Chemosphere 104 (2014) 262–264

Contents lists available at ScienceDirect

Chemosphere

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

Short Communication

Optimal descriptor as a translator of eclectic data into endpoint prediction: Mutagenicity of fullerene as a mathematical function of conditions

Andrey A. Toropov *, Alla P. Toropova

IRCCS – Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, Milano 20156, Italy

HIGHLIGHTS

- Quasi-QSAR for fullerene C60 nanoparticles is suggested.
- The model of mutagenicity is a mathematical function of conditions.
- The statistical quality of the quasi-QSAR is quite good.

ARTICLE INFO

Article history:
Received 12 September 2013
Received in revised form 21 October 2013
Accepted 26 October 2013
Available online 15 November 2013

Keywords:
Fullerene C60
Bacterial reverse mutation test
Quasi-QSAR
Optimal descriptor

ABSTRACT

The experimental data on the bacterial reverse mutation test on C60 nanoparticles (TA100) is examined as an endpoint. By means of the optimal descriptors calculated with the Monte Carlo method a mathematical model of the endpoint has been built up. The model is the mathematical function of (i) dose (g/plate); (ii) metabolic activation (i.e. with S9 mix or without S9 mix); and (iii) illumination (i.e. dark or irradiation). The statistical quality of the model is the following: $n = 10$, $r^2 = 0.7549$, $q^2 = 0.5709$, $s = 7.67$, $F = 25$ (Training set); $n = 5$, $r^2 = 0.8987$, $s = 18.4$ (Calibration set); and $n = 5$, $r^2 = 0.6968$, $s = 10.9$ (Validation set).

1. Introduction

The development of the systematical representation for various nanomaterials is complex task since the molecular architecture of these substances is very untypical in comparison with architecture of organic (Toropova et al., 2011a), inorganic, organometallic (Toropova et al., 2011b) substances which can be represented by the molecular graph (Toropov and Toropova 2002; Toropov and Toropova 2003; Toropov and Roy 2004; Castillo-Garit et al., 2007; Fourches et al., 2010; Afantitis et al., 2011; Furtula and Gutman, 2011), simplified molecular input-line entry systems (García et al., 2011; Garro Martinez et al., 2011; Mullen et al., 2011; Toropov et al., 2011; Ibezim et al., 2012) (SMILES) or International Chemical Identifier (InChI) (Toropov et al., 2009; Toropov et al., 2010). Even in the case of majority of polymers their molecular structure can be systematized via the architecture of monomers (Toropov et al., 1999). Quantitative structure–property/activity relationships (QSPRs/QSARs) give the possibility to define preferable substances to solve various practical tasks (Puzyn et al., 2009; Leszczynski, 2010). This definition is based on the mathematical comparison of variety of molecular structures characterized by the measure of their ability to solve the given practical task. These measures are obtained from the relevant experiment. In fact the selection of molecular structure is the selection of the substance with the most appropriate above-mentioned measure.

"Classic" QSPR/QSAR analyses for nanomaterials are limited by the absence of standardized databases on their structure together with physicochemical and biomedical endpoints. However, in the case of the nanomaterials various technological conditions as well as conditions of exposure can influence an effect of these substances upon biological objects. Under such circumstances the quasi-QSAR where instead of the representation of a substance by molecular structure the representation of this by the above-mentioned available eclectic information becomes attractive alternative of "classic" QSAR (Toropov et al., 2007; Toropova et al., 2013).

Optimal descriptors calculated with so-called correlation weights of various attributes of nanomaterials may be a tool to build up the quasi-QSAR. The aim of the present work is the estimation of the optimal descriptors as possible approach to build

* Corresponding author. Tel.: +39 02 39014595; fax: +39 02 39014735.
E-mail address: andrey.toropov@marionegri.it (A.A. Toropov).

0045-6535/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.chemosphere.2013.10.079

up predictive model of mutagenicity of fullerene C60 nanoparticles under various conditions (dose, illumination, presence/absence of S9 mix).

