

Quasi-SMILES and nano-QFPR: The predictive model for zeta potentials of Metal Oxide Nanoparticles

Andrey A. Toropov¹, P. Ganga Raju Achary², Alla P. Toropova^{*,1}

¹*IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, 20156 Milan, Italy*

²*Department of Chemistry, Institute of Technical Education and Research(ITER), Siksha 'O'
Anusandhan University,Bhubaneswar,Odisha-751030(India)*

Published version of this paper could be find here <https://doi.org/10.1016/j.cplett.2016.08.018>

Chemical Physics Letters, Volume 660, 1 September 2016, Pages 107-110

Abstract

Building up of the predictive quantitative structure–property/activity relationships (QSPRs/QSARs) for nanomaterials usually are impossible owing to the complexity of the molecular architecture of the nanomaterials. Simplified molecular input-line entry system (SMILES) is a tool to represent the molecular architecture of “traditional” molecules for "traditional" QSPR/QSAR. The quasi-SMILES is a tool to represent features (conditions and circumstances), which accompany the behavior of nanomaterials. Having, the training set and validation set, so-called quantitative feature–property relationships (QFPRs), based on the quasi-SMILES, one can build up model for zeta potentials of metal oxide nanoparticles for situations characterized by different features.

Keywords: QSPR/QSAR; QFPR; quasi-SMILES; Monte Carlo method; CORAL software

^{*)} Corresponding author

Alla P. Toropova

Laboratory of Environmental Chemistry and Toxicology,
IRCCS - Istituto di Ricerche Farmacologiche Mario Negri,
Via La Masa 19, 20156 Milano, Italy

Tel: +39 02 3901 4595

Fax: +39 02 3901 4735

Email: alla.toropova@marionegri.it

1. Introduction

Quantitative structure–property/activity relationships (QSPR/QSAR) based on different descriptors are a tool to build up predictive model for endpoints of different substances as a mathematical function of their molecular structure represented by the molecular graph [1]. Simplified molecular input-line entry system (SMILES) is a possible alternative of molecular graph for representation of the molecular structure for the QSPR/QSAR [2,3]. The CORAL software [4] gives possibility to build up QSPR/QSAR models where the molecular structure is represented by SMILES [5-7].

Intensive research work on the nanomaterials stimulates the search for approaches aimed to predict physicochemical and biochemical behavior of nanomaterials [8].

However, the “traditional” QSPR/QSAR analysis can solve not all tasks related to nanomaterials, because

- (i) The limited number of “more or less” regular sources of data on nanomaterials are available for praxis;
- (ii) Very complex molecular structure of nanomaterials, as a rule, cannot be represented by graph or SMILES; and
- (iii) Usually, a physicochemical and biochemical experiments are based on analysis of conditions (dose, irradiation, time of exposure, etc.), in other words, the molecular structure of nanomaterials sometimes has no influence on an experimental result.

The quasi-SMILES [9-14] is possible way to build up predictive models for nanomaterials. In contrast to traditional SMILES, quasi-SMILES are representations of conditions. It is to be noted the molecular structure in principle can be examined as a special kind of conditions, if it is expedient [15].

The development of quantitative feature-property relationships (QFPRs), based on quasi-SMILES, for zeta potentials of metal oxide nanoparticles is aim of this study.

2. Method

2.1. Data

The numerical data on the zeta potential [mV] of metal oxide nanoparticles are taken from the literature [16]. The model is a mathematical function of different features of nanoparticles. The features are first, fifteen metal oxides, and second, four circumstances: (i) Acid (pH 5.6); (ii) Basic (pH 7.4); (iii) Serum corona; and (iv) Surfactant corona. Table 1 contains the list of the features above and their representations in quasi-SMILES. The quasi-SMILES were randomly split into the training ($\approx 70\%$), calibration ($\approx 15\%$), and external validation ($\approx 15\%$) sets. Three different splits are examined in this work.

2.2. Optimal descriptors

The model for zeta potentials is the following one-variable correlation

$$ZP[mV] = C_0 + C_1 \times DCW(T^*, N^*) \quad (1)$$

where

The $DCW(T^*, N^*)$ is optimal descriptor obtained by the Monte Carlo method with threshold T^* and the number of epochs of the optimization N^* . These values threshold and the number of epochs give best statistical quality of the model for the calibration set. The C_0 and C_1 are regression coefficients (intercept and slope). The threshold is a number in order to define rare (noise) components of quasi-SMILES.

