INSILI 1 (2025) 100105

Contents lists available at ScienceDirect

In Silico Research in Biomedicine

journal homepage: www.sciencedirect.com/journal/in-silico-research-in-biomedicine

Chaos, gambling, simplicity: simulation of nanoparticle delivery efficiency to organs of mice

Alla P. Toropova [a,*], Andrey A. Toropov [a], Ivan Raska Jr [b], Maria Raskova [b], Emilio Benfenati [a], Danuta Leszczynska [c], Jerzy Leszczynski [d]

[a] Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Science, Istituto di Ricerche Pharmacologic Mario Negri IRCCS, Via Mario Negri 2, 20156 Milano, Italy
[b] 3rd Medical Department, 1st Faculty of Medicine, Charles University in Prague, Prague 2, Czech Republic
[c] Department of Civil and Environmental Engineering, Jackson State University, 132 Lynch St, Jackson, MS 39217-0510, USA
[d] Department of Chemistry, Physics and Atmospheric Sciences, Jackson State University, 1400 J. R. Lynch Street, PO Box 17910, Jackson, MS 39217, USA

ARTICLE INFO

Keywords:
CORAL software
Las Vegas algorithm
Monte Carlo method
Nanoparticle
Quasi-SMILES
Simulation

ABSTRACT

Nanoparticles can be designed for targeted delivery to treat cancer. However, low efficiency of delivery to the tumor site is the rule rather than the exception. Understanding the influence of experimental properties on the distribution of nanoparticles to the target tumor site can help improve the design of drugs. Computational methods can assist in selecting conditions for the efficient delivery of nanoparticles to organs in the body. Analogies are an important element of research findings. The process of calculating the so-called correlation weights for building models of different endpoints is similar to the questionnaires used in sociological surveys. However, unlike sociological surveys, this modeling processes poll quasi-SMILES, representing different nanoparticles observed under various experimental conditions. The models obtained using the above questionnaires have poor statistical quality in the training sets, but satisfactory statistical quality in the calibration and validation sets. There is reproducibility of the results. However, reproducibility is observed only for special splits of data obtained by the Las Vegas algorithm. These splits are not identical. The above computer experiments demonstrated the potential for predicting the efficiency of nanoparticle delivery to tumors and organs in mice. Most likely, this is one of the possible ways of modeling the processes of delivery of therapeutic agents to various organs of living organisms, which can become the basis for new methods of treating cancer.

Introduction

There are various novel approaches to make drug deliveries more efficient. Among such methods, applications of nanoparticles have been gaining importance in recent years. Cancer treatment is a vital area of medicine, and drug delivery by nanoparticles is a promising approach to improving cancer therapy. Investigations of applications of nanoparticles in cancer diagnostics and treatment have increased dramatically in recent years [1,2]. Research is underway on the use of nanoparticles in the treatment of tumors in the liver [3], lung [4], kidney [5], and other organs.

Simulation the biochemical properties of nanomaterials, taking into account experimental conditions, is an attractive alternative compared to developing models based solely on the structure of nanomaterials. For this purpose, artificial neural networks are used [6]; groups of antibacterial profiles of nanoparticles are studied under various experimental conditions [7]; research is being conducted aimed at developing models of the cytotoxicity of groups of various nanoparticles (silicon, nickel, nickel oxide and others) [8]; classification models of nanoparticles are being developed using Random Forest technique [9]; data fusion algorithms are being developed using discriminant analysis [10]. Nevertheless, despite the large number of computational approaches to the simulation of nanomaterials developed in recent times [6–10], the simulation of the biochemical behavior of nanoparticles under various experimental conditions remains a quite actual task of modern natural science.

Simulation of physicochemical and biochemical behavior of nanoparticles represents an important task for their medical application.

* Corresponding author at: Laboratory of Environmental Chemistry and Toxicology, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via Mario Negri 2, 20156 Milano, Italy.
E-mail address: alla.toropova@marionegri.it (A.P. Toropova).

https://doi.org/10.1016/j.insi.2025.100105
Received 28 August 2025; Received in revised form 29 September 2025; Accepted 13 October 2025
Available online 14 October 2025

Table 1
Codes for the representation of experimental conditions of measurement of endpoints related to nanoparticles.

