Nano-QSAR: Model of mutagenicity of fullerene as a mathematical function of different conditions

Alla P. Toropova[a], Andrey A. Toropov[a,*], Aleksandar M. Veselinović[b], Jovana B. Veselinović[b], Emilio Benfenati[a], Danuta Leszczynska[c], Jerzy Leszczynski[d]

[a]*IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, Milano, Italy*

[b]*University of Niš, Faculty of Medicine, Department of Chemistry, Niš, Serbia*

[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, P.O. Box 17910, Jackson, MS 39217, USA*

Abstract

The experimental data on the bacterial reverse mutation test (under various conditions) on C60 nanoparticles for the cases (i) TA100, and (ii) WP2uvrA/pkM101 are examined as endpoints. By means of the optimal descriptors calculated with the Monte Carlo method a mathematical model of these endpoints has been built up. The models are a mathematical function of eclectic data such as (i) dose (g/plate); (ii) metabolic activation (i.e. with mix S9 or without mix S9); and (iii) illumination (i.e. darkness or irradiation). The eclectic data on different conditions were represented by so-called quasi-SMILES. In contrast to the traditional SMILES which are representation of molecular structure, the quasi-SMILES are representation of conditions by sequence of symbols. The calculations were carried out with the CORAL software, available on the Internet at http://www.insilico.eu/coral. The main idea of the suggested descriptors is the accumulation of all available eclectic information in the role of logical and digital basis for building up a model. The computational experiments have shown that the described approach can be a tool to build up models of mutagenicity of fullerene under different conditions.

Key words: Fullerene nanoparticle; Mutagenicity; quasi-SMILES; nano-QSAR

*) Corresponding author
Andrey A. Toropov
Laboratory of Environmental Chemistry and Toxicology,
IRCCS - Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, 20159 Milano, Italy
Tel: +39 02 3901 4595
Fax: +3902 3901 4735
Email: andrey.toropov@marionegri.it

1. Introduction

During last decades, a list of traditional substances has expanded and a new group of species – nanomaterials has been introduced both in research laboratories and in the everyday life. Nanomaterials already play significant role in many aspects of modern life, such as chemical and pharmacological industries, medicine, ecology, toxicology. For the case of the majority of organic compounds the estimation of their dangerous and attractive properties (endpoints) can be done by means of well-known quantitative structure – property / activity relationships (QSPRs/QSARs) (Salahinejad and Ghasemi, 2014; Wang et al., 2015; Peric et al., 2015). Thus, endpoints related to nanomaterials are waiting for predictive QSPR/QSAR models (Fourches et al., 2010; Leszczynski, 2010).

However, the simply application of the classic approaches based on the representation of various substances by the structure of their molecules is impossible for majority of nanomaterials. Moreover, the classic QSPR/QSAR models are based on manipulation with considerable large databases, where the number of available substances can be hundreds or even thousands. In the case of nanomaterials as rule their synthesis and identification are very difficult actions and as result, the number of nanomaterials which are examined in some standardized conditions is very small (units or tens, in the best case).

Therefore, in the case of nanomaterials one cannot expect large databases to be available in the near future. Consequently, predictive models for endpoints related to nanomaterials should focus and be developed using small data sets. Moreover, molecular structure of majority of nanomaterials has no convenient representations (for the QSPR/QSAR analyses) such as molecular graph (Toropov and Toropova, 2002; Toropov and Toropova, 2003; Toropov and Roy, 2004) or simplified molecular input-line entry system (SMILES) (García et al., 2010, Garro Martinez et al., 2011; Ibezim et al., 2011).

Thus, for the case of nanomaterials, the traditional classic QSPR/QSAR theory and praxis should be cardinally modified. International organizations such as, OECD (Organisation for Economic Co-operation and Development) (OECD, 2006-2010) and REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) (REACH, 2011) pay attention on necessity of the further evolution of knowledge related to nanomaterials. Multiple toxicity of nanomaterials have been carried out in the last decade. However, most of these works used non-standardized testing protocols leading to unreproducible and poorly comparable results, which therefore are insufficient for estimation of hazard and risk assessment (Burello and Worth, 2011; Richarz et al., 2015). In addition, the unavailability of consistent physico-chemical characterization data in the same studies makes it difficult to identify which material

characteristics determine the documented toxic effects. In other words, the need for standardization and new approaches in the nanosafety area has been recognized (Salahinejad, 2015).

