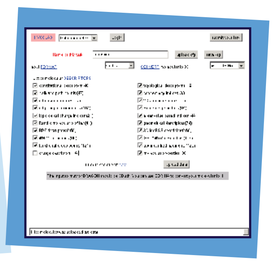
Virtual Computational Chemistry Laboratory http://vcclab.org

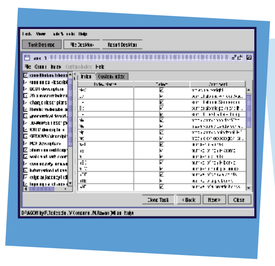
Igor V. Tetko,1,* J. Gasteiger,2 R. Todeschini,3 A. Mauri,3 D.J. Livingstone,4 P. Ertl,5
V.A. Palyulin,6 E.V. Radchenko,6 N.S. Zefirov,6 A.S. Makarenko,7 V.Yu. Tanchuk,8 V.V. Prokopenko,8

1- GSF – Centre for Environment and Health, Institute for Bioinformatics (MIPS), 2-University of Erlangen-Nürnberg, Computer-Chemie-Centrum, Erlangen, Germany,
3-Milano Chemometrics and QSAR Research Group, Department of Environmental Sciences, Università di Milano - Bicocca, Milano, Italy,
4-Centre for Molecular Design, Portsmouth and ChemQuest, UK, 5-Novartis Institute for BioMedical Research, Cheminformatics, Switzerland,
6-Moscow State University, Department of Chemistry, Moscow, Russia, 7-Institute of Applied System Analysis, Kyiv, Ukraine,
8- Institute of Bioorganic & Petroleum Chemistry, National Ukrainian Academy of Sciences, Kyiv, Ukraine

Calculation of Descriptors

E-DRAGON is the electronic remote version of the DRAGON, which is an application for the calculation of molecular descriptors developed by the Milano Chemometrics and QSAR Research Group of Prof. Todeschini. DRAGON provides more than 1,600 molecular descriptors that are divided into 20 logical blocks. The user can calculate not only the simplest atom type, functional group and fragment counts, but also several topological and geometrical descriptors. If the 3D atom coordinates are not available for molecules, the user can calculate 3D coordinates using CORINA, provided by the group of Prof. Gasteiger.

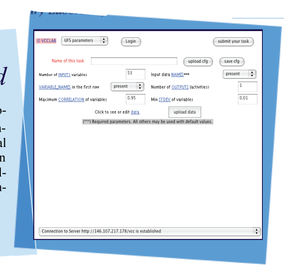




Parameter Client (PCLIENT) is an extension of E-Dragon and encloses three different index generation programs, namely DRAGON, atom-type and bond-type E-state indices and Fragment-based indices. The conversion of molecules to 3D structure is provided using CORINA. This software makes it possible to generate more than 3000 indices.

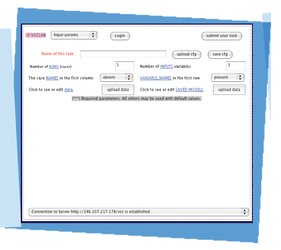
Dimensionality Reduction

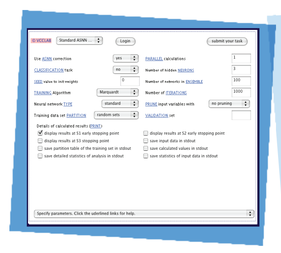
Unsupervised Forward Selection (UFS) is a data reduction algorithm that selects from a data matrix a maximal linearly independent set of columns with a minimal amount of multiple correlation. This software can be used to process indices generated by index calculation programs in order to decrease data redundancy.



Data Analysis

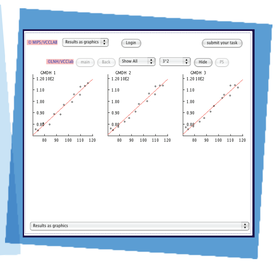
Partial Least Squares (PLS) uses a two-step descriptor selection procedure to significantly increase the predictive ability of the obtained models. The first step eliminates low-variable descriptors and the second step optimises the descriptor subset using a Q2-guided descriptor selection by means of a genetic algorithm. The computational experiments demonstrate the stability and good prediction accuracy of models.

