure was calculated for each conformation of the subsets. Table B lists the smallest RMS_{XYZ} deviations ($RMS_{XYZ}{}^{best}$) which were found for the conformations in the different subsets. The conformation with the smallest RMS_{XYZ} deviation is the best matching conformation to the X-ray structure in the receptor bound state.

Table A Statistics on the ROTATE runs processing 2-L-benzylsuccinate (performed on an AMD Athlon XP1800+, Linux workstation).

Number of all possible conformations: 7567 (6^5);
Number of unique conformations: 1334; CPU time [s]: 11

Subsets obtained by classification in Cartesian space (RMS_{XYZ})			Subsets obtained by classification in torsion space (RMS_{TA})		
RMS_{XYZ} threshold [Å]	number of classes	CPU time [s]	RMS_{TA} threshold [°]	number of classes	CPU time [s]
0.15	223	137	15	323	9
0.30	61	42	30	161	8
0.45	37	25	45	100	7
0.60	18	17	60	50	7
0.90	10	12	90	11	7
1.20	5	8	120	4	6

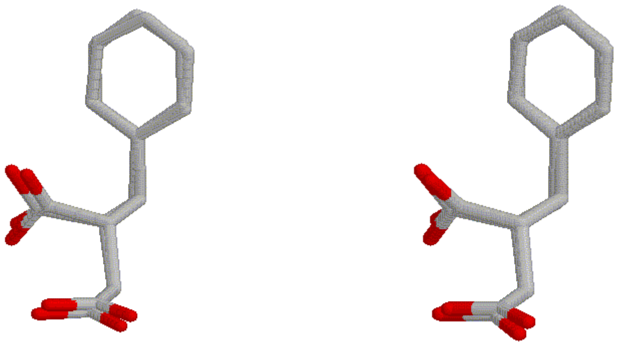
Table B The Smallest RMS_{XYZ} deviations (RMS_{XYZ}^{best}) to the experimentally determined structure which were found for the conformations in the different subsets.

Subsets obtained by classification in Cartesian space (RMS_{XYZ})		Subsets obtained by classification in torsion space (RMS_{TA})	
RMS_{XYZ} threshold [Å] for classification	RMS_{XYZ}^{best} [Å]	RMS_{TA} threshold [°] for classification	RMS_{XYZ}^{best} [Å]
0.15	0.28	15	0.28
0.30	0.28	30	0.36
0.45	0.56	45	0.36
0.60	0.89	60	0.58
0.90	0.97	90	0.58
1.20	1.23	120	0.64

As described above, two conformations are considered to be identical if their RMS_{XYZ}

deviation is less than 0.3 Å. For both classification methods the best matching conformation has a RMS_{XYZ} deviation to the X-ray structure which lies below this empirically derived value. Thus, the structures can be regarded as one and the same conformation. The best matching conformations which are obtained by classification in torsion space and a RMS_{TA} threshold value of 30° and 45° are also in a range to be considered as very similar to the experimental structure.

Figure 10 shows the superimpositions of the experimentally determined structure of 2-L-benzylsuccinate with the two conformations which have the smallest RMS_{XYZ} deviation to the X-ray geometry (RMS_{XYZ}^{best} = 0.28 Å for classification in Cartesian space and torsion space, respectively).



RMS_{XYZ}^{best} = 0.28 Å
classification in Cartesian space
(RMS_{XYZ}: 0.15 Å, 0.30 Å)

RMS_{XYZ}^{best} = 0.28 Å
classification in torsion space
(RMS_{TA}: 15°)

Figure 10 Superimpositions of the experimentally determined structure of 2-L-benzylsuccinate with the two conformations which have the smallest RMS_{XYZ} deviation to the X-ray geometry (RMS_{XYZ}^{best}).

In the following section the question how well the conformational subspaces, which are described by the different subsets of conformations, sample the conformational space of all possible conformations which can be obtained by ROTATE (complete conformational space) is addressed. For this purpose, an approach applying Kohonen neural networks and atom radial distribution functions as molecular (conformational) descriptor is used.

