Prediction of toxicity for metal oxide nanoparticles

I. Kopernyk[1], V. Kovalishyn[1], L. Charochkina[1], N. Abramenko (Golovina) [2], W. Peijnenburg[3], L. Metelitsya[1]

[1] - Institute of Bioorganic Chemistry & Petroleum Chemistry, Murmanskaya, 1, 02660, Kyiv, Ukraine;
[2] - Chemistry Department, Moscow State University, Leninskie Gory 1, bldg 3, 119991 Moscow, Russia;
[3] - Institute of Environmental Sciences (CML), Leiden University, 2300 RA, Leiden, The Netherlands.
lapaldina@gmail.com

Toxicity of nanomaterials is one of the most attractive scientific areas of research and taking one of the first places during recent years. There are numerous examples of already established and possible applications of using nanoparticles such as textile, cosmetics, optical, pharmacy, electronics, etc.

Although the nanotechnology field is growing rapidly, the potential harmful effects of nanomaterials on human health or the environment have not yet been identified. The goal of the present study is the development of robust QSPR-based models for predicting nanoparticles toxicity (1).

In this study, we used the Online Chemical Modelling Environment (OCHEM) (2) to develop a high accuracy model for predicting nanotoxicity. Data set of nanoparticles with known LC50 values were collected from different published papers and were uploaded into OCHEM (2).

The main priorities were given to toxicity of metal and metal oxides nanoparticles (Fe, Ag, Pd, Ni, TiO2, ZnO, CuO, etc). About 300 data points were collected. The basic characteristics of nanoparticles such as material of nanoparticles, average particle size (APS), shape and information about experimental species were used as obligatory condition for all properties in OCHEM. Thus each record was required to incorporate information about these the most important parameters of nanoparticles.

In a preprocessing step using Chemaxon Standardizer, all structures were standardized and optimized with Corina (3). Unsupervised filtering of descriptors was applied to each descriptor set before using it as a machine learning input. The overall best performance was attained by Associative Neural Network (ASNN) and k-Nearest Neighbor Method (*k*NN) methods. The accuracy of all individual models was estimated using cross-validation procedures. The QSPR models were developed solely based on training sets and the resulting models were validated through predicting the toxicity of the NP in the respective test sets (4).

The commonly used measures of a regression model performance are the root mean square error (RMSE), the mean absolute error (MAE), the squared correlation coefficient R^2 and cross-validated coefficient q^2. The OCHEM system calculates these statistical parameters for both the training and the validation sets (2).

Based on previously suggested recommendations, QSPR models with $q^2 > 0.5$ are considered to have an acceptable predictive power (4). The q^2 coefficients for the training sets were in the range 0.58-0.81. q^2 coefficients obtained for independent test prediction sets were in the range 0.66-0.73 (Table 1). The limitations and advantages of the proposed approach are discussed.

Table 1. Statistical parameters

M.	Set	Amount	Descr.	MLT[a]	R^2	q^2	RMSE	MAE
1	Training set	100	47	ASNN	0.79 ± 0.03	0.79 ± 0.03	0.7 ± 0.04	0.58 ± 0.04
	Test set	43			0.82 ± 0.05	0.82 ± 0.06	0.56 ±0.07	0.45 ± 0.05
2	Training set	100	132	*k*NN	0.59 ± 0.07	0.58 ± 0.07	1.04 ±0.07	0.85 ± 0.06
	Test set	43			0.68 ± 0.08	0.66 ± 0.09	0.8 ± 0.07	0.67 ± 0.06

[a]MLT – machine learning technique

Acknowledgment.

We would like to acknowledge the NATO Science for Peace and Security Programme (grant EAP.SFPP 984401) for funding this research.

References.

1. Chatterjee R. The challenge of regulating nanomaterials. Environ.Sci.Technol, 2008, 42, 339–343.
2. http://ochem.eu/
3. https://www.molecular-networks.com/products/corina
4. Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. Mol. Inf.2010, 29, 476–488.