2020

Advancement of predictive modeling of zeta potentials (ζ) in metal oxide nanoparticles with correlation intensity index (CII)

Andrey A. Toropov

Natalia Sizochenko

Alla P. Toropova

See next page for additional authors

Authors
Andrey A. Toropov, Natalia Sizochenko, Alla P. Toropova, Danuta Leszczynska, and Jerzy Leszczynski

Advancement of predictive modeling of zeta potentials (ζ) in metal oxide nanoparticles with correlation intensity index (CII)

Andrey A. Toropov[1], Natalia Sizochenko[2,3], Alla P. Toropova[1,*],

Danuta Leszczynska[4], Jerzy Leszczynski[5]

[1]Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Science, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, via Mario Negri 2, 20156 Milano, Italy

andrey.toropov@marionegri.it (A.A.T.); alla.toropova@marionegri.it (A.P.T.)

[2]Department of Informatics, Postdoctoral Institute for Computational Studies, Enfield, NH 03748, USA

natalia.sizochenko@picomps.org (N.S.)

[3]Blanchardstown Campus, Technological University Dublin, YV78 Blanchardstown, Ireland

[4]Department of Civil and Environmental Engineering, Jackson State University, Jackson, MS 39217, USA

danuta@icnanotox.org (D.L.)

[5]Interdisciplinary Center for Nanotoxicity, Department of Chemistry, Physics, and Atmospheric Sciences, Jackson State University, Jackson, MS 39217, USA

jerzy@icnanotox.org (J.L.)

Abstract

It was expected that index of the ideality of correlation (*IIC*) and correlation intensity index (*CII*) could be used as possible tools to improve the predictive power of the quantitative model for zeta potential of nanoparticles. In this paper, we test how the statistical quality of quantitative structure-activity models for zeta potentials (ζ, a common measurement that reflects surface charge and stability of nanomaterial) could be improved with the use of these two indexes. Our hypothesis was tested using the benchmark data set that consists of 87 measurements of zeta potentials in water. We used quasi-SMILES molecular representation to take into consideration the size of nanoparticles in water and calculated optimal descriptors and predictive models based on the Monte Carlo method. We observed that the models developed with utilization of *CII* are statistically more reliable than models obtained with the *IIC*. However, the described approach gives an improvement of the statistical quality of these models for the external validation sets to the detriment for the training sets. Nevertheless, this circumstance is rather an advantage than a disadvantage.

Keywords: zeta potential; nano-QSPR; metal oxide nanoparticles; quasi-SMILES; Index of Ideality of Correlation; Correlation Intensity Index

[*)] Corresponding author

Alla P. Toropova

Laboratory of Environmental Chemistry and Toxicology,

Istituto di Ricerche Farmacologiche Mario Negri IRCCS

Via Mario Negri 2, 20156 Milano, Italy

Tel: +39 02 3901 4595

Fax: +39 02 3901 4735

Email: alla.toropova@marionegri.it

1.Introduction

Zeta potential can serve a measure of surface charge of the nanoparticle as well as to be the measure of the stability of nanoparticle. Under such circumstances, the data on zeta potential becomes the basis of physicochemical and biochemical analysis [1-3]. Nanoparticles have several advantages as medical materials for various human diseases including brain and retinal diseases [1]. The central nervous system remains an area where drug access and delivery are difficult clinically due to the blood-brain barrier. By means of nanotechnology, many researchers have designed and produced nanoparticle-based systems to solve this problem. Data on the zeta potentials is an important component of these studies [1,2]. Besides, the zeta potential of nanoparticles is an indicator of the ability of nanoparticle to interact with cell membranes [2,3].

Unfortunately, the repetitive experimental and theoretical endpoints studies are often time-consuming and inefficient. Significant progress tends to require a combination of large databases with the knowledge of how to combine available facts in order to reach progress [4-9]. The databases are a relatively new paradigm of applying and developing knowledge. Quantitative structure-property/activity relationships (QSPRs/QSARs) are the majority of applying of databases for chemistry, biochemistry, and medicinal chemistry. The current period of evolution of natural sciences characterized also by the development of databases on nanomaterials, which however remain far from to be perfect [10]. QSPR/QSAR analysis is a tool of interpretation and prognosis of phenomena in the above fields of natural sciences. Hence, most likely, the impact of the QSPR/QSAR on the natural sciences as a whole will be increasing.

Predictive modeling of zeta potential values for untested nanoparticles comes handy, as a massive synthesis and experimental evaluation of every possible nanomaterial is an expensive and time-consuming process. For this purpose, QSPR modeling is a convenient way to estimate the data on zeta potentials [4,5]. However, there is not a direct translation of traditional QSPR into methods appropriate for nanomaterials (so-called nano-QSPR). Many attempts to develop nano-descriptors similarly to traditional descriptors faced the major problem: the molecular structure of nanomaterials is complex. Therefore, the applications of the molecular graph [6-8] or utilization of a simplified molecular input-line entry system (SMILES) [9] to build up the nano-descriptors are impossible or at least extremely limited. Nonetheless, so-called quasi-SMILES [11-18] address this limitation, and such techniques can be applied to develop the nano descriptors [4].

The key component of the traditional QSPR is a predictive potential. There are established criteria for the development of predictive potential for QSPR. Apparently, such criteria are also necessary for the nano-QSPR.

At the same time, there is room to improve the predictive potential for previously delivered nano-QSPR models [4]. To reach this aim, we suggested applying additional predictive potential criteria: index of the ideality of correlation (*IIC*) [19,20], and Correlation Contradictions Index (*CCI*) [21,22]. In addition, the new Correlation Intensity Index (*CII*) is applying here.

2.Method

2.1 Data

Data on zeta potentials measurements for 87 metal oxide nanoparticles in water, along with quasi-SMILES representation were taken from our previous publication [4]. As was discussed, quasi-SMILES descriptors describe nanoparticles using available eclectic data, encoding the type of metal oxide, and discrete representations of nominal size and size in H_2O. Table 1 contains the definition of the quasi-SMILES elements.

