Chemosphere 92 (2013) 31–37

Contents lists available at SciVerse ScienceDirect

Chemosphere

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

QSAR as a random event: Modeling of nanoparticles uptake in PaCa2 cancer cells

Andrey A. Toropov [a,*], Alla P. Toropova [a], Tomasz Puzyn [b], Emilio Benfenati [a], Giuseppina Gini [c], Danuta Leszczynska [d], Jerzy Leszczynski [e]

[a] Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via La Masa 19, 20156 Milano, Italy
[b] Faculty of Chemistry, Laboratory of Environmental Chemometrics, University of Gdansk, ul. Sobieskiego 18/19, Gdansk 80-952, Poland
[c] Department of Electronics and Information, Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 Milano, Italy
[d] Interdisciplinary Nanotoxicity Center, Department of Civil and Environmental Engineering, Jackson State University, 1325 Lynch St., Jackson, MS 39217-0510, USA
[e] Interdisciplinary Nanotoxicity Center, Department of Chemistry and Biochemistry, Jackson State University, 1400 J. R. Lynch Street, P.O. Box 17910, Jackson, MS 39217, USA

HIGHLIGHTS

- QSAR analysis of cellular uptake in PaCa2 cancer cells of nanoparticles is carried out.
- The concept of QSAR as a random event is suggested.
- Five distributions 109 nanoparticles into the training and test sets are studied.

ARTICLE INFO

Article history:
Received 10 January 2013
Received in revised form 21 February 2013
Accepted 6 March 2013
Available online 6 April 2013

Keywords:
Nanoparticle
QSAR
CORAL software
Optimal descriptor

ABSTRACTS

Quantitative structure–property/activity relationships (QSPRs/QSARs) are a tool to predict various endpoints for various substances. The "classic" QSPR/QSAR analysis is based on the representation of the molecular structure by the molecular graph. However, simplified molecular input-line entry system (SMILES) gradually becomes most popular representation of the molecular structure in the databases available on the Internet. Under such circumstances, the development of molecular descriptors calculated directly from SMILES becomes attractive alternative to "classic" descriptors. The CORAL software (http://www.insilico.eu/coral) is provider of SMILES-based optimal molecular descriptors which are aimed to correlate with various endpoints. We analyzed data set on nanoparticles uptake in PaCa2 pancreatic cancer cells. The data set includes 109 nanoparticles with the same core but different surface modifiers (small organic molecules). The concept of a QSAR as a random event is suggested in opposition to "classic" QSARs which are based on the only one distribution of available data into the training and the validation sets. In other words, five random splits into the "visible" training set and the "invisible" validation set were examined. The SMILES-based optimal descriptors (obtained by the Monte Carlo technique) for these splits are calculated with the CORAL software. The statistical quality of all these models is good.

1. Introduction

During the last 20 years, there has been considerable increase of interest in nanostructures. It obviously facilitates surge of new directions in the basic research and generates many novel experimental projects. New materials have been developed, tested, and fast forwarded into production lines. Nanomanufacturing becomes a substantial part of the 21st Century industry. However, this also might create adverse effects. In particular, some of nanomaterials can be harmful to the environment and humans.

Quantitative structure–property/activity relationships (QSPR/QSAR) are a tool for prediction of various endpoints (García et al., 2011; Garro Martinez et al., 2011; Ojha et al., 2011; Mullen et al., 2011; Ibezim et al., 2012) using molecular descriptors (Furtula and Gutman, 2011; Afantitis et al., 2011) calculated with molecular graph (Toropov and Roy, 2004; Castillo-Garit et al., 2007), quantum-chemical descriptors (Petrova et al., 2011) as well as with simplified molecular input-line entry system (SMILES) (Toropov et al., 2008, 2012).

The understanding and predicting of the biological effects of the manufactured nanoparticles represents an important task of modern natural sciences. The experimental analysis of these substances is expensive. Theoretical investigations of such phenomena could provide an efficient approach to evaluation of nano-bio interactions (Leszczynski, 2010). Consequently, the development of the

* 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.2013.03.012

QSPR/QSAR for nanoparticles is useful from point of view the both praxis and theory. The recent review provides discussion of such challenges in developing Nano-QSAR methods (Puzyn et al., 2009).

