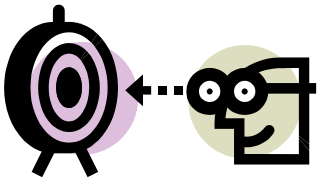
# USER CUSTOMIZED MOLECULAR DIVERSITY ANALYSIS USING THE MOLDIA SOFTWARE

Jean-Pierre Doucet, Ana G. Maldonado, Michel Petitjean and Bo-Tao Fan†

ITODYS, Institut de Topologie et de Dynamique des Systèmes, CNRS UMR-7086, Université Paris-7
1, rue Guy de la Brosse, 75005 Paris, France.

*† This poster is dedicated to the memory of Prof. Bo Tao Fan.*

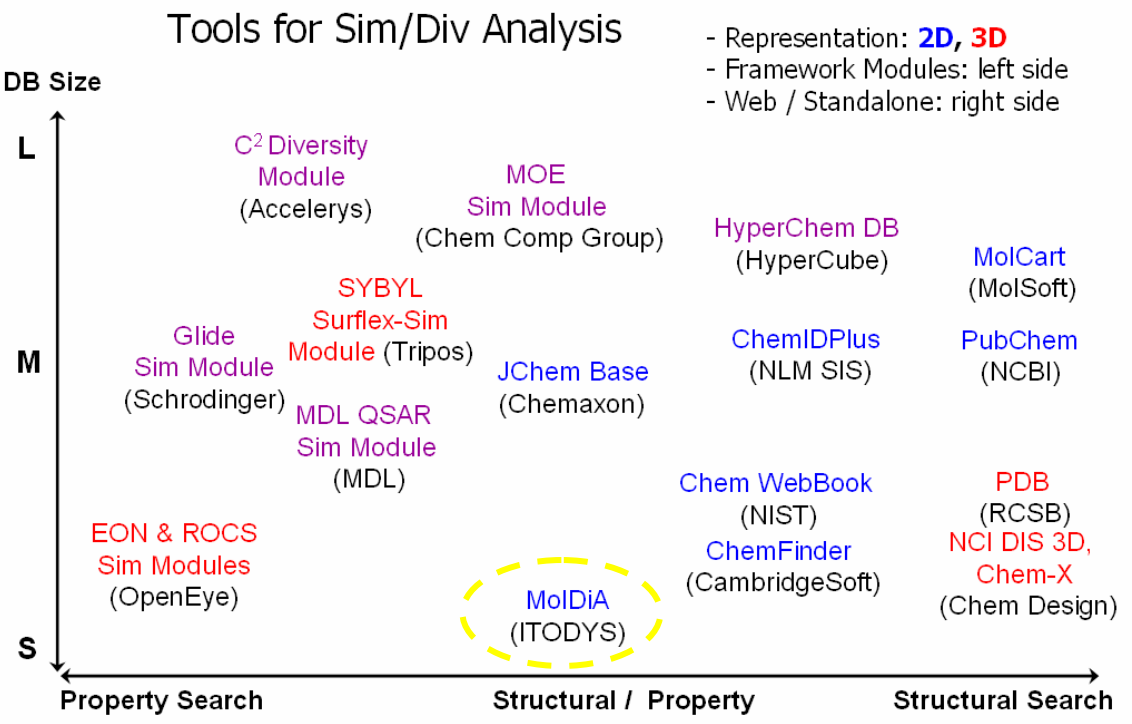
# Problematic



**`Today's chemist problems:`**
* retrieval and analysis of known molecules
* search and synthesis of *new* materials, drugs, catalysts, additives, etc.

**`BUT...`**
Commercial, academic or industrial molecular databases available are *too large* (CAS~32mill) and *not standardized* (>80 formats).