2. Method

2.1. Data

Data on the bacterial reverse mutation test that was conducted using *Salmonella typhimurium* strain TA100 in the presence and absence of metabolic activation under dark conditions and irradiation are taken in the literature (Shinohara et al., 2009). The data are split into the training, calibration, and validation sets according to the following principles: (i) the split is random; and (ii) the ranges of endpoint for the above-mentioned sets are similar.

2.2. Optimal descriptor

The optimal descriptors used in this study are calculated as the following:

$$DCW(T, N_{Epoch}) = \Sigma CW\ (CI_k) \quad (1)$$

where CI_k is a code of an attribute of fullerene C60 nanoparticle (Table 1).

The $CW(CI_k)$ are correlation weights of various attributes. The correlation weights of attributes are calculated with optimization by the Monte Carlo technique. The correlation weights should provide maximal value of the correlation coefficient between the $DCW(T,N_{Epoch})$ and experimental TA100. The T and the N_{epoch} are parameters of the optimization: the T (threshold) is coefficient for classification of impacts into two categories rare and not rare. Correlation weight for rare impact is fixed equal to zero. Therefore rare attributes are not involved in a model. The N_{epoch} is the number of epochs of the Monte Carlo optimization.

Having data on optimal correlation weights, one can

(i) calculate $DCW(T, N_{epoch})$ for all fullerene C60 nanoparticles;

(ii) calculate (with data on the training set) a model for TA100:

$$TA100 = C_0 + C_1 \times DCW(T, N_{epoch}) \quad (2)$$

The model should give preferable statistical quality for the calibration set, i.e. best quality for a preliminary external test set (Toropov et al., 2008); and (iii) predictive potential of the model should be checked up with an external validation set. The fullerene C60 nanoparticles of the validation set are not involved in building up model.

Table 1
List of attributes of fullerene C60 nanoparticles exposure and their codes.

Attribute	Codes of attributes (CI_k) and their meaning
Dark or Irradiation	0 = Dark 1 = Irradiation
Mix S9	+ = with Mix S9 – = without Mix S9
Dose (g/plate)	*A* = 50 *B* = 100 *C* = 200 *D* = 400 *E* = 1000

Table 2
The list of fullerene C60 nanoparticles and data on the bacterial reverse mutation test TA100.

No.	Set	Dark or Irradiation	Mix S9	Dose	TA100
1	Validation	0	+	A	146
2	Training	0	+	B	141
3	Training	0	+	C	159
4	Validation	0	+	D	160
5	Training	0	+	E	177
6	Calibration	0	–	A	143
7	Training	0	–	B	139
8	Validation	0	–	C	169
9	Training	0	–	D	168
10	Training	0	–	E	152
11	Calibration	1	+	A	129
12	Training	1	+	B	131
13	Validation	1	+	C	138
14	Training	1	+	D	137
15	Calibration	1	+	E	160
16	Validation	1	–	A	136
17	Training	1	–	B	136
18	Training	1	–	C	138
19	Calibration	1	–	D	164
20	Calibration	1	–	E	172

Table 3
Experimental and calculated with Eq. (3) values of the TA100 for different C60 nanoparticles.

ID	Code	*DCW* (1,3)	$TA100_{expr}$	$TA100_{calc}$	$\Delta = TA100_{expr} - TA100_{calc}$
Training set					
2	0+B	2.93450	141.0	145.657	4.657
3	0+C	3.37500	159.0	157.105	1.895
5	0+E	3.68850	177.0	165.252	11.748
7	0–B	2.87200	139.0	144.033	5.033
9	0–D	3.49600	168.0	160.249	7.751
10	0–E	3.62600	152.0	163.628	11.628
12	1+B	2.30950	131.0	129.415	1.585
14	1+D	2.93350	137.0	145.631	8.631
17	1–B	2.24700	136.0	127.791	8.209
18	1–C	2.68750	138.0	139.239	1.239
Calibration set					
6	0–A	2.43550	143.0	132.690	10.310
11	1+A	1.87300	129.0	118.072	10.928
15	1+E	3.06350	160.0	149.010	10.990
19	1–D	2.87100	164.0	144.007	19.993
20	1–E	3.00100	172.0	147.386	24.614
Validation set					
1	0+A	2.49800	146.0	134.314	11.686
4	0+D	3.55850	160.0	161.873	–1.873
8	0–C	3.31250	169.0	155.481	13.519
13	1+C	2.75000	138.0	140.863	–2.863
16	1–A	1.81050	136.0	116.448	19.552

