Ecotoxicology and Environmental Safety 108 (2014) 203–209

Contents lists available at ScienceDirect

Ecotoxicology and Environmental Safety

journal homepage: www.elsevier.com/locate/ecoenv

Optimal descriptor as a translator of eclectic information into the prediction of membrane damage: The case of a group of ZnO and TiO_2 nanoparticles

Alla P. Toropova [a], Andrey A. Toropov [a,*], Emilio Benfenati [a], Tomasz Puzyn [b], Danuta Leszczynska [c], Jerzy Leszczynski [d]

[a] IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, 20156, Via La Masa 19, Milano, Italy
[b] Faculty of Chemistry, Laboratory of Environmental Chemometrics, University of Gdansk, ul. Sobieskiego 18/19, Gdansk 80-952, Poland
[c] Interdisciplinary Nanotoxicity Center, Department of Civil and Environmental Engineering, Jackson State University, 1325 Lynch St, Jackson, MS 39217-0510, USA
[d] Interdisciplinary Nanotoxicity Center, Department of Chemistry and Biochemistry, Jackson State University, 1400 J. R. Lynch Street, PO Box 17910, Jackson, MS 39217, USA

ARTICLE INFO

Article history:
Received 10 June 2014
Received in revised form
3 July 2014
Accepted 4 July 2014
Available online 1 August 2014

Keywords:
QSAR
QFAR
Monte Carlo method
Nanoparticle
Membrane damage

ABSTRACT

The development of quantitative structure—activity relationships for nanomaterials needs representation of molecular structure of extremely complex molecular systems. Obviously, various characteristics of nanomaterial could impact associated biochemical endpoints. Following features of TiO_2 and ZnO nanoparticles ($n=42$) are considered here: (i) engineered size (nm); (ii) size in water suspension (nm); (iii) size in phosphate buffered saline (PBS, nm); (iv) concentration (mg/L); and (v) zeta potential (mV). The damage to cellular membranes (units/L) is selected as an endpoint. Quantitative features—activity relationships (QFARs) are calculated by the Monte Carlo technique for three distributions of data representing values associated with membrane damage into the training and validation sets. The obtained models are characterized by the following average statistics: $0.78 < r^2_{\text{training}} < 0.92$ and $0.67 < r^2_{\text{validation}} < 0.83$.

1. Introduction

To characterize nanoparticles (NPs) and their potential hazards sufficiently, empirical data are necessary. Since the early days of the REACH (2006) proposals, it has been agreed by all partners that the number of animals used to gain toxicity information on chemicals should be kept to an absolute minimum. There is evidence that in vitro and in silico methods for acute chemical toxicity are able to provide sufficient data to permit classification and labelling. However, for those substances with no available toxicity data a read-across and quantitative structure–activity relationship techniques (QSAR) are not possible, therefore in vivo testing is required to identify rapidly hazardous substances (Drobne et al., 2009).

The development of quantitative structure–property/activity relationships (QSPRs/QSARs) that links property (activity) of series of molecules with their structural characteristics is one of the important tasks of modern natural sciences (Afantitis et al., 2011; Furtula and Gutman, 2011; García et al., 2011; Garro Martinez et al., 2011; Ibezim et al., 2012; Mullen et al., 2011). The main aim of QSPR/QSAR analyses is design of predictive models which provide tools for estimation of various properties (endpoints) of chemical compounds from analysis of their molecular structures. In the case of "classic" QSPR/QSAR analysis the paradigm can be described by the following statement: endpoint is a mathematical function of molecular structure. In the case of nanomaterials (Fourches et al., 2010; Petrova et al., 2011; Puzyn et al., 2009; Toropova and Toropov, 2013), this paradigm encounters a fundamental problem. More information including size and shape, in additon to the molecular structure are necessary to characterize fully a nanomaterial. Consequently, classic descriptors cannot be a used as a tool to design predictive models for endpoints related to various nanomaterials which are widely involved in the modern everyday life (food packaging, forestry and paper, drugs, cosmetics, plastics and paints, microelectronics, etc.). Optimal descriptors (Toropov and Toropova, 2002, 2003; Toropov et al., 2008, 2011, 2012, 2013; Toropova et al., 2011) provide an approach that offers an alternative to QSPR/QSAR analyses based on the molecular structure. It could be used for evaluation of nanomaterials. This approach involves application of all available eclectic information related

* Corresponding author.
E-mail address: andrey.toropov@marionegri.it (A.A. Toropov).

http://dx.doi.org/10.1016/j.ecoenv.2014.07.005
0147-6513/

to genesis and various features of nanomaterials, in order to develop predictive models that links these information with desired endpoint (Toropov, et al., 2008; Toropova and Toropov, 2013).

Two NPs types are used in such large amounts that they are particularly likely to accumulate in the environment: nanoparticulate titanium dioxide (nano-TiO_2) and zinc oxide (nano-ZnO) (Ge et al., 2012). These NPs are used in applications as diverse as sunscreens, cosmetics, and coatings, and therefore are increasingly being introduced into the environment (Gottschalk et al., 2009; Robichaud et al., 2009). Both of these NPs can alter the composition of the soil bacterial community, and the effects increase with dose (Ge et al., 2011).