The aim of the present study is the evaluation of the CORAL software as a possible tool of the QSAR analysis for cellular uptake of nanoparticles in PaCa2 cancer cells. This study was carried out for the nanoparticles involving the same metal core but various surfaces modified by different small molecules.

2. Method

2.1. Data

We have examined 109 nanoparticles. They have the same nano-core, but various surfaces modifiers (small organic molecules). The cellular uptake in PaCa2 cancer cells of above-mentioned nanoparticles was studied. The selected endpoint (cellular uptake) is defined as minus decimal logarithm of the concentration (pM) of nanoparticles per cell (Fourches et al., 2010). The data for these 109 nanoparticles were randomly split into the sub-training, calibration, test, and validation sets. The roles of these sets are different: sub-training set is the "developer" of the model since correlation weights of compounds from the set are used to build up the model; calibration set is the "critic" of the model since data from this set are used to check whether model is working for compounds which are absent in the sub-training set; the test set is "estimator" of the model in cases of various threshold values; finally, the "invisible" validation set is used for the final estimation of the model with threshold value which gives the best statistical quality for the test set. These splits are random and various. In fact, the sub-training set, the calibration set, and test set are "visible" during of building up the model; hence, they should be qualified as a structured training set. The validation set is "invisible" during building up a model; hence, statistical characteristics of this set are true measure of predictability of this approach. The distribution into the structured training set and validation set as rule has apparent influence upon statistical characteristics of the model. Consequently, one should examine several distributions into the training and validation in order to obtain realistic estimation of ability of this approach. The QSAR for each distribution is a random event and statistical characteristics of each model are random coefficients. We deem that average values of these coefficients (e.g. correlation coefficient, root-mean-square error, mean absolute error (MAE), and others) as well as their ranges should be qualified as more robust information than data on these coefficients for only one distribution into the training and validation sets.

2.2. Descriptors

We have formulated the following principles of building up a model of an endpoint with the CORAL software (Benfenati et al., 2008):

(a) Molecular structure of each compound can be represented by molecular features which are extracted from SMILES.
(b) There are local and global molecular features which can be extracted from SMILES. The local features are some fragments (or individual atoms). The global features are some indices which characterized molecules in whole.
(c) Building up of QSPR/QSAR model for an arbitrary split into the training and validation sets can be examined as a random event.
(d) The statistical quality of each QSPR/QSAR model is a mathematical function of split into the sub-training, calibration, test, and validation sets.
(e) The average statistical quality of QSPR/QSAR models that is obtained for several splits into training and test sets is more robust criterion for the estimation of an approach than statistical quality for solely one split.
(f) The average statistical quality of a model for external "invisible" validation sets is more significant data than the average statistical quality for "visible" (i.e. substances involved in building up model) sub-training and calibration sets.

The SMILES-based optimal descriptors are calculated as the following:

$$\mathrm{DCW}(\mathrm{Threshold},N_{\mathrm{epoch}}) = \sum \mathrm{CW}(S_k) + \sum \mathrm{CW}(SS_k) + \mathrm{CW}(\mathrm{NOSP}) \quad (1)$$

where S_k, SS_k are local SMILES attributes which are extracted from the SMILES; the extraction of S_k and SS_k can be represented as (Toropov et al., 2012):

$$\text{"ABCDE"} : \text{"A"}, \text{"B"}, \text{"C"}, \text{"D"}, \text{"E"}\,(S_k);$$

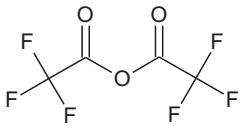
$$\text{"ABCDE"} : \text{"AB"}, \text{"BC"}, \text{"CD"}, \text{"DE"}\,(SS_k); \quad (2)$$

NOSP are global (Toropov and Toropova, 2002, 2003) SMILES attributes which are extracted from the SMILES (Benfenati et al., 2008; Toropov et al., 2011; Toropova et al., 2011a,b). The NOSP is an indicator of presence (absence) of four chemical elements: nitrogen, oxygen, sulfur, and phosphorus (Toropov et al., 2011). Table 1 contains the definition of the NOSP according to presence

Table 1
Calculation of the NOSP index according to presence (absence) of nitrogen (N), oxygen (O), sulfur (S), and phosphorus (P).