3. Results and discussion

Quasi-QSAR for the bacterial reverse mutation test TA100 is the following:

$$TA100 = 69.3980(\pm 7.2079) + 25.9872(\pm 2.5129) * DCW(1,3) \quad (3)$$

$$n = 10,\ r^2 = 0.7549, q^2 = 0.5709,\ s = 7.67,\ F = 25 \text{ (Training set)}$$

$$n = 5,\ r^2 = 0.8987,\ s = 18.4 \text{ (Calibration set)}$$

$$n = 5,\ r^2 = 0.6968,\ s = 10.9 \text{ (Validation set)}$$

The statistical quality of the model is quite good. However, the number of substances is small. Under such circumstances,

Table 4
Correlation weights for various attributes of fullerene C60 nanoparticles calculated by the Monte Carlo method.

Code of attribute, CI_k	Correlation weight, $CW(CI_k)$
+	1.0020
–	0.9395
0	1.4960
1	0.8710
A	0.0
B	0.4365
C	0.8770
D	1.0605
E	1.1905

Table 5
An example of calculation of optimal descriptor with correlation weights for fullerene C60 nanoparticle represented by code "0+B"; $DCW(1,3) = 2.9345$; TA100 = 69.3980 + 25.9872∗2.9345 = 145.6574.

CI_k	$CW(CI_k)$
0	1.4960
+	1.0020
B	0.4365

additional check up is desirable. The Y-randomization (Ojha and Roy, 2011) gives ${}^{c}R_p^2$ (for training set) = 0.645 and ${}^{c}R_p^2$ (for calibration set) = 0.759. A correlation is not random if ${}^{c}R_p^2 > 0.5$ (Ojha and Roy, 2011). Consequently, correlations between experimental and predicted TA100 are not random ones. Metrics of predictability (Roy et al., 2013) for the calibration set are $\overline{r_m^2} = 0.7297$, and $\Delta r_m^2 = 0.1245$. According to the literature a model has predictive potential if $\overline{r_m^2} > 0.5$, $\Delta r_m^2 < 0.2$ (Roy et al., 2013). Consequently, the model calculated with Eq. (3) has predictive potential. It is to be noted, the software to calculate the above mentioned statistical characteristics is available on the Internet (http://dtclab.webs.com/software-tools and http://aptsoftware.co.in/rmsquare/) (see Tables 2–5).

4. Conclusions

Optimal descriptors can be a tool for the quasi-QSAR analysis of the mutagenicity of fullerene C60 nanoparticles characterized by various exposures, such as (i) various doses, (ii) the presence/absence of metabolic activation (S9 mix), and (iii) the presence/absence of irradiation.

Acknowledgements

We thank the EC project NANOPUZZLES (Project Reference: 309837) and EC project PreNanoTox (contract 309666). We also express our gratitude to Dr. L. Cappellini, Dr. G. Bianchi and Dr. R. Bagnati for valuable consultations on the computer science.