Here, we attempt to identify a set of properties of two specific metal oxide nanomaterials TiO_2 and ZnO that could be used to characterize and predict the induced cellular membrane damage of immortalized human lung epithelial cells (Sayes and Ivanov, 2010). It is to be noted that the approach gives good results for the case of separated TiO_2 nanoparticles (Toropova and Toropov, 2013).

The aim of the present study is evaluation of optimal descriptors based on such approach, for two nanometal oxides. The calculations are performed using the Monte Carlo method. We study influence of TiO_2 and ZnO nanoparticles, chracterized by various sets of physicochemical features, upon the membrane damage.

2. Method

2.1. Data

Experimental data on the physicochemical features of TiO_2 and ZnO nanoparticles and their influence on the membrane damage (Table 1) are taken from the literature (Sayes and Ivanov, 2010). The considered features include: (i) engineered size (nm); (ii) size in water suspension (nm); (iii) size in phosphate buffered saline (PBS, nm); (iv) concentration (mg/L); and (v) zeta potential (mV). The above-mentioned characteriscics of nanoparticles as well as presence of TiO_2 and/or ZnO in process that impacts on cells are the features which are applied in order to build up the quantitative features – activity relationships (QFARs). The SMILES (Weininger, 1988, 1990; Weininger et al., 1989) are using to represent molecular structure for the "classic" QSPR/QSAR (Veselinović et al., 2013; Achary, 2014a, 2014b; Worachartcheewan et al., 2014). The quasi-SMILES (Table 2) are used in this work to represent data on various nanoparticles under different physicochemical conditions. Membrane damage values related to TiO_2 and ZnO nanoparticles (characterized by different physicochemical features) are examined as the endpoint. Analysis of the results of three random splits of the data into the sub-training, calibration, and test set and invisible validation sets is performed for the developed QFAR approach

2.2. Optimal descriptors

The optimal descriptors are calculated as the following (Toropova and Toropov, 2013):

$$DCW(Threshold, N_{epoch}) = \sum CW(CI_k) \quad (1)$$

where CI_k is code of kth impact; $CW(CI_k)$ is the correlation weight for CI_k; The *Threshold* and N_{epoch} are parameters of the Monte Carlo optimization. The threshold is a tool to define two classes of impacts: rare (noise) and not rare, i.e. active. The optimal descriptors are calculated with the correlation weights related to the active impacts. Correlation weights for rare impacts are fixed equal to zero, i.e. these are not involved in building up model; The N_{epoch} is the number of epochs of the Monte Carlo optimization.

Considered here codes of impacts are defined by the following scheme:

1. Normalization of impact X_k according to formula

$$Norm(X_k) = \frac{\min X_k + X_k}{\min X_k + \max X_k} \quad (2)$$

Table 1 contains experimental data on the impacts; Table 2 contains SMILES-like representation of nanoparticles (NPs) according to the coding defined below.

Table 1
Experimental data on features (impacts) of TiO_2 and ZnO nanoparticles, and their denotations.

ID	NPs	Enginered size (nm) A	Size in water (nm) B	Size in PBS (nm) C	Concentration (mg/L) D	-Zeta potential (mV) E
1.	TiO_2	30	125	1250	25	10
2.	–	30	102	987	25	12
3.	–	30	281	1543	50	15
4.	–	30	101	1045	50	9
5.	–	30	299	1754	100	11
6.	–	30	134	961	100	11
7.	–	30	600	1876	200	12
8.	–	30	298	1165	200	12
9.	–	45	129	2567	25	9
10.	–	45	129	2309	25	10
11.	–	45	201	2431	50	9
12.	–	45	201	2987	50	11
13.	–	45	451	2941	100	11
14.	–	45	451	1934	100	9
15.	–	45	876	1965	200	11
16.	–	45	876	2109	200	10
17.	–	125	136	3215	25	11
18.	–	125	136	2667	25	10
19.	–	125	149	3782	50	10
20.	–	125	149	2144	50	15
21.	–	125	343	3871	100	12
22.	–	125	343	2890	100	9
23.	–	125	967	3813	200	9
24.	–	125	967	2671	200	8
25.	ZnO	50	55	158	25	55
26.	–	60	68	208	25	45
27.	–	70	71	198	25	50
28.	–	50	56	258	50	50
29.	–	60	78	386	50	50
30.	–	70	95	279	50	50
31.	–	50	168	314	100	25
32.	–	60	151	385	100	30
33.	–	70	172	354	100	29
34.	–	1000	1245	1319	25	44
35.	–	1200	1268	1325	25	33
36.	–	1500	1198	1381	25	25
37.	–	1000	1268	1459	50	30
38.	–	1200	1301	1587	50	32
39.	–	1500	1283	1523	50	29
40.	–	1000	1243	1925	100	20
41.	–	1200	1124	1805	100	21
42.	–	1500	1269	2109	100	21