[Table 1 around here]

The data set was split randomly into four subsets of equal size (%25): active training set, passive training set, calibration set, and validation set [23]. Each set has aimed to solve its' task. The task of the active training set is building up the model (i.e. the definition of correlation weights for molecular features extracted from quasi-SMILES). The task of the passive training set is inspection: whether the current model is satisfactory for quasi-SMILES which are not involved to building up model? The task of the calibration set is to detect starting of the overfitting. The task of the validation set is the final checkup of the predictive potential of the model.

2.2 Model

The model of zeta-potential suggested here is the following:

$$\zeta = C_0 + C_1 \times DCW(T^*, N^*) \qquad (1)$$

In other words, the model is one variable correlation of the descriptor of correlation weights (*DCW*). The correlation weights are calculated with the Monte Carlo method. The C_0 and C_1 are regression coefficients. The T and N are parameters of the Monte Carlo optimization (described below).

2.3 Optimal descriptor

The optimal descriptor calculated as the following:

$$DCW(T^*, N^*) = \sum_{k=1}^{NA} CW(S_k) + \sum_{k=1}^{NA-1} CW(SS_k) \qquad (2)$$

The S_k is a quasi-SMILES atom; SS_k is a pair of connected quasi-SMILES atoms. In other words, if a quasi-SMILES is the sequence of quasi-SMILES atoms: “ABCD”, the S_k are [A,B,C,D]; and the SS_k are [AB,BC,CD]. The NA is the number of quasi-SMILES atoms.

2.4 Monte Carlo optimization

Each quasi-SMILES attribute A_k (i.e. S_k or SS_k) is characterized by the correlation weight $CW(A_k)$. The numerical data on the $CW(A_k)$ is calculated by the Monte Carlo optimization.

Three target functions of the optimization are compared in this work. The Monte Carlo calculations are aimed to provide maximal value for the target functions.

The first target function is defined as the following:

$$TF_1 = \mathrm{R}+\mathrm{R}'-|\mathrm{R}-\mathrm{R}'|\times 0.1 \tag{3}$$

where the R and R' are correlation coefficients for the active training set and the passive training set, respectively.

The second target function is calculated as the following:

$$TF_2 = TF_1 + \mathrm{IIC}\times 0.2 \tag{4}$$

where the *IIC* is the index of ideality of correlation [19,20] that is calculated with data on observed and calculated values of endpoint for the calibration set:

$$\begin{gathered} IIC_{CLB} = r_{CLB}\frac{\min({}^{-}MAE_{\ CLB}, {}^{+}MAE_{\ CLB})}{max({}^{-}MAE_{\ CLB}, {}^{+}MAE_{\ CLB})} \\ {}^{-}MAE_{\ CLB} = \frac{1}{{}^{-}N}\sum_{k=1}^{-N}|\,\Delta_k|, \quad {}^{-}N\ is\ the\ number\ of\ \Delta_k < 0 \\ {}^{+}MAE_{\ CLB} = \frac{1}{{}^{+}N}\sum_{k=1}^{+N}|\,\Delta_k|, \quad {}^{+}N\ is\ the\ number\ of\ \Delta_k \geq 0 \\ \Delta_k = observed_k - calculated_k \end{gathered} \tag{5}$$

The third target function is defined as the following:

$$TF_3 = TF_1 + \mathrm{CII}\times 0.2 \tag{6}$$

where the *CII* is the correlation intensity index calculated as follows:

$$CII = 1 - \sum \Delta R_j^2 > 0 \quad (7)$$

where $\Delta R_j^2 = R_j^2 - R^2$

An example of the calculation of *CII* is presented in Table 2. The *CII* can be calculated with the correlation contradiction index (*CCI*) as follows [21,22]:

$$CII = 1 - CCI \quad (8)$$

[Table 2 around here]

The *T* is an integer to divide the quasi-SMILES atoms into two classes: (i) rare, if the frequency of the quasi-SMILES in the active training set is less than *T*; and (ii) non-rare if the frequency of the quasi-SMILES in the active training set is larger or equal to *T*. The *N* is the number of epochs of the Monte Carlo optimization. The *T*=*T** and *N*=*N** are values of the above parameters which provide the best statistical quality for the calibration set.

Finally, we calculated a set of statistical metrics to assess the predictive potential of developed nano-QSPR models (Table 3). In addition to that, we calculated commonly used statistical metrics (R^2, *CCC*, *RMSE*, and *MAE*).

[Table 3 around here]

3. Results and Discussion

To provide reliable results, we developed three models for each type of target function. Three models calculated with three random splits are the following:

The Monte Carlo optimization target function with TF_1 resulted in following QSPR equations:

$$\zeta = -8.95(\pm 0.74) + 8.20(\pm 0.11) * DCW(1,3) \quad (9)$$

$$\zeta = 48.22 (\pm 1.50) + 13.38 (\pm 0.47) * DCW(1,5) \quad (10)$$

$$\zeta = -14.45(\pm 0.70) + 27.54(\pm 0.61) * DCW(1,6) \quad (11)$$

The Monte Carlo optimization with target function TF_2

$$\zeta = -22.92(\pm 0.93) + 8.56(\pm 0.18) * DCW(1,15) \quad (12)$$

$$\zeta = -25.04(\pm 0.92) + 8.07(\pm 0.23) * DCW(1,15) \quad (13)$$

$$\zeta = -61.18(\pm 2.16) + 13.63(\pm 0.54) * DCW(1,15) \quad (14)$$

The Monte Carlo optimization with target function TF_3

$$\zeta = 11.51(\pm 0.62) + 16.61(\pm 0.34) * DCW(1,15) \quad (15)$$

$$\zeta = 33.72(\pm 1.50) + 12.86(\pm 0.37) * DCW(1,15) \quad (16)$$

$$\zeta = 31.92(\pm 1.02) + 9.82 (\pm 0.18) * DCW(1,15) \quad (17)$$

Table 4 contains the information about used splits of the dataset: active training set (denoted as +), passive training set (denoted as -), calibration set (denoted as #), and validation set (denoted as *) as well as experimental and calculated (with Eqs. 14-16) values of zeta potentials.