Categorical questions (circumstances)	Comment	Code
Type	Inorganic nanomaterial	[INM]
	Organic nanomaterial	[ONM]
Materials of nanoparticles (MAT)	Dendrimer	[Dendr]
	Gold	[Gold]
	Hybrid Hydrogel	[Hydro]
	Iron	[Iron]
	Liposome	[Lipos]
	Polymer	[Polum]
	Silica	[Silic]
Targeting strategy	Active	[Activ]
	Passive	[Passi]
Shape	Spherical	[Spher]
	Rod	[Rod]
	Plate	[Plate]
	Others	[Other]
Tumour model	Allograft Heterotopic	[AH]
	Allograft Orthotopic	[AO]
	Xenograft Heterotopic	[XH]
	Xenograft, Orthotopic	[XO]
Cancer type	Brain	[Brain]
	Breast	[Breas]
	Cervix	[Cervi]
	Colon	[Colon]
	Glioma	[Gliom]
	Liver	[Liver]
	Lung	[Lung]
	Others	[Other]
	Ovary	[Ovary]
	Pancreas	[Pancr]
	Prostate	[Prost]
	Sarcoma	[Sarco]
	Skin	[Skin]

Table 2
Codes for the representation of continuous characteristics of measurement of endpoints related to nanoparticles.

Continuous characteristics	Comment	Code
Size	lg Hydrodynamic diameter (nm)	[S0,43]..[S3,08]
Zeta potential	Zeta potential in mV	[Z-65,12]..[Z1.3]
Administration dose	Dose (mg/kg) $\times 10^{-9}$	[A1,35]..[A1292]

Table 3
Delivery efficiency of nanoparticles, percentage of injection dose (%ID) for mice.

Object exposed to nanoparticles	Range of experimental measurement	Denoted
Tumour	0.0001 – 43.52	DE1
Heart	0.0001 – 18.95	DE2
Liver	0.008 – 427.68	DE3
Lung	0.0002 – 41.06	DE4
Spleen	0.0009 – 302.30	DE5
Kidney	0.0001 – 51.74	DE6

Related research using various computational techniques has been performed by various groups to assist the medical enterprise [11–15].

In this work, a study is carried out to develop a method that can be called "questionnaire of experimental conditions". We applied here quasi-SMILES approach, using a special code describing experimental conditions for the determination of the properties and biochemical potential of nanoparticles using the CORAL software (http://www.insilico.eu/coral).

Quasi-SMILES is an extension of the classical SMILES (Simplified Molecular Input-Line Entry System) [16] method. SMILES can be used as a basis for representing molecular structure when developing quantitative structure-property/activity relationships (QSPR/QSAR) for different molecular systems and various endpoints [17–23]. The use of quasi-SMILES instead of traditional SMILES may also be a source of QSPR/QSAR approach [24–28]. Quasi-SMILES variants are possible that contain fragments of traditional SMILES. However, in this study, quasi-SMILES are used that encode only experimental conditions, without fragments conveying molecular structure.

Searching for analogies is a part of research work. Sometimes researchers try to push the noticed analogies forward, making them more noticeable and convenient to grasp for the readers in their published papers. However, no analogy can claim to be complete; otherwise, one should talk not about analogy, but about identity. Sometimes, authors try to hide analogies that form the basis of the developed method. The reason for such an approach may be too strong simplicity, which, in the opinion of the authors, can reduce the significance of the theoretical foundations of the research work for the reader.

In this work, an analogy is used that unites the behavior of groups of people (experts) taking part in a social questionnaire [29–31] and the behavior of the statistical objects denoted as "quasi-SMILES", which are used in stochastic simulation of endpoints related to nanoparticles. One can claim Quasi-SMILES to be “experts” because they contain information on the experiments being studied. Therefore, there is a chance that their recommendations may be useful. The results of such a "survey" may lead to non-trivial (unexpected) heuristic hypotheses. However, in some cases, the results obtained may be difficult to interpret [29–31].

One could consider the expansion of the information involved in computational simulations. If the codes (symbols) representing the molecular structure are supplemented with codes representing the various experimental conditions that occur when determining a certain endpoint, the resulting information construction is a richer basis for searching for QSPR/QSAR models. Such construction has been named quasi-SMILES. The practical need to develop endpoint models related to nanomaterials is one of the reasons that determined the feasibility of developing quasi-SMILES [25,26].

There are various approaches applied in support of computational simulations. The Monte Carlo method is a random process. On the other hand, the Las Vegas algorithm is a "casino winning method", which can be defined as follows: "if I play again and again, I will win someday". In the reality of Las Vegas casinos, such a strategy is most likely counterproductive, but in a computer search for a satisfied distribution that guarantees good statistics for a calibration set, this approach gives acceptable results [32].

