QSAR MODEL OF TOXICITY TOWARDS E. COLI BACTERIA FOR NANOSIZED OXIDES BY SMILES-BASED OPTIMAL DESCRIPTORS

Andrey A. Toropov[1,2], Alla P. Toropova[1,2] Xiao-Ke Hu[3], Huey-Min Hwang[3], Bakhtiyor F. Rasulev[4], Tomasz Puzyn[4], Emilio Benfenati[2], Danuta Leszczynska[4] and Jerzy Leszczynski[4]

[1] Institute of Geology and Geophysics - Khodzhibaev st. 49 - 100041 Tashkent - Uzbekistan
[2] Istituto di Ricerche Farmacologiche Mario Negri - Via Giuseppe La Masa 19 - 20156 Milano - Italy
[3] Department of Biology, Jackson State University - 1400 J. R. Lynch Street - P.O. Box 18540, Jackson, MS 39217 - USA
[4] Nanotoxicity Center, Department of Chemistry, Jackson State University - 1400 J. R. Lynch Street - P.O. Box 17910, Jackson, MS 39217 - USA

Abstract

Statistical quality of quantitative structure – activity relationship obtained by the optimal descriptors calculated with simplified molecular input line entry system for toxicity towards *E. Coli bacteria* (-pLD50) of nanosized oxides are the following: n=7, r2=0.9908, s=0.053, F=539 (training set); and n=7, r2=0.8226, s=0.241, F=23 (test set).
External predictability of the SMILES-based models has been checked with three random splits into training and test set.

Introduction

Nanoscience and nanotechnology are one of the fastest growing research and industrial areas in the 21st Century. Nanotechnology has gained a great deal of public interest due to the needs and applications of nanomaterials in almost all areas of human endeavors including industry, agriculture, business, medicine and public health. One of the issues that have to be addressed in the near future, before massive fabrication of nanomaterials, is their toxicity to humans and the environment. While there have been significant advances in nanoscience and nanotechnology, there have been concerns that the wide production and utilization of nanomaterials is rapidly overtaking efforts to evaluate their toxicity to humans and the environment. To date, very few studies have focused on the evaluation of the impact of nanomaterials on human health. The quantitative structure – activity relationships (QSAR) have been indicated as an efficient way to provide information on toxicity, as a screening tool, and also as an alternative way to experimental studies. There is an interest in regulations for these approaches: in USA decision makers use QSAR for many decades, and recently the REACH legislation promotes this approach also in Europe. Bacteria can serve as a good indicator of ecotoxicity of many chemical contaminants due to their important roles in biogeochemical cycling of elements. Among them, E. coli has been used as a biota model to test the toxicity of various nanomaterials.

Materials & Methods

The measurement of toxicity towards E. Coli bacteria of metal oxides listed in Table 1 has been carried out in Jackson State University.
Optimal descriptors have been calculated as the following

$$DCW = \sum CW(SAk) \quad (1)$$

where the SAk is a SMILES attribute, i.e., one symbol (e.g., 'O', '=', 'V', etc.) or two symbols (e.g., 'Al', 'Bi', 'Cu', etc.) in the SMILES notation. Numbers of double bonds have been used as global SMILES attributes. They are denoted as '=001' and '=002'. The '=001' is indicator of one double bond, the '=002' is indicator of two double bonds.
The available set of 14 oxides has been randomly split into training (n=7) and test (n=7) sets. Three splits have been examined (*Table 1*).

Table 1:
Splits into the training set and test set: nano oxides of the training set are indicated by '+', nano oxides of the test set are indicated by '#'

Split 1	Split 2	Split 3
+1	+1	+1
+2	+2	#2
#3	#3	#3
#4	#4	+4
#5	#5	#5
+6	+6	+6
#7	+7	+7
+8	+8	+8
#9	#9	#9
+10	+10	+10
#11	#11	#11
+12	+12	+12
#13	#13	#13
+14	#14	#14

Results

By the Monte Carlo optimization of correlation weights of the above SMILES attributes [1-3] one-variable models for the toxicity have been obtained (*Table 2*). Statistical quality of these models is satisfactory (*Table 3*). There is reproducibility of statistical characteristics of these models in three runs of the optimization. First run of the Monte Carlo optimization gave the following model:

$$-pLD50 = 1.3231 (\pm 0.0305) + 0.2742 (\pm 0.0080) * DCW \quad (2)$$

n=7, r2=0.9908, s=0.053, F=539 (training set);
n=7, r2=0.8226, s=0.241, F=23 (test set).

