Chemical Education via **MOLGEN**

C. Benecke, R. Grund, A. Kerber, R. Laue, T. Wieland

Department of Mathematics,
University of Bayreuth,
D-95440 Bayreuth,
Germany

Abstract

The program system MOLGEN is a software tool for chemical education and research. It allows to generate and to visualize all the *constitutional formulae* (=connectivity isomers) corresponding to a given gross formula. In order to restrict the usually very big number of such isomers, the user can prescribe or forbid certain substructures and the class of the substance can be chosen. Moreover, MOLGEN also allows to compute all *configurational isomers* to a given structural isomer as well as their realizations in space.

Introduction

A main topic at certain school classes, university lectures or seminars is the very large variety of hydrocarbon molecules such as the well known benzene. Therefore it is important to demonstrate to the students that there is not only the famous benzene ring that corresponds to the gross formula C_6H_6 but that there are many constitutions of that particular form. Few students probably know of the existence of 217 constitutional formulae, and they will be astonished to hear that about 50 of them have been observed in nature or laboratory up to now. It is quite clear that this fact and the particular example requires demonstration on the screen of a computer monitor.
MOLGEN provides a solution to that problem. It was awarded the German-Austrian University Software Prize 1993 as an excellent educational software in chemistry. Several schools, universities and chemistry companies in Germany use this program which is the result of a research project supported by the Deutsche Forschungsgemeinschaft for several years.
We will describe **MOLGEN** by presenting short examples taken from chemical education and research.

Learning with MOLGEN

Consider, for example, benzene. The 217 constitutional formulae corresponding to its chemical formula C_6H_6 are obtained by **MOLGEN** within a fraction of a second.
After generation a visualization of the result is needed. In setting up the display a two-dimensional placement function is called for drawing the molecules. Having done this for the 217 isomers of benzene, you will see the well-known benzene ring showing up as number 214 (see fig. 1). The molecule no. 215 is benzvalene, which was in fact discovered just 25 years ago, while its existence had been postulated by E. Hückel already in 1937.

Teaching with the help of MOLGEN

There are many cases, when we do *not* want to show or even to calculate *all* cases, but only the cases belonging to a given class of chemical substances. For example: The chemical formula for alkanoles is $C_nH_{2n+1}OH$, where the structural element OH, the hydroxyl group, is emphasized. In this particular case we want to compute or to see only such constitutional formulae that consist of n carbon atoms together with $2n+2$ hydrogen atoms and a single oxygen atom which has to belong to a hydroxyl group. The interested user can easily enter the chemical formula C_3H_8O together with the substructure OH and check that there are in fact two connectivity isomers, whereas there are three of them if the existence of the hydroxyl group is not prescribed.
Next is a corresponding exercise taken from an A-level examination in chemistry at Bavarian high schools in 1992 which nicely allows to demonstrate the use of **MOLGEN**: In a first question the brutto formula for tartaric acid was derived as $C_4H_6O_6$. The exercise continues:

1.2 During an acid-base titration it was discovered that 1 mol of tartaric acid is equivalent to 2 mol of sodium hydroxide. Derive all possible constitutional structures compatible with this result and the chemical formula derived in 1.1. ...

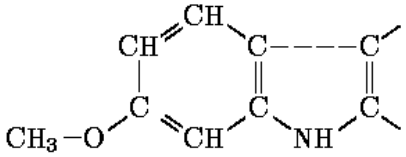
For the demonstration with **MOLGEN** the hint 1.2 should be interpreted as the existence of two substructures of the form COOH, which have to be prescribed. After entering all these together with the gross formula, **MOLGEN** produces 22 solutions. Their examination shows several of them containing substructures that are quite uncommon. For example, the solutions numbered 17 and 18 (among others) contain two bonded oxygens (peroxo groups).
If you want to abandon such cases, you may proceed as follows: Besides the two substructures COOH prescribe two hydroxyl groups OH in addition. Thus, after another run of the generator, only 4 constitutional formulae remain (all solutions containing two oxygen atoms between two carbon atoms were skipped). Two of these remaining 4 solutions show the hydroxyl groups attached to the same carbon atom, and so we may cancel them by an application of the Erlenmeyer Rule, which is still used in under-graduate chemistry. Finally we arrive at the two constitutional formulae suggested by the people who designed that exercise.
We recall that we started using MOLGEN from altogether 22 isomers, the final two solutions of the problem are the isomers numbered 1 and 4 (see figure 2).
In practice, the chemist will see by inspection that several further restrictions must be imposed. Therefore the built-in graphical editor allows to isolate and to define substructures which are either prescribed or forbidden. Another run of the generator can respect these conditions, and so, in an interactive way, the total amount of isomers can be reduced to a significantly smaller set of candidates which must be examined.