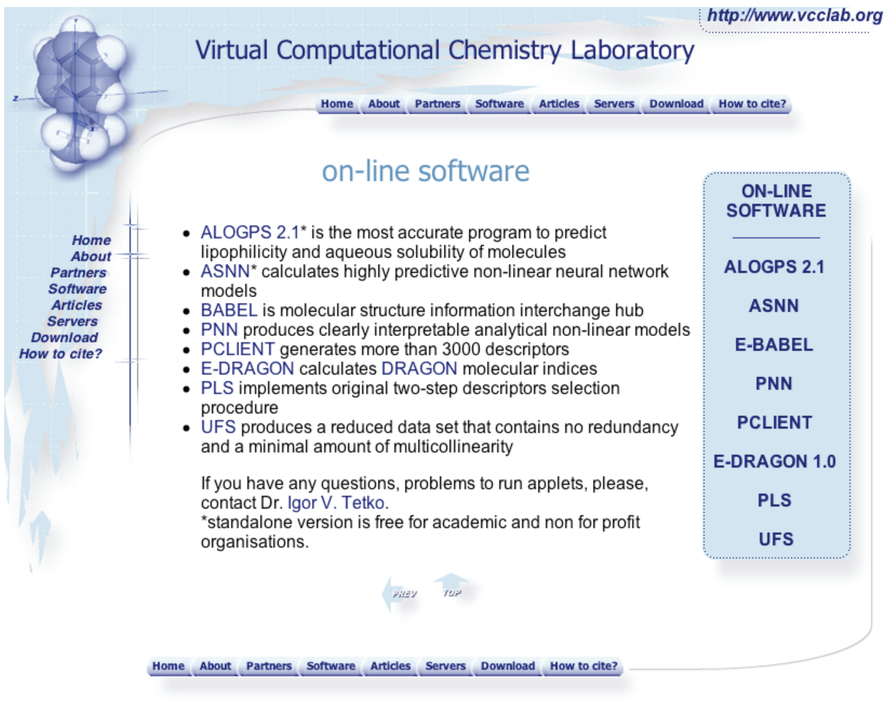


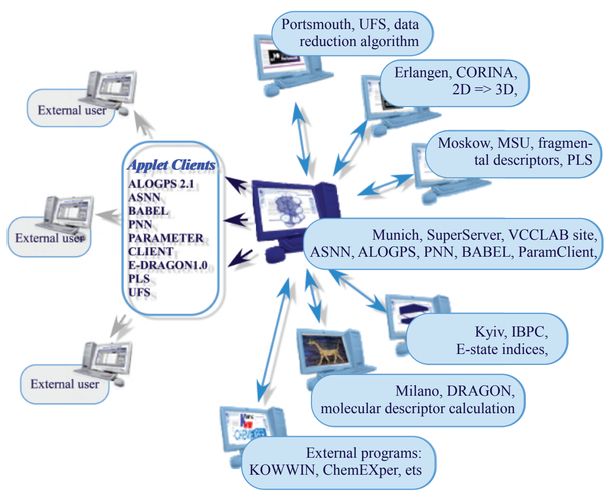


Associative Neural Network (ASNN) represents an innovative method to calculate non-linear models between indices and molecular properties. The method represents a combination of an ensemble of feed-forward neural networks and the k-nearest neighbour technique. If new data become available, the network further improves its predictive ability and provides a reasonable approximation of the unknown function without a need to retrain the neural network ensemble.

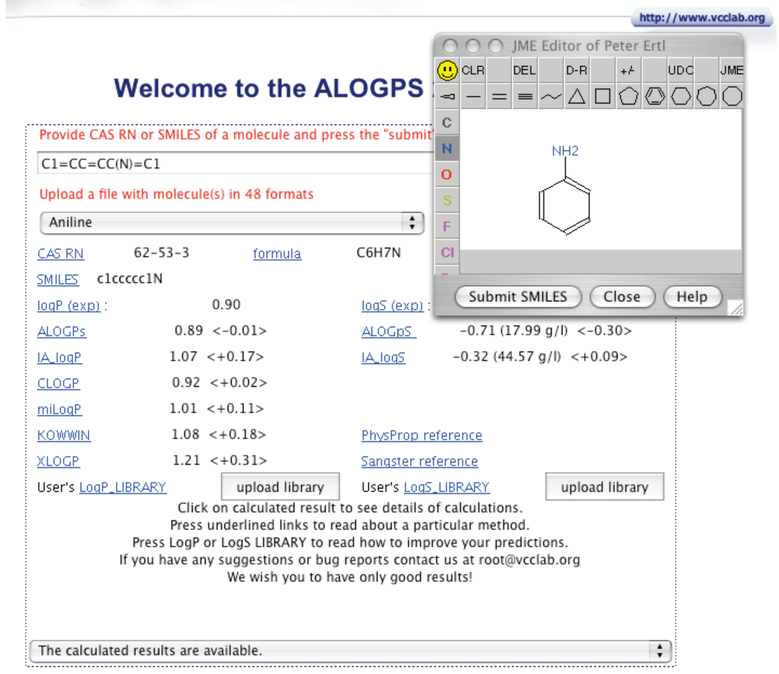
Polynomial Neural Network (PNN) correlates input and target variables using (non) linear regression. In this particular software the user can define the desired properties of the solution such as the number of terms and the maximum degree of polynoms. The PNN calculates analytical models that could be easily interpreted. This is a substantial advantage of this method over other neural network approaches. Both approaches were recently compared to several other neural network methods using several QSAR datasets.



