

An application of the structure generator MOLGEN to patents in chemistry

A. Kerber, R. Laue, M. Meringer*

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

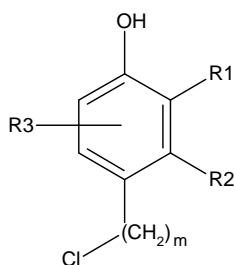
1 Introduction

MOLGEN[2, 3, 4] is a software package for the fast and redundancy free generation of structural formulae corresponding to prescribed data such as molecular formula, ring sizes, required or forbidden substructures, hydrogen distribution, hybridization etc. A particular version, MOLGEN-COMB[5], generates combinatorial libraries, starting from a given central molecule.

The generation of a library corresponding to a given Markush formula, as encountered in chemistry patents is a quite similar task, and so MOLGEN applies to this problem as well.

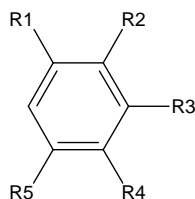
2 The Comparison of two Markush Formulae

Consider two Markush formulae, say the following, taken from [1]



$R1$: methyl or ethyl (substituent variation)
 $R2$: alkyl (homology variation)
 $R3$: amino (position variation)
 m : 1-3 (frequency variation)

and the following, composed by us,



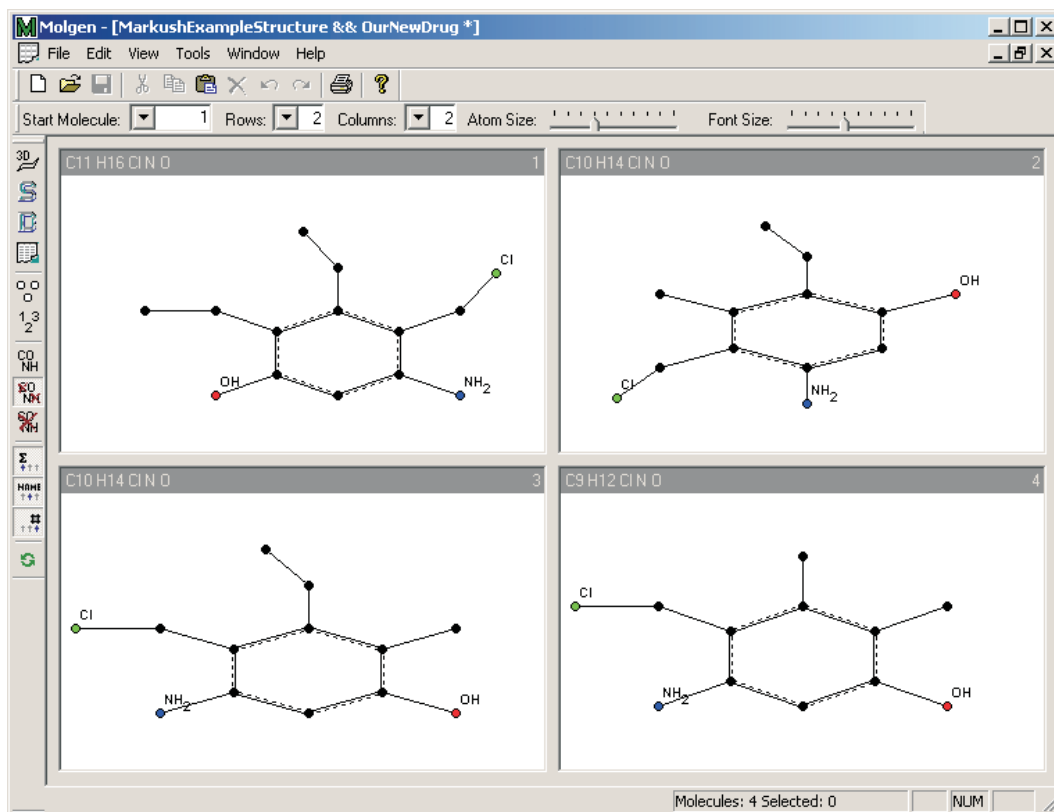
$R1$: methyl, ethyl, OH
 $R2$: alkyl (1-6 carbon atoms)
 $R3$: OH, OCH_3 , OC_2H_5 , Methyl, Ethyl
 $R4$: OH, CH_2Cl , amino
 $R5$: H, methyl, ethyl, amino

*Financial support by the BMBF and the DFG is gratefully acknowledged

which may be claimed in two patents. We ask whether there are patent infringements, i.e. structural formulae occurring in both libraries. In order to solve this problem, MOLGEN can be used as follows:

1. We restrict attention in $R2$ of the first example to maximal 6 carbon atoms. In this case, MOLGEN generates 33 alkyl residues for $R2$. These 33 structures are stored in a separate library for $R2$.
2. MOLGEN produces (on a PC with a P4 of 2.5GHz) for the first case a library of 396 structural formulae within 4.9 seconds, and for the second Markush formula we get 5939 structural formulae, in 44 seconds. The interested reader should note that the purely combinatorial argument would give 5940 structural formulae in the second case, but since the data are in canonical form in MOLGEN, the doublette is skipped automatically.
3. Comparing these two libraries of 396+5939 structural formulae, MOLGEN finds, within ten seconds, that there is an intersection of exactly 4 structural formulae, and so there are exactly 4 potential patent violations.

Here is the intersection:



References

- [1] J. Barnard and G. Downs. Use of markush structure techniques to avoid enumeration in diversity analysis of large combinatorial libraries. Oral presentation at the MSI Combinatorial Chemistry Consortium Meeting, 1997,

and the Daylight Chemical Information Systems MUG97 meeting, 1997.
<http://www.daylight.com/meetings/mug97/Barnard/970227JB.html>.

- [2] C. Benecke, R. Grund, R. Hohberger, R. Laue, A. Kerber, and T. Wieland. MOLGEN+, a generator of connectivity isomers and stereoisomers for molecular structure elucidation. *Analytica Chimica Acta*, 141–147, 1995.
- [3] C. Benecke, T. Grüner, A. Kerber, R. Laue, and T. Wieland. Molecular structure generation with molgen, new features and future developments. *Fresenius J. Anal. Chem.*, 358:23–32, 1997.
- [4] T. Grüner, A. Kerber, R. Laue, and M. Meringer. Molgen 4.0. *MATCH*, 37:205–208, 1998.
- [5] R. Gugisch, A. Kerber, R. Laue, M. Meringer, and J. Weidinger. MOLGEN-COMB, a software package for combinatorial chemistry. *MATCH*, 41:189–203, 2000.