The optimal descriptors are calculated with so-called correlation weights of the features (Table 1)

$$DCW(T^*, N^*) = \sum CW(F_k) \quad (2)$$

The $T=T^*$ and $N=N^*$ are defining in this work by means of the Monte Carlo calculations with ranges $T = (1, 2)$, and $N = (1, 100)$. The CORAL software after calculations with the range above gives $T=T^*$ and $N=N^*$. Having these data one can calculate the model using the Eq. 1.

3. Results and Discussion

Table 2 contains the models of zeta potentials obtained with three different splits into the training, calibration, and validation sets. One can see, the statistical characteristics of models for different splits are comparable and enough good (Table 2).

Table 3 contains the correlation weights for three random splits into the training, calibration, and validation sets, which are utilized to calculate the $DCW(T^*, N^*)$.

Table 4 contains the numerical data on the experimental and calculated zeta potentials of examined metal oxide nanoparticles together with splits into the training, calibration, and validation sets.

According to OECD principles [17], a predictive model should provide the following information:

(i) a defined endpoint; (ii) an unambiguous algorithm; (iii) a defined domain of applicability; (iv) appropriate measures of goodness-of-fit, and robustness and predictivity; and (v) a mechanistic interpretation, if possible.

Endpoint. The models suggested in this work aimed to predict the zeta potential of metal oxide nanoparticles.

Unambiguous algorithm. The algorithm of building up model with quasi-SMILES is described [9-15] and moreover, available on the Internet [4].

Domain of applicability. The domain of applicability for model based on quasi-SMILES is described in the literature [16]. It is to be noted, according the criteria [16] all compounds for split 1 and split 2 fall into domain of applicability, but in the case of split 3 the number of suspected compounds is twenty four. It is paradox, but statistical characteristics of this model are the best.

Appropriate measures of goodness-of-fit. The statistical measure of the goodness-of-fit used in this work accumulate traditional criteria [9-15] (i) the n , i.e. the number of quasi-SMILES in training, calibration, and validation sets; (ii) determination coefficient (r^2); (iii) cross-validated discrimination coefficient (q^2); (iv) Y-randomization ($^C R_p^2$); (v) mean absolute error (MAE); and (vi) Fischer F-ratio.

Mechanistic interpretation. The mechanistic interpretation for suggested model is available via multifold runs of the Monte Carlo optimization [16] in form of the list of promoters of increase (all correlation weights are positive) or decrease (all correlation weights are negative) of the endpoints (Table 5). Thus, the suggested models built up according to the OECD principles [17].

Conclusions

The quasi-SMILES (Table 1) gives possibility to build up predictive model for zeta potential of metal oxide nanoparticles in the form of quantitative feature-property relationships (QFPRs) similar to described early QFAR. The approach gives models for zeta potential of metal oxide nanoparticles according to OECD principles.

Acknowledgments

AAT and APT thank the EC project PeptiCAPS (Project reference: 686141).

References

- [1] E. Estrada, N. Guevara, I. Gutman, Extension of edge connectivity index. Relationships to line graph indices and QSPR applications, *J. Chem. Inform. Comput. Sci.* 38 (3) (1998) 428-431.
- [2] L.M.A. Mullen, P.R. Duchowicz, E.A. Castro, QSAR treatment on a new class of triphenylmethyl-containing compounds as potent anticancer agents, *Chemom. Intell. Lab. Syst.* 107 (2) (2011) 269-275.
- [3] J. García, P.R. Duchowicz, M.F. Rozas, J.A. Caram, M.V. Mirífico, F.M. Fernández, E.A. Castro, 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 (2011) 10-19.
- [4] CORAL, <http://www.insilico.eu/coral>, accessed July 5, 2016
- [5] A. Worachartcheewan, C. Nantasenamat, C. Isarankura-Na-Ayudhya, V. Prachayasittikul, QSAR study of H1N1 neuraminidase inhibitors from influenza a virus, *Lett. Drug Des. Discovery* 11 (4) (2014) 420-427.
- [6] P.G.R. Achary, Simplified molecular input line entry system-based optimal descriptors: QSAR modelling for voltage-gated potassium channel subunit Kv7.2, *SAR QSAR Environ. Res.* 25 (1) (2014) 73-90.
- [7] P.G.R. Achary, QSPR modelling of dielectric constants of π -conjugated organic compounds by means of the CORAL software, *SAR QSAR Environ. Res.* 25 (6) (2014) 507-526.
- [8] J. Leszczynski, Bionanoscience: Nano meets bio at the interface, *Nat. Nanotechnol.* 5 (9) (2010) 633-634.
- [9] A.A. Toropov, A.P. Toropova, Optimal descriptor as a translator of eclectic data into endpoint prediction: Mutagenicity of fullerene as a mathematical function of conditions, *Chemosphere* 104 (2014) 262-264.
- [10] A.A. Toropov, A.P. Toropova, Quasi-QSAR for mutagenic potential of multi-walled carbon-nanotubes, *Chemosphere*, 124 (1) (2015) 40-46.
- [11] A.P. Toropova, A.A. Toropov, Mutagenicity: QSAR -quasi-QSAR -nano-QSAR, *Mini Rev. Med. Chem.* 15 (8) (2015) 608-621.