[Table 4 around here]

Table 5 provides the details of the statistical quality of models calculated with Eqs. 8-16.

[Table 5 around here]

For the validation set (a set that reflects the predictive power of model) we observed the R^2 in a range of 0.6793 - 0.9336 and *RMSE* variated from 6.6 to 28.0. Based on these two parameters we can conclude that models with TF_1 (Eqs. 6-9) had the worst predictive power (Table 4). Moreover, these three models were less robust compared to models reported in the original paper [4]. Values of *RMSE* and R^2 for models optimized with target function TF_2 were comparable to the values of these parameters in models reported in the original paper [4].

However, models with TF_2 have lower stability (predictive power in the training set) comparing to models from the original paper. Finally, models optimized with TF_3 have statistically overperformed all other models, which makes them the most reliable source for future predictions.

The same observation about TF_1, TF_2, and TF_3 models could be done based on Monte Carlo optimizations progression through epochs (Figure 1). One can see from Table 4 and Figure 1 that the optimization with target function TF_1 delivers models with a defined number of epochs, and further optimization results in the overtraining (i.e. reduction of the statistical quality for the calibration and validation set). At the same time, the optimization with TF_2 or TF_3 has blocked the overtraining. As discussed above, optimization with TF_3 resulted in increased reliability of models. This directly correlates with the fact, the progression through epochs for TF_3 models is smoother than progression for TF_2 models. As a result, we can conclude that the correlation intensity index (TF_3 derives from *CII*) improves the predictive potential of models for metal oxide nanoparticles' zeta potentials.

[Figure 1 around here]

4. Conclusions

In this article, we have demonstrated that weighting quasi-SMILES parameters (descriptors that take into account size-dependent behavior of nanoparticles) with correlation intensity index (*CII*) or index of the ideality of correlation (*IIC*) improves the quality of structure-property models for zeta potentials in metal oxide nanoparticles. We have demonstrated that the inclusion of either *CII* and *IIC* into model blocks overtraining in the Monte Carlo simulations. Developed models had reasonable statistical characteristics, and *CII* overperformed previously reported models for the same dataset. The presented approach does not require complex calculations and noticeably improves the quality of nano-QSPR models for zeta potentials. We suggest that this approach could be successfully transferred to predictive modeling of other physicochemical properties and biological activities of nanomaterials.

Author contributions

The authors contributed equally to this work.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgment

J.L. and DL would like to thank the NSF-CREST program for the support (grant HRD #154774). A.A.T. and A.P.T. express gratitude for the project LIFE-CONCERT (LIFE17 GIE/IT/000461) for the support.

References

[1] D.H. Jo, J.H. Kim, T.G. Lee, J.H. Kim, Size, surface charge, and shape determine therapeutic effects of nanoparticles on brain and retinal diseases, Nanomedicine 11 (7) (2015) 1603-1611. DOI: 10.1016/j.nano.2015.04.015

[2] A. Mikolajczyk, A. Gajewicz, B. Rasulev, N. Schaeublin, E. Maurer-Gardner, S. Hussain, J. Leszczynski, T. Puzyn, Zeta potential for metal oxide nanoparticles: A predictive model developed by a nano-quantitative structure-property relationship approach, Chem. Mater. 27 (7) (2015) 2400-2407. DOI: 10.1021/cm504406a

[3] S.S. Teske, C.S. Detweiler, The Biomechanisms of metal and metal-oxide nanoparticles' interactions with cells, Int. J. Environ. Res. Public Health 12 (2) (2015) 1112-1134. DOI: 10.3390/ijerph120201112

[4] A.A. Toropov, N. Sizochenko, A.P. Toropova, J. Leszczynski, Towards the development of global nano-quantitative structure-property relationship models: Zeta potentials of metal oxide nanoparticles, Nanomaterials 8 (4) (2018) 243. DOI: 10.3390/nano8040243

[5] N. Sizochenko, J. Leszczynski, Review of Current and Emerging Approaches for Quantitative Nanostructure-Activity Relationship Modeling – the Case of Inorganic Nanoparticles, Journal of Nanotoxicology and Nanomedicine (JNN) 1(1) (2016) 1-16. DOI: 10.4018/JNN.2016010101

[6] S. Marković, I. Gutman, Spectral moments of the edge adjacency matrix in molecular graphs. Benzenoid hydrocarbons, J. Chem. Inf. Comput. Sci. 39 (2) (1999) 289-293. https://doi.org/10.1021/ci980032u

[7] A. Mercader, E.A. Castro, A.A. Toropov, QSPR modeling of the enthalpy of formation from elements by means of correlation weighting of local invariants of atomic orbital molecular graphs, Chem. Phys. Lett. 330(5-6) (2000) 612-623. DOI: 10.1016/S0009-2614(00)01126-X

[8] H. González-Díaz, S. Arrasate, A.G.-S. Juan, N. Sotomayor, E. Lete, A. Speck-Planche, J.M. Ruso, F. Luan, M.N.D.S. Cordeiro, Matrix trace operators: From spectral moments of molecular graphs and complex networks to perturbations in synthetic reactions, micelle nanoparticles, and drug ADME processes, Curr. Drug Metab. 15(4) (2014) 470-488. DOI: 10.2174/1389200215666140908101604

[9] D. Weininger, SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules, J. Chem. Inf. Comput. Sci. 28(1) (1988) 31-36. DOI: 10.1021/ci00057a005