**`One solution is:`** computer assisted molecular design.

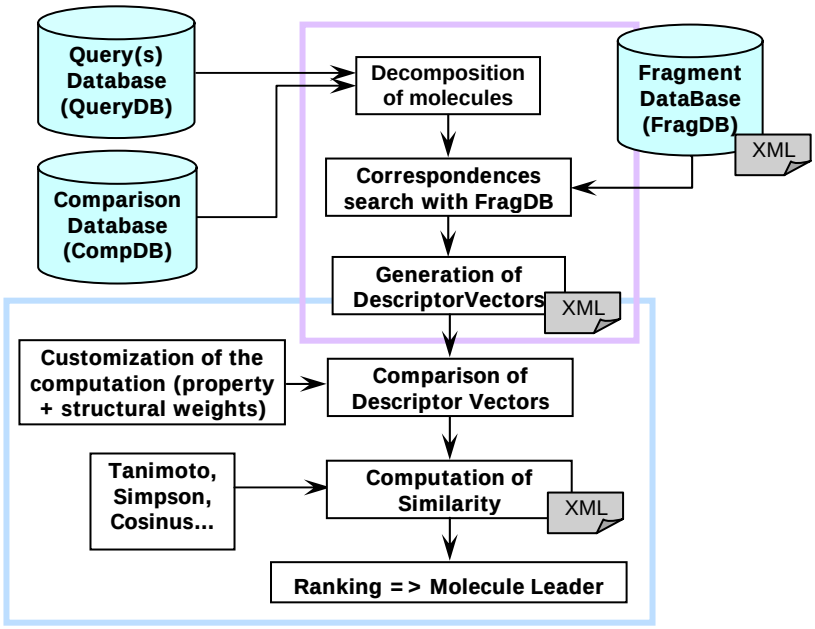


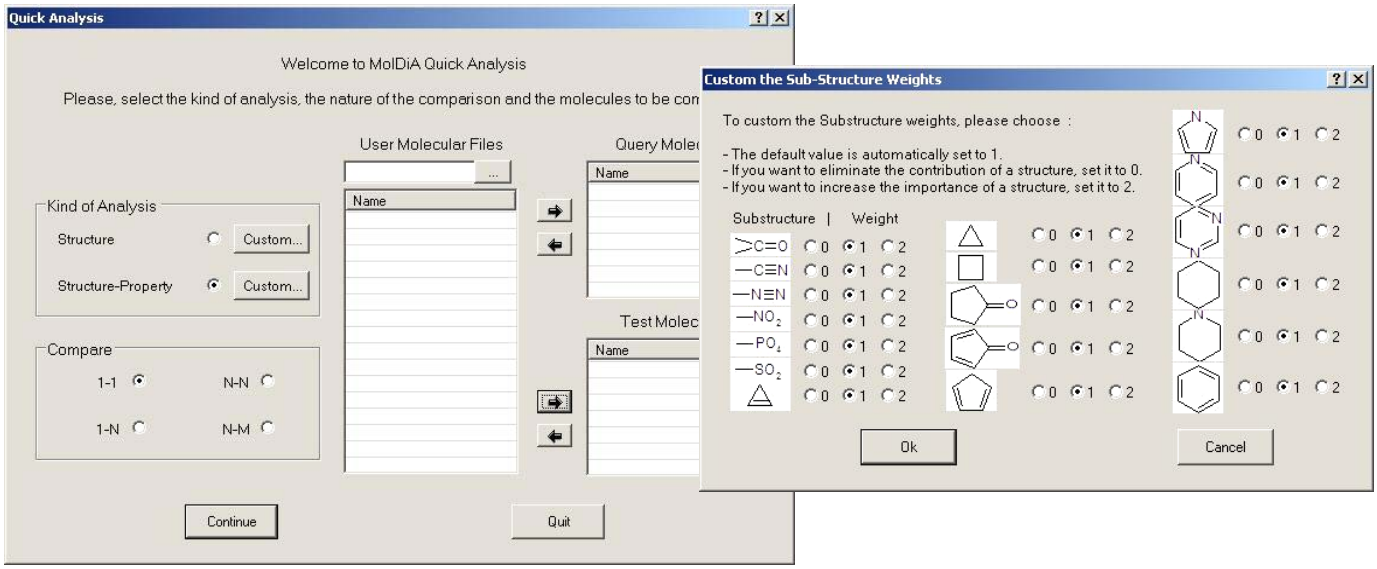

**`BUT...`**
*We*, as chemists have clear notions about the features that we're looking for (structure, size, functional groups, properties, etc):
* good parametrization/customization tools are necessary
* a comprehensible GUI (graphic user interface) is as important as the calculus features or the power of computation of the software