[12] A.A. Toropov, A.P. Toropova, Quasi-SMILES and nano-QFAR: United model for mutagenicity of fullerene and MWCNT under different conditions, *Chemosphere* 139 (2015) 18-22.

[13] A.A. Toropov, R. Rallo, A.P. Toropova, Use of Quasi-SMILES and Monte Carlo optimization to develop quantitative feature property/activity relationships (QFPR/QFAR) for nanomaterials, *Curr. Top. Med. Chem.* 15 (18) (2015) 1837-1844.

[14] A.P. Toropova, A.A. Toropov, Optimal descriptor as a translator of eclectic information into the prediction of membrane damage by means of various TiO₂ nanoparticles, *Chemosphere* 93 (10) (2013) 2650-2655.

[15] A.P. Toropova, A.A. Toropov, R. Rallo, D. Leszczynska, J. Leszczynski, Optimal descriptor as a translator of eclectic data into prediction of cytotoxicity for metal oxide nanoparticles under different conditions, *Ecotoxicol. Environ. Saf.* 112 (2015) 39-45.

[16] W.-S. Cho, R. Duffin, F. Thielbeer, M. Bradley, I.L. Megson, W. MacNee, C.A. Poland, C.L. Tran, K. Donaldson, Zeta potential and solubility to toxic ions as mechanisms of lung inflammation caused by metal/metal oxide nanoparticles, *Toxicol. Sci.* 126 (2) (2012) 469-477.

[17] OECD, <http://www.oecd.org/dataoecd/55/35/38130292.pdf>, 2007

Table 1

List of features utilized to develop model of zeta potentials

Features	Components of quasi-SMILES, F_k
AgNP	%11
Al ₂ O ₃ NP	%12
CeO ₂ NP _a	%13
CeO ₂ NP _b	%14
Co ₃ O ₄ NP	%15
Cr ₂ O ₃ NP	%16
CuONP _a	%17
CuONP _b	%18
MgONP	%19
NiONP	%20
SiONP ₂	%21
TiO ₂ NP _a	%22
TiO ₂ NP _b	%23
ZnONP _a	%24
ZnONP _b	%25
Acid (pH 5.6)	%31
Basic (pH 7.4)	%32
Serum corona	%33
Surfactant corona	%34

a) Particles, which show no endotoxin contamination [16].

b) Particles, for which, sizes were measured by a transmission electron microscopy [16].

Table 2

The statistical characteristics of models for zeta potentials of metal oxide nanoparticles

Split 1						
$ZP = -23.51(\pm 0.2275) + 9.857(\pm 0.1265) * DCW(2,38)$						
	N	r^2	q^2	$C_{R_p}^2$	MAE*	F
Training set	40	0.7671	0.7451	0.7573	6.54	125
Calibration set	10	0.8495		0.7979	11.6	
Validation set	10	0.7806			7.57	
Split 2						
$ZP = -32.55 (\pm 0.2687) + 9.020 (\pm 0.1421) * DCW(2,11)$						
	N	r^2	q^2	$C_{R_p}^2$	MAE*	F
Training set	38	0.7463	0.7186	0.7370	7.34	106
Calibration set	11	0.8454		0.7967	11.2	
Validation set	11	0.8996			6.95	
Split 3						
$ZP = -23.10 (\pm 0.2896) + 8.660 (\pm 0.1343) * DCW(2,21)$						
	N	r^2	q^2	$C_{R_p}^2$	MAE*	F
Training set	37	0.7479	0.7207	0.7227	7.70	104
Calibration set	12	0.9017		0.8726	7.29	
Validation set	11	0.9270			3.73	