The possible way, to estimate endpoints related to nanomaterials with using so-called optimal descriptors has been suggested (Toropova and Toropov, 2013; Toropova et al., 2013; Toropova et al., 2015). The advantage of the approach is the possibility of standardization of eclectic data. In the case of traditional QSPR/QSAR analyses, the paradigm "*Endpoint = F(structure)*" is the basis to build up a predictive model. In this case, SMILES is the representation of the molecular structure. In the case of the predictive modelling for nanomaterials, an alternative paradigm can be defined as "*Endpoint = F(eclectic data)*". The eclectic data can be represented by sequence of symbols similar to SMILES, but where symbols are a representation of a list of eclectic conditions. These sequences of symbols should be denoted as "quasi-SMILES" in order to avoid their false interpretation as the "traditional SMILES". Since these models are based on list of eclectic conditions (not on structure), instead of QSPR/QSAR one should use some other terminology for these researches, e.g. "quantitative *conditions* - property / activity relationships" (QCPR/QCARs).

The aim of the present work is to estimate QCAR based on data (conditions) related to (i) dose (g/plate); (ii) metabolic activation (with or without mix S9); and (iii) illumination (dark or irradiation) as a tool to predict two endpoints. The endpoint-1 is the bacterial reverse mutation test that was conducted using *Salmonella typhimurium* strains TA100; and the endpoint-2 is mutagenic effect of fullerene for *Escherichia coli* strain WP2 uvrA/pKM101 (Shinohara et al., 2009).

2. Method

2.1. Data

Data on (i) the bacterial reverse mutation test that was conducted using *Salmonella typhimurium* strains TA100 in the presence and absence of metabolic activation under dark conditions and irradiation are taken in the literature, and (ii) the numerical data on the bacterial reverse mutation test that was conducted using *Escherichia coli* strain WP2 uvrA/pKM101 in the presence and absence of metabolic activation under dark condition and irradiation are taken in the literature (Shinohara et al., 2009). Twenty quasi-SMILES have been defined for the above data. These twenty quasi-SMILES were randomly split into the training, calibration, and validation sets. The training set is basis to build up a model. The calibration set is used to avoid the overtraining (situation where the excellent statistical quality of a model for the training set

is accompanied by poor statistical quality of the model for external quasi-SMILES). The validation set is utilized for the final estimation of the predictive potential of the model. For each endpoint three random splits were examined.

2.2. Optimal descriptor

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

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

where S_k is a symbol for representation of conditions listed in Table 1.

The $CW(S_k)$ are correlation weights of a symbol from quasi-SMILES. The correlation weights of symbols 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 data on the TA100 or WP2 uvrA/pKM101. 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 the 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 quasi-SMILES;

(ii) calculate (with data on the training set) a model for TA100 or for WP2 uvrA/pKM101:

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

$$WP2uvrA/\ pKM101 = C_0 + C_1 \times DCW(T, N_{epoch}) \quad (3)$$

These models should give preferable statistical quality for the calibration set, i.e. the best quality for a preliminary external set: quasi-SMILES from the calibration set are visible but their role is passive checking up "whether overtraining happens?". If yes, the process should be stopped and restart of the optimization till N_{epoch}-1, where overtraining is not reached should be carried out (Toropova et al., 2011);

and

(iii) predictive potential of the model should be checked up with an *external* validation set. The quasi-SMILES of the validation set are not involved in building up model.

