QSAR models for nanoparticles: linear and non-linear approaches

PAOLA GRAMATICA[1], ESTER PAPA[1,2], JEAN PIERRE DOUCET[2], ANNICK DOUCET-PANAYE[2],

[1] QSAR Research Unit in Environmental Chemistry and Ecotoxicology, Department of Theoretical and Applied Sciences, University of Insubria, 2100, Varese, paola.gramatica@uninsubria.it
[2] ITODYS, UMR 7086 Université Paris Diderot, 75013, Paris

The correct assessment of the potential hazard and risks associated to nanoparticles is a topic of primary importance due to their increased use in different aspects of daily life. The application of in silico methods to "nano-environments" is limited by scarce data available characterizing the structures and the effects/properties of nanomaterials. Work still needs to be done to fill the experimental gaps related to lacking data and to refine theoretical calculations and statistical methods commonly used to study traditional "molecular-environments" for their best applications to NPs. In this poster we show the results form 2 case studies (CS1 and CS2) dealing with the modelling of surface modified magnetic nanoparticles (CLIO-NPs), and nano-oxides (TiO_2 and ZnO) (WEISSLEDER et al. 2005, SAYES AND IVANOV 2000). These case studies illustrate the application of several linear and non-linear approaches (GRAMATICA et al. 2013, GRAMATICA et al. 2014, R 2008) (i.e. Multiple Linear Regression by Ordinary Least Sqares (MLR-OLS), Support Vector Machines (SVM Linear and Radial), Radial Basis Function (RBFNN) and General Regression (GRegNN) Neural Networks) to predict NPs activities (i.e. Cellular Uptake in PaCa2 and HUVEC human cell lines, in CS1, and cellular membrane disruption, in CS2) (WEISSLEDER et al. 2005, SAYES AND IVANOV 2000). QSAR calculations started from the identification of X variables, which were both classical theoretical descriptors calculated for surface ligands for CLIO-NPs with the same metallic core (CS1), or properties of the experimental environment (CS2).

CS1: Results show that MLR-OLS, SVM (Linear) and RBFNN have better performance for robustness and external predictivity than SVM (radial) and GRegNN. The best results obtained by the different approaches show comparable ranges of RMSE in the training and prediction sets for the two modeled responses (PaCa2 RMSE Tr. 0.19 – 0.23, RMSE Pred. 0.15 – 0.33; HUVEC RMSE Tr 0.28 – 0.32, RMSE Pred. 0.26 – 0.41). These results are comparable or better than literature models (OKSEL et al. 2015)] at the same level of complexity or higher (here we used up to 8 molecular descriptors).

CS2: Good performances were obtained by using different linear and non-linear methods and up to 3 simple descriptors. The model developed including both TiO_2 and ZnO NPs had the following ranges of RMSE Tr. and RMSE Pred.: 0.02 – 0.12; 0.09 – 0.14. Ranges of RMSE Tr. 0.03 – 0.11 and 0.06- 0.08 were calculated for models developed on the small datasets available for TiO_2 and ZnO separately. MLR, SVM-Linear and RBFNN showed slightly better results than SVM radial and GRegNN. Finally a simple J48 Regression tree was calculated by WEKA to classify the potential for membrane disruption on the basis of 2 descriptors, with Overall Accuracy, Specificity, and Sensitivity > 80%.

Observations from this study may support the prediction of properties and activities of NPs at hazard screening level, the design of safer NPs, and provide additional information useful for the future improvement of the current QSARs.