PREDICTIVE MODELS FOR HALOGEN BOND BASICITY SCALE pKI2

Glavatskikh M.V.,[a,b] Madzhidov T.I.,[b] Solov'ev V.P.,[c] Horvath D.,[a] Marcou G.,[a] Varnek A.A.[a,b]

[a]University of Strasbourg, Strasbourg, 67008, France
[b]Kazan (Volga region) Federal University, Kremlyovskaya, 4/5, Kazan, 420111, Russia, e-mail: mvglavatskikh@gmail.com
[c]A.N. Frumkin Institute of Physical Chemistry and Electrochemistry, RAS, Leninsky prospekt, 31, Moscow, 119071, Russia

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 0 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. At this stage, developed models were used to predict $\log K_{I2}$ of the acceptor sites followed by an estimation of the predicted 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.

References

1. Laurence C., Graton J., Berthelot M., El Ghomari M.J. *Chemistry-a European Journal* 2011, *17*, 10431-10444
2. Varnek A., Fourches D., Hoonakker F., Solov'ev V.P. J. Computer-Aided Mol. Design, *2005, 19, 693-703*
3. Ruggiu F., Marcou G., Varnek A., Horvath D. *Mol. Informatics,* 2010*,* 29, 855 – 868
4. Ruggiu F., Solov'ev V., Marcou G., Horvath D., Graton J., Le Questel J.-Y., Varnek A. *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.