N	O	S	P	Comments	Representation for the CORAL calculations
0	0	0	0	Nitrogen, oxygen, sulfur, and phosphorus are absent	NOSP00000000
0	0	0	1	The molecule only contains phosphorus	NOSP00010000
0	0	1	0	The molecule only contains sulfur	NOSP00100000
0	0	1	1	The molecule contains sulfur and phosphorus	NOSP00110000
0	1	0	0	The molecule only contains oxygen	NOSP01000000
0	1	0	1	The molecule contains oxygen and phosphorus	NOSP01010000
0	1	1	0	The molecule contains oxygen and sulfur	NOSP01100000
0	1	1	1	The molecule contains oxygen, sulfur, and phosphorus	NOSP01110000
1	0	0	0	The molecule only contains nitrogen	NOSP10000000
1	0	0	1	The molecule contains nitrogen and phosphorus	NOSP10010000
1	0	1	0	The molecule contains nitrogen and sulfur	NOSP10100000
1	0	1	1	The molecule contains nitrogen, sulfur, and phosphorus	NOSP10110000
1	1	0	0	The molecule contains nitrogen and oxygen	NOSP11000000
1	1	0	1	The molecule contains nitrogen, oxygen and phosphorus	NOSP11010000
1	1	1	0	The molecule contains nitrogen, oxygen, and sulfur	NOSP11100000
1	1	1	1	The molecule contains nitrogen, oxygen, sulfur, and phosphorus	NOSP11110000

Table 2
Example of calculation DCW(2,48) with Eq. (1) in the case of split 1.

Structure				
SMILES	FC(F)(F)C(=O)OC(=O)C(F)(F)F			
DCW(2,48)	11.64175			
Structural attribute (SA)	Correlation weight of SA, CW (SA)	The number of SA in sub-training set	The number of SA in calibration set	The number of SA in test set
S_k				
F.....	1.0000	2	2	0
C.....	−1.2158	34	42	18
(.....	−1.8085	29	39	15
F.....	1.0000	2	2	0
(.....	−1.8085	29	39	15
(.....	−1.8085	29	39	15
F.....	1.0000	2	2	0
(.....	−1.8085	29	39	15
C.....	−1.2158	34	42	18
(.....	−1.8085	29	39	15
=.....	0.2138	23	33	14
O.....	−0.4728	25	34	14
(.....	−1.8085	29	39	15
O.....	−0.4728	25	34	14
C.....	−1.2158	34	42	18
(.....	−1.8085	29	39	15
=.....	0.2138	23	33	14
O.....	−0.4728	25	34	14
(.....	−1.8085	29	39	15
C.....	−1.2158	34	42	18
(.....	−1.8085	29	39	15
F.....	1.0000	2	2	0
(.....	−1.8085	29	39	15
(.....	−1.8085	29	39	15
F.....	1.0000	2	2	0
(.....	−1.8085	29	39	15
F.....	1.0000	2	2	0
SS_k				
F...C.....	1.8710	2	0	0
C...(.....	1.7550	28	37	15
F...(.....	0.8428	2	2	0
F...(.....	0.8428	2	2	0
(...(.....	1.1230	3	2	0
F...(.....	0.8428	2	2	0
F...(.....	0.8428	2	2	0
C...(.....	1.7550	28	37	15
C...(.....	1.7550	28	37	15
=...(.....	2.0968	23	33	14
O...=.....	0.6240	23	33	14
O...(.....	0.9355	25	34	14
O...(.....	0.9355	25	34	14
O...C.....	1.0888	19	25	8
C...(.....	1.7550	28	37	15
=...(.....	2.0968	23	33	14
O...=.....	0.6240	23	33	14
O...(.....	0.9355	25	34	14
C...(.....	1.7550	28	37	15
C...(.....	1.7550	28	37	15
F...(.....	0.8428	2	2	0
F...(.....	0.8428	2	2	0
(...(.....	1.1230	3	2	0
F...(.....	0.8428	2	2	0
F...(.....	0.8428	2	2	0
F...(.....	0.8428	2	2	0
NOSP				
NOSP01000000	1.6290	17	22	7

