Extending of QSPR/QSAR-algorithms for the prediction of the behavior of nanomaterials

Alla P. Toropova* and Andrey A. Toropov*

Istituto di Ricerche Farmacologiche Mario Negri IRCCS, 20156, Via Mario Negri 2, Milano, Italy

Abstract

To represent the molecular structure of most nanomaterials by a molecular graph or even simplified molecular input-line entry system (SMILES) is very problematic. Databases that contain a large number of different nanomaterials are not established up to now. Nevertheless, large collections of the behavior of individual nanomaterials under different experimental conditions are available. Namely, the experimental conditions are a tool to define how to influence nanomaterial in order to obtain some attractive effect on different targets such as cells, organisms, or chemical-technological processes. Traditional SMILES provide special codes related to the molecular structure which can be used to build up traditional QSPR/QSAR models. Quasi-SMILES is an extension of the traditional SMILES by means of considering additional codes that reflect experimental conditions. The quasi-SMILES were applied to build up models for different endpoints related to nanomaterials such as mutagenic potential of multiwalled carbon nanotubes (MWCNTs) [1]; cytotoxicity for metal oxide nanoparticles [2,3]; cytotoxicity of MWCNTs [4]; solubility of fullerenes C60 and C70 in various solvents [5]; cell viability of human lung and skin cells exposed to different metal oxide nanomaterials [6]; mutagenic potential of silver nanoparticles [7]. Most probably, quasi-SMILES will find many others applications in the nearest future, e.g. by using the CORAL software [1-7] (http://www.insilico.eu/coral).

Keywords: Nano-QSPR; Nano-QSAR; quasi-SMILES; Monte Carlo method; CORAL software.

*) 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; andrey.toropov@marionegri.it

1. Introduction

Nanomaterials become an important component of modern everyday life. These become component of clearer, drugs, cosmetics, food, etc. Under such circumstances, the prediction of physicochemical, biochemical, behavior of these substances becomes a quite important task. However, the naive applying of classical quantitative stricture – property/activity relationships (QSPRs/QSARs) approaches for nanomaterials often is failed.

There are several reasons why it is not successful. Probably, the first cause is the impossibility of the representation of the molecular structure of nanomaterials by a molecular graph or simplified molecular input-line entry system (SMILES). The second cause is the absence of available databases on nanomaterials. The third reason is the weak interdependence of the molecular structure of nanomaterials and their endpoints.

So-called quasi-SMILES is probably quite effective technology to build up predictive models for endpoints related to various nanomaterials. The basic idea of the approach is including for the development of predictive models of nanomaterials all available eclectic data. Paradox, but probably, descriptors of quantum mechanics remain the most popular approach to the development of predictive models for nanomaterials.

Nevertheless, so-called optimal descriptors calculated with the Monte Carlo technique are an attractive alternative for the above-mentioned approaches, because (i) these descriptors can be calculated from arbitrary eclectic information [8, 9]; and (ii) these descriptors can be easily modified for fresh experimental data if these will become available [10, 11].

2. Method

The quasi-SMILES can be represented as a sequence of special codes (x_{ij}) and experimental endpoint values (e_j):

$$I = \left\{ \begin{matrix} x_{11} & x_{21} \dots & x_{1n} \\ x_{21} & x_{22} \dots & x_{2n} \\ \dots & \dots & \dots \\ x_{m1} & x_{m2\dots} & x_{mn} \end{matrix} \middle| \begin{matrix} e_1 \\ e_2 \\ \dots \\ e_n \end{matrix} \right\} \quad (1)$$

The matrix I (input) should be translated into matrix M (model):

$$M = \left\{ \begin{matrix} C(x_{11}) & C(x_{21}) \dots & C(x_{1n}) \\ C(x_{21}) & C(x_{22}) \dots & C(x_{2n}) \\ \dots & \dots & \dots \\ C(x_{m1}) & C(x_{m2}) \dots & C(x_{mn}) \end{matrix} \middle| \begin{matrix} e_1 \\ e_2 \\ \dots \\ e_n \end{matrix} \right\} \quad (2)$$

