Accessing the Three-Dimensional World of Molecules

3D Structure Generator CORINA™

Molecular Networks GmbH Computerchemie, Nägelsbachstr. 25, D-91052 Erlangen, Germany
WWW: http://www.mol-net.de, Email: info@mol-net.de

Molecular Networks

Overview

CORINA™ - Fast and Efficient Generation of High-Quality Three-Dimensional Molecular Models

The three-dimensional structure of a molecule is closely related to a large variety of chemical, physical and biological properties. The need for computer generated 3D molecular structures has clearly been recognized in drug design and related disciplines. Since the number of experimentally determined molecular geometries is limited - about 250.000 X-ray structures in the Cambridge Crystallographic Database compared to approximately 22 millions of known compounds - one needs a method to predict 3D atomic coordinates of molecules fast and efficiently.

The rule- and databased program system **CORINA**™ (**CO**o**R**d**INA**tes)[1] automatically generates three-dimensional atomic coordinates from the constitution of a molecule as expressed by a connection table or linear string. It can handle massive volumes of structures such as those in company databases and combinatorial chemistry experiments. The scheme on the right side shows the general principles of **CORINA**™.

Features

- Applicable to the entire range of classical organic chemistry, including stable molecules, radicals, reactive intermediates, multi-fragment records (e.g. salts), fused rings, spiro, polycyclic, macrocyclic and polymacrocyclic systems
- Generation of low-energy conformations
- Consideration of stereochemical information
- Processing of structures with atoms having up to six neighbors
- Conversion of organo-metallic complexes
- No upper limits to the size of the molecules or size of ring systems[2]
- Generation of multiple ring conformations for rings consisting of less than ten atoms
- Processing of a large variety of structural data file fomats (e.g. MDL SDFile, MDL RDFile, SYBYL MOLFILE and MOL2, SMILES linear code and PDB format)
- Interface to the flexible ligand docking program FlexX™ for the generation of multiple ring conformations during the docking process
- Several options to influence the generation process, e.g. addition of implicitly given hydrogen atoms in the input file, neutralization of charged molecules, or orientation of the 3D structures according to their moments of inertia
- Available for common UNIX platforms (SGI, Sun, DEC, and Linux) as well as for MS Windows systems

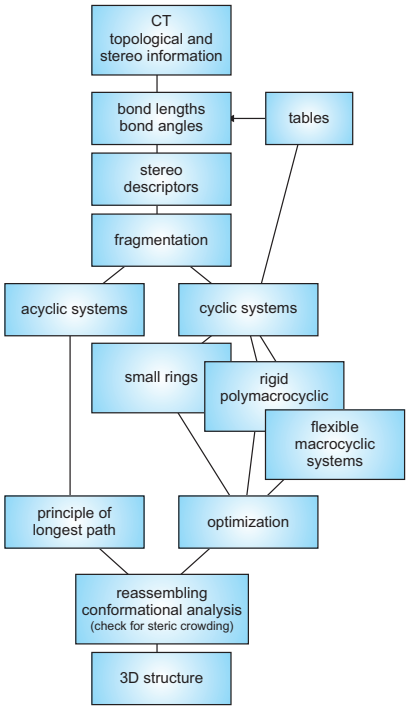
Performance

- High-quality molecular models: Lowest *RMS* deviations from published X-ray structures among all commercially available model builders
- Fast: Conversion in less than one second for medium-sized organic molecules on a common UNIX workstation
- Robust: Conversion of the open part of the NCI Database with a rate of more than 99% without intervention and program crash
- General: Processing of a database with over six million compounds and a conversion rate of more than 99%

References

[1] J. Sadowski, J. Gasteiger *Chem. Reviews* **1993**, *93*, 2567.
[2] Schönberger, H.; Schwab, C.H.; Hirsch, A.; Gasteiger, J. *J. Mol. Model.* **2000**, *6*, 379-395.
[3] J. Sadowski, J. Gasteiger, G. Klebe *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000.

Method

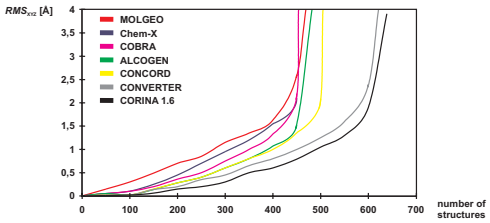


Performance

High-Quality Molecular Models

A dataset of 639 X-ray structures taken from the Cambridge Crystallographic Database was used to evaluate in 1994/95 several commercially available 3D model builders.[3]

The following diagram characterizes the relationships between the quantity (conversion rate) and the quality (the degree of reproduction of the X-ray structures), i.e. the efficiency of the different conversion programs. For each program the ordered *RMS* deviations (X-ray structure against predicted model) of the non-hydrogen atoms are plotted *vs.* the number of converted structures.



Fast and Efficient Conversion

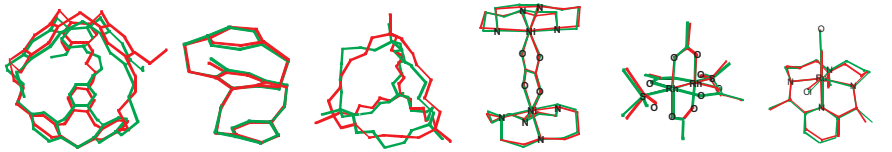
CORINA™ is applicable to the entire range of organic chemistry. This has been shown by converting large databases, such as the National Cancer Institute (NCI) file. In the following some statistics on the **CORINA**™ run processing the NCI database are given:

- Number of structures in database: 249,081 compounds
- Structures read: 249,081 - 100%
- Structures converted: 247,791 - 99.48%
- Structures excluded due to structural insufficiencies: 823 - 0.33%
- Structures not converted due to internal program problems: 468 - 0.19%
- CPU time: 7,920 sec. - 0.03 sec/molecule
  (**CORINA**™ Version 2.6, Intel Pentium III, 600 MHz, Linux workstation)

Results

High-Quality Molecular Models

The following figure shows the X-ray structures of three polymacrocycles and three organo-metallic compounds with the corresponding **CORINA**™ models and the *RMS* deviation between them.



DIJJUB: *RMS* = 0.69 Å | FEFZIZ: *RMS* = 0.14 Å | CISZUZ: *RMS* = 0.95 Å | VADTOJ: *RMS* = 0.15 Å | PVARHC: *RMS* = 0.24 Å | DAJKAI: *RMS* = 0.20 Å