2. Discrimination of impact according to scale (Fig. 1) into one of categories from 1,2,…,9. Following symbols: A, B, C, D, E are used to define engineered size (nm); size in water suspension (nm); size in phosphate buffered saline (PBS, nm); concentration (mg/L); and (v) zeta potential (mV), respectively. Table 2 contains representation of different TiO_2 and ZnO nanoparticles using corresponding codes. TiO_2 and ZnO are denoted by T and Z characters in the strings of NPs representation (coding). Defining such representations of nanoparticles, one can carry out calculation analogous to QSPR/QSAR analysis with SMILES-based optimal descriptors. However, SMILES-like (quasi-SMILES) strings examined in this work (Table 2) are not the traditionally used SMILES (Weininger, 1988, 1990; Weininger et al., 1989), since elements of their strings are representations of various physicochemical conditions, not the molecular structure. In fact, each quasi-SMILES can contain: (i) type of NP, i.e. "T" for TiO_2 and "Z" for ZnO; (ii) two symbols for representation of the engineered size (i.e. "A0", "A1", …, "A9"); (iii) size in water suspension (i.e. "B0", "B1", …,"B9"); (iv) size in PBS (i.e. "C0", "C1", …, "C9"); (v) concentration (i.e. "D0", "D1", …, "D9"); and (vi) zeta potential (i.e. "E0", "E1", …, "E9"). One can see, some of these binary combinations are absent in the available dataset (Table 3).

Having the representation of available data by list of the quasi-SMILES, one can carry out calculation by means of the CORAL software (http://www.insilico.eu/coral) described in the literature (Toropov et al., 2013). The essence of these calculations is the following (i) the definition of the threshold and the number of epochs of the Monte Carlo optimization which give the best statistical quality of the model for the test set; and (ii) the checking up of the predictive potential of the model with invisible (during building up the model) validation set. This approach

Table 2
The normalization of data on physicochemical properties (Eq. (2)) and build up codes of impacts for the quasi-SMILES.

ID	NPs	A	B	C	D	E
1	T	A0	B1	C3	D2	E2
2	T	A0	B1	C2	D2	E3
3	T	A0	B2	C4	D3	E3
4	T	A0	B1	C2	D3	E2
5	T	A0	B2	C4	D5	E3
6	T	A0	B1	C2	D5	E3
7	T	A0	B4	C5	D9	E3
8	T	A0	B2	C3	D9	E3
9	T	A0	B1	C6	D2	E2
10	T	A0	B1	C6	D2	E2
11	T	A0	B1	C6	D3	E2
12	T	A0	B1	C7	D3	E3
13	T	A0	B3	C7	D5	E3
14	T	A0	B3	C5	D5	E2
15	T	A0	B6	C5	D9	E3
16	T	A0	B6	C5	D9	E2
17	T	A1	B1	C8	D2	E3
18	T	A1	B1	C7	D2	E2
19	T	A1	B1	C9	D3	E2
20	T	A1	B1	C5	D3	E3
21	T	A1	B2	C9	D5	E3
22	T	A1	B2	C7	D5	E2
23	T	A1	B7	C9	D9	E2
24	T	A1	B7	C7	D9	E2
25	Z	A0	B0	C0	D2	E9
26	Z	A0	B0	C0	D2	E8
27	Z	A0	B0	C0	D2	E9
28	Z	A0	B0	C1	D3	E9
29	Z	A0	B0	C1	D3	E9
30	Z	A0	B1	C1	D3	E9
31	Z	A0	B1	C1	D5	E5
32	Z	A0	B1	C1	D5	E6
33	Z	A0	B1	C1	D5	E5
34	Z	A6	B9	C3	D2	E8
35	Z	A8	B9	C3	D2	E6
36	Z	A9	B9	C3	D2	E5
37	Z	A6	B9	C4	D3	E6
38	Z	A8	B9	C4	D3	E6
39	Z	A9	B9	C4	D3	E5
40	Z	A6	B9	C5	D5	E4
41	Z	A8	B8	C4	D5	E4
42	Z	A9	B9	C5	D5	E4
Minimum		30	55	158	25	8
Maximum		1500	1301	3871	200	55

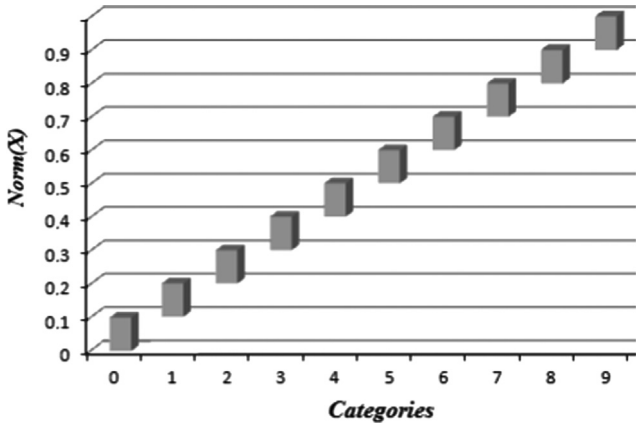

Fig. 1. Discrimination of normalized physicochemical features (A, B, …, E) into categories 0, 1, 2, …, 9 according to its value (Eq. (2)).

can be used to build up predictive models of endpoints related to nanomaterials in general (Toropova et al., 2013; Toropov and Toropova, 2014) as well as for membrane damage by nanoparticles in particular (Toropova and Toropov, 2013). In fact, the list of quasi-SMILES (Table 3) are extended version of the previous data

Table 3
The coding (quasi-SMILES) of representation of NPs and experimental membrane damage values.