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 Counterpropagation Artificial Neural Networks. Eur. J. Med. Chem. 46, 497–508.
Castillo-Garit, J.A., Marrero-Ponce, Y., Torrens, F., Rotondo, R., 2007. Atom-based stochastic and non-stochastic 3D-chiral bilinear indices and their applications to central chirality codification. J. Mol. Graphics Model. 26, 32–47.
Fourches, D., Pu, D., Tassa, C., Weissleder, R., Shaw, S.Y., Mumper, R.J., Tropsha, 2010. A quantitative nanostructure–activity relationship modeling. ACS Nano 4, 5703–5712.
Furtula, B., Gutman, I., 2011. Relation between second and third geometric–arithmetic indices of trees. J. Chemom. 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 Model. 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.
Ibezim, E., Duchowicz, P.R., Ortiz, E.V., Castro, E.A., 2012. QSAR on aryl-piperazine derivatives with activity on malaria. Chemom. Intell. Lab. Syst. 110, 81–88.
Leszczynski, J., 2010. Nano meets bio at the interface. Nat. Nanotech. 5, 633–634.
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.
Ojha, P.K., Roy, K., 2011. Comparative QSARs for antimalarial endochins: importance of descriptor-thinning and noise reduction prior to feature selection. Chemom. Intell. Lab. Syst. 109, 146–161.
Puzyn, T., Leszczynska, D., Leszczynski, J., 2009. Towards the development of "Nano-QSARs": advances and challenges. Small 5, 2494–2509.
Roy, K., Chakraborty, P., Mitra, I., Ojha, P.K., Kar, S., Das, R.N., 2013. Some case studies on application of "r_m^2" metrics for judging quality of quantitative structure-activity relationship predictions: emphasis on scaling of response data. J. Comput. Chem. 34, 1071–1082.
Shinohara, N., Matsumoto, K., Endoh, S., Maru, J., Nakanishi, J., 2009. In vitro and in vivo genotoxicity tests on fullerene C60 nanoparticles. Toxicol. Lett. 191, 289–296.
Toropov, A.A., Roy, K., 2004. QSPR modeling of lipid-water partition coefficient by optimization of correlation weights of local graph invariants. J. Chem. Inf. Comput. Sci. 44, 179–186.
Toropov, A.A., Toropova, A.P., 2002. QSAR modeling 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 modeling of alkanes properties based on graph of atomic orbitals. J. Mol. Struct. THEOCHEM 637, 1–10.
Toropov, A.A., Voropaeva, N.L., Ruban, I.N., Rashidova, S.S., 1999. Quantitative structure–property relationships for binary polymer–solvent systems: correlation weighing of the local invariants of molecular graphs. Polym. Sci. – Ser. A 41, 975–985.
Toropov, A.A., Leszczynska, D., Leszczynski, J., 2007. Predicting thermal conductivity of nanomaterials by correlation weighting technological attributes codes. Mater. Lett. 61, 4777–4780.
Toropov, A.A., Rasulev, B.F., Leszczynski, J., 2008. QSAR modeling of acute toxicity by balance of correlations. Bioorg. Med. Chem. 16, 5999–6008.
Toropov, A.A., Toropova, A.P., Benfenati, E., Leszczynska, D., Leszczynski, J., 2009. Additive InChI-based optimal descriptors: QSPR modeling of fullerene C60 solubility in organic solvents. J. Math. Chem. 46, 1232–1251.
Toropov, A.A., Toropova, A.P., Benfenati, E., Leszczynska, D., Leszczynski, J., 2010. InChI-based optimal descriptors: QSAR analysis of fullerene [C60]-based HIV-1 PR inhibitors by correlation balance. Eur. J. Med. Chem. 45, 1387–1394.
Toropov, A.A., Toropova, A.P., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2011. SMILES-based QSAR approaches for carcinogenicity and anticancer activity: comparison of correlation weights for identical SMILES attributes. Anti-Cancer Agents Med. Chem. 11, 974–982.
Toropova, A.P., Toropov, A.A., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2011a. 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., Benfenati, E., Gini, G., 2011b. Co-evolutions of correlations for QSAR of toxicity of organometallic and inorganic substances: an unexpected good prediction based on a model that seems untrustworthy. Chemom. Intell. Lab. Syst. 105, 215–219.
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.