The Kohonen procedure is an unsupervised learning method, which is able to project the information of a high-dimensional descriptor space into a two-dimensional map (Kohonen map) by retaining the topology of the information [22]. Thus, neighboring points in the descriptor space occupy neighboring neurons in the Kohonen map. One of the main prerequisites of using neural network techniques is, that each object (e.g. conformations of a molecule) has to be represented by a vector of fixed length (descriptor). In this approach atom radial distribution functions (ARDF) are used for the coding of the conformers [23]. These distance functions, which have been derived from X-ray crystallography, are very sensitive towards small variations of the spatial arrangement of the atoms of a molecule and are able to distinguish between different conformations by describing each conformer as a vector of constant length.

First of all, the ARDF codes of all possible conformations which can be generated with ROTATE are calculated and are used for the training of a Kohonen neural network. The resulting network is called a template network and contains the information of the complete conformational space which can be described with ROTATE. Then, the ARDF codes of the various subsets of conformations are calculated and presented to the template network, without any further training of the network. The resulting Kohonen maps visualize the distribution of the conformations of the various subsets in the complete conformational space and show whether the subspaces sample the conformational space in a balanced manner or if they tend to describe local regions of the conformational space (formation of clusters). Figure 11 illustrates this procedure.

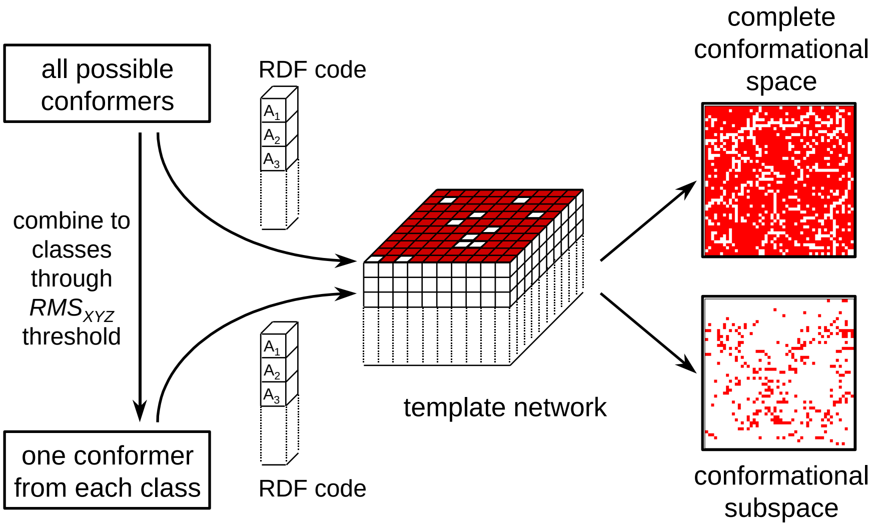

all possible conformers
RDF code
complete conformational space
A_1
A_2
A_3
combine to classes through RMS_{XYZ} threshold
template network
one conformer from each class
RDF code
conformational subspace

Figure 11 General procedure for visualizing conformational spaces by using Kohonen neural networks and atom radial distribution functions (ARDF).

Figure 12 shows the Kohonen map for the complete conformational space which can be described by ROTATE. The dimension of the network is 37 x 37 neurons and each conformation is represented by 100 ARDF coefficients. From the original set of 7567 possible conformation all duplicate conformations have been removed. For the training of the network – the template network – the set of 1334 unique conformations represented by their ARDF codes were used. Each dot (neuron) in the Kohonen map marks at least one single conformation.