ID	The coding (quasi-SMILES)	Membrane damage (units/L)
1	TA0B1C3D2E2	0.90
2	TA0B1C2D2E3	1.00
3	TA0B2C4D3E3	0.75
4	TA0B1C2D3E2	0.70
5	TA0B2C4D5E3	1.04
6	TA0B1C2D5E3	1.09
7	TA0B4C5D9E3	1.15
8	TA0B2C3D9E3	1.20
9	TA0B1C6D2E2	0.90
10	TA0B1C6D2E2	0.85
11	TA0B1C6D3E2	0.75
12	TA0B1C7D3E3	0.78
13	TA0B3C7D5E3	1.40
14	TA0B3C5D5E2	1.50
15	TA0B6C5D9E3	1.35
16	TA0B6C5D9E2	1.40
17	TA1B1C8D2E3	1.25
18	TA1B1C7D2E2	1.17
19	TA1B1C9D3E2	1.00
20	TA1B1C5D3E3	1.10
21	TA1B2C9D5E3	1.50
22	TA1B2C7D5E2	1.42
23	TA1B7C9D9E2	1.60
24	TA1B7C7D9E2	1.65
25	ZA0B0C0D2E9	1.10
26	ZA0B0C0D2E8	1.03
27	ZA0B0C0D2E9	1.08
28	ZA0B0C1D3E9	1.00
29	ZA0B0C1D3E9	0.92
30	ZA0B1C1D3E9	0.99
31	ZA0B1C1D5E5	1.12
32	ZA0B1C1D5E6	1.25
33	ZA0B1C1D5E5	1.19
34	ZA6B9C3D2E8	1.58
35	ZA8B9C3D2E6	1.69
36	ZA9B9C3D2E5	1.59
37	ZA6B9C4D3E6	0.92
38	ZA8B9C4D3E6	0.95
39	ZA9B9C4D3E5	0.84
40	ZA6B9C5D5E4	1.25
41	ZA8B8C4D5E4	1.39
42	ZA9B9C5D5E4	1.45

examined in the recent work (Toropova and Toropov, 2013) which are prepared by the same scheme.

3. Results and discussion

Table 3 contains list of TiO_2 and ZnO nanoparticles examined in this work. The QFAR models for the membrane damage (MD), calculated with three random splits are the following:

Split 1

$$MD = -0.3708(\pm 0.0499) + 0.2378(\pm 0.0081) \times DCW(1,5) \quad (3)$$

$n=17$, $r^2=0.8131$, $q^2=0.7582$, $s=0.114$, $F=65$ (sub-training set)
$n=14$, $r^2=0.8062$, $s=0.256$ (calibration set)
$n=5$, $r^2=0.9952$, $s=0.241$, $\overline{R_m^2}=0.7354$, $\Delta R_m^2=0.0669$ (test set)
$n=6$, $r^2=0.8362$, $s=0.244$, (validation set)

Split 2

$$MD = 0.5784(\pm 0.0187) + 0.1596(\pm 0.0051) \times DCW(3,19) \quad (4)$$

Table 4
Correlation weights of codes of impacts (*CI*) for calculation of the optimal descriptor for splits 1, 2, and 3.

CI_k	CW(CI_k)	Frequency of CI_k in training set	Frequency of CI_k in calibration set	Frequency of CI_k in test set
Split 1				
A0	0.49800	6	11	3
A1	1.36250	5	1	1
A6	1.18950	3	0	0
A8	1.70300	1	1	1
A9	0.92600	2	1	0
B0	0.68250	1	2	0
B1	0.50500	6	4	3
B2	0.54600	2	1	1
B3	0.0	0	2	0
B4	0.0	0	1	0
B6	1.50400	1	1	0
B7	2.00500	1	1	0
B8	0.0	0	1	0
B9	1.07100	6	1	1
C0	0.90400	1	1	0
C1	0.54900	1	3	0
C2	0.0	0	2	0
C3	2.00200	4	0	1
C4	0.55200	3	2	0
C5	1.45300	3	4	0
C6	0.0	0	0	3
C7	1.55500	3	2	0
C9	1.27900	2	0	1
D2	1.18450	5	2	3
D3	0.50300	6	2	1
D5	2.00300	3	7	1
D9	1.52100	3	3	0
E2	1.00100	5	3	3
E3	0.93000	4	5	1
E4	1.09500	1	2	0
E5	1.45400	3	1	0
E6	0.87400	2	0	1
E8	1.05100	1	1	0
E9	1.27800	1	2	0
T	1.10950	9	8	4
Z	0.98450	8	6	1
Split 2				
A0	−0.13550	11	8	3
A1	2.05500	4	2	2
A6	0.0	2	0	0
A8	0.0	0	2	0
A9	0.0	1	2	0
B0	0.0	2	1	1
B1	−0.90000	7	5	3
B2	0.0	2	2	1
B3	0.0	1	1	0
B6	0.0	2	0	0
B7	0.0	1	1	0
B8	0.0	0	1	0
B9	0.40400	3	3	0
C0	0.0	1	0	1
C1	0.63650	3	2	1
C2	0.0	1	1	1
C3	0.99150	3	1	0
C4	0.0	1	4	0
C5	0.74500	4	2	0
C6	0.0	1	1	0
C7	0.0	2	3	0
C8	0.0	0	0	1
C9	0.0	2	0	1
D2	1.79700	7	1	2
D3	0.05950	4	5	1
D5	2.53950	4	6	2
D9	2.84700	3	2	0
E2	0.90100	8	2	1
E3	0.98750	3	6	2
E4	0.0	1	2	0
E5	0.0	2	1	1
E6	0.0	0	2	0
E8	0.0	2	0	0
E9	0.0	2	1	1
T	0.40100	11	8	3
Z	1.91150	7	6	2