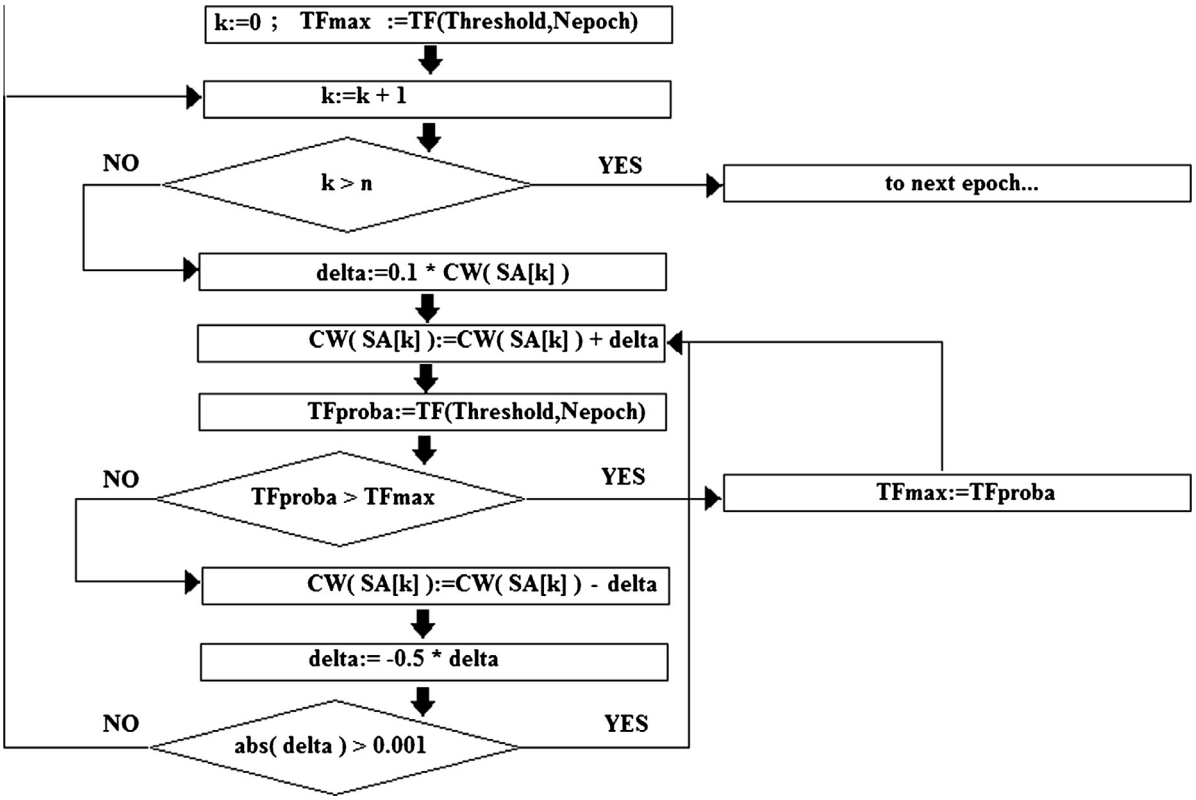


Fig. 1. The flowchart of one epoch of the Monte Carlo optimization of correlation weights (n is the number of correlation weights involved in building up model).

(absence) of nitrogen, oxygen, sulfur, and phosphorus. Table 2 contains an example of the SMILES attributes with their correlation weights which were obtained in the case of the split 1.

The threshold and Nepoch are parameters of the Monte Carlo optimization. The threshold is a tool to define two classes of molecular features: rare (noise) and not rare, i.e. active. The optimal descriptors are calculated with the correlation weights of active molecular features (attributes). Correlation weights for rare attributes are fixed equal to zero, i.e. these are not involved in building up model. For example, the threshold is defined as three. In this case, all SMILES attributes (S_k, SS_k, NOSP) which have presence lesser than in four SMILES from the sub-training set will be classified as rare.

The N_{epoch} is the number of epochs of the Monte Carlo optimization. The target function (TF) of the optimization is defined as the following:

$$TF = R + R' - W_R \cdot |R - R'| - W_C \cdot (|C_0| + |C'_0| + |C_1 - C'_1|) \quad (3)$$

where R and R' are correlation coefficient between the optimal descriptor and an endpoint (EP) for sub-training and calibration sets, respectively; C_0, C_1, C'_0, and C'_1 are coefficients from equations obtained by the Least squares method:

$$EP = C_0 + C_1 \cdot DCW(Threshold, N_{epoch}) \quad \text{for sub-training set;} \quad (4)$$

$$EP = C'_0 + C'_1 \cdot DCW(Threshold, N_{epoch}) \quad \text{for calibration set;} \quad (5)$$

$W_R = 0.1$ and $W_C = 0.01$ are empirical parameters. Fig. 1 shows the flowchart of one epoch of the Monte Carlo optimization.