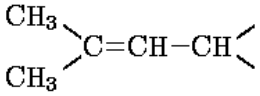
Research using MOLGEN

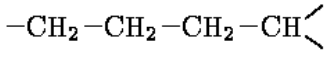
MOLGEN allows to handle cases which were impossible before efficient generators became available. Here are a few remarks concerning applications in research:

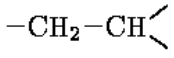
- **MOLGEN** provides *all mathematically possible solutions*, several of which may chemically be plausible but unknown yet and therefore may correspond to *new substances*.
- The *identification* of chemical substances by using spectroscopic data is a central problem in chemical laboratories. The wish for automatization may arise since many of such problems occur daily, for example in environmental research. This identification clearly needs a generator that provides all the possible cases from which the correct ones have to be taken! A major task is to reduce the amount of candidates by sophisticated theoretical and practical methods.
- In the first approximation, molecules are described by their gross formula. The second approximation is its constitutional formula. MOLGEN allows to do this second step in an efficient way. Moreover, another module of the MOLGEN program system allows to generate the corresponding *stereo isomers*, as well as three-dimensional realizations of them, using Allinger's MM2 energy model (a simplified version). See tetra-methyl-cyclo-butane in fig. 3 as an example.
- The central part of an expert system for *molecular structure elucidation* therefore is a *generator* like **MOLGEN**. It generates all mathematically possible constitutional formulae and in addition the possible stereo isomers compatible with a given set of chemical conditions. Molecular structure elucidation thus may proceed in the following way: Spectroscopic data, obtained from MS-, IR- or NMR-spectra, are translated to a chemical formula together with a set of conditions which can be used as an input for **MOLGEN**. After a run of the generator the resulting set of candidates is examined by a chemist. Usually additional restrictions are found for a next run of the generator. This procedure is iterated until the set of candidates is small enough to find the solution by visual inspection.

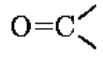
If you enter, for example, the gross formula $C_{12}O_2Cl_4H_4$ together with the single substructure of the dioxin skeleton (substitutes removed), within a second you obtain from **MOLGEN** the 22 constitutional isomers of the dioxin, three of which are shown in figure 4.
Here is, finally, a complex problem from molecular structure elucidation in industry which we solved by using **MOLGEN**. The gross formula given was $C_{22}H_{25}N_3O_3$, and we were told that the molecule in question contains the following 6 non-overlapping substructures (macro atoms, in terms of **MOLGEN**) F_1, ..., F_6:



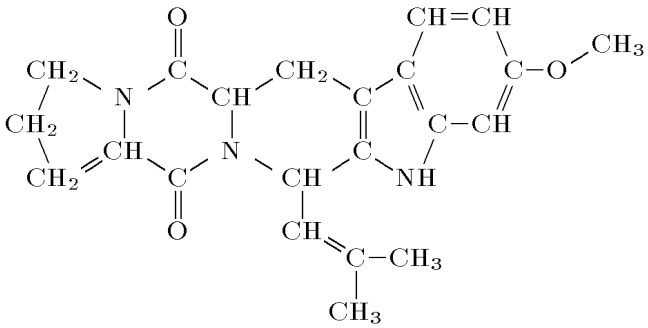








Using these substructures, the gross formula reduces to $F_1F_2F_3F_4F_5F_6N_2$. **MOLGEN** obtains (after several steps which we won't describe in detail here) 2,337 constitutional formulae for that reduced gross formula, and after expansion it constructs 8,916 isomers. Now we used the following restrictions: no triple bond and ring sizes between 5 and 6. Only 201 candidates remained from which an expert easily obtained the correct solution:



MOLGEN runs under DOS on a PC, versions for Windows, OS/2, Macintosh and various workstations are in preparation. There exists a limited version for education and a full version for science and research. For more information contact the address given below.

Authors:

Prof. Dr. A. Kerber, Prof. Dr. R. Laue, Dipl.-Math. R. Grund,
Dipl.-Math. C. Benecke, Dipl.-Math. T. Wieland, R. Hohberger.