Table 4 (*continued*)

CI_k	CW(CI_k)	Frequency of CI_k in training set	Frequency of CI_k in calibration set	Frequency of CI_k in test set
Split 3				
A0	−0.20300	13	4	6
A1	0.0	0	6	0
A6	1.49900	3	0	0
A8	2.57600	1	0	1
A9	1.75200	2	1	0
B0	0.10400	4	0	1
B1	−0.30300	5	6	3
B2	0.0	0	3	1
B3	3.04900	2	0	0
B4	1.10300	1	0	0
B6	2.38450	1	0	1
B7	0.0	0	1	0
B8	2.33350	1	0	0
B9	1.21550	5	1	1
C0	0.34900	3	0	0
C1	2.05400	4	0	2
C2	−0.29700	1	1	0
C3	0.72400	1	3	0
C4	−0.29600	3	0	2
C5	1.80100	5	1	1
C6	0.64250	1	0	2
C7	1.40300	1	3	0
C8	0.0	0	1	0
C9	0.0	0	2	0
D2	1.72500	5	3	1
D3	−0.21250	5	3	4
D5	3.50200	7	3	1
D9	3.20300	2	2	1
E2	1.52500	4	4	2
E3	1.14900	2	6	2
E4	−0.22100	3	0	0
E5	0.68650	2	1	1
E6	1.14900	2	0	1
E8	3.02900	2	0	0
E9	2.52900	4	0	1
T	0.85400	6	10	4
Z	1.69700	13	1	3

Example of DCW(1,5) calculation for Split 1; Quasi-SMILES=" T.A0.B1.C3.D2.E2"

$$DCW(Threshold, N_{epoch}) = \sum CW(CI_k) =$$
$$CW(T)+CW(A0)+CW(B1)+CW(C3)+CW(D2)+CW(E2) =$$
$$1.1095+0.4980+0.5050+2.0020+1.1845+1.0010 = 6.3000$$

MD(calculated)$= -0.3708+0.2378 \times 6.3000 = 1.1273$ (units/L)

$n=18$, $r^2=0.7845$, $q^2=0.7308$, $s=0.129$, $F=58$ (sub-training set)
$n=14$, $r^2=0.8197$, $s=0.110$ (calibration set)
$n=5$, $r^2=0.9873$, $s=0.056$, $\overline{R_m^2}=0.8839$, $\Delta R_m^2=0.0228$ (test set)
$n=5$, $r^2=0.7159$, $s=0.157$, (validation set)

Split 3

$$MD = 0.4544(\pm 0.0209)+0.0974(\pm 0.0027) \times DCW(1,13) \quad (5)$$

$n=19$, $r^2=0.9255$, $q^2=0.8932$, $s=0.068$, $F=211$ (sub-training set)
$n=11$, $r^2=0.6693$, $s=0.354$ (calibration set)
$n=7$, $r^2=0.8866$, $s=0.102$, $\overline{R_m^2}=0.6710$, $\Delta R_m^2=0.1513$ (test set)
$n=5$, $r^2=0.6763$, $s=0.318$, (validation set)

In Eqs. (3)–(5), n is the number of nanoparticles in a set (training, test, and validation); r is correlation coefficient; q is leave-one-out cross-validated correlation coefficient; RMSE is root-mean-square error; $\overline{R_m^2}$ and ΔR_m^2 are metrics of predictability (Ojha et al., 2011; Roy et al., 2012). According to the rules of QSAR/QSPR approaches a developed model has predictability if $\overline{R_m^2} > 0.5$ and $\Delta R_m^2 < 0.2$.

Table 4 contains the correlation weights for various codes of impacts which are involved in building up model. Fig. 2 contains the graphical representation of these models.

Having data on several runs of the Monte Carlo optimization, the impacts of four kinds can be detected. The first, impacts with positive correlation weights for all runs (these are promoters of endpoint increase). The second, impacts with negative correlation weights for all runs (these are promoters of endpoint decrease). The third, impacts with mixed correlation weights: there are both positive and negative values of the correlation weights in several runs of the Monte Carlo optimization. The role of these impacts is undefined. Finally, blocked (rare) impacts, which are not involved in building up model since they are "noise" which can lead to "overtraining". The analysis of this information gives possibility to extract impacts which are promoters of increase of the endpoint or vice versa which are promoters of decrease of the endpoint.

