communications in mathematical
match
and in computer chemistry

no. 37, March 1998

ISSN 0340-6253

MATCDY (37) 205-208 (1998)

MOLGEN 4.0

A. Kerber, R. Laue, T. Grüner[1], M. Meringer[1]

Lehrstuhl II für Mathematik,
Universität Bayreuth,
D-95440 Bayreuth, Germany

The central mathematical tool for the generation of combinatorial libraries[2] as well as for the automatic spectra interpretation within molecular structure elucidation is a generator of molecular graphs which is able to compute in an efficient and redundancy free way all the connectivity-isomers that fulfill certain conditions like prescribed molecular formula, bounds for ring sizes, multiplicities of bonds and other things that can be deduced from the observation of spectra or that come from information on the medical or biological effect of particular substructures in molecules, say.

There exist different computer programs for that purpose[3]. One of them is the **MOLGEN** generator that has been developed at the University of Bayreuth. The current version 3.5 is available for Windows NT and Windows 95.

During the use of **MOLGEN** in research and in industry it turned out that the mathematical concept to use prescribed conditions (the so-called goodlist of prescribed substructures and the badlist of forbidden ones) as a filter *after generation* considerably limits the use of this software package, since sometimes there are billions of isomers that have to be constructed in contrast to the few ones that fulfill all the conditions.

[1]Financially supported by BMBF under contract KE7BA 1-4

[2]see the detailed description in T. Wieland: *Konstruktionsalgorithmen bei molekularen Graphen und deren Anwendung,* MATCH **36** (1997), 7-155

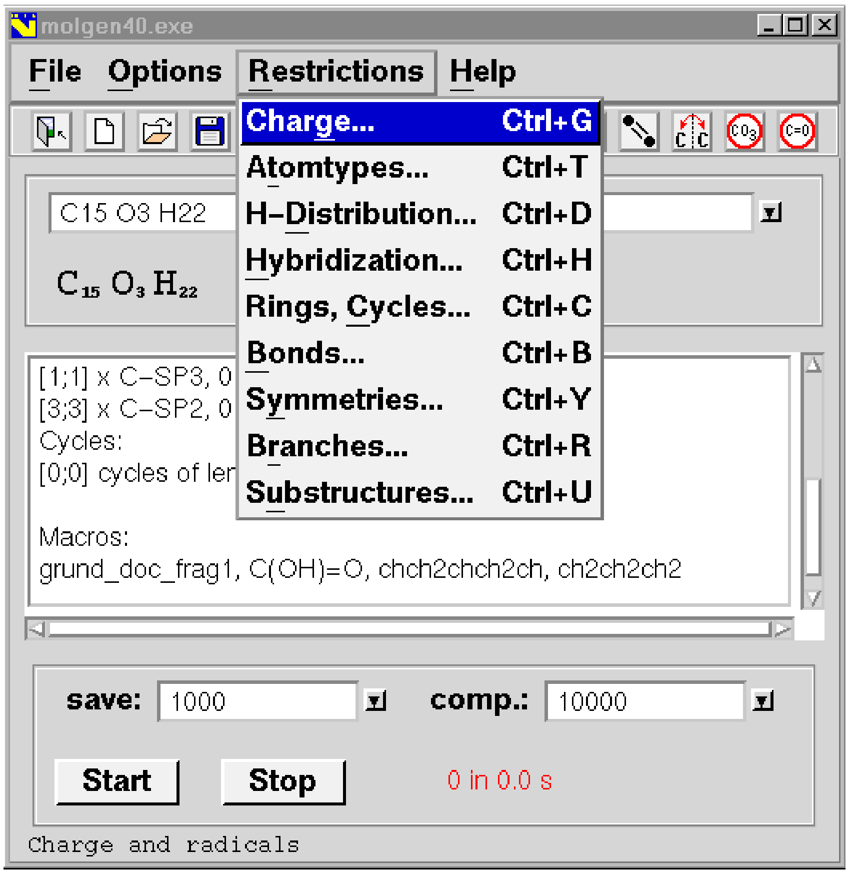
[3]see MATCH **27** for descriptions of several generators

Therefore, in order to improve the efficiency of our generator, we developed a new algorithm to

- allow much more detailed restrictions, so that (hopefully) all information that is known from experiments can be used in order to reduce the number of solutions and
- to use nearly all of the restrictions **during** the construction process in order to reduce the time-complexity of our algorithm.