3. Results and Discussion

The utilization of the optimal descriptors calculated according to scheme suggested in the literature (Toropova et al., 2013) gives the following models:

TA100 = 69.4243 + 25.9161*DCW(2,2) (4)

n= 10, r^2=0.7551, s=7.67, F=25 (sub-training set)

n=5, r^2=0.9032, s=19.0 (Calibration set)

n=5, r^2=0.6826, s=14.4 (Validation set)

TA100 = 94.0379 + 14.8826 * DCW(2,12) (5)

n= 10, r^2=0.6318, s=9.00, F=14 (sub-training set)

n=5, r^2=0.7867, s=18.7 (Calibration set)

n=5, r^2=0.7108, s=25.8 (Validation set)

TA100 = 117.813 + 12.3159 * DCW(2,3) (6)

n= 10, r^2=0.6810, s=9.78, F=17 (sub-training set)

n=5, r^2=0.9396, s=7.91 (Calibration set)

n=5, r^2=0.7884, s=7.79 (Validation set)

WP2uvrA/pKM101 = -30.1488 + 91.6017 * DCW(3,3) (7)

n=9, r^2=0.7877, s=16.2, F=26 (sub-training set)

n=6, r^2=0.8729, s=12.6 (calibration set)

n=5, r^2=0.8138, s=23.2 (validation set)

WP2uvrA/pKM101 = -11.3863 + 46.37* DCW(2,3) (8)

n=9, r^2=0.8206, s=15.4, F=32 (sub-training set)

n=7, r^2=0.8767, s=36.7 (calibration set)

n=5, r^2=0.8185, s=39.7 (validation set)

WP2uvrA/pKM101 = 84.9481 + 16.1111 * DCW(3,6) (9)

n=10, r^2=0.6805, s=12.1, F=17 (sub-training set)

n=5, r^2=0.7480, s=16.5 (calibration set)

n=5, r^2=0.8367, s=25.7 (validation set)

Table 1 contains definition of symbols used in quasi-SMILES. Table 2 contains numerical data on TA100 together with the general scheme of construction of quasi-SMILES. Table 3 contains the QCAR model for TA100. Table 4 contains correlation weights for

calculation of the optimal descriptors used in Eq.4, Eq.5 and Eq.6. Table 5 contains an example of the calculation of TA100. Table 6 contains the QCAR model for WP2uvrA/pKM101. Table 7 contains the correlation weights for calculation with Eq. 5.

Mechanistic interpretation: Having data on several runs of the Monte Carlo procedure one can classify the conditions in accordance with their correlation weights: if every time the correlation weight is large in all probes, the condition should be estimated as a promoter of endpoint increase; if every time the correlation weight is small in all probes, the condition should be estimated as a promoter of endpoint decrease; if correlation weight has unstable value in several probes of the runs of the Monte Carlo procedure, one should estimate the condition as undefined. One can see, that in the case of TA100 (Table 4), darkness is promoter of TA100 increase, irradiation is promoter of TA100 decrease; doses 200, 400, and 100 g/plate are promoter TA100 increase; and presence or absence of Mix S9, and impacts of doses 50 and 100 g/plate are undefined (i.e. these are not stable promoter of increase or decrease of the endpoint). In the case of WP2uvrA/pKM101 (Table 7), presence of Mix S9 is promoter of decrease and absence of Mix S9 is promoter of increase of the mutagenicity; darkness is promoter of decrease and irradiation is promoter of increase of the endpoint; dose 1000 g/plate is promoter of the endpoint increase; and impact of doses 50, 100, 200, and 400 g/plate are undefined.

Domain of applicability. The defect of quasi-SMILES (Toropova et al., 2015) is a criterion to detect outliers. Table 3 contains data on quasi-SMILES which are recognized as outliers for TA100 models according to the above statistical defect of quasi-SMILES (Toropova et al., 2015). Table 6 contains the similar data for WP2uvrA/pKM101 models.

The comparison of dispersion of the experimental definition and standard error of estimation of models which are calculated with Eqs. 4-9 indicates that accuracy of calculated mutagenicity is lower than the accuracy of experimental definition, but these are comparable. Thus the possibility of the prediction the numerical data on TA100 and WP2uvrA/pKM101 for fullerene nanoparticles under different conditions is demonstrated. In our previous work (Toropov and Toropova, 2014) only one split into the training, calibration, and validation sets has been examined. The analysis of models represented here has shown that the statistical quality of the models significantly depend on the distribution into visible training and calibration sets and invisible validation set. However all examined models (Eqs. 4-9) have predictive potentials. In the above-mentioned work (Toropov and Toropova, 2014) the statistical characteristics of the model for mutagenicity of fullerene (the same data on TA100) are the following: n=10, r^2=0.7549, s=7.7 (training set); n=5, r^2=0.8987, s=18.7 (calibration