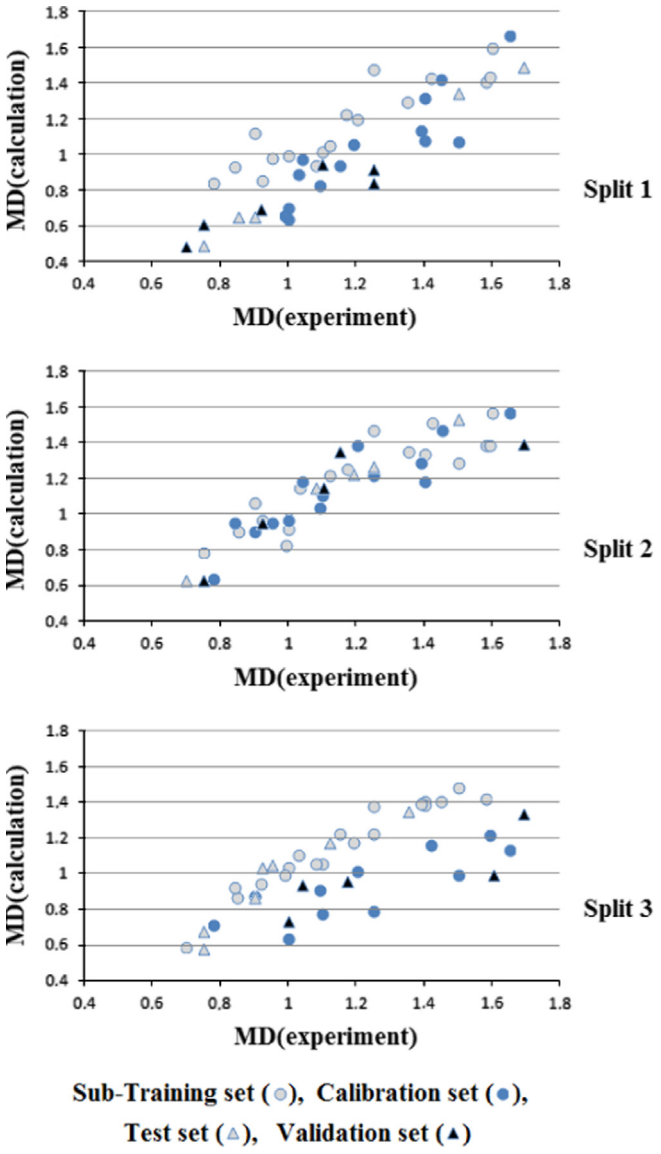

Fig. 2. Graphical representation of models for splits 1, 2, and 3.

In the case of examined data (Table 2) the following statistically significant promoters of increase for membrane damage were extracted (Table S4 in Supplementary materials): "D2", "D3", and "E2". The physical meaning of the result is the following: (i) the range of concentration of nanoparticles from 25 to 50 mg/L is promoter of the endpoint increase; and (ii) the range of zeta potential from −15 to −10 is promoter of the endpoint increase. Vice versa, the "A0", "B0", "B1" and "C4" are statistically significant promoters of the endpoint decrease. The physical meaning of the result is the following: (i) the range of engineering size of nanoparticles from 30 to 70 nm is the promoter of the endpoint decrease; (ii) the range of size in water from 55 to 299 nm is the promoter of the endpoint decrease; and (iii) the range of size in PBS from 1523 to 1805 nm also is the promoter of decrease for the membrane damage by ZnO and TiO_2 nanoparticles. In fact, the suggested QFAR models can have mechanistic interpretation like in the case of "classic QSAR models". The domain of applicability for the QFAR models can be defined as the quasi-SMILES containing codes which are involved in building up the model. Thus, suggested approach gives models built up in accordance with OECD principles (OECD, 2007).

Statistical quality of regression model for membrane damage described in the literature for the same group of 24 nanoparticles is $0.15 < r^2 < 0.7$, without distribution of the considered data into the training and test sets (Sayes and Ivanov, 2010). Thus, we concluded that developed by us models, calculated with Eqs. (3)–(5) can be estimated as quite good. The proposed approach has established fundation for novel way of evaluation properties of nanoparticles. We do believe that it could be succesfully applied to a larger pool of nanomaterials and endpoints.

The Supplementary materials section contains the details of three splits into the sub-training, calibration, test and validation sets and correlation weights of physicochemical features of NPs studied here for models calculated with Eqs. (3)–(5) (Tables S1–S4); contains the extraction of features (impacts) which are statistically significant promoters of increase or decrease for the membrane damage.

4. Conclusions

The response of a biological system to nanoparticles appeared to be unique and depends on the physico-chemical characteristics of nanoparticles, dose and duration of exposure. The data from biological tests should therefore be interpreted and processed differently from data for chemicals. This is in line with the recommendations provided by the European Commission Scientific Committee on Emerging and Newly-Identified Health Risks (SCENIHR, 2007). At the present state of knowledge comparative information on the biological activity of nanoparticles would serve best for characterization of hazard and prioritization of nanosized material.

The suggested QFARs give possibility to solve the task. Our study confirmed that optimal descriptors can be an efficient tool for modelling of membrane damage caused by TiO_2 and ZnO nanoparticles. The models developed here successfully apply the optimal descriptors based on representation of both metal oxide nanoparticles by its physicochemical features to evaluate membrane damage. The results of this work (Eqs. (3)–(5)) indicate that the distribution of available data into the training and test sets influences the statistical quality of models: the average correlation coefficient is equal to 0.743 ($\pm$0.068) and the average root mean standard error is equal to 0.240 ($\pm$0.066). In other words the best prediction (Split 1) characterized by $r^2_{validation}$=0.8362 and the worst prediction (Split 3) characterized by $r^2_{validation}$=0.6763 are significantly different.

Thus, the optimal descriptors calculated by the Monte Carlo method which are many times checked up in the role of a tool for the "classic" QSPR/QSAR analyses (Toropova et al., 2011; García et al., 2011; Garro Martinez et al., 2011; Ibezim et al., 2012; Mullen et al., 2011; Veselinović et al., 2013; Achary, 2014a,b; Comelli et al., 2014; Deng et al., 2014a,b; Toropova et al., 2014; Toropova and Toropov, 2014; Singh and Gupta, 2014; Worachartcheewan et al., 2014), can be gradually involved as a tool of QFAR analysis for nanomaterials when the large databases on these substances will become available.