[10] S. Panneerselvam, S. Choi, Nanoinformatics: Emerging databases and available tools, Int. J. Mol. Sci. 15 (5) (2014) 7158-7182. DOI: 10.3390/ijms15057158

[11] 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. DOI: 10.1016/j.chemosphere.2015.05.042

[12] 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. DOI: 10.1016/j.ecoenv.2014.10.003

[13] T.X. Trinh, J.-S. Choi, H. Jeon, H.-G. Byun, T.-H. Yoon, J. Kim, Quasi-SMILES-Based Nano-Quantitative Structure-Activity Relationship Model to Predict the Cytotoxicity of

Multiwalled Carbon Nanotubes to Human Lung Cells, Chem. Res. Toxicol. 31(3) (2018) 183-190. DOI: 10.1021/acs.chemrestox.7b00303

[14] A.P. Toropova, A.A. Toropov, Nano-QSAR in cell biology: Model of cell viability as a mathematical function of available eclectic data, J. Theor. Biol. 416 (2017) 113-118. DOI: 10.1016/j.jtbi.2017.01.012

[15] J.-S. Choi, T.X. Trinh, T.-H. Yoon, J. Kim, H.-G. Byun, Quasi-QSAR for predicting the cell viability of human lung and skin cells exposed to different metal oxide nanomaterials, Chemosphere 217 (2019) 243-249. DOI: 10.1016/j.chemosphere.2018.11.014

[16] S. Ahmadi, Mathematical modeling of cytotoxicity of metal oxide nanoparticles using the index of ideality correlation criteria, Chemosphere 242 (2020) 125192. DOI: 10.1016/j.chemosphere.2019.125192

[17] K. Jafari, M.H. Fatemi, Application of nano-quantitative structure-property relationship paradigm to develop predictive models for thermal conductivity of metal oxide-based ethylene glycol nanofluids, J. Therm. Anal. Calorim. (2020) In press. DOI: 10.1007/s10973-019-09215-3

[18] R. Qi, Y. Pan, J. Cao, Z. Jia, J. Jiang, The cytotoxicity of nanomaterials: Modeling multiple human cells uptake of functionalized magneto-fluorescent nanoparticles via nano-QSAR, Chemosphere 249 (2020) 126175. DOI: 10.1016/j.chemosphere.2020.126175

[19] A.A. Toropov, A.P. Toropova, The index of ideality of correlation: A criterion of predictive potential of QSPR/QSAR models? Mutat. Res. Genet. Toxicol. Environ. Mutagen. 819 (2017) 31-37. DOI: 10.1016/j.mrgentox.2017.05.008

[20] A.P. Toropova, A.A. Toropov, The index of ideality of correlation: A criterion of predictability of QSAR models for skin permeability? Sci. Total. Environ. 586 (2017) 466-472. DOI: 10.1016/j.scitotenv.2017.01.198

[21] A.A. Toropov, A.P. Toropova, QSAR as a random event: criteria of predictive potential for a chance model, Struct. Chem. 30 (5) (2019) 1677-1683. DOI: 10.1007/s11224-019-01361-6

[22] A.A. Toropov, A.P. Toropova, The Correlation Contradictions Index (CCI): Building up reliable models of mutagenic potential of silver nanoparticles under different conditions using quasi-SMILES, Sci. Total. Environ. 681(2019) 102-109. DOI: 10.1016/j.scitotenv.2019.05.114

[23] A.P. Toropova, A.A. Toropov, Does the Index of Ideality of Correlation Detect the Better Model Correctly? Mol. Inf. 38 (2019) 1800157. https://doi.org/10.1002/minf.201800157

[24] N. Chirico, P. Gramatica, Real external predictivity of QSAR models: How to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient, J. Chem. Inf. Model. 51(9) (2011) 2320-2335. DOI: 10.1021/ci200211n

[25] K. Roy, S. Kar, The rm2 metrics and regression through origin approach: Reliable and useful validation tools for predictive QSAR models (Commentary on 'Is regression through origin useful in external validation of QSAR models?'), Eur. J. Pharm. Sci. 62 (2014)111-114. DOI: 10.1016/j.ejps.2014.05.019

[26] L.I.-K. Lin, Assay validation using the concordance correlation coefficient, Biometrics 48 (2) (1992) 599-604. DOI: 10.2307/2532314

Table 1.

The definition of quasi-SMILES elements.

Definition of the attribute of quasi-SMILES for nominal size

Range (in nm)		Number of samples in range	Quasi-SMILES element
From	To		
3.590	17.470	33	%11
17.470	31.351	18	%12
31.351	45.231	9	%13
45.231	59.111	7	%14
59.111	72.992	9	%15
72.992	86.872	1	%16
86.872	100.752	1	%17
100.752	114.633	3	%18
114.633	128.513	2	%19
128.513	142.393	1	%20
142.393	156.274	1	%21
156.274	170.154	0	%22
170.154	184.034	0	%23
184.034	197.915	1	%24
197.915	211.795	0	%25
211.795	225.675	0	%26
225.675	239.556	0	%27
239.556	253.436	0	%28
253.436	267.316	0	%29
267.316	281.197	0	%30
281.197	295.077	0	%31
295.077	308.957	0	%32
308.957	322.838	0	%33
322.838	336.718	0	%34
336.718	350.598	0	%35
350.598	364.479	0	%36
364.479	378.359	0	%37
378.359	392.239	0	%38
392.239	406.120	0	%39
406.120	420.000	1	%40

Definition of the attribute of quasi-SMILES for size in H_2O

Range (in nm)		Number of samples in range	Quasi-SMILES element
From	From		
28.900	227.937	37	%51
227.937	426.973	23	%52