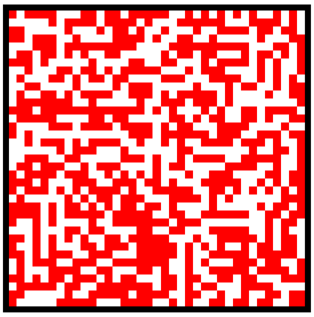
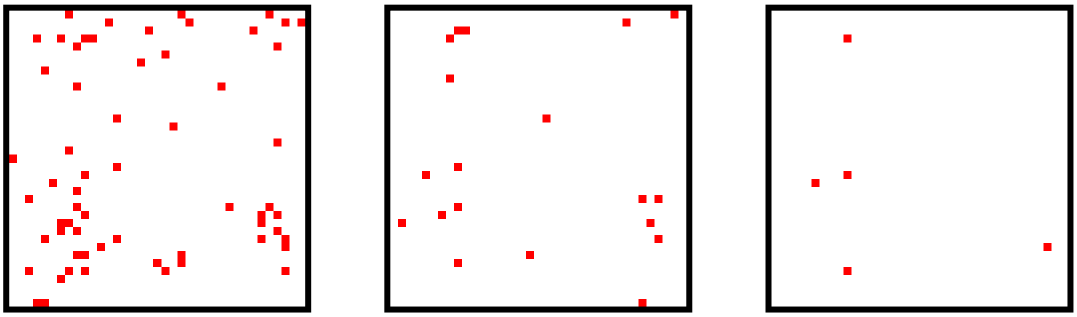


Figure 12 Visualization of the complete conformational space of 2-L-benzylsuccinate described by ROTATE (template network: 37 x 37 neurons, 1334 conformations, 100 ARDF coefficients).

After the training process the different subsets are presented to the template network. The projection of the high-dimensional descriptor space into the two-dimensional Kohonen map visualizes the conformational subspaces and the distribution of the conformations of the subsets in the entire conformational space. The resulting Kohonen maps are shown in Figure 13 and in Figure 14 for various subsets, classification methods, and *RMS* thresholds.

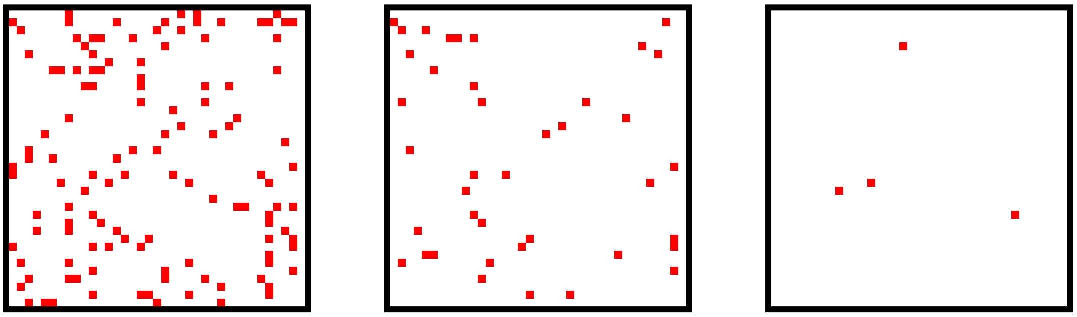


RMS_{XYZ} = 0.3 Å
61 conformers

RMS_{XYZ} = 0.6 Å
18 conformers

RMS_{XYZ} = 1.2 Å
5 conformers

Figure 13 Kohonen maps (37 x 37 neurons) visualizing the conformational subspaces which have been obtained by classifying the conformers in Cartesian space with different RMS_{XYZ} threshold values. Each dot (neuron) in the Kohonen map marks at least one single conformation.



RMS_{TA} = 30°
161 conformers

RMS_{TA} = 60°
50 conformers

RMS_{TA} = 120°
4 conformers

Figure 14 Kohonen maps (37 x 37 neurons) visualizing the conformational subspaces which have been obtained by classifying the conformers in torsion angle space with different RMS_{TA} threshold values. Each dot (neuron) in the Kohonen map marks at least one single conformation.

The Kohonen maps shown in Figure 13 and Figure 14 of the conformational subspaces obtained by the different classification methods and various *RMS* threshold values indicate a quite even distribution of the conformations in the complete conformational space. Cluster formation in the maps and the description of preferred local regions of the conformational space cannot be observed. Thus, a balanced sampling of the conformational space can be expected by applying either classification method, either in Cartesian space or in torsion space.