Acknowledgments

We thank EC project PreNanoTox (contract 309666), the EC project NanoPUZZLES (Project Reference: 309837), the National Science Foundation (NSF/CREST HRD-0833178, and EPSCoR (Award #: 362492-190200-01/NSFEPS-090378) for financial support. We also express our gratitude to Dr. L. Cappellini, Dr. G. Bianchi and Dr. R. Bagnati for valuable consultations on the computer science.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ecoenv.2014.07.005.

References

Afantitis, A., Melagraki, G., Koutentis, P.A., Sarimveis, H., Kollias, G., 2011. Ligand—based virtual screening procedure for the prediction and the identification of novel b-amyloid aggregation inhibitors using Kohonen maps and Counter-propagation Artificial Neural Networks. Eur. J. Med. Chem. 46, 497–508.

Achary, P.G.R., 2014a. Simplified molecular input line entry system-based optimal descriptors: QSAR modelling for voltage-gated potassium channel subunit Kv7.2. SAR QSAR Environ. Res 25, 73–90.

Achary, P.G.R., 2014b. QSPR modelling of dielectric constants of π-conjugated organic compounds by means of the CORAL software. SAR QSAR Environ. Res 25, 507–526.

Comelli, N.C., Ortiz, E.V., Kolacz, M., Toropova, A.P., Toropov, A.A., Duchowicz, P.R., Castro, E.A., 2014. Conformation-independent QSAR on c-Src tyrosine kinase inhibitors. Chemom. Intell. Lab. Syst. 134, 47–52.

Deng, F.-F., Xie, M.-H., Li, P.-Z., Tian, Y.-L., Zhang, X.-Y., Zhai, H.-L., 2014a. Study on the antagonists for the orphan G protein-coupled receptor GPR55 by quantitative structure-activity relationship. Chemom. Intell. Lab. Syst. 131, 51–60.

Deng, F., Xie, M., Zhang, X., Li, P., Tian, Y., Zhai, H., Li, Y., 2014b. Combined molecular docking, molecular dynamics simulation and quantitative structure-activity relationship study of pyrimido[1,2-*c*][1,3]benzothiazin-6- imine derivatives as potent anti-HIV drugs. J. Mol. Struct. 1067, 1–13.

Drobne, D., Jemec, A., Pipan Tkalec, Z., 2009. In vivo screening to determine hazards of nanoparticles: nanosized TiO_2. Environ. Pollut. 157, 1157–1164.

Fourches, D., Pu, D., Tassa, C., Weissleder, R., Shaw, S.Y., Mumper, R.J., Tropsha, 2010. A quantitative nanostructure–activity relationship modelling. ACS Nano 4, 5703–5712.

Furtula, B., Gutman, I., 2011. Relation between second and third geometric–arithmetic indices of trees. J. Chem 25, 87–91.

García, J., Duchowicz, P.R., Rozas, M.F., Caram, J.A., Mirífico, M.V., Fernández, F.M., Castro, E.A., 2011. A comparative QSAR on 1,2,5-thiadiazolidin-3-one 1,1-dioxide compounds as selective inhibitors of human serine proteinases. J. Mol. Graphics Modell 31, 10–19.

Garro Martinez, J.C., Duchowicz, P.R., Estrada, M.R., Zamarbide, G.N., Castro, E.A., 2011. QSAR study and molecular design of open-chain enaminones as anticonvulsant agents. Int. J. Mol. Sci. 12, 9354–9368.

Ge, Y., Schimel, J.P., Holden, P.A., 2011. Evidence for negative effects of TiO_2 and ZnO nanoparticles on soil bacterial communities. Environ. Sci. Technol. 45, 1659–1664.

Ge, Y., Schimel, J.P., Holdena, P.A., 2012. Identification of soil bacteria susceptible to TiO_2 and ZnO nanoparticles. Appl. Environ. Microbiol. 78, 6749–6758.

Gottschalk, F., Sonderer, T., Scholz, R.W., Nowack, B., 2009. Modeled environmental concentrations of engineered nanomaterials (TiO_2, ZnO, Ag, CNT, fullerenes) for different regions. Environ. Sci. Technol. 43, 9216–9222.

Ibezim, E., Duchowicz, P.R., Ortiz, E.V., Castro, E.A., 2012. QSAR on arylpiperazine derivatives with activity on malaria. Chemom. Intell. Lab. Syst. 110, 81–88.

Mullen, L.M.A., Duchowicz, P.R., Castro, E.A., 2011. QSAR treatment on a new class of triphenylmethyl-containing compounds as potent anticancer agents. Chemom. Intell. Lab. Syst. 107, 269–275.

OECD, 2007. Guidance Document on the Validation of Quantitative Structure-activity Relationships [(q)sar] Models, ⟨http://www.oecd.org/dataoecd/55/35/38130292.pdf⟩.

Ojha, P.K., Mitra, I., Das, R.N., Roy, K., 2011. Further exploring rm 2 metrics for validation of QSPR models. Chemom. Intell. Lab. Syst. 107, 194–205.