206 set); and n=5, r^2=0.6968, s=10.9 (validation set). Consequently, the predictive potential of
207 models for splits 2 and 3 (Eqs. 5 and 6) slightly better or at least comparable with predictive
208 potential of model which has been suggested in the above work.
209 Thus, the checking up of the suggested approach (using the CORAL software,
210 http://www.insilico.eu/coral) with other endpoints related to substances with complex and
211 unclear molecular structure e.g. other nanomaterials and/or peptides is quite attractive task. It is
212 to be noted that the local QCAR models for small datasets, can be integrated in joint QCAR
213 model if corresponding experimental data will become available (Toropov and Toropova,
214 2015; Toropova et al., 2014; Toropova et al., 2015).
215 **4. Conclusions**
216 Due to a fact that large datasets are not available for nanomaterials one needs to develop
217 methodology that provides predictions based on small datasets. The QCAR approach developed
218 here provides reasonable good prediction for the mutagenicity (TA100 and
219 WP2uvrA/pKM101) of fullerene nanoparticels under different conditions.
220
221 **Acknowledgements**
222 We are gratitude to EC project NanoPUZZLES (Project Reference: 309837); EU FP7
223 project PreNanoTox **(**contract 309666); and EU project PROSIL funded under the LIFE
224 program (project LIFE12ENV/IT/000154). D.L. and J.L. acknowledge support from the
225 National Science Foundation (NSF/CREST HRD-0833178).
226
227
228
229 **References**
230 Burello, E., Worth, A.P., 2011. QSAR modeling of nanomaterials. Wiley Interdiscip. Rev.
231 Nanomed. Nanobiotechnol. 3, 298–306.
232 Fourches, D., Pu, D., Tassa, C., Weissleder, R., Shaw, S.Y., Mumper, R.J., Tropsha, A., 2010.
233 A quantitative nanostructure_activity relationship modeling. ACS nano 4, 5703-5712.
234 García, J., Duchowicz, P.R., Rozas, M.F., Caram, J.A., Mirífico, M.V., Fernández, F.M.,
235 Castro, E.A., 2010. A comparative QSAR on 1,2,5-thiadiazolidin-3-one 1,1-dioxide
236 compounds as selective inhibitors of human serine proteinases. J. Mol. Graph. Model. 31,
237 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., 2011. QSAR on aryl-piperazine derivatives with activity on malaria. Chemometr. Intell. Lab. Syst. 110, 81–88.

Leszczynski,J., 2010. Nano meets bio at the interface. Nat. Nanotech. 5,633-634.

OECD 2006-2010 http://www.oecd.org/env/ehs/nanosafety/47104296.pdf

Peric, B., Sierra, J., Martí, E., Cruañas, R., Garau, M.A., 2015. Quantitative structure–activity relationship (QSAR) prediction of (eco)toxicity of short aliphatic protic ionic liquids. Ecotoxicol. Environ. Saf. 115, 257-262

Richarz, A.N., Madden, J.C., Robinson, R.L.M., Lubiński, L., Mokshina, E., Urbaszek, P., Kuz′min, V.E., Puzyn, T., Cronin, M.T.D., 2015. Development of computational models for the prediction of the toxicity of nanomaterials. Perspect. Sci. 3, 27-29.

REACH 2011 http://ec.europa.eu/environment/chemicals/nanotech/pdf/report_ripon1.pdf

Salahinejad, M., Ghasemi, J.B., 2014. 3D-QSAR studies on the toxicity of substituted benzenes to Tetrahymena pyriformis: CoMFA, CoMSIA and VolSurf approaches, Ecotoxicol. Environ. Saf. 105, 128-134.

Salahinejad, M., 2015. Nano-QSPR modelling of carbon-based nanomaterials properties Curr. Top. Med. Chem. 15, 1868-1886.

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., 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., 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., 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.

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

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

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., 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., Benfenati, E., Puzyn, T., Leszczynska, D., Leszczynski, J., 2014. Optimal descriptor as a translator of eclectic information into the prediction of membrane damage: The case of a group of ZnO and TiO2 nanoparticles, Ecotoxicol. Environ. Saf. 108, 203-209.