426.973	626.010	7	%53
626.010	825.047	5	%54
825.047	1024.083	1	%55
1024.083	1223.120	0	%56
1223.120	1422.157	1	%57
1422.157	1621.193	5	%58
1621.193	1820.230	1	%59
1820.230	2019.267	1	%60
2019.267	2218.303	0	%61
2218.303	2417.340	1	%62
2417.340	2616.377	1	%63
2616.377	2815.413	1	%64
2815.413	3014.450	0	%65
3014.450	3213.487	0	%66
3213.487	3412.523	0	%67
3412.523	3611.560	0	%68
3611.560	3810.597	0	%69
3810.597	4009.633	1	%70
4009.633	4208.670	1	%71
4208.670	4407.707	0	%72
4407.707	4606.743	0	%73
4606.743	4805.780	0	%74
4805.780	5004.817	0	%75
5004.817	5203.853	0	%76
5203.853	5402.890	0	%77
5402.890	5601.927	0	%78
5601.927	5800.963	0	%79
5800.963	6000.000	1	%80

Table 2.

An example of calculation of the *CII* (R^2= 0.8290; *CII*= 0.8956)

Numbers	R_j^2	ΔR_j^2	$\sum \Delta R_j^2 > 0$
1	0.8157	-0.0133	0.0000
2	0.8155	-0.0135	0.0000
3	0.8278	-0.0012	0.0000
4	0.8244	-0.0045	0.0000
5	0.8455	0.0165	0.0165
6	0.8842	0.0552	0.0717
7	0.8277	-0.0013	0.0717
8	0.8288	-0.0001	0.0717
9	0.8160	-0.0130	0.0717
10	0.8246	-0.0044	0.0717
11	0.8236	-0.0053	0.0717
12	0.8212	-0.0078	0.0717
13	0.8191	-0.0099	0.0717
14	0.8428	0.0138	0.0855
15	0.8355	0.0065	0.0921
16	0.8390	0.0100	0.1021
17	0.8291	0.0001	0.1022
18	0.8030	-0.0260	0.1022
19	0.8305	0.0015	0.1037
20	0.8296	0.0007	0.1044
21	0.8280	-0.0010	0.1044

Table 3.

A collection of criteria of predictive potential of models

The criterion of the predictive potential	Reference
$Q^2 = 1 - \frac{\sum(y_k - \acute{y}_k)^{\ 2}}{\sum(y_k - \bar{y}_k)^{\ 2}}$ $Q^2_{F1} = 1 - \frac{\left[\sum_{i=1}^{N_{EXT}}(\acute{y}_i - y_i)^{\ 2}\right]/N_{EXT}}{\left[\sum_{i=1}^{N_{EXT}}(y_i - \bar{y}_{\ TR})^{\ 2}\right]/N_{EXT}}$ $Q^2_{F2} = 1 - \frac{\left[\sum_{i=1}^{N_{EXT}}(\acute{y}_i - y_i)^{\ 2}\right]/N_{EXT}}{\left[\sum_{i=1}^{N_{EXT}}(y_i - \bar{y}_{\ EXT})^{\ 2}\right]/N_{EXT}}$ $Q^2_{F3} = 1 - \frac{\left[\sum_{i=1}^{N_{EXT}}(\acute{y}_i - y_i)^{\ 2}\right]/N_{EXT}}{\left[\sum_{i=1}^{N_{TR}}(y_i - \bar{y}_{\ TR})^{\ 2}\right]/N_{TR}}$	[24]
$\overline{R_m^2} = \frac{R_m^2(x,y) + R_m^2(y,x)}{2}$	[25]
$CCC = \frac{2\sum\ (\mathrm{x} - \bar{\mathrm{x}})(\mathrm{y} - \bar{\mathrm{y}})}{\sum(\mathrm{x} - \bar{\mathrm{x}})^{\ 2} + \sum(\mathrm{y} - \bar{\mathrm{y}})^{\ 2} + \mathrm{n}(\bar{\mathrm{x}} - \bar{\mathrm{y}})^{\ 2}}$	[26]

Table 4.

Experimental and calculated values of zeta potential and quasi-SMILES used for the representation of corresponding nanoparticles [4]