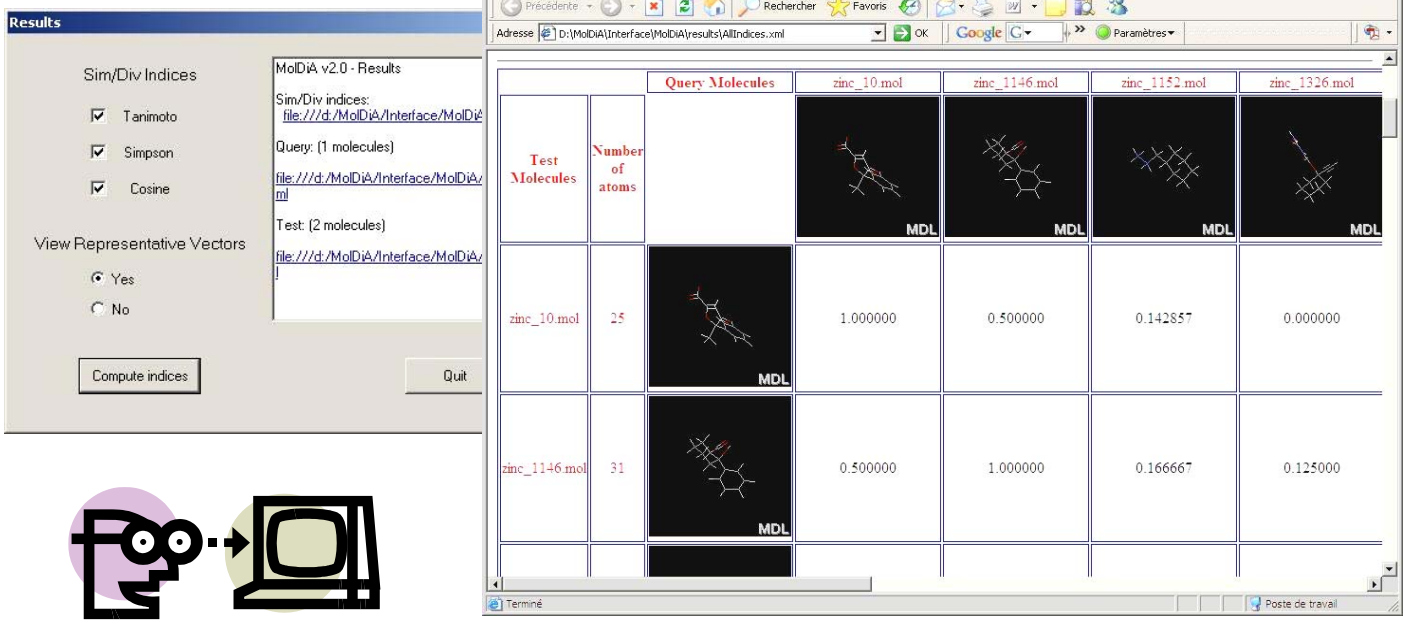
# MolDiA

* Molecular Databases:
FragDB, CompDB, QueryDB
* Molecular Representations:
2D <DescriptorVector>
* Similarity Computation:
Different indices (T, S, C)