*) MAE = mean absolute error

Table 3

Correlation weights for F_k calculated with the Monte Carlo method

F_k	$CW(F_k)$ for split 1	$CW(F_k)$ for split 2	$CW(F_k)$ for split 3
%11	-1.37474	-1.06411	-1.43676
%12	0.05999	0.81741	0.68916
%13	1.12672	0.0	0.0
%14	0.0	0.0	-1.12771
%15	0.43488	0.0	0.56367
%16	-0.06176	0.44144	0.0
%17	0.31610	0.68796	0.68661
%18	-1.12092	-1.43917	-1.25086
%19	-0.93269	-0.55774	-1.12432
%20	0.99532	1.68975	1.50188
%21	-0.06738	0.19224	-0.00415
%22	0.0	-0.93714	0.0
%23	0.0	0.37740	0.18268
%24	-0.68621	-0.68748	-0.68836
%25	-1.81062	-1.49800	-1.94168
%31	2.81455	3.62677	3.37730
%32	-0.56032	-0.00004	-0.87314
%33	1.12792	1.62002	1.12691
%34	0.81510	1.55771	0.75275

Table 4

Experimental and predicted values of zeta potential for metal oxide nanoparticles

ID	Split1	Split2	Split3	Quasi-SMILES	Zeta potential experiment, [mV]	Split1	Split2	Split3
1.	T*	T	T	%11%31	-13.7	-9.3187	-9.4306	-6.2999
2.	C	T	T	%12%31	16.5	4.8227	7.5410	12.1099
3.	T	T	C	%13%31	17.5	15.3370	0.1678	6.1420
4.	C	C	T	%14%31	10.6	4.2314	0.1678	-3.6237
5.	C	C	C	%15%31	14.8	8.5178	0.1678	11.0232
6.	T	T	V	%16%31	3.2	3.6227	4.1497	6.1420
7.	C	V	T	%17%31	16.5	7.3471	6.3733	12.0878
8.	T	C	T	%18%31	5.9	-6.8169	-12.8137	-4.6901
9.	T	T	T	%19%31	-23.6	-4.9617	-4.8631	-3.5943
10.	T	T	T	%20%31	19.8	14.0418	15.4097	19.1478
11.	T	T	T	%21%31	-8.4	3.5673	1.9019	6.1060
12.	T	V	V	%22%31	6.8	4.2314	-8.2853	6.1420
13.	V	C	T	%23%31	9.2	4.2314	3.5720	7.7239
14.	T	V	T	%24%31	5.5	-2.5322	-6.0334	0.1810
15.	T	T	T	%25%31	-9.2	-13.6149	-13.3444	-10.6724
16.	T	T	T	%11%32	-36.7	-42.5832	-42.1449	-43.1074
17.	T	T	C	%12%32	-33.2	-28.4417	-25.1734	-24.6976
18.	C	C	C	%13%32	-44.1	-17.9275	-32.5465	-30.6655
19.	C	C	T	%14%32	-50.3	-29.0330	-32.5465	-40.4312
20.	T	V	T	%15%32	-30.1	-24.7467	-32.5465	-25.7843
21.	T	T	C	%16%32	-38.7	-29.6417	-28.5646	-30.6655
22.	T	T	T	%17%32	-37.2	-25.9173	-26.3410	-24.7197
23.	T	T	T	%18%32	-44.3	-40.0814	-45.5280	-41.4976
24.	T	T	T	%19%32	-25.0	-38.2261	-37.5774	-40.4018
25.	T	T	T	%20%32	-18.7	-19.2226	-17.3047	-17.6597
26.	T	T	T	%21%32	-21.2	-29.6971	-30.8125	-30.7015
27.	V	T	V	%22%32	-38.6	-29.0330	-40.9996	-30.6655
28.	C	T	T	%23%32	-39.5	-29.0330	-29.1423	-29.0836
29.	T	T	T	%24%32	-25.4	-35.7966	-38.7477	-36.6265
30.	T	T	V	%25%32	-49.7	-46.8794	-46.0587	-47.4799
31.	T	T	V	%11%33	-29.0	-25.9430	-27.5318	-25.7876
32.	T	T	T	%12%33	-25.7	-11.8015	-10.5602	-7.3778
33.	C	V	C	%13%33	-9.3	-1.2873	-17.9333	-13.3457
34.	V	T	T	%14%33	-27.7	-12.3929	-17.9333	-23.1113
35.	T	C	V	%15%33	-7.1	-8.1065	-17.9333	-8.4645
36.	T	T	T	%16%33	-3.5	-13.0016	-13.9515	-13.3457
37.	T	T	T	%17%33	-2.0	-9.2772	-11.7279	-7.3999
38.	V	T	C	%18%33	-28.5	-23.4412	-30.9149	-24.1778
39.	T	T	T	%19%33	-19.4	-21.5859	-22.9643	-23.0819
40.	T	T	V	%20%33	-9.0	-2.5825	-2.6915	-0.3399
41.	V	V	T	%21%33	-8.9	-13.0570	-16.1993	-13.3817
42.	V	V	T	%22%33	-20.0	-12.3929	-26.3865	-13.3457
43.	C	C	C	%23%33	-5.6	-12.3929	-14.5292	-11.7638
44.	T	V	V	%24%33	-20.5	-19.1565	-24.1346	-19.3067
45.	T	T	T	%25%33	-25.9	-30.2392	-31.4456	-30.1600