Table 4 contains the correlation weights for the DCW calculation with Eq. 1

Table 2:
Example of the QSAR model of – pLD50 of nanosized oxides

ID	Formula	SMILES	DCW	Expr	Calc	Expr-Calc
		Training set				
1	**ZnO**	O=[Zn]	7.4964167	3.450	3.379	0.071
2	**TiO_2**	O=[Ti]=O	1.5785166	1.742	1.756	-0.014
6	**Fe_2O_3**	O=[Fe-]O[Fe-]=O	3.9376291	2.405	2.403	0.002
8	**Y_2O_3**	O=[Y]O[Y]=O	5.9178614	2.868	2.946	-0.078
10	**ZrO_2**	O=[Zr]=O	2.8084706	2.151	2.093	0.058
12	**In_2O_3**	O=[In]O[In]=O	5.5731328	2.807	2.851	-0.044
14	**Sb_2O_3**	O=[Sb]O[Sb]=O	4.7980500	2.642	2.639	0.003
		Test set				
3	**SnO_2**	O=[Sn]=O	2.5685177	2.006	2.027	-0.021
4	**CuO**	[Cu]=O	5.5183250	3.203	2.836	0.367
5	**La_2O_3**	O=[La]O[La]=O	5.0833134	2.873	2.717	0.156
7	**Al_2O_3**	O=[Al]O[Al]=O	5.0875610	2.495	2.718	-0.223
9	**Bi_2O_3**	O=[Bi]O[Bi]=O	5.0827454	2.818	2.717	0.101
11	**SiO_2**	O=[Si]=O	2.5672319	2.199	2.027	0.172
13	**V_2O_3**	O=[V-]O[V-]=O	5.4808394	3.142	2.826	0.316

Table 3:
Statistical characteristics of the models in three probes of runs of the Monte Carlo method optimization

	Training set, n=7			Test set, n=7		
Run	**r^2**	**s**	**F**	**r^2**	**s**	**F**
Split 1						
1	0.9908	0.053	539	0.8226	0.241	23
2	0.9906	0.053	525	0.8248	0.240	24
3	0.9904	0.054	518	0.8212	0.242	23
Split 2						
1	0.9922	0.048	638	0.9016	0.229	46
2	0.9942	0.042	860	0.9138	0.222	53
3	0.9942	0.042	863	0.9126	0.222	52
Split 3						
1	0.9592	0.092	117	0.8881	0.265	40
2	0.9527	0.099	101	0.8832	0.263	38
3	0.9584	0.093	115	0.8852	0.266	39

Table 4:
Correlation weights for calculation with Eq. 1.

SA_k	$CW(SA_k)$ in run 1	$CW(SA_k)$ in run 2	$CW(SA_k)$ in run 3
Local SMILES attributes			
-	0.1966747	0.2087473	0.2006224
=	0.0130870	0.0120459	0.0144845
Al	1.0007154	0.9993486	1.0031185
Bi	0.9983076	0.9983822	1.0006130
Cu	1.0029309	0.9964202	1.0019729
Fe	0.2290747	0.2257543	0.2270241
In	1.2435013	1.2446089	1.2401183
La	0.9985916	1.0039874	0.9951787
O	0.0071878	0.0073432	0.0079173
Sb	0.8559599	0.8583021	0.8581188
Si	1.0016039	1.0008090	1.0007846
Sn	1.0028897	0.9984757	0.9992503
Ti	0.0128886	0.0128195	0.0139781
V	1.0006799	1.0013360	0.9975716
Y	1.4158656	1.4191345	1.4225923
[	0.7566572	0.7578458	0.7568752
Zn	2.9810226	2.9756969	2.9812541
Zr	1.2428426	1.2424288	1.2408108
Global SMILES attributes			
=001	2.9818049	2.9837389	2.9766216
=002	0.0117640	0.0078640	0.0077726

Conclusions

- **The SMILES-based optimal descriptors can be used as a tool for prediction of the toxicity nanosized oxides.**
- **Reproducing of statistical quality of these models takes place.**

References

[1] A. A. Toropov, E. Benfenati Cur. Drug. Disc. Tech, 4 (2007) 77-116
[2] A. A. Toropov, D. Leszczynska, J. Leszczynski Mat. Lett. 61 (2007) 4777-4780
[3] A.A.Toropov, E. Benfenati Computat. Biol. Chem. 31 (2007) 57-60

Acknowledgements

The authors thank
the ***Marie Curie Fellowships***
for financial support through the contract
MIF1-CT-2006-039036 - CHEMPREDICT