Graphic Interface:
- User customization
- Web compatibility



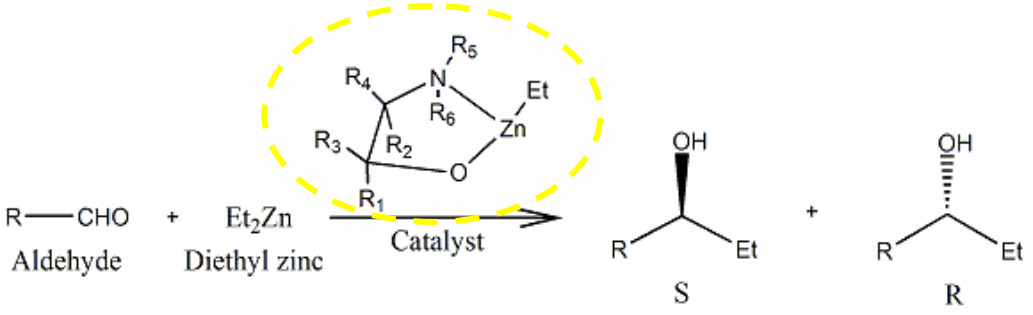



# Similarity/Diversity Analysis

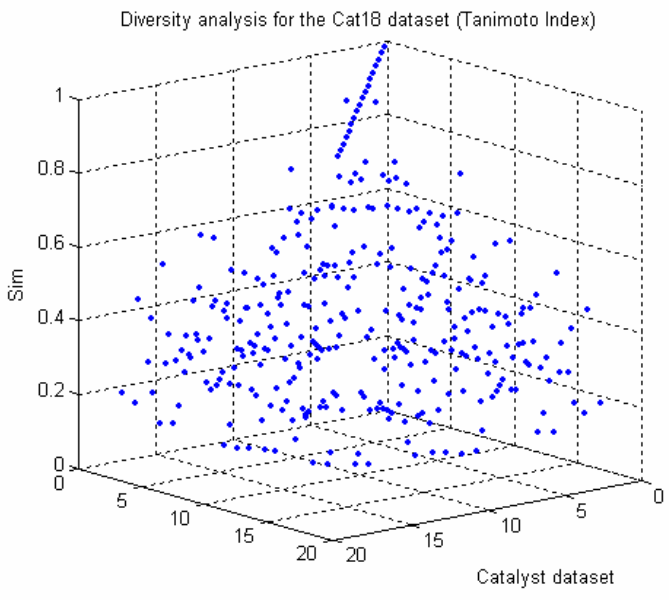
**Some parametrization examples:**

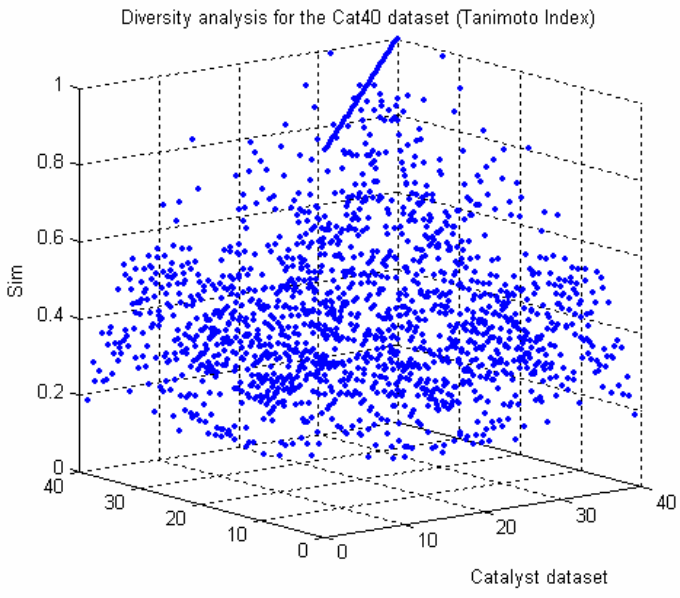
I) Comparative diversity analysis of a Zn-aminoalcohols catalysts dataset and his structural extension, using the Tanimoto index




Asymmetric alkylation of aldehydes catalyzed by Zn-aminoalcohols

* Basis set: 18 catalysts from literature. By changing the groups $R_1$-$R_6$ a structurally extended set of 40 catalysts is obtained



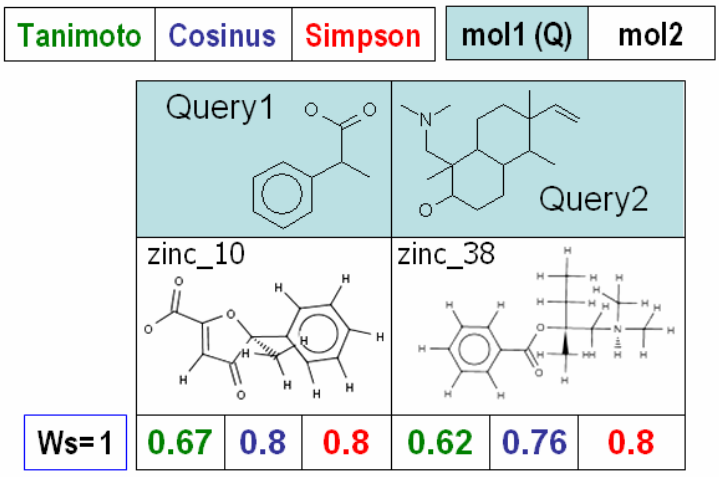

A quick diversity analysis shows that:

=> increasing the fragment diversity does not necessarily increase the catalyst diversity. Both sets are similarly diverse

=> Using MolDiA is a simple and fast way to analyze datasets

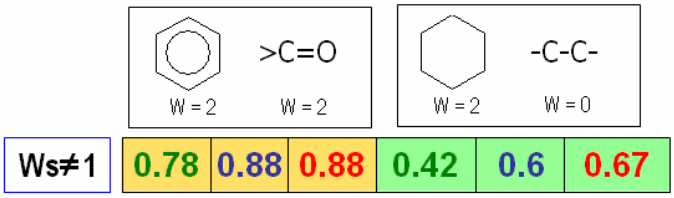
II) Similarity indices using structural weight analysis (Ws) for some 'ZINC database' molecules http://www.blaster.docking.org/zinc

* Three different indices used:




* Use of default Ws values
- poor analysis for general situations

=> Use of customized MolDiA Ws:
- similarity **increases** or **decreases** when the relevant fragments are taken into account




# Conclusions

* Similarity and Diversity analysis using several levels of comparison, different kinds of weights and indices
=> easy user-customization of the analysis
=> user can effectuate data fusion and extrapolation
=> comparison analysis fast and simple

* User interface and output using XML stylesheets
=> no need to program or execute scripts
=> easy export of results
=> good Web compatibility

Comments? Questions? Further information? Please contact:
Prof. Jean-Pierre Doucet, doucet@paris7.jussieu.fr
Dr. Ana Maldonado, maldonad@science.uva.nl