Toropova, A.P., Toropov, A.A., Rallo, R., Leszczynska, D., Leszczynski J., 2015. Optimal descriptor as a translator of eclectic data into prediction of cytotoxicity for metaloxide nanoparticles under different conditions. Ecotoxicol. Environ. Saf. 112, 39–45.

Wang, C., Wei, Z., Wang, L., Sun, P., Wang, Z., 2015. Assessment of bromide-based ionic liquid toxicity toward aquatic organisms and QSAR analysis. Ecotoxicol. Environ. Saf. 115, 112-118.

Table 1

The list of conditions which have impact upon mutagenicty of fullerene C60 nanoparticles and which were utilized to build up models.

Conditions	**Symbols for quasi-SMILES**
Dark or Irradiation	"0" = Darkness "1" = Irradiation
Mix S9	"+" = with Mix S9 "- "= without Mix S9
Dose (g/plate)	"A" = 50 "B" = 100 "C" = 200 "D" = 400 "E" = 1000

300 Table 2

301 The numerical data on endpoints and dispersion of the experimental definition and the scheme

302 of construction of quasi-SMILES.

303

ID	Darkness or Irradiation	Mix S9	dose	TA100	WP2uvrA/pKM101
1	0	+	A	146±4	113±7
2	0	+	B	141±24	106±19
3	0	+	C	159±8	112±15
4	0	+	D	160±21	115±15
5	0	+	E	177±8	145±13
6	0	-	A	143±14	160±15
7	0	-	B	139±22	162±21
8	0	-	C	169±20	174±26
9	0	-	D	168±7	179±15
10	0	-	E	152±15	220±6
11	1	+	A	129±6	114±10
12	1	+	B	131±4	105±12
13	1	+	C	138±12	113±16
14	1	+	D	137±14	110±18
15	1	+	E	160±6	123±8
16	1	-	A	136±11	127±25
17	1	-	B	136±6	133±12
18	1	-	C	138±7	121±15
19	1	-	D	164±16	117±29
20	1	-	E	172±14	138±19

304

Table 3

Experimental and calculated with Eq. 4,5, and 6 values of the TA100 for fullerene nanoparticles impact under different conditions

ID	1*	2	3	Quasi-SMILES	Expriment	Eq.4	Eq.5	Eq.6	1*	2	3
1	V	V	V	0+A	146	132.8046	115.3775	143.2652	Y	Y	Y
2	T	C	V	0+B	141	145.8861	121.3559	144.0029	N	Y	Y
3	T	C	C	0+C	159	157.1709	136.5559	157.8729	N	Y	Y
4	V	C	V	0+D	160	162.0685	142.5034	161.6878	Y	Y	Y
5	T	V	T	0+E	177	165.3027	143.7145	165.5465	Y	N	Y
6	C	C	C	0-A	143	130.9643	134.7925	147.8862	Y	Y	Y
7	T	T	C	0-B	139	144.0458	140.7708	148.6238	N	Y	N
8	V	T	T	0-C	169	155.3307	155.9708	162.4939	N	Y	Y
9	T	V	C	0-D	168	160.2283	161.9183	166.3087	Y	Y	Y
10	T	T	T	0-E	152	163.4625	163.1294	170.1675	Y	N	N
11	C	V	T	1+A	129	116.4477	113.3044	130.1540	Y	Y	Y
12	T	C	T	1+B	131	129.5292	119.2827	130.8917	N	Y	Y
13	V	T	T	1+C	138	140.8141	134.4827	144.7618	N	Y	Y
14	T	T	T	1+D	137	145.7117	140.4303	148.5766	Y	Y	Y
15	C	V	T	1+E	160	148.9459	141.6413	152.4354	Y	N	N
16	V	T	T	1-A	136	114.6075	132.7193	134.7750	Y	Y	Y
17	T	T	V	1-B	136	127.6890	138.6977	135.5127	N	Y	Y
18	T	T	C	1-C	138	138.9739	153.8977	149.3827	N	Y	Y
19	C	T	T	1-D	164	143.8715	159.8452	153.1976	Y	Y	Y
20	C	T	V	1-E	172	147.1057	161.0562	157.0563	Y	N	N

*) Split 1, 2, and 3; T=training set; C=calibration set; and V=validation set; Y=quasi-SMILES falls into Domain of applicability (otherwise "N").