ID	Set*	Set	Set	Quasi-SMILES	Experiment	Eq. 14	Eq. 15	Eq. 16
1.	*	#	#	O=[Al]O[Al]=O%11%51	39.2	59.5854	55.4805	39.7962
2.	-	#	*	O=[Al]O[Al]=O%15%54	33.1	12.9653	30.5613	23.0576
3.	+	+	+	O=[Al]O[Al]=O%11%52	38.0	27.2658	19.7513	17.0003
4.	#	-	*	O=[Al]O[Al]=O%12%51	43.0	61.5090	54.0057	52.3818
5.	+	+	#	O=[Al]O[Al]=O%13%52	36.2	36.6891	41.2958	47.4311
6.	*	*	-	O=[Al]O[Al]=O%14%52	30.3	31.7833	7.9330	5.7311
7.	*	*	*	O=[Bi]O[Bi]=O%21%71	-16.5	-21.2014	-4.7394	-16.2568
8.	+	-	+	O=[Ce][Ce]=O%11%51	41.2	28.3292	31.5044	25.0026
9.	#	#	*	O=[Ce][Ce]=O%11%51	26.5	28.3292	31.5044	25.0026
10.	+	#	+	O=[Ce][Ce]=O%12%51	21.4	30.2527	30.0296	37.5882
11.	#	*	*	O=[Ce][Ce]=O%11%63	15.0	19.4829	4.2426	14.0535
12.	#	+	-	[Co]=O^O=[Co]O[Co]=O%11%51	23.0	30.8860	22.5773	2.7684
13.	+	-	+	[Co]=O^O=[Co]O[Co]=O%11%51	24.6	30.8860	22.5773	2.7684
14.	+	*	#	[Co]=O%15%51	21.6	4.8870	43.0765	18.0431
15.	#	+	-	[Co]=O%14%52	17.5	-3.5979	11.1079	-1.4868
16.	-	#	+	O=[Cr]O[Cr]=O%24%52	-32.6	-24.3981	-12.0014	-32.7718
17.	#	#	+	O=[Cr]O[Cr]=O%14%52	-12.0	-29.7866	-9.4556	-10.8888
18.	-	+	*	[Cu]=O%12%51	37.4	18.4691	39.6837	40.9596
19.	*	*	#	[Cu]=O%11%51	17.0	16.5456	41.1586	28.3740
20.	*	+	*	[Cu]=O%11%52	7.6	-15.7740	5.4294	5.5781
21.	#	*	-	[Cu]=O%12%52	24.4	10.6026	37.0502	0.0540
22.	+	-	-	O=[Dy]O[Dy]=O%11%53	50.6	50.6458	38.7161	11.2430
23.	#	-	#	O=[Fe]O[Fe]=O%12%55	-22.8	-26.1493	-15.1870	-10.2410
24.	-	#	*	O=[Fe]O[Fe]=O%12%58	-11.2	-14.6342	12.4539	-10.2410
25.	-	#	+	O=[Fe]O[Fe]=O%11%51	-2.1	-1.9844	13.0566	1.5650
26.	*	#	#	O=[Fe]O[Fe]=O%15%80	-6.3	-48.6045	-5.5220	-15.1736
27.	-	-	-	O=[Fe]^O=[Fe]O[Fe]=O%11%51	22.1	2.0586	11.9652	0.4439
28.	*	+	#	O=[Fe]^O=[Fe]O[Fe]=O%12%54	-17.7	-22.1063	-15.8117	-11.3621
29.	-	+	+	O=[Fe]^O=[Fe]O[Fe]=O%19%51	8.3300	-3.5935	1.9282	-4.5212
30.	#	-	*	O=[Fe]^O=[Fe]O[Fe]=O%11%51	-2.1	2.0586	11.9652	0.4439
31.	#	#	#	O=[Gd]O[Gd]=O%13%51	6.5	4.6263	38.6164	16.8080
32.	+	-	+	O=[Hf]=O%12%52	33.5	33.7313	13.0555	33.2282
33.	+	-	+	O=[In]O[In]=O%13%51	57.2	48.1012	45.0083	44.6085
34.	-	+	-	O=[In]O[In]=O%15%51	61.9	22.1732	28.9050	12.3294
35.	-	-	+	O=[In]O[In]=O%15%52	22.6	5.0341	18.6433	21.4170
36.	+	+	+	O=[In]O[In]=O%11%52	-31.6	9.1709	8.7545	4.0686
37.	+	-	-	O=[La]O[La]=O%12%51	54.3	44.2685	36.6170	11.6497
38.	-	*	-	O=[La]O[La]=O%15%53	-3.6	-8.7750	22.0952	-12.3970
39.	*	#	*	O=[Mg]%11%60	6.9	8.3151	21.3346	13.6341
40.	+	+	-	O=[Mn]O[Mn]=O%14%52	-46.1	-35.2464	-37.3765	-35.0010
41.	#	+	-	O=[Mn]O[Mn]O[Mn]=O%11%52	-14.4	-41.0522	-18.5907	-16.4719
42.	*	-	-	O=[Ni]O[Ni]=O%20%52	32.2	44.4924	23.2079	3.5890
43.	+	*	+	[Ni]=O%11%51	48.9	50.9909	71.6341	41.8253
44.	#	#	#	[Ni]=O%12%59	13.3	26.8260	43.3905	30.0193
45.	*	+	*	[Ni]=O%11%52	27.6	18.6713	35.9049	19.0294
46.	*	*	*	[Ni]=O%11%52	26.0	18.6713	35.9049	19.0294