The increase in the threshold leads to decrease of correlation coefficient (between experimental and calculated values of endpoint) for the sub-training and calibration sets, but as the rule, there is a maximum of the correlation coefficient for the test set. The increase in the number of epochs of the Monte Carlo optimization leads to increase in the correlation coefficient for sub-training and calibration sets, but again, as the rule, there is the maximum of the correlation coefficient for the test set. Thus, it is necessary to define preferable values of the threshold (T^*) and the number of epochs (N^*) which are providing maximum of correlation coefficient for the test set (Fig. 2). Finally, the model should be checked up with data on the validation set (substances which were not involved in the modeling process). However, even this checking up can be improved if these actions will be carried out for several distributions into the training and validation sets.

3. Results and discussion

Table 3 contains the statistical quality of models which were built up with the CORAL software. The data were obtained according to the above-mentioned scheme (Fig. 2). Fig. 3 shows the models for five random splits graphically. One can see (Table 3) that all these models are statistically satisfactory, but each model contains good predictions i.e. dots near diagonal together with poor predictions i.e. groups of dots remote from the diagonal (Fig. 3).

If one carried out several runs of the Monte Carlo optimization, the molecular features of three kinds will be obtained. The first, molecular features with positive correlation weights for all runs (these are promoters of endpoint increase). The second, molecular features with negative correlation weights for all runs (these are promoters of endpoint decrease). The third, molecular features with mixed correlation weights: there are both positive and negative values of the correlation weights in several runs of the Monte Carlo optimization. The role of these features is undefined. We have detected that for all five splits: (i) branching of carbon skeleton and presence of oxygen atoms are promoters of pM increase; and (ii) –N–C– fragments and branching of the aromatic carbon skeleton are promoters of pM decay. Lists of stable promoters of