The translation is the following: Each of the above codes can be provided by correlation weights $C(x_{ij})$ that is a special coefficient. The numerical data on the correlation coefficients are calculated with the Monte Carlo method. The simplest version of the approach is target function equal to the Pearson correlation coefficient. In other words, the numerical data on the correlation weights provide the maximal value of the correlation coefficient between the endpoint and sum of correlation weights for the training set:

$$R[\sum C(x_{ij}), e_j] \rightarrow Maximum \quad (3)$$

However, the criterion can be improved. For instance, the index of ideality of correlation gives possibility to build up a paradox model, where the statistical quality on the validation set can be better that the statistical quality of the model for the training set.

The generalized form of predictive model is calculated by the least square method:

$$e_j = Intercept + Slope \times \sum C(x_{ij}) \quad (4)$$

The model is calculated with data on the training set. The validation of the model should be checked up with external validation set.

3. Results

3.1. Mutagenic potential of multi-walled carbon-nanotubes

Optimal descriptors calculated with eclectic data gave a statistically significant model for mutagenic potentials of MWCNTs under various conditions. These models (quasi-QSARs) are built up in accordance with the OECD principles for validation of a QSAR. The basis of the model is (i) process with preincubation or process without preincubation; (ii) process with metabolic activation or process without metabolic activation; and (iii) dose (mg/plate) [1].

3.2. Cytotoxicity of metal oxide nanoparticles for metal oxide nanoparticles

Quasi-SMILES contain two components: (i) traditional SMILES represented metal oxide nanoparticle, and (ii) special symbol indicated photoinduction of the process interaction between metal oxide nanoparticle and cells [2].

3.3. Cytotoxicity of metal oxide nanoparticles using the index of ideality correlation criteria

Quasi-SMILES contain the following components: (i) cell line (i.e. MCF-7, HT-1080, HepG-2, HT-29, and PC-12); (ii) time exposition (i.e. 24h, 56h, 72h); (iii) concentration (mg/mL); (iv) normalized nanoparticle size; (v) metal oxide type (i.e. SnO_2, MnO_2, ZnO, Bi_2O_3, NiO, CeO_2, SiO_2, and TiO_2) [3].

3.4. Cytotoxicity of Multiwalled Carbon Nanotubes to Human Lung Cells

The quasi-SMILES applied to build the optimal descriptors were containing codes of (i) diameter; (ii) length; (iii) surface area; (iv) in vitro toxicity assay; (v) cell line; (vi) exposure time; and (vii) dose. The model confirms potential for use in the estimation of human lung cell viability after exposure to MWCNTs with the following properties: diameter, 12–74 nm; length, 0.19–20.25 μm; surface area, 11.3–380.0 m2/g; and dose, 0–200 ppm [4].

3.5. Solubility of fullerene C60 and C70 (mole fraction) in different organic solvents

The quasi-SMILES applied to build up the model were containing (i) traditional SMILES for representation of molecular structure of a solvent; (ii) graph invariants of hydrogen suppressed graph of a solvent; and (iii) special symbol to separate C60 ('x') and C70 ('y') [5].

Table 1 contains the statistical quality of models listed in sections 3.1; 3.2; 3.3; 3.4; and 3.5.

Table 1

The statistical quality of models for endpoints related to nanomaterials described in the literature.

Ref.	Number of quasi-SMILES in training set	Correlation coefficient between experimental and predicted endpoint values for training set	Number of quasi-SMILES in validation set	Correlation coefficient between experimental and predicted endpoint values for validation set
[1]	13	0.80	5	0.91
[2]	9+6=15	(0.90+0.99)/2=0.94	6	0.98
[3]	29+29+12=70	(0.92+0.92+0.92)/3=0.92	13	0.94
[4]	-	-	-	0.81-0.88
[5]	67+67+39=173	(0.89+0.91+0.86)/3=0.89	39	0.92

Thus, there is a diversity of versions for the quasi-SMILES. One can see, the list of the above versions can be extended [6-11]. The above-mentioned index of the ideality of correlation is factually a new possibility to modify the statistical quality of QSPR/QSAR models, in general, and predictive models for nanomaterials, in particular. Recently, checking up the ability of the index of the ideality of correlation was carried out [12].

