

## Abstract

*Source Code for Biology and Medicine*, 7 [12] 1-4 (2012)

Grabner M., Varmuza K., Dehmer M.:

### **RMol: a toolset for transforming SD/Molfile structure information into R objects.**

**Background:** The graph-theoretical analysis of molecular networks has a long tradition in chemoinformatics. As demonstrated frequently, a well designed format to encode chemical structures and structure-related information of organic compounds is the Molfile format. But when it comes to use modern programming languages for statistical data analysis in Bio- and Chemoinformatics, R as one of the most powerful free languages lacks tools to process Molfile data collections and import molecular network data into R.

**Results:** We design an R object which allows a lossless information mapping of structural information from Molfiles into R objects. This provides the basis to use the RMol object as an anchor for connecting Molfile data collections with R libraries for analyzing graphs. Associated with the RMol objects, a set of R functions completes the toolset to organize, describe and manipulate the converted data sets. Further, we bypass R-typical limits for manipulating large data sets by storing R objects in bz-compressed serialized files instead of employing RData files.

**Conclusions:** By design, RMol is a R toolset without dependencies to other libraries or programming languages. It is useful to integrate into pipelines for serialized batch analysis by using network data and, therefore, helps to process sdf-data sets in R efficiently. It is freely available under the BSD license. The script source can be downloaded from <http://sourceforge.net/p/rmol-toolset>.