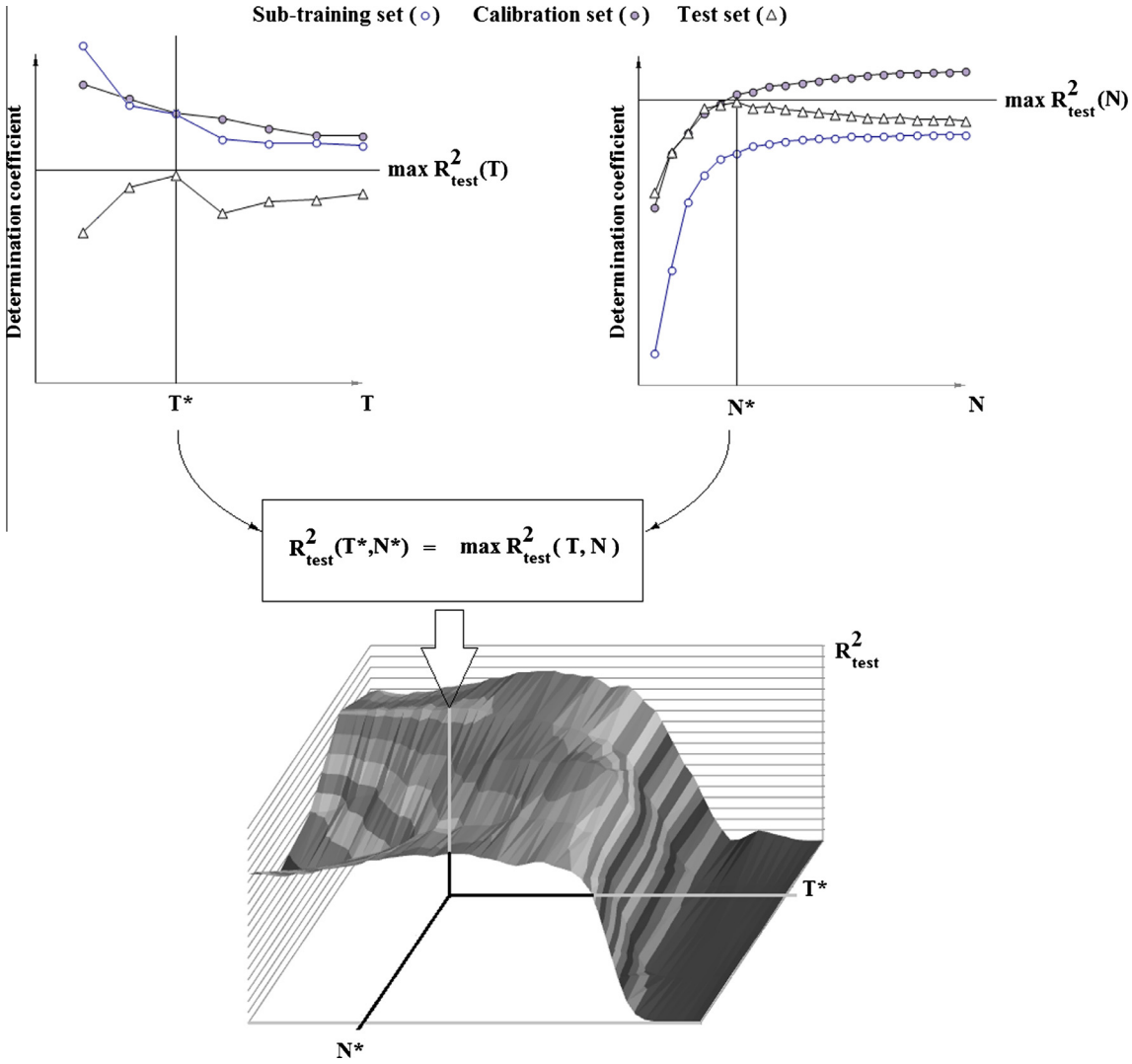


Fig. 2. Scheme of definition of the preferable CORAL model. *T* is threshold; *N* is the number of epochs of the Monte Carlo optimization; T^* and N^* are values which give maximum for the correlation coefficient between experimental and calculated endpoint values for the test set.

Table 3
The statistical characteristics for QSAR models of pM calculated with T^* and N^* values (Fig. 2).

			Training										Validation		
			Sub-training set			Calibration set			Test set				Validation set		
Split	T^*	N^*	*n*	r^2	MAE	*n*	r^2	MAE	*n*	r^2	R^2_m	MAE	*n*	r^2	MAE
1	2	20	34	0.6913	0.186	42	0.6886	0.259	18	0.9280	0.7011	0.149	15	0.9341	0.129
2	2	28	32	0.6972	0.185	37	0.6444	0.300	23	0.8431	0.8360	0.133	17	0.8043	0.143
3	1	26	40	0.7287	0.173	34	0.9032	0.228	15	0.8379	0.7358	0.160	20	0.8723	0.148
4	1	25	37	0.7557	0.173	38	0.7555	0.204	15	0.7642	0.6275	0.150	19	0.8232	0.112
5	3	15	32	0.6504	0.215	39	0.6389	0.237	21	0.8859	0.8683	0.097	18	0.8429	0.153

T^* and N^* are preferable threshold and the number of epochs of the Monte Carlo method optimization (i.e. values of threshold and number of epochs which give best statistics for the test set); *n*, r^2, and MAE are the number of compounds in the set, the square correlation coefficient, and mean absolute error, respectively; R^2_m is metric of predictability (Roy et al., 2012; Ojha et al., 2011), $R^2_m = r^2 \times (1 - \sqrt{r^2 - r^2_0})$: a model has predictability if $R^2_m > 0.5$. (Additional information on R^2_m is available via link http://aptsoftware.co.in/rmsquare/).

increase or decrease for the cellular uptake in PaCa2 cancer cells can be used to search for the mechanistic interpretation of the model. E.g. the presence in nanoparticle (with the same core) of modifiers which contain the majority of stable promoters of increase for the endpoint can be interpreted as high probability that this nanoparticle will be characterized by high value of the cellular uptake in PaCa2 cancer cells.

The domain of applicability for CORAL models can be defined as nanoparticles (with the same core) which do not contain SMILES attributes absent in the sub-training set.