4. Conclusions

Quasi-SMILES is convenient conception of building up predictive models for nanomaterials. The conception is like a traditional QSPR/QSAR scheme. The approach provides the user with the possibility to define a list of potential conditions with further analysis of results aimed to separate significant and non-significant conditions applied to develop a predictive model for endpoints related to nanomaterials. The index of the ideality of correlation is a measure of the

predictive potential of a model. Software to calculate this index is available on the Internet (http://www.insilico.eu/coral).

Acknowledgements

The authors gratefully thank Prof. González-Díaz for invitation to participation in MOL2NET 2020, International Conference on Multidisciplinary Sciences.

References

1. Toropov AA, Toropova AP. Quasi-QSAR for mutagenic potential of multi-walled carbon-nanotubes. *Chemosphere* 2015; 124 (1): 40-46. DOI: 10.1016/j.chemosphere.2014.10.067

2. Toropova AP, Toropov AA, Rallo R, Leszczynska D, Leszczynski J. Optimal descriptor as a translator of eclectic data into prediction of cytotoxicity for metal oxide nanoparticles under different conditions. *Ecotoxicol Environ Saf.* 2015; 112: 39-45.

DOI: 10.1016/j.ecoenv.2014.10.003

3. Ahmadi S. Mathematical modeling of cytotoxicity of metal oxide nanoparticles using the index of ideality correlation criteria. *Chemosphere* 2020; 242: 125192. DOI: 10.1016/j.chemosphere.2019.125192

4. Trinh TX, Choi J-S, Jeon H, Byun H-G, Yoon T-H, Kim J. Quasi-SMILES-Based Nano-Quantitative Structure-Activity Relationship Model to Predict the Cytotoxicity of Multiwalled Carbon Nanotubes to Human Lung Cells. *Chem. Res. Toxicol.* 2018; 31 (3): 183-190. DOI: 10.1021/acs.chemrestox.7b00303

5. Toropova AP, Toropov AA. QSPR and nano-QSPR: What is the difference? *J Mol Struct*. 2019; 1182: 141-149. DOI: 10.1016/j.molstruc.2019.01.040

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

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

8. Toropova AP, Toropov AA. Optimal descriptor as a translator of eclectic information into the prediction of membrane damage by means of various TiO_2 nanoparticles. *Chemosphere* 2013; 93: 2650–2655. https://doi.org/10.1016/j.chemosphere.2013.09.089

9. Toropov AA, Toropova AP. Optimal descriptor as a translator of eclectic data into pTA100 prediction: mutagenicity of fullerene as a mathematical function of conditions. *Chemosphere* 2014; 104: 262–264. https://doi.org/10.1016/j.chemosphere.2013.10.079

10. Toropov AA, Toropova AP, Benfenati E, Leszczynska D, Leszczynski J. SMILES-based optimal descriptors: QSAR analysis of fullerene-based HIV-1 PR inhibitors by means of balance of correlations. *J Comput Chem.* 2010; 31: 381–392. https://doi.org/10.1002/jcc.21333

11. Toropova AP, Toropov AA, Benfenati E, Leszczynska D, Leszczynski J. QSAR modeling of measured binding affinity for fullerene-based HIV-1 PR inhibitors by CORAL. *J Math Chem*. 2010; 48: 959–987. https://doi.org/10.1007/s10910-010-9719-x

12. Toropova AP, Toropov AA. Does the Index of Ideality of Correlation Detect the Better Model Correctly? *Mol Inf*. 2019; 38(8-9): 1800157. DOI: 10.1002/minf.201800157