Book of Abstracts
and Scientific Program

International School-Seminar on Computer-Aided Molecular Design

May 18-20, 2016
Kazan, Russia

M. Glavatskikh[1,2]
T. Madzhidov[2]
V. Solov'ev[3]
D. Horvath[1]
G. Marcou[1]
A. Varnek[1,2]

PREDICTIVE MODELS FOR DIIODINE BASICITY SCALE OF HALOGEN BOND

[1] Laboratory of Chemoinformatics, University of Strasbourg, Blaise Pascal Str, 4, Strasbourg, France;
[2] Laboratory of Chemoinformatics and Molecular Modeling, Kazan Federal University, Kremlyovskaya Str, 18, Kazan, Russia;
[3] Frumkin Institute of Physical Chemistry and Electrochemistry RAS, Leninsky prospect Str, 31, Moscow, Russia;
mvglavatskikh@gmail.com

This work is devoted to QSPR (Quantitative Structure-Property Relationship) models building of Halogen Bond basicity scale pK_{I2}. The scale is based on the experimental 1:1 (B:I_2) complexation constant $\log K_{I2}$ of organic compounds (B) with diiodine (I_2) as a reference halogen-bond donor in alkanes at 298 K. Models based on ISIDA local descriptors were performed using Support Vector Machine (SVM) and Multiple Linear Regression (eMLR) methods on a set of 598 organic compounds. A consensus model returning the mean of values predicted by the most successful individual SVM models, based on various ISIDA fragmentation schemes, and including applicability domain assessment strategies (bounding box, standard deviation of consensus prediction). This model has then been challenged on the external test set of 11 *polyfunctional* compounds, for which unambiguous assignment of the measured effective complexation constant could not be assigned to either of halogen acceptor site. Developed consensus model was used to predict individual $\log K_{I2}$ of each acceptor sites followed by the prediction of an effective complexation constant with the help of the ChemEqui program. The best consensus models perform well both in cross-validation (root mean squared error *RMSE* = 0.45-0.56 $\log K_{I2}$ units) and external (*RMSE* = 0.55) set. The models are implemented on our website (*http://infochim.u-strasbg.fr/webserv/VSEngine.html*) together with the estimation of their applicability domain and an automatic detection of potential Halogen bond acceptors.

1. Laurence, C.; Graton, J.; Berthelot, M.; El Ghomari, M. J. *Chemistry-a European Journal* 2011, *17*, 10431-10444
2 A. Varnek, D. Fourches, F. Hoonakker, V. P. Solov'ev J. Computer-Aided Mol. Design, *2005, 19, 693-703*
3. Fiorella Ruggiu, Gilles Marcou, Alexandre Varnek and Dragos Horvath. *Mol. Informatics*, 2010, 29, 855 – 868
4. Fiorella Ruggiu, Vitaly Solov'ev, Gilles Marcou, Dragos Horvath, Jerome Graton, Jean-Yves Le Questel, Alexandre Varnek *Mol. Informatics*, 2014, 33, 477 – 487
5. Solov'ev, V. P.; Tsivadze, A. Y. *Protection of Metals and Physical Chemistry of Surfaces* 2015, *51*, 1-35.