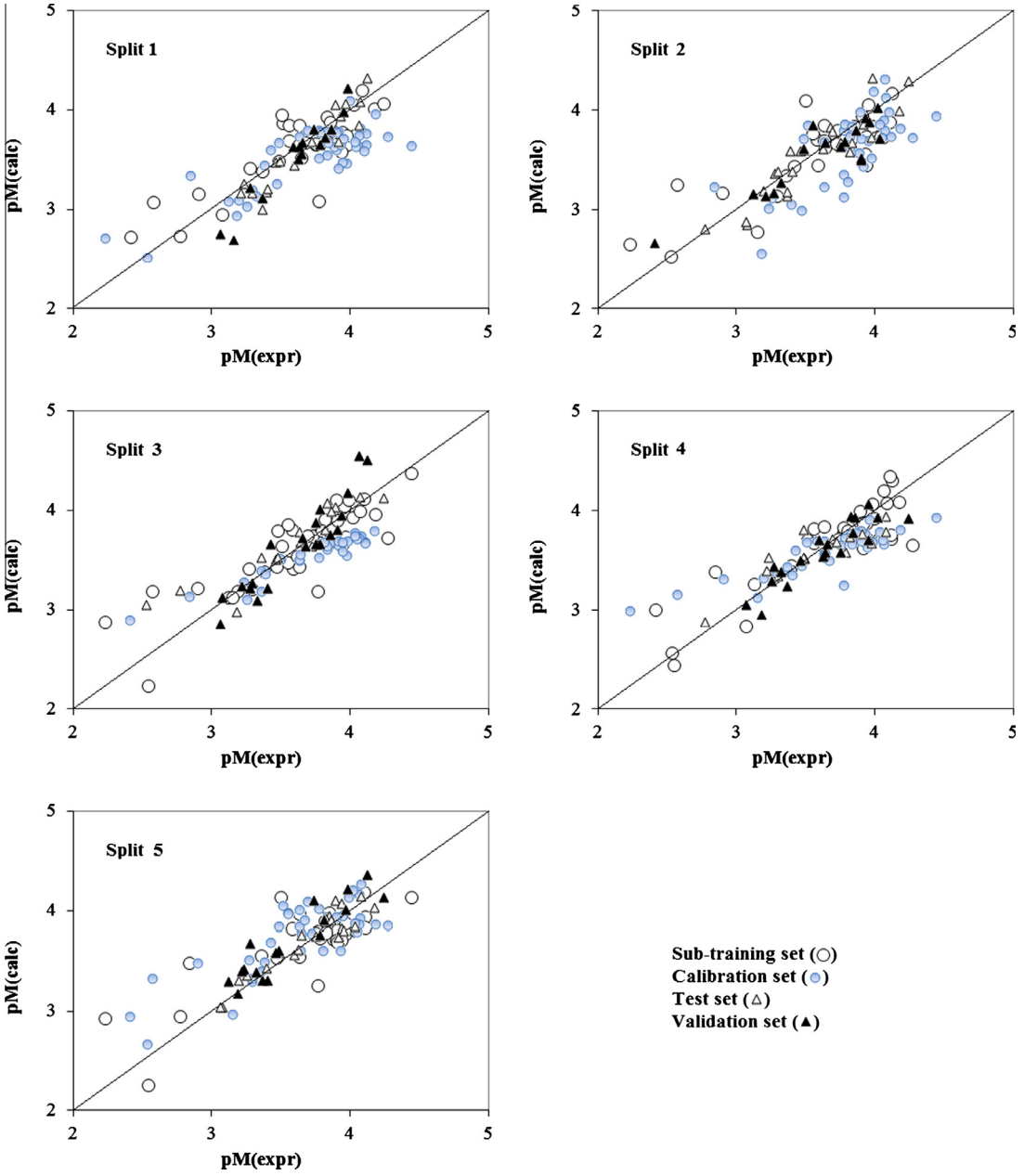


Fig. 3. Models of the cellular uptake in PaCa2 cancer cells for 109 nanoparticles which were obtained for five various splits into the training (sub-training, calibration, and test sets) and validation sets.

The reproducibility of the satisfactory statistical quality of a model for external validation set in several runs of the Monte Carlo method optimization should be considered as the measure of reliability of a CORAL model. However, even this criterion should be checked up for several splits into the sub-training, calibration, test, and validation sets.

Thus, the CORAL software provides models according to OECD principles (Gramatica, 2007).

The statistical characteristics of models described in the literature (Fourches et al., 2010) where four model with 87 and 22 substances in training and test sets, respectively, and one model with 88 substances in the training set and 21 substance in the test set are characterized by r^2 values from 0.67 to 0.90 and MAE ranging from 0.13 to 0.21 log units. In other words, statistical characteristics of the CORAL models are similar to the statistics of the above-mentioned models. However, the CORAL models were derived without inclusion of the data on van der Waals surface and lipophilicity (Fourches et al., 2010).