47.	-	*	+	O=[Sb]O[Sb]=O%12%51	-24.2	-23.7448	-28.5352	-18.9018
48.	+	-	#	O=[Sb]O[Sb]=O%11%51	-35.3	-25.6684	-27.0604	-31.4874
49.	+	+	#	O=[Sb]O[Sb]=O%16%53	-20.7	-20.5144	-21.0619	-41.5308
50.	+	+	+	O=[Si]=O%11%52	-29.2	-54.8367	-53.9468	-49.9792
51.	#	+	#	O=[Si]=O%11%51	-33.5	-22.5171	-18.2176	-27.1833
52.	#	*	*	O=[Si]=O%18%51	-43.0	-28.1692	-50.4802	-40.3008
53.	-	#	#	O=[Si]=O%11%51	-31.8	-22.5171	-18.2176	-27.1833
54.	+	-	+	O=[Si]=O%13%51	-23.1	-15.9064	-17.6930	-9.4394
55.	-	*	*	O=[Si]=O%14%51	-30.1	-20.8726	-38.8681	-39.8353
56.	-	*	#	O=[Si]=O%18%51	-33.1	-28.1692	-50.4802	-40.3008
57.	*	#	*	O=[Si]=O%40%54	-39.0	-41.7341	-67.3896	-42.5042
58.	*	*	#	O=[Si]=O%12%57	-29.8	-46.6820	-46.4612	-38.9893
59.	-	-	-	O=[Sn]=O%15%51	-38.8	-22.7432	1.5851	-22.7310
60.	*	-	#	O=[Sn]=O%11%70	-21.1	-12.2722	-10.0979	-19.1449
61.	#	+	#	O=[Ti]=O%12%52	-16.5	-9.3689	3.7231	-23.4008
62.	*	-	+	O=[Ti]=O%19%51	-13.5	-9.0780	-2.2055	-0.0459
63.	-	*	*	O=[Ti]=O%14%53	-18.9	-19.8463	-20.8059	-4.6588
64.	-	+	-	O=[Ti]=O%11%51	47.0	-3.4259	7.8315	4.9191
65.	#	*	#	O=[Ti]=O%18%51	-4.64	-9.0780	-24.4311	-8.1984
66.	-	+	+	O=[Ti]=O%11%51	-19.4	-3.4259	7.8315	4.9191
67.	*	#	*	O=[Ti]=O%11%51	15.0	-3.4259	7.8315	4.9191
68.	-	*	-	O=[Ti]=O%11%58	7.09	-0.7571	8.2106	-6.0299
69.	#	#	-	O=[Ti]=O%17%58	4.07	-11.1278	-7.3588	-10.4018
70.	#	#	-	O=[Ti]=O%14%58	1.77	-3.8313	4.2533	-9.9363
71.	*	#	*	O=[Ti]=O%11%64	-3.75	-12.2722	-19.4303	-6.0299
72.	#	*	#	O=[Ti]=O%13%54	-10.7	0.2244	-8.5534	-4.3130
73.	*	#	#	O=[W](=O)=O%11%51	-45.2	-61.6097	-49.5739	-64.7132
74.	+	-	+	O=[W](=O)=O%11%51	-61.3	-61.6097	-49.5739	-64.7132
75.	*	+	+	O=[W](=O)=O%11%53	-54.4	-49.2022	-48.9497	-52.5343
76.	-	-	-	O=[Y]O[Y]=O%13%52	42.7	15.5184	23.9072	6.6991
77.	+	-	#	O=[Y]O[Y]=O%13%52	16.3	15.5184	23.9072	6.6991
78.	+	#	-	O=[Yb]O[Yb]=O%15%52	9.9	9.1256	12.2514	-6.3834
79.	*	*	*	[Zn]=O%12%51	16.4	5.6240	25.9991	15.2806
80.	#	+	+	[Zn]=O%12%53	-46.8	-24.9644	-45.5657	-47.0129
81.	#	#	-	[Zn]=O%12%54	0.017	-20.4645	-0.3030	-9.1111
82.	-	-	-	[Zn]=O%13%53	20.3	2.8509	20.0115	-1.2596
83.	-	#	+	[Zn]=O%12%51	28.8	5.6240	25.9991	15.2806
84.	#	*	*	[Zn]=O%11%52	-15.0	-28.6191	-8.2553	-20.1010
85.	+	+	*	[Zn]=O%15%58	-20.9	-20.5075	-21.8737	-14.0437
86.	*	*	+	O=[Zr]=O%13%52	-12.8	-25.3938	2.9793	3.0667
87.	+	-	-	O=[Zr]=O%12%62	-6.9	-6.9728	-11.0796	-16.3743

*) Active training set (+), passive training set (-), calibration set (#), validation set (*).

Table 5. The statistical characteristics for developed models

The Monte Carlo optimization with target function TF_1													
split	**Set**[*]	**n**	R^2	***CCC***	***IIC***	***CII***	Q^2	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	$\overline{R^2_m}$	**RMSE**	**MAE**
1 (Eq. 9)	AT	22	0.8757	0.9337	0.6479	0.9109	0.8537					12.8	8.28
	PT	22	0.8618	0.6677	0.6776	0.8787	0.8454					19.1	15.8
	C	22	0.4830	0.6886	0.6585	0.7701	0.3851	0.3915	0.3115	0.6789	0.4401	19.0	15.2
	V	21	0.7999									15.4	12.5
2 (Eq. 10)	AT	22	0.5709	0.7269	0.7556	0.7556	0.4994					22.2	16.8
	PT	22	0.6603	0.7221	0.3788	0.8377	0.6031					23.1	17.5
	C	22	0.6453	0.6569	0.2970	0.7781	0.5849	0.2134	0.1968	0.6426	0.2783	20.7	17.3
	V	21	0.5770									17.6	14.0
3 (Eq. 11)	AT	23	0.8475	0.9175	0.7082	0.9010	0.8229					13.5	10.4
	PT	22	0.9416	0.5530	0.9704	0.9622	0.9250					30.6	28.5
	C	21	0.6779	0.7498	0.6298	0.7788	0.6299	0.3936	0.1505	0.5195	0.5769	22.7	18.6
	V	21	0.7267									28.0	22.7
The Monte Carlo optimization with target function TF_2													
split	**Set**	**n**	R^2	***CCC***	***IIC***	***CII***	Q^2	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	$\overline{R^2_m}$	**RMSE**	**MAE**
1 (Eq. 12)	AT	22	0.7868	0.8807	0.7392	0.8618	0.7522					18.4	14.3
	PT	22	0.7872	0.5787	0.3631	0.8949	0.7258					24.6	21.9
	C	22	0.7961	0.8536	0.8918	0.9007	0.7530	0.6657	0.6077	0.8043	0.6954	15.0	12.3
	V	21	0.8005									15.6	11.8
2 (Eq. 13)	AT	22	0.6256	0.7697	0.6591	0.7741	0.5666					20.7	15.4
	PT	22	0.6258	0.7118	0.3369	0.8410	0.5713					23.6	17.8
	C	22	0.7339	0.7728	0.8562	0.8488	0.6696	0.5620	0.5528	0.8010	0.4245	15.4	13.1
	V	21	0.6793									14.2	10.9
3 (Eq. 14)	AT	23	0.5631	0.7205	0.6879	0.7485	0.4832					22.7	18.5
	PT	22	0.4909	0.2470	0.0703	0.7236	0.3528					45.7	41.8
	C	21	0.6874	0.7303	0.8282	0.7753	0.6238	0.4011	0.3304	0.5893	0.3929	20.1	15.9
	V	21	0.7973									15.7	12.1
The Monte Carlo optimization with target function TF_3													
split	**set**	**n**	R^2	***CCC***	***IIC***	***CII***	Q^2	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	$\overline{R^2_m}$	**RMSE**	**MAE**
1 (Eq. 15)	AT	22	0.8751	0.9334	0.7796	0.9023	0.8479					12.8	7.91
	PT	22	0.8493	0.6860	0.3925	0.8783	0.8289					19.8	14.9
	C	22	0.6937	0.8281	0.6720	0.8596	0.6364	0.6647	0.6206	0.8230	0.6665	14.1	11.8
	V	21	0.8290									14.9	11.0
2 (Eq. 16)	AT	22	0.7138	0.8330	0.8449	0.8274	0.6617					18.1	12.4
	PT	22	0.8595	0.8877	0.7240	0.8919	0.8348					14.2	11.3
	C	22	0.7536	0.8263	0.5884	0.8665	0.6990	0.5448	0.5352	0.7932	0.6336	15.7	11.8
	V	21	0.8396									15.4	13.0
3 (Eq. 17)	AT	23	0.8098	0.8949	0.6922	0.8642	0.7807					15.0	11.4
	PT	22	0.9598	0.4951	0.5854	0.9809	0.9427					25.8	21.9
	C	21	0.8920	0.9278	0.7321	0.9417	0.8687	0.8731	0.8222	0.8995	0.8373	10.4	8.64
	V	21	0.9336									6.6	5.2
Previously reported modes, validation sets [4]													
	V	19	0.6707									17.2	14.7
	V	16	0.8213									15.8	11.6
	V	21	0.7268									13.1	11.7