8 File Formats

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

Table C Input/output file formats.

Format	Input	Output	Reference
MDL SDF	x	x	[1]
SYBYL MOLFILE	x	x	[2]
SYBYL MOL2	x	x	[2]
PDB		x	[4]
CTX	x	x	[3]

For each conformation a new record is written to the output file containing all information which was read in except the newly generated 3D coordinates.

8.1 MDL Structural Data File (SDF)

The MDL SDFile format was implemented following the description in the literature [1]. ROTATE 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 SDFile. This is a great advantage for database purposes since all information except the 3D coordinates 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 ROTATE 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.

8.2 SYBYL File Formats

Both SYBYL MOLFILE and MOL2 have been implemented following the SYBYL program manual [2]. Since both formats are based on rather special atom types, their applicability as a standard database format is limited and many cases can occur where no meaningful atom type can be assigned. Atoms with lacking atom types are assigned a dummy atom type. MOL2 files are written by using the keywords @<TRIPOS>MOLECULE, @<TRIPOS>ATOM, and @<TRIPOS>BOND.

8.3 Brookhaven Protein Data Bank Format (PDB)

The PDB format was implemented following the literature [4]. The following keywords are used: COMPND, REMARK, HETAM, CONECT, and END. The compound name is written to the COMPND statement. The atomic symbols and the 3D coordinates are written to HETATM statements. The bond graph is reflected by CONECT statements.

Two special suboption **pdbatom** forces ROTATE to replace the keyword HETATM by ATOM.

8.4 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, STEREO, HIGEOM, INTCOORD, 3DCOORD, and END. These keywords, except 3DCOORD, and all additional information are directly piped to the output when the output file type also was set to CTX.

9 Error Messages

9.1 General Errors

ROTATE contains several routines which are build in to prevent the program of an uncontrolled exit (program crash). These routines always print an error message to standard out or to the trace file and are not completely listed below. Please restart the program or, if reoccurring, your system.

ERROR licence(): Licence time out.
The license time has expired. You are not allowed to use the program any more.

ERROR pfopen(): Can’t open file *filename (path)*.
The program is unable to open the specified file (path in parentheses).

ERROR ioopen(): Can’t open trace file.
The trace file *rotate.trc* can’t be opened.

ERROR ioopen(): Identical input/output files.
Identical file names for the input and output files are not allowed.

ERROR ioopen(): Can’t open input file *filename*.
Error opening the specified input file.

ERROR ioopen(): Can’t open output file *filename*.
Error opening the specified output file.

ERROR rotate(): Too many non-option parameters.
The command line must not contain more than two non-option parameters (the input and output file names). An options start with an ”-”. After the first non-option parameter no more options are allowed.

ERROR rotate(): Input file type not supported.
The input file type which is specified is not supported.

ERROR rotate(): Output file type not suported.
The output file type which is specified is not supported.

ERROR rotate(): Specifying both classification methods not allowed.
(*see next error message*.)

ERROR generate(): Specifying both classification methods not allowed.

Only one method for the classification of the conformers can be specified. To define both driver suboptions **rms_xyz** and **rms_ta** is not allowed.

ERROR rotate(): No bonds to rotate specified >rot=0< !

The driver suboption **rot** requires the specification of a value (**rot**=<value>) The option **rot** is set without a value.

9.2 Input File Format Errors

9.2.1 MDL SDFile

The error messages given below are completed by the line number in the MDL SDFile and the record number where the error occurred.

ERROR rmdl(): Can’t read counts line.

ERROR rmdl(): Can’t read atom block.

ERROR rmdl(): Can’t read bond block.

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

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

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

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

ERROR rmdl(): Unknown element symbol.

ERROR rmdl(): Bond atoms out of range.

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

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

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

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

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

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

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

Errors with missing columns:

ERROR rmdl(): No atomic charge column specified.