4. Conclusions

The concept of QSAR as a random event is suggested as the alternative to build up QSAR with sole distribution of available data into the subset of the training and subset of validation. The CORAL software gives satisfactory and stable predictions of the cellular uptake of nanoparticles in PaCa2 cancer cells for five random splits (i.e. for five described random events).

Supplementary materials

(i) SMILES of examined substances together with the numerical data on the endpoint; (ii) details of the five splits into sub-training,

calibration, test, and validation sets; and (iii) correlation weights (for three runs of the Monte Carlo optimization) of the most significant promoters of increase and decay of the endpoint for five splits which are examined in this work. It is to be noted that aforementioned data represented in the Supplementary materials section give possibility to reproduce the described QSAR models using the CORAL software available on the Internet (http://www.insilico.eu/coral).

Acknowledgements

We thank the EC project NANOPUZZLES (Project Reference: 309837) and the National Science Foundation (NSF/CREST HRD-0833178, and EPSCoR Award #:362492-190200-01/NSFEPS-090378) for financial support.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.chemosphere.2013.03.012.

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.

Benfenati, E., Toropov, A.A., Toropova, A.P., Monganaro, A., Gonela Diaza, R., 2008. IRFMN, Milan. <http://www.insilico.eu/coral> (Reference Manual).

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 Modell. 26, 32–47.

Fourches, D., Pu, D., Tassa, C., Weissleder, R., Shaw, S.Y., Mumper, R.J., Tropsha, 2010. A quantitative nanostructure–activity relationship modelling. ACS Nano 4, 5703–5712.

Furtula, B., Gutman, I., 2011. Relation between second and third geometric–arithmetic indices of trees. J. Chem. 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 Modell. 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.

Gramatica, P., 2007. Principles of QSAR models validation: internal and external. QSAR Comb. Sci. 26, 694–701.

Ibezim, E., Duchowicz, P.R., Ortiz, E.V., Castro, E.A., 2012. QSAR on arylpiperazine derivatives with activity on malaria. Chemometr. Intell. Lab. Syst. 110, 81–88.

Leszczynski, J., 2010. Nano meets bio at the interface. Nat. Nanotechnol. 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. Chemometr. Intell. Lab. Syst. 107, 269–275.

Ojha, P.K., Mitra, I., Das, R.N., Roy, K., 2011. Further exploring rm 2 metrics for validation of QSPR models. Chemometr. Intell. Lab. Syst. 107, 194–205.

Petrova, T., Rasulev, B.F., Toropov, A.A., Leszczynska, D., Leszczynski, J., 2011. Improved model for fullerene C60 solubility in organic solvents based on quantum-chemical and topological descriptors. J. Nanopart. Res. 13, 3235–3247.

Puzyn, T., Leszczynska, D., Leszczynski, J., 2009. Towards the development of "Nano-QSARs": advances and challenges. Small 5, 2494–2509.

Roy, K., Mitra, I., Kar, S., Ojha, P.K., Das, R.N., Kabir, H., 2012. Comparative studies on some metrics for external validation of QSPR models. J. Chem. Inf. Model. 52, 396–408.

Toropov, A.A., Roy, K., 2004. QSPR modelling 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., Rasulev, B.F., Leszczynski, J., 2008. QSAR modelling of acute toxicity by balance of correlations. Bioorg. Med. Chem. 16, 5999–6008.

Toropov, A.A., Toropova, A.P., 2003. QSPR modelling of alkanes properties based on graph of atomic orbitals. J. Mol. Struct. (THEOCHEM) 637, 1–10.

Toropov, A.A., Toropova, A.P., 2002. QSAR modelling of toxicity on optimization of correlation weights of Morgan extended connectivity. J. Mol. Struct. (THEOCHEM) 578, 129–134.

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. Chemometr. Intell. Lab. Syst. 105, 215–219.

Toropov, A.A., Toropova, A.P., Benfenati, E., Gini, G., Puzyn, T., Leszczynska, D., Leszczynski, J., 2012. Novel application of the CORAL software to model cytotoxicity of metal oxide nanoparticles to bacteria Escherichia coli. Chemosphere 89, 1098–1102.