Petrova, T., Rasulev, B.F., Toropov, A.A., Leszczynska, D., Leszczynski, J., 2011. Improved model for fullerene C60 solubility in organic solvents based on quantum-chemical and topological descriptors. J. Nanopart. Res. 13, 3235–3247.

Puzyn, T., Leszczynska, D., Leszczynski, J., 2009. Towards the development of "Nano- QSARs": advances and challenges. Small 5, 2494–2509.

REACH, 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 Concerning the Registration, Evaluation, Authorization and Restriction of Chemicals. ⟨http://ec.europa.eu/environment/chemicals/reach/reach_intro.htm⟩.

Robichaud, C.O., Uyar, A.E., Darby, M.R., Zucker, L.G., Wiesner, M.R., 2009. Estimates of upper bounds and trends in nano-TiO_2 production as a basis for exposure assessment. Environ. Sci. Technol 43, 4227–4233.

Roy, K., Mitra, I., Kar, S., Ojha, P.K., Das, R.N., Kabir, H., 2012. Comparative studies on some metrics for external validation of QSPR models. J. Chem. Inf. Model. 52, 396–408.

Sayes, C., Ivanov, I., 2010. Comparative study of predictive computational models for nanoparticle-induced cytotoxicity. Risk Anal. 30, 1723–1734.

SCENIHR (Scientific Committee on Emerging and Newly-Identified Health Risks), The Appropriateness of the Risk Assessment Methodology in Accordance with the Technical Guidance Documents for New and Existing Substances for Assessing the Risks of Nanomaterials, 21–22 June 2007. ⟨http://ec.europa.eu/health/ph_risk/committees/04_scenihr/docs/scenihr_o_010.pdf⟩.

Singh, K.P., Gupta, S., 2014. Nano-QSAR modeling for predicting biological activity of diverse nanomaterials. RSC Adv. 4, 13215–13230.

Toropov, A.A., Toropova, A.P., 2002. QSAR modelling of toxicity on optimization of correlation weights of Morgan extended connectivity. J. Mol. Struct. THEOCHEM 578, 129–134.

Toropov, A.A., Toropova, A.P., 2003. QSPR modelling of alkanes properties based on graph of atomic orbitals. J. Mol. Struct. THEOCHEM 637, 1–10.

Toropov, A.A., Rasulev, B.F., Leszczynski, J., 2008. QSAR modelling of acute toxicity by balance of correlations. Bioorg. Med. Chem. 16, 5999–6008.

Toropov, A.A., Toropova, A.P., Benfenati, E., Gini, G., Puzyn, T., Leszczynska, D., Leszczynski, J., 2012. Novel application of the CORAL software to model cytotoxicity of metal oxide nanoparticles to bacteria *Escherichia coli*. Chemosphere 89, 1098–1102.

Toropov, A.A., Toropova, A.P., Puzyn, T., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2013. QSAR as a random event: modeling of nanoparticles uptake in $PaCa_2$ cancer cells. Chemosphere 92 (2013), 31–37.

Toropov, A.A., Toropova, A.P., 2014. Optimal descriptor as a translator of eclectic data into endpoint prediction: mutagenicity of fullerene as a mathematical function of conditions. Chemosphere 104, 262–264.

Toropova, A.P., Toropov, A.A., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2011. CORAL: quantitative structure–activity relationship models for estimating toxicity of organic compounds in rats. J. Comput. Chem. 32, 2727–2733.

Toropova, A.P., Toropov, A.A., Puzyn, T., Benfenati, E., Leszczynska, D., Leszczynski, J., 2013. Optimal descriptor as a translator of eclectic information into the prediction of thermal conductivity of micro-electro-mechanical systems. J. Math. Chem. 51, 2230–2237.

Toropova, A.P., Toropov, A.A., 2013. Optimal descriptor as a translator of eclectic information into the prediction of membrane damage by means of various TiO_2 nanoparticles. Chemosphere 93, 2650–2655.

Toropova, A.P., Toropov, A.A., Veselinović, J.B., Miljković, F.N., Veselinović, A.M., 2014. QSAR models for HEPT derivates as NNRTI inhibitors based on Monte Carlo method. Eur. J. Med. Chem 77, 298–305.

Toropova, A.P., Toropov, A.A., 2014. CORAL software: prediction of carcinogenicity of drugs by means of the Monte Carlo method. Eur. J. Pharm. Sci. 52, 21–25.

Veselinović, A.M., Milosavljević, J.B., Toropov, A.A., Nikolić, G.M., 2013. SMILES-based QSAR model for arylpiperazines as high-affinity 5-HT1A receptor ligands using CORAL. Eur. J. Pharm. Sci 48, 532–541.

Weininger, D., 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 28, 31–36.

Weininger, D., Weininger, A., Weininger, J.L., 1989. SMILES. 2. Algorithm for generation of unique SMILES notation. J. Chem. Inf. Comput. Sci 29, 97–101.

Weininger, D., 1990. Smiles. 3. Depict. Graphical depiction of chemical structures. J. Chem. Inf. Comput. Sci 30, 237–243.

Worachartcheewan, A., Nantasenamat, C., Isarankura-Na-Ayudhya, C., Prachayasittikul, V., 2014. QSAR study of H1N1 neuraminidase inhibitors from influenza a virus. Lett. Drug Des. Discov. 11, 420–427.