Department of Mathematics, University of Bayreuth
D-95440 Bayreuth, Germany
Telephone (0921) 55-3388, Fax (0921) 55-3385
E-mail: kerber@btm2x2.mat.uni-bayreuth.de

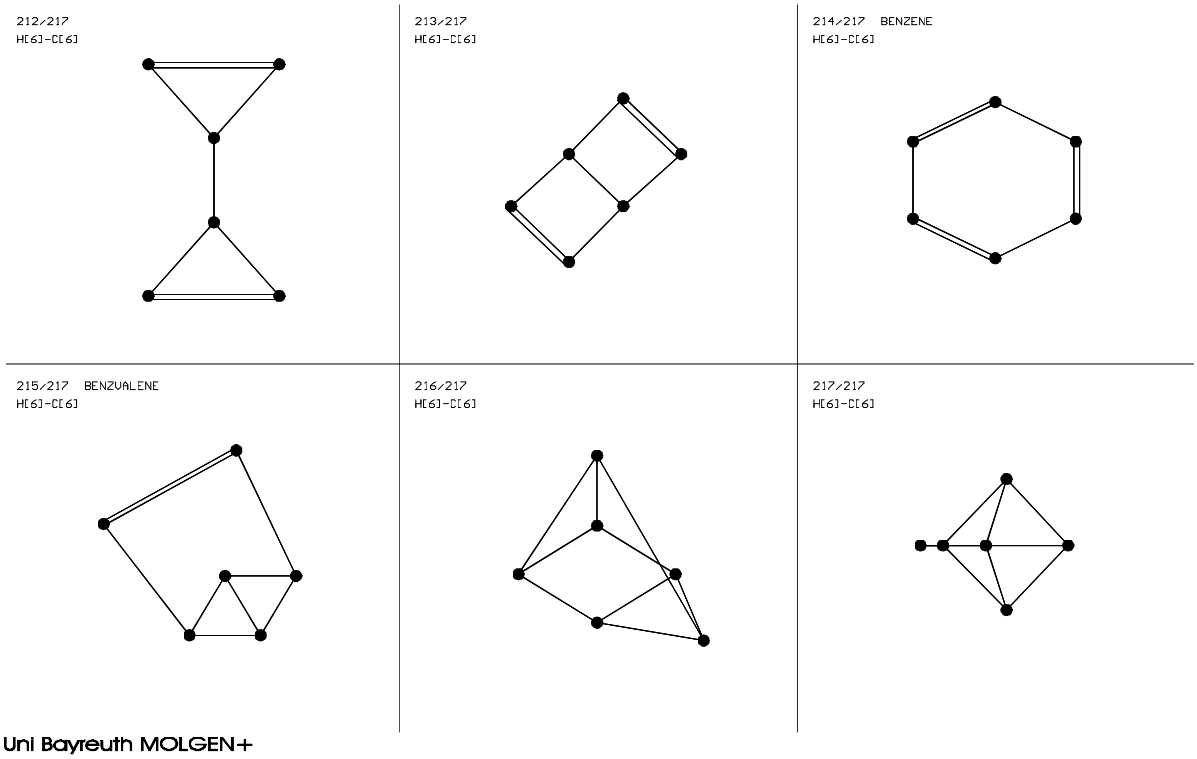


Figure 1: Some of the connectivity isomers of benzene