In June 1998 the new **MOLGEN 4.0** generator will be available for Silicon Graphics, Sun Solaris, Dec Alpha and Linux. In Oct. 1998 **MOLGEN 4.0** will also be available for Windows NT.

Here is the main window of MOLGEN 4.0.



We combined a new construction strategy with modular design so that it is easy to control the sequence of the constructive steps by the input conditions.

Some of the *new* features of **MOLGEN 4.0** are:

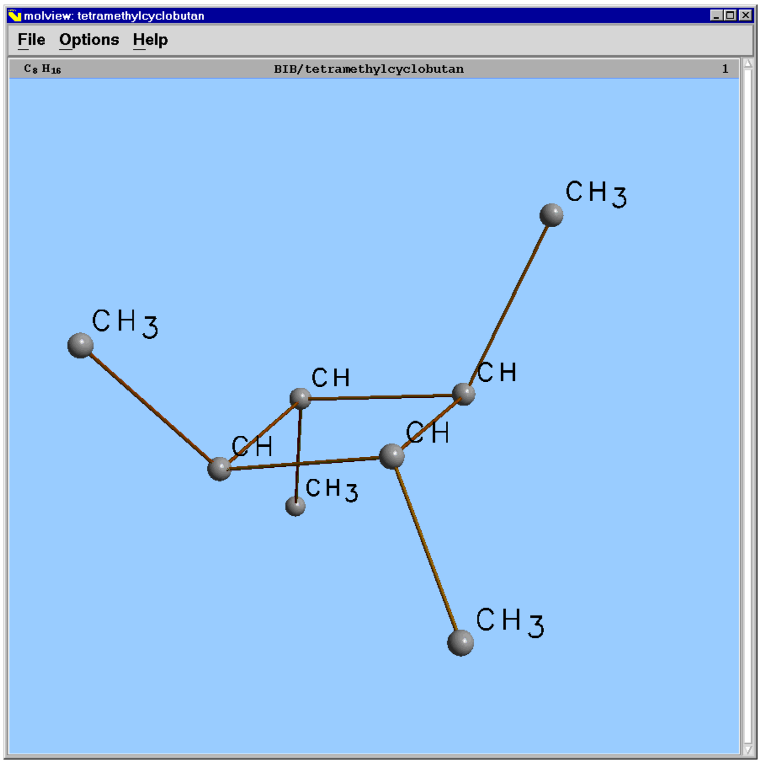
- *Extended data structures* allow the use of atom-charges and radical sites.
- The user can choose *intervals for the number of atoms* in the molecular formula, e.g. C_6H_{6-8} in order to run through several molecular formulae in the case when the exact formula is unknown.
- The user can now prescribe the H-distribution over the atoms in the hydrogen free skeleton.
- The hybridization states of some atoms can be prescribed in advance, too.
- The number of C-NMR signals can be given. The generator **MOLGEN 4.0** will then construct only such isomers, the (topological!) symmetry group of which has at most that many orbits on the set of carbon atoms. (Note that this number of orbits ia a *lower bound* for the possible number of C-NMR-signals.)
- Prescribed and forbidden substructures of the unknown with a very variable description of the *neighbourhood* of the substructures.
- Possible conflicts between some of the restrictions can automatically be prevented: you can switch the restrictions on and off independently.

Here is a table that shows how useful the impact of further conditions like hydrogen distribution and hybridization is. It shows clearly that the number of isomers that have to be constructed can enormously be reduced:

Restrictions	no. of isomers	CPU-time
Chemical formula $C_6H_8O_6$ only	2,558,517	838 s
no triple bonds	2,434,123	703 s
hydrogen distribution 1CH_2,2CH_1,3C,4OH	79,831	25 s
no substructure -O-O-	35,058	97 s
hybridization 1Csp3-2H,2Csp3-1H,3Csp2-OH,1Osp2-OH	990	8 s
minimal size of rings =5	348	5 s
contains at least one CO_3 branch	15	11 s

Besides these features

- there is a new OpenGL-based 3-dimensional graphics for our molecule viewer:



- We have also an improved molecule-editor with user-friendly editing functions.

More details can be found on the home page of **MOLGEN**

http://www.mathe2.uni-bayreuth.de/molgen

The email number is

molgen@btm2x2.mat.uni-bayreuth.de