ERROR rmdl(): No atom stereo column specified.

ERROR rmdl(): No mass difference column specified.

ERROR rmdl(): Too many additional lines.
Only a limited number of additional lines in MDL SDFile is allowed (5,000).

ERROR rmdl(): Line too long.
Only 80 characters per line are allowed in MDL SDFiles.

Errors in charge, radical, and attachment point lines:

ERROR rmdl(): CHG atom out of range.

ERROR rmdl(): RAD atom out of range.

ERROR rmdl(): APO atom out of range.

9.3 Errors in the Generation and Classification of Conformations

ROTATE contains several routines which are build in to prevent the program of an uncontrolled exit (program crash). These routines always print an error message to standard out or to the trace file and are not completely listed below. Please restart the program or, if reoccurring, your system.

ERROR confanalys(): No conformation left after classification.
A fatal error occurred during the classification of the generated conformers. This error cannot be interpreted at once. Please restart the program, or, if reoccurring, your system.

ERROR confanalys(): Violation of classification method.
A fatal error occurred during initializing the specified classification method. This error cannot be interpreted at once. Please restart the program, or, if reoccurring, your system.

ERROR comparco(): Different number of atoms to match: #, #.
During the classification in Cartesian space two conformers with a different number of atoms (#, #) have been encoutered.

ERROR comparco(): Ambiguous list size.
During the classification in Cartesian space some distance matrices differ in their size.

ERROR confanalys(): Fatal error during initializing classification routine!
A fatal error occurred during the initialization of the specified classification method. This error cannot be interpreted at once. Please estart the program, or, if reoccurring, your system.

ERROR confanalys(): Drive mode ambiguous.

A fatal error occurred during the initialization of the conformation analysis routine. This error cannot be interpreted at once. Please restart the program, or, if reoccurring, your system.

10 Warning Messages

10.1 General Warnings

WARNING rotate(): No 3D coordinates found, record discarded!

The current structure record contains no information about 3D atom coordinates. No conformations can be generated and the record is discarded.

10.2 Warnings during the Generation and Classification of Conformations

WARNING generate(): No classification method specified: Writing out all possible conformers!

Since no classification method is specified, ROTATE generates all possible conformations and writes them out (including duplicate conformations). This might lead to huge output files.

WARNING generate(): Specified bond # not rotatable or not existing! Determining and processing most central rotatable bond.

The bond to rotate which was specified by its bond label (driver suboption bnd=<value>) is not a rotatable bond or is not existing. ROTATE now determines and processes the most central rotatable bond.

WARNING contact(): Bad contacts detected for conf.#, conformation rejected.

Unfavourable non-bonded interactions (crowded atoms) are detected for conformation #. The conformation is rejected.

WARNING matcon(): Timeout reached.

During the classification in Cartesian space two conformations could not be matched within a certain time limit. The next conformation is generated.

11 Acknowledgments

ROTATE was developed in the research group of Prof. Dr. Johann Gasteiger at the University of Erlangen-Nürnberg since 1996. The program development was originated by Dr. Christof H. Schwab and Dr. Jens Sadowski.

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

ROTATE is now maintained for general usage by Molecular Networks GmbH Computerchemie.

12 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] SYBYL, Tripos Associates Inc.: St. Louis, MO, USA.

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

[4] Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535-542.

[5] Hiller, C.; Gasteiger, J. Ein automatisierter Molekülbaukasten. In *Software-Entwicklung in der Chemie*, Vol 1; Gasteiger, J., Ed.; Springer: Berlin, 1987; pp 53-66.

[6] Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic Generation of 3D Atomic Coordinates for Organic Molecules. *Tetrahedron Comp. Method.* **1990**, *3*, 537-547.

[7] Sadowski, J.; Rudolph, C.; Gasteiger, J. The Generation of 3D Models of Host-guest Complexes. *Anal. Chim. Acta* **1992**, *265*, 233-241.

