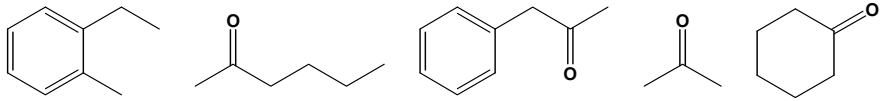
Software SubMat Application: Substructure search

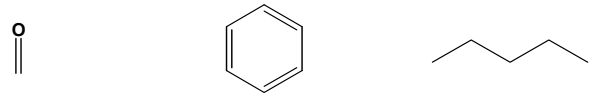
A tiny demo example

Input file (*.SDF) with 5 molecular structures



1 2 3 4 5

Input file (*.SDF) with 3 substructures



1 2 3

Output file (*.TXT) with the generated binary substructure descriptors (matrix 5 × 3).

0	1	0
1	0	1
1	1	0
1	0	0
1	0	1

Each molecular structure corresponds to a row
Each substructure corresponds to a column
1, **0** substructure present, absent

Info: Kurt Varmuza
kurt.varmuza@tuwien.ac.at, www.lcm.tuwien.ac.at/vk/
c/o Vienna University of Technology (Austria)

Remarks

SubMat does not consider 3D or stereo data.

SubMat produces a result file in text format that can be easily imported by other software.

Binary substructure descriptors are used for a variety of tasks in chemoinformatics, e. g.,

+ Structure similarity searches
+ Cluster analysis of chemical structures
+ Characterization of the diversity of chemical structure libraries
+ Modeling structure property/activity relationships - preferably together with other molecular descriptors

Example with more structures

6458 molecular structures (substances used for Ames mutagenicity tests) [1];
1365 substructures [2].
Computing time 10 s (3.4 GHz processor, MS Windows 10). Result file (*.TXT, comma-separated) 17 MB (matrix 6458 × 1365)

[1] Hansen K. et al.: J. Chem. Inf. Model. **49**, 2077 (2009)
[2] Varmuza K. et al.: Croat. Chem. Acta, **78**, 141 (2005)