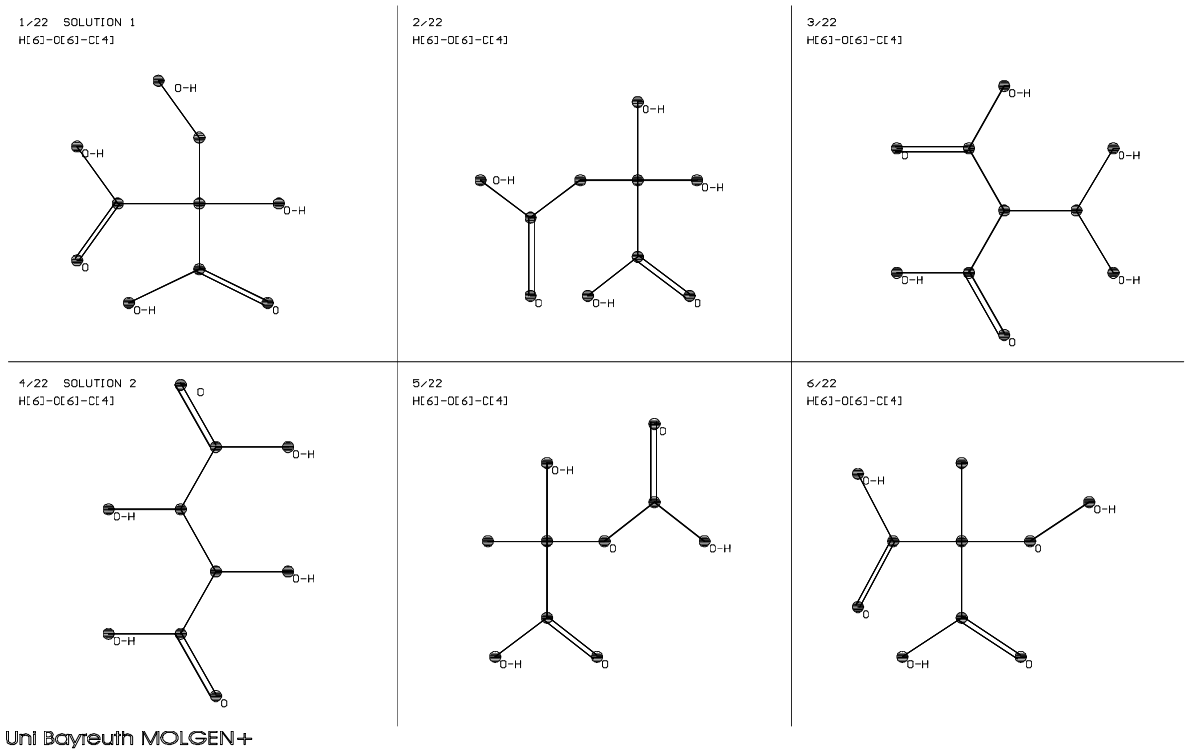


Figure 2: The solutions 1-6 of the high school exercise

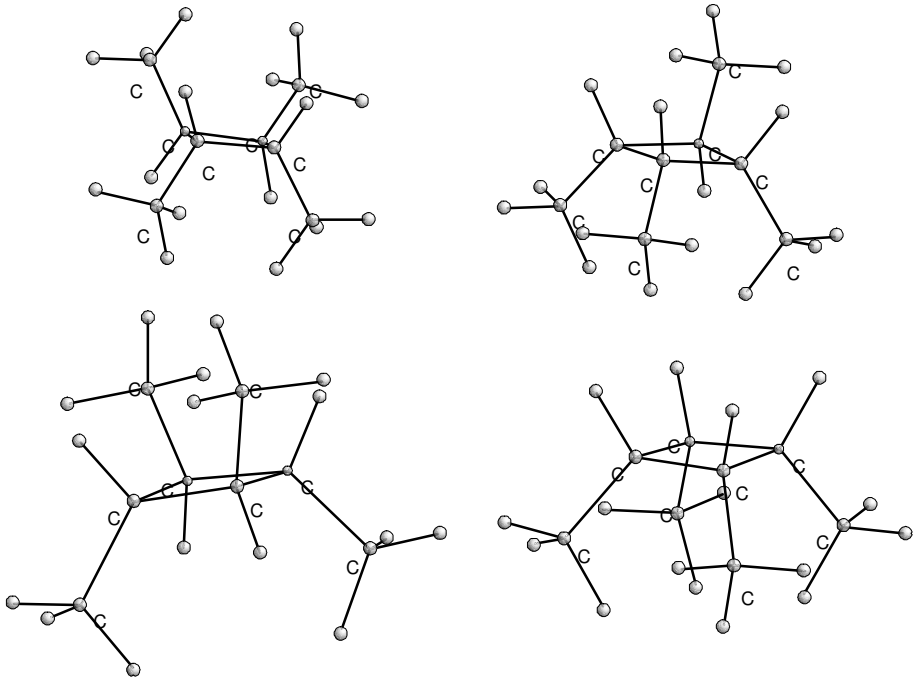


Figure 3: Stereo-isomers of 1,2,3,4-tetra-methyl-cyclo-butane

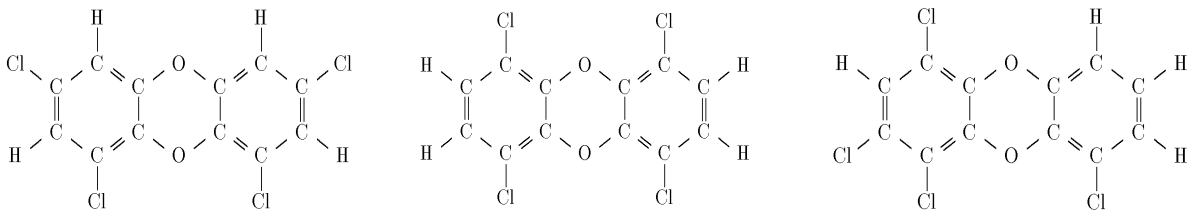


Figure 4: Three isomers of dioxin