[8] Sadowski, J.; Gasteiger, J. Polygon Patterns for the Generation of Conformations of Large Rings. In *Software Development in Chemistry*, Vol 7; Ziessow, D., Ed.; Gesellschaft Deutscher Chemiker: Frankfurt am Main, 1993; pp 65-76.

[9] Sadowski, J.; Gasteiger, J. From Atoms and Bonds to Three-dimensional Atomic Coordinates: Automatic Model Builders. *Chemical Reviews* **1993**, *93*, 2567-2581.

[10] Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-Ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000-1008.

[11] 3D Structure Generator CORINA, Version 2.4, Program Description; Molecular Networks GmbH, 1998.

[12] (a) Gillespie, R.J. *J. Chem. Educ.* **1963**, *40*, 295-301. (b) Gillespie, R.J. Angew. Chem. 1967, 79, 885-896. (c) Gillespie, R.J. *J. Chem. Educ.* **1970**, *47*, 18-23.

[13] Pauling, L. Die Natur der chemischen Bindung; Verlag Chemie: Weinheim, 1962.

[14] Morgan, H. L. The Generation of Unique Machine Description for Chemical Structures—a Technique developed at Chemical Abstract Service. *J. Chem. Doc.* **1965**, 107-113. (See also: Zupan, J. *Algorithms for Chemists*; John Wiley & Sons: Chichester, New York, 1989; pp 227-230.)

[15] Cambridge Structural Database: a) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, D. S.; Rodgers, J. R.; Watson, D. G. *Acta Cryst. Sect. B: Struct. Crystallogr. Cryst. Chem.* 1979, B35, 2331. b) Allen, F. H.; Kennard, O. *Acc. Chem. Res.*, 1983, 16, 146.

[16] Klebe, G.; Mietzner, T. A Fast and Efficient Method to Generate Biologically Relevant Conformations. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 583-606.

[17] Klebe, G.; Mietzner, T.; Weber, F. Methodological Developments and Strategies for a Fast Flexible Superposition of Drug-Size Molecules. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 35-49.

[18] Klebe, G. Structure Correlation and Ligand/Receptor Interaction. In *Structure Correlation, Volume 2*; Bürgi, H.-B. and Dunitz, J. D., Ed.; VCH Verlagsgesellschaft: Weinheim, 1994; pp 543-603.

[19] Murray-Rust, P. How Useful are X-Ray Studies of Conformation? In *Molecular Structure and Biological Activity*; Griffin, J. F. and Duax, W. L., Ed.; Elsevier Biomed.: New York, 1982; pp 117-133.

[20] Mangani, S.; Carloni, P.; Orioli, P. Crystal Structure of the Complex between Carboxopeptidase A and the Biproduct Analog Inhibitor 2-L-Benzylsuccinate at 2.0 Resolution. *J. Mol. Biol.* **1992**, *223*, 573-578.

[21] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Databank. *Nucleic Acid Research* **2000**, *28*, 235-242. The Protein Data Bank can be accessed via the Internet: http://www.rcsb.org/pdb.

[22] Zupan, J.; Gasteiger, J. Neural Networks in Chemistry and Drug Design, 2nd Edition. Wiley-VCH: Weinheim, 1999. (See also: KMAP 4.0, Program Manual.)

[23] Hemmer, C. H.; Steinhauer, V.; Gasteiger, J. Deriving the 3D Structure of Organic Molecules from their Infrared Spectra. *Vibrational Spectroscopy* **1999**, *19*, 151-164.

13 Report Form

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

Molecular Networks GmbH Computerchemie
Neumeyerstr. 28
90411 Nuremberg, Germany
FAX: +49-911-597 424 09

User:

ROTATE program and version number ("rotate -v"):

Command line to run ROTATE:

Error and warning messages by ROTATE:

System messages:

Short description:

Please include the input file, output file and trace file (*rotate.trc*) generated by ROTATE and send it to us by email to support@mn-am.com. These files will help us to analyze your problems. All data will be treated confidentially.