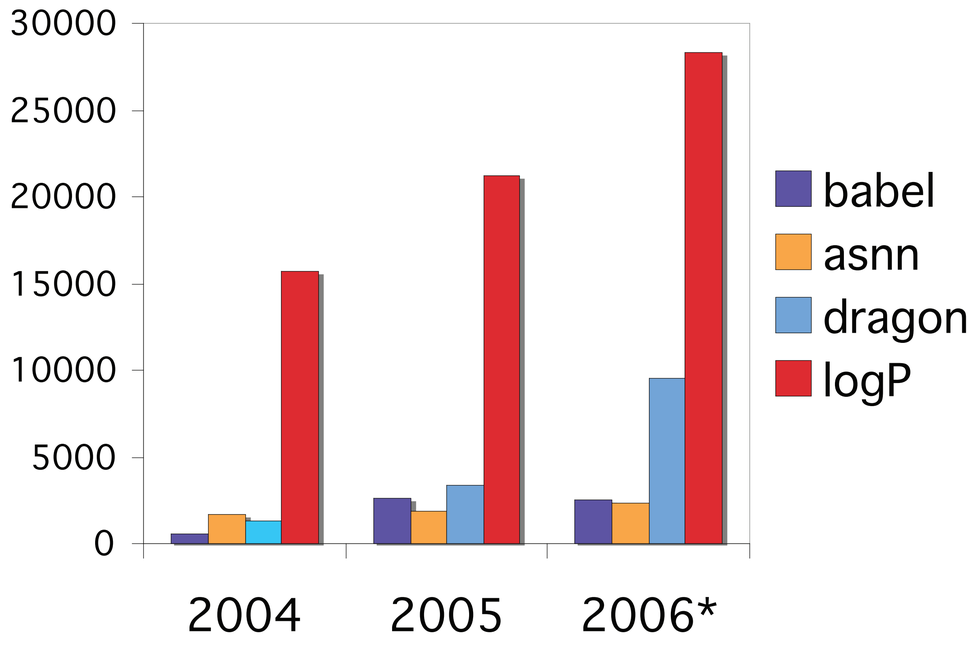




Lipophilicity, logP and aqueous solubility, logS

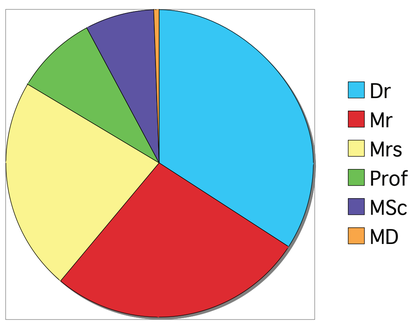


Popular VCCLAB tasks

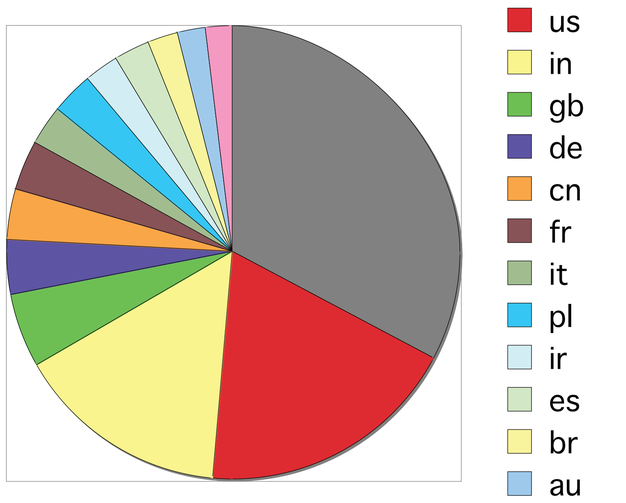


Who are the VCCLAB users?

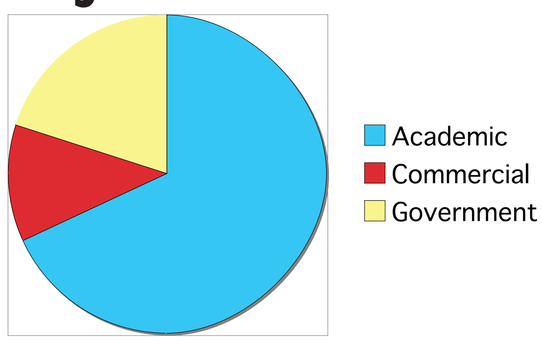
Education



Countries



Organisations



Future plans

Exporting of some tasks as Web Services
Estimation of Applicability Domain of logP/logS predictions
Similarity search in logP/logS spaces
Calculation of new ADME/T properties
HTML interface for non-Java users

Acknowledgement This work was partially supported with INTAS-INFO 00-0363 grant.