46.	T	T	T	%11%34	-28.7	-29.0263	-28.0938	-29.0277
47.	T	T	T	%12%34	3.6	-14.8848	-11.1222	-10.6179
48.	T	C	C	%13%34	-6.3	-4.3706	-18.4953	-16.5858
49.	V	V	V	%14%34	-23.2	-15.4761	-18.4953	-26.3514
50.	T	C	T	%15%34	-7.6	-11.1898	-18.4953	-11.7046
51.	V	C	V	%16%34	-12.4	-16.0848	-14.5134	-16.5858
52.	T	V	T	%17%34	-8.8	-12.3604	-12.2898	-10.6400
53.	T	T	T	%18%34	-34.8	-26.5245	-31.4768	-27.4179
54.	T	T	V	%19%34	-20.8	-24.6692	-23.5263	-26.3220
55.	V	T	C	%20%34	-1.7	-5.6657	-3.2535	-3.5800
56.	T	V	C	%21%34	-13.6	-16.1402	-16.7613	-16.6218
57.	V	T	C	%22%34	-29.2	-15.4761	-26.9484	-16.5858
58.	C	T	T	%23%34	-5.8	-15.4761	-15.0912	-15.0039
59.	T	T	T	%24%34	-38.9	-22.2397	-24.6965	-22.5468
60.	T	T	T	%25%34	-39.7	-33.3225	-32.0075	-33.4001

*) T=training set; C=calibration set; V=validation set

Table 5

Correlation weights of features (Table 1) which are obtained in three probes of the Monte Carlo optimization

No.	F_k	$CW(F_k)$ in Probe 1	$CW(F_k)$ in Probe 2	$CW(F_k)$ in Probe 3
Split 1				
1	%31	4.00353	3.30951	2.93382
2	%34	0.87901	0.93549	0.87030
3	%33	1.37451	1.31519	1.18470
4	%12	0.06225	0.06393	0.18479
5	%15	0.62977	0.49563	0.56698
6	%17	0.43362	0.37803	0.44076
7	%20	1.50042	1.18796	1.12709
8	%13	1.75228	1.37365	1.31131
1	%32	-1.25333	-0.68317	-0.56528
2	%11	-2.25073	-1.68453	-1.30761
3	%19	-1.43879	-1.06181	-0.81231
4	%24	-1.06229	-0.80945	-0.56572
5	%25	-2.87036	-2.19226	-1.75214
6	%18	-1.75174	-1.31161	-1.00255
7	%21	-0.19072	-0.12197	-0.00238
Split 2				
1	%33	1.12124	1.74898	1.06399
2	%34	1.05943	1.68784	1.00199
3	%31	3.06563	3.62603	2.93972
4	%12	0.87204	0.87513	0.87252
5	%20	1.68824	1.62807	1.62436
6	%16	0.50118	0.44181	0.49511
7	%17	0.74632	0.69002	0.68734
8	%21	0.24753	0.25055	0.24846
9	%23	0.24882	0.37951	0.37564
1	%11	-0.93648	-0.94031	-0.87801
2	%19	-0.44036	-0.43387	-0.44209
3	%25	-1.44174	-1.31208	-1.31515
4	%18	-1.31530	-1.18998	-1.18656
5	%22	-0.75164	-0.75459	-0.74502
6	%24	-0.62495	-0.56340	-0.56007
Split 3				
1	%31	3.18329	3.00050	3.05922
2	%33	1.31239	0.94179	0.99510
3	%34	1.00352	0.62345	0.62961
4	%17	0.56672	0.62267	0.68879
5	%12	0.56091	0.68818	0.68335
6	%23	0.19051	0.18962	0.25349
7	%15	0.44093	0.49692	0.56317
8	%20	1.24744	1.37315	1.43275
1	%32	-0.31231	-0.81464	-0.81510
2	%11	-1.18693	-1.31569	-1.30757
3	%14	-0.94045	-1.00222	-1.00036
4	%18	-0.99790	-1.12345	-1.06515

5	%19	-0.93493	-1.05971	-0.99604
6	%24	-0.56076	-0.62248	-0.56177
7	%25	-1.62786	-1.75160	-1.75303