[*)]AT – active training set, PT – passive training set, C – calibration set, V – validation set.

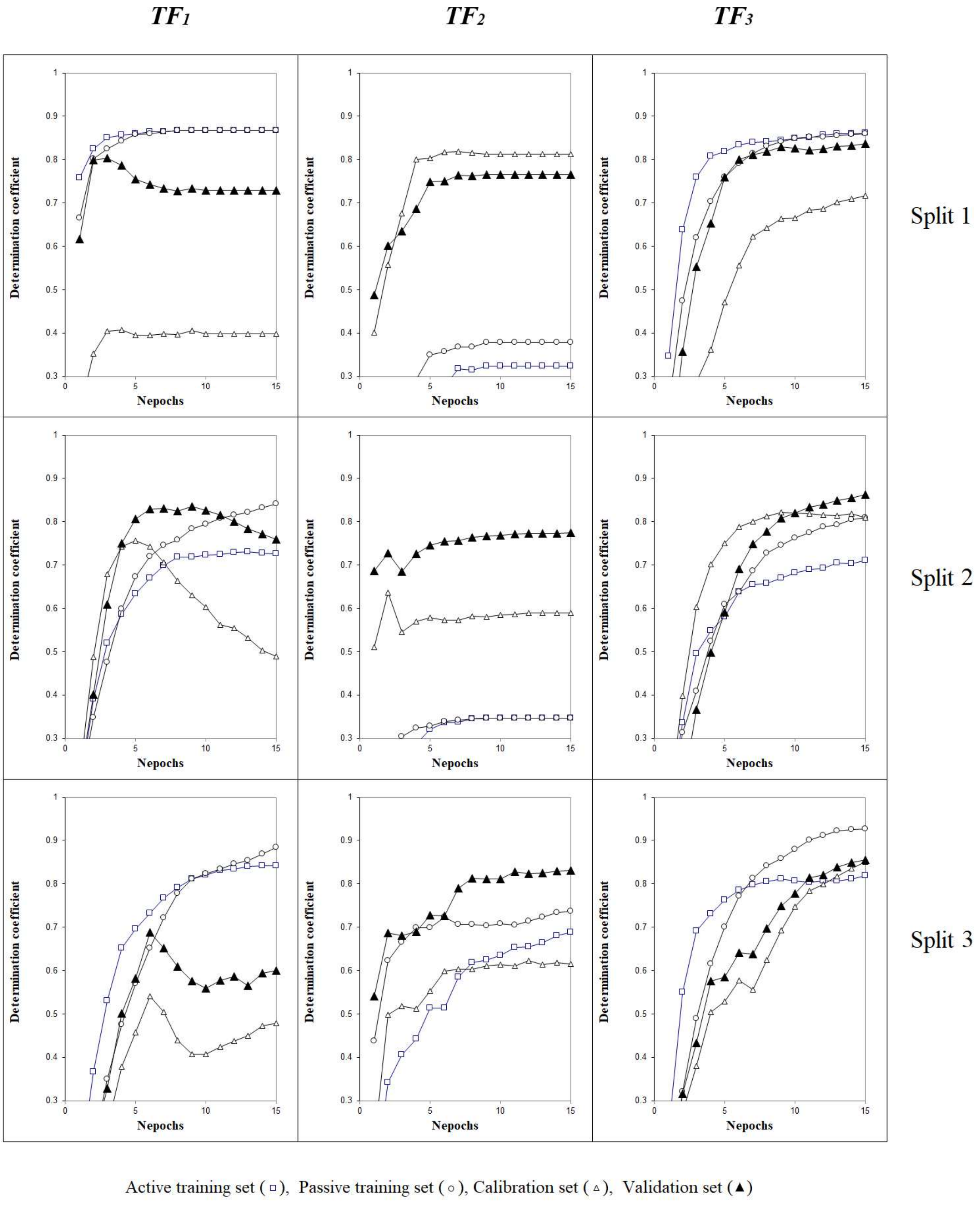
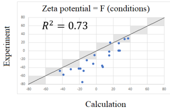
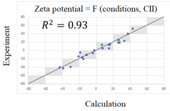


Figure 1.

The comparison of histories of the Monte Carlo optimization with target functions TF_1, TF_2, and TF_3


Zeta potential = F (conditions)
$R^2 = 0.73$
Experiment
Calculation
80
60
40
20
0
-20
-40
-60
-80
-80
-60
-40
-20
0
20
40
60
80


Zeta potential = F (conditions, CII)
$R^2 = 0.93$
Experiment
Calculation
80
60
40
20
0
-20
-40
-60
-80
-80
-60
-40
-20
0
20
40
60
80