Electronic Supplementary Material (ESI) for RSC Advances.

Causal inference methods to assist in mechanistic interpretation of classification nano-SAR models

Natalia Sizochenko[1,2], Bakhtiyor Rasulev[2,4], Agnieszka Gajewicz[1], Elena Mokshyna[3], Victor Kuz'min[3], Jerzy Leszczynski[2], Tomasz Puzyn[1,*]

[1]*Laboratory of Environmental Chemometrics, Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, 80-952, Gdansk, Poland;*

[2]*Interdisciplinary Center for Nanotoxicity, Department of Chemistry, Jackson State University, 1400 J. R. Lynch Street, P. O. Box 17910, Jackson, MS 39217, USA;*

[3]*A.V. Bogatsky Physical–Chemical Institute National Academy of Sciences of Ukraine, 65080, Odessa, Ukraine;*

[4]*Center for Computationally Assisted Science and Technology (CCAST), North Dakota State University, 1805 NDSU Research Park Dr, PO Box 6050, Fargo, ND 58108, USA*

*corresponding author: t.puzyn@qsar.eu.org

Originally measured data were characterized as slopes of the dose-response based on doses of nanoparticles (in μg/mL) and responses (the decimal logarithm of EC50). In current study the final endpoints' data set was transformed to binary rank scale: slopes with negative values were marked as "0", while positive slopes were marked as "1". Thus, "0" means non-toxic compound and "1" means toxic compound in given conditions. Original experimental details related to sizes of nanoparticles, the conduction band energy (Ec) and point of zero zeta-potential (PZZP) are provided in Table 1.

Supplementary table 1. Experimental data of nano-sized metal oxides

Compound name	BEAS-2B	RAW 264.7	E_c (eV)	PZZP	Size of individual nanoparticle, nm	Size of aggregate in DMEM, nm
Al_2O_3	0	0	-1.51	7.4	14.7	230.5
CuO	1	1	-5.17	7.9	12.8	313.8
CeO_2	1	0	-3.80	7.8	18.3	248.4
Co_3O_4	1	1	-4.59	9.4	10.0	230.3
CoO	1	1	-4.42	9.2	71.8	191.6
Cr_2O_3	1	1	-4.44	5.3	193.0	318.7
Fe_2O_3	0	0	-4.99	7.2	12.3	196.6
Fe_3O_4	1	0	-5.00	5.0	12.0	243.6
Gd_2O_3	0	0	-2.83	8.0	43.8	222.0
HfO_2	0	1	-2.96	8.1	28.4	280.6
In_2O_3	0	0	-3.63	9.2	59.6	244.5
La_2O_3	0	0	-2.38	9.4	24.6	162.6
Mn_2O_3	1	1	-4.65	3.7	51.5	291.7
NiO	1	1	-3.57	11.4	13.1	189.5
Ni_2O_3	1	1	-4.31	8.3	140.6	322.6
Sb_2O_3	0	0	-3.64	1.0	11.8	196.5
SiO_2	1	0	.2.02	1.0	13.5	46.4
SnO_2	0	0	-4.01	4.0	62.4	198
TiO_2	0	0	-4.16	6.4	12.6	55.9
Y_2O_3	1	0	-5.53	0.3	32.7	351.6
Yb_2O_3	0	1	-2.35	9.6	61.7	235.6
ZnO	1	1	-2.83	8.2	22.6	69.4
ZrO_2	0	0	-3.89	9.6	40.1	312.3
WO_3	0	0	-3.19	5.8	16.6	198.5

Descriptors

Simplex Representation of Molecular Structure (SiRMS)

In the current study we utilized a 2D level of structure representation to generate simplex fragments (simplexes). Simplexes are two-, tri- and tetra-atomic molecular fragments of fixed structure and symmetry. In the SiRMS approach, the connectivity of atoms in a simplex, atom type, and bond nature are considered.

In case of each property, the continuous scale was transformed into the ranges (typically 4-7 intervals). All atoms were divided into five (A_{el}-F_{el}) groups corresponding to their electronegativity (A_{el}<1.5<B_{el}<2.0<C_{el}<2.5<D_{el}<3.5<F_{el}), four (A_{depth}-D_{depth}) groups corresponding to the depth of the Lennard-Jones potential (A_{depth}<0.01<B_{depth}<0.02<C_{depth}<0.4< D_{depth}), and five (A_{dist}-F_{dist}) groups according to the distance, at which the potential reaches minimum (A_{dist}<2.5<B_{dist}<3<C_{dist} <3.5<D_{dist}<4<F_{dist}). Particular simplex vertices (atoms) have been then assigned to groups (A_i-F_i) defined by each of i^{th} properties mentioned above.

After differentiating atoms based on the properties, all molecules have been fragmentized into all possible types of simplexes. Finally, simplexes of each type (e.g., A_{el}-B_{el}-D_{el}-A_{el}; A_{depth}-A_{depth}-A_{depth}-B_{depth}) have been counted. The numbers of simplexes of given types in a single molecule were used as descriptors.

Metal-ligand Binding (MLB) characteristics

In the current study, two ion characteristics were used to describe metal ion's affinity of binding to biochemical ligands: covalent index (CI) and cation polarizing power (CPP).