Table 4

Correlation weights for symbols which represent conditions of impact of fullerene C60 nanoparticles (Eq. 4,5, and 6 for TA100)

Symbols of quasi-SMILES, S_k	***CW*(S_k) *run 1***	***CW*(S_k) *run 2***	***CW*(S_k) *run 3***
+	1.06698	0.47336	0.19086
-	0.99597	1.77789	0.56606
0	1.37861	0.96051	1.93566
1	0.74747	0.82121	0.87109
A	0.0	0.0	-0.05990
B	0.50476	0.40170	0.0
C	0.94020	1.42303	1.12619
D	1.12918	1.82266	1.43594
E	1.25398	1.90403	1.74925

Table 5

An example of calculation of optimal descriptor with correlation weights for fullerene C60 nanoparticle represented by code “0+B” ; DCW(2,2) = 2.9504;

TA100 = 69.4243 + 25.9161*2.95036 = 145.8861

S_k	$CW(S_k)$
0	1.3786
+	1.0670
B	0.5048

Table 6

Experimental and calculated with Eq. 7,8, and 9 values of the WP2uvrA/pKM101 for fullerene nanoparticles impact under different conditions

ID	1*	2	3	Quasi-SMILES	Expriment	Eq.7	Eq.8	Eq.9	1*	2	3
1	T	C	T	0+A	113	118.9590	95.2449	133.2322	Y	Y	Y
2	T	V	C	0+B	106	118.9590	127.8341	126.7746	Y	N	Y
3	V	T	V	0+C	112	118.9590	127.9124	126.7746	Y	Y	Y
4	T	T	C	0+D	115	118.9590	95.2449	126.7746	Y	Y	Y
5	T	C	T	0+E	145	152.9472	169.4336	144.8871	Y	Y	Y
6	C	T	T	0-A	160	159.2997	136.8401	162.2816	Y	Y	Y
7	V	T	C	0-B	162	159.2997	169.4294	155.8240	Y	N	Y
8	C	V	C	0-C	174	159.2997	169.5076	155.8240	Y	Y	Y
9	V	V	T	0-D	179	159.2997	136.8401	155.8240	Y	Y	Y
10	T	T	V	0-E	220	193.2879	211.0289	173.9365	Y	Y	Y
11	V	C	T	1+A	114	84.2194	53.6913	104.3638	Y	Y	Y
12	C	V	V	1+B	105	84.2194	86.2806	97.9062	Y	N	Y
13	V	C	V	1+C	113	84.2194	86.3588	97.9062	Y	Y	Y
14	T	V	C	1+D	110	84.2194	53.6913	97.9062	Y	Y	Y
15	C	T	T	1+E	123	118.2076	127.8801	116.0187	Y	Y	Y
16	T	V	T	1-A	127	124.5601	95.2866	133.4132	Y	Y	Y
17	C	T	T	1-B	133	124.5601	127.8759	126.9556	Y	N	Y
18	T	T	V	1-C	121	124.5601	127.9541	126.9556	Y	Y	Y
19	C	C	T	1-D	117	124.5601	95.2866	126.9556	Y	Y	Y
20	T	C	T	1-E	138	158.5483	169.4754	145.0681	Y	Y	Y

*) Split 1, 2, and 3; T=training set; C=calibration set; and V=validation set; Y=quasi-SMILES falls into Domain of applicability (otherwise "N").

335

336 Table 7

337 Correlation weights for symbols which represent conditions of impact of fullerene C60

338 nanoparticles (Eq. 5, for WP2uvrA/pKM101)

Symbols of quasi-SMILES, S_k	***CW*(S_k) *run 1***	***CW*(S_k) *run2***	***CW*(S_k) *run 3***
+	0.50000	0.69675	0.39982
-	0.94039	1.60115	2.20288
0	1.12778	1.60305	2.19630
1	0.74854	0.69727	0.40447
A	0.0	0.0	0.40082
B	0.0	0.99558	0.0
C	0.0	0.70300	0.0
D	0.0	0.0	0.0
E	0.37104	1.59979	1.12422

339