(CI) reflects the relative influence of covalent interactions in the binding process. This is represented by a combination of the electronegativity (χ) and the Pauling radius (r):

$$(CI) = \chi^2 r \qquad (1)$$

(CPP) indirectly reflects relative importance of covalent interaction with bio-ligands. CPP represents the ratio of ion charge (Z) and Pauling radius (r):

$$(CPP) = Z^2/r \quad (2)$$

«Liquid Drop» Model (LDM)

To describe the geometric and volume features of studied metal oxide nanoparticles, the physical model of "liquid drop" was used. Within this model a nanoparticle is represented as a spherical drop, whereas the nanoparticle's density is equal to the density of bulk state.

The minimum radius of the interactions between elementary particles in the cluster is described by the Wigner-Seitz radius (r_w):

$$r_w = \left(\frac{3M}{4\pi\rho}\right)^{1/3} \quad (3)$$

where M - Molecular weight, ρ - mass density.

Based on spherical shape of nanoparticle, the number of molecules in the nanocluster (n) is defined as follows:

$$n = \left(\frac{r_0}{r_w}\right)^3, \quad (4)$$

where r_0 represents the radius of the nanoparticle.

The ratio of surface molecules (F) to molecules in volume is supposed to be significant:

$$F = 4n^{-1/3} \quad (5)$$

The ratio of surface molecules to molecules in volume:

$$(SV) = \left(\frac{\text{surface molecules}}{\text{molecules in volume}}\right) = \frac{F}{1-F} \quad (6)$$

The aggregation parameter (AP) represents the ratio of the aggregate's size to the size of individual particles – might be defined:

$$AP = \left(\frac{size\ of\ aggregate}{size\ of\ individual\ nanoparticle}\right) \quad (7)$$

Supplementary table 2. Ranking of metal oxide nanoparticles using model for BEAS-2B

Compound name	Set	Observed	Predicted	Predicted Rank 0	Predicted Rank 1	Prox. measure 0	Prox. measure 1
Al_2O_3	training set	0	0	4	1	1.80	0.28
CuO	training set	1	1	1	4	0.32	0.76
CeO_2	training set	0	0	5	-	2.12	0.12
Co_3O_4	training set	1	1	1	4	0.16	0.28
CoO	test set	1	1	-	5	0.16	1.92
Cr_2O_3	test set	1	1	-	5	0.00	1.80
Fe_2O_3	test set	0	0	4	1	2.56	0.32
Fe_3O_4	test set	0	0	3	2	1.28	0.48
Gd_2O_3	training set	0	0	5	-	1.92	0.68
HfO_2	training set	1	1	-	5	0.00	0.16
In_2O_3	training set	0	0	5	-	1.92	0.68
La_2O_3	test set	0	0	5	-	2.44	0.40
Mn_2O_3	training set	1	1	-	5	0.00	0.80
NiO	training set	1	1	-	5	0.16	0.80
Ni_2O_3	training set	1	1	-	5	0.00	1.08
Sb_2O_3	training set	0	0	5	-	1.40	0.36
SiO_2	training set	0	0	4	1	1.68	0.20
SnO_2	training set	0	0	5	-	0.52	0.04
TiO_2	training set	0	0	4	1	2.48	0.32
Y_2O_3	training set	0	0	5	-	0.72	0.04
Yb_2O_3	training set	1	0	4	1	1.84	0.04
ZnO	training set	1	1	1	4	0.52	0.64
ZrO_2	test set	0	1	1	4	0.04	1.12
WO_3	training set	0	0	4	1	0.96	0.28

Supplementary table 3. Ranking of metal oxide nanoparticles using model for RAW 264.7

Compound name	Set	Observed	Predicted	Predicted Rank 0	Predicted Rank 1	Prox. measure 0	Prox. measure 1
Al_2O_3	training set	0	0	3	2	0.2	0.60
CuO	training set	1	1	1	4	0.48	1.76

CeO_2	training set	1	1	1	4	0.88	0.84
Co_3O_4	training set	1	1	2	3	0.88	1.12
CoO	test set	1	1	-	5	0.00	2.00
Cr_2O_3	test set	1	1	1	4	0.44	1.28
Fe_2O_3	test set	0	0	4	1	1.56	0.64
Fe_3O_4	test set	1	1	2	3	0.88	2.12
Gd_2O_3	training set	0	0	4	1	1.84	0.60
HfO_2	training set	0	0	4	1	1.64	0.60
In_2O_3	training set	0	0	4	1	1.44	0.76
La_2O_3	test set	0	1	1	4	0.76	2.76
Mn_2O_3	training set	1	1	-	5	0.32	1.56
NiO	training set	1	1	1	4	0.48	1.76
Ni_2O_3	training set	1	1	-	5	0.12	0.84
Sb_2O_3	training set	0	0	4	1	0.92	0.36
SiO_2	training set	1	1	-	5	0.20	0.60
SnO_2	training set	0	0	5	-	1.32	0.16
TiO_2	training set	0	0	5	-	0.44	0.04
Y_2O_3	training set	1	1	-	5	0.20	1.72
Yb_2O_3	training set	0	0	5	-	1.36	0.28
ZnO	training set	1	1	-	5	0.00	1.00
ZrO_2	test set	0	0	3	2	0.80	0.32
WO_3	training set	0	0	4	1	0.92	0.16