Chaos and mysticism are often intricately intertwined. The observer notes the complete victory of chaos and calms down: there is no point in expecting any logic in the events. But if suddenly some of the observed events fit into some pattern, the observer, not being able to rationally interpret the observation, concludes that some paranormality or, more simply, mysticism is taking place here.

In comparing the statistical quality of models of the efficiency of nanoparticle delivery in various organs constructed based on sophisticated descriptors used in highly intelligent algorithms and the rather primitive optimal descriptors presented here, it is observed a certain correlation between bad and good forecasts obtained with the involvement of artificial intelligence and such (bad and good) forecasts obtained based on primitive descriptors calculated using quasi-SMILES. The common issue between a sociological survey and the selection of correlation weights is discussed below. Developing computational methods to model efficient delivery of a variety of nanoparticles to the organs of mice under different experimental conditions could aid in the discovery of new medical procedures and drug discovery.

The aim of this study was to evaluate the feasibility of using quasi-SMILES to predict the efficiency of nanoparticle delivery to various organs of mice.

Table 4
The statistical quality of the model for DE1.

Sets	n	D	CCC	IIC	CII	Q^2	RMSE	F	Na
A	148	0.0781	0.1448	0.2249	0.7964	0.0487	0.322	12	
P	149	0.1000	0.1439	0.2661	0.7827	0.0768	0.313	16	
C	40	0.5642	0.7137	0.7500	0.8020	0.5190	0.064	49	
V	40	0.5672	-	-	-	-	0.08	-	29
A	150	0.0751	0.1397	0.2096	0.7970	0.0477	0.314	12	
P	152	0.0801	0.1328	0.2421	0.8144	0.0574	0.317	13	
C	36	0.7121	0.7852	0.8433	0.8166	0.6754	0.066	84	
V	39	0.6484	-	-	-	-	0.07	-	28
A	150	0.1467	0.2559	0.3263	0.7589	0.1242	0.308	25	
P	150	0.0237	0.0941	0.1313	0.8903	0.0000	0.325	4	
C	39	0.5929	0.7587	0.7699	0.7921	0.5065	0.089	54	
V	38	0.6947	-	-	-	-	0.07	-	30

Table 5
The statistical quality of the model for DE2.

	n	D	CCC	IIC	CII	Q^2	RMSE	F	Na
A	154	0.0962	0.1755	0.2653	0.7964	0.0731	0.314	16	
P	151	0.0452	0.1367	0.1584	0.8464	0.0183	0.319	7	
C	35	0.8483	0.9011	0.9210	0.9007	0.8258	0.052	185	
V	37	0.6705	-	-	-	-	0.08	-	31
A	148	0.4444	0.6153	0.5366	0.7819	0.4219	0.244	117	
P	153	0.4484	0.6346	0.6068	0.7867	0.4301	0.219	123	
C	38	0.8140	0.8899	0.9018	0.8730	0.7947	0.079	158	
V	38	0.7851	-	-	-	-	0.11	-	30
A	151	0.4706	0.6400	0.5327	0.7826	0.4510	0.249	132	
P	149	0.3602	0.5453	0.5455	0.7548	0.3336	0.226	83	
C	38	0.8355	0.8848	0.9130	0.9097	0.8083	0.090	183	
V	39	0.7627	-	-	-	-	0.09	-	30

Table 6
The statistical quality of the model for DE3.

	n	D	CCC	IIC	CII	Q^2	RMSE	F	Na
A	151	0.3106	0.4739	0.3882	0.7302	0.2909	12.1	67	
P	153	0.3274	0.4985	0.4040	0.7393	0.3086	10.8	73	
C	36	0.8889	0.9403	0.9407	0.9429	0.8744	2.58	272	
V	37	0.7589	-	-	-	-	4.89	-	30
A	148	0.3719	0.5422	0.4646	0.7546	0.3551	11.5	86	
P	149	0.2876	0.4706	0.4408	0.7604	0.2664	11.4	59	
C	39	0.7915	0.8885	0.8892	0.8819	0.7703	3.19	140	
V	41	0.7116	-	-	-	-	5.77	-	30
A	149	0.3286	0.4947	0.3974	0.7173	0.3103	11.3	72	
P	149	0.3031	0.4384	0.3373	0.7448	0.2827	12.0	64	
C	40	0.8588	0.9221	0.9267	0.9182	0.8444	2.58	231	
V	39	0.8153	-	-	-	-	4.77	-	28

Table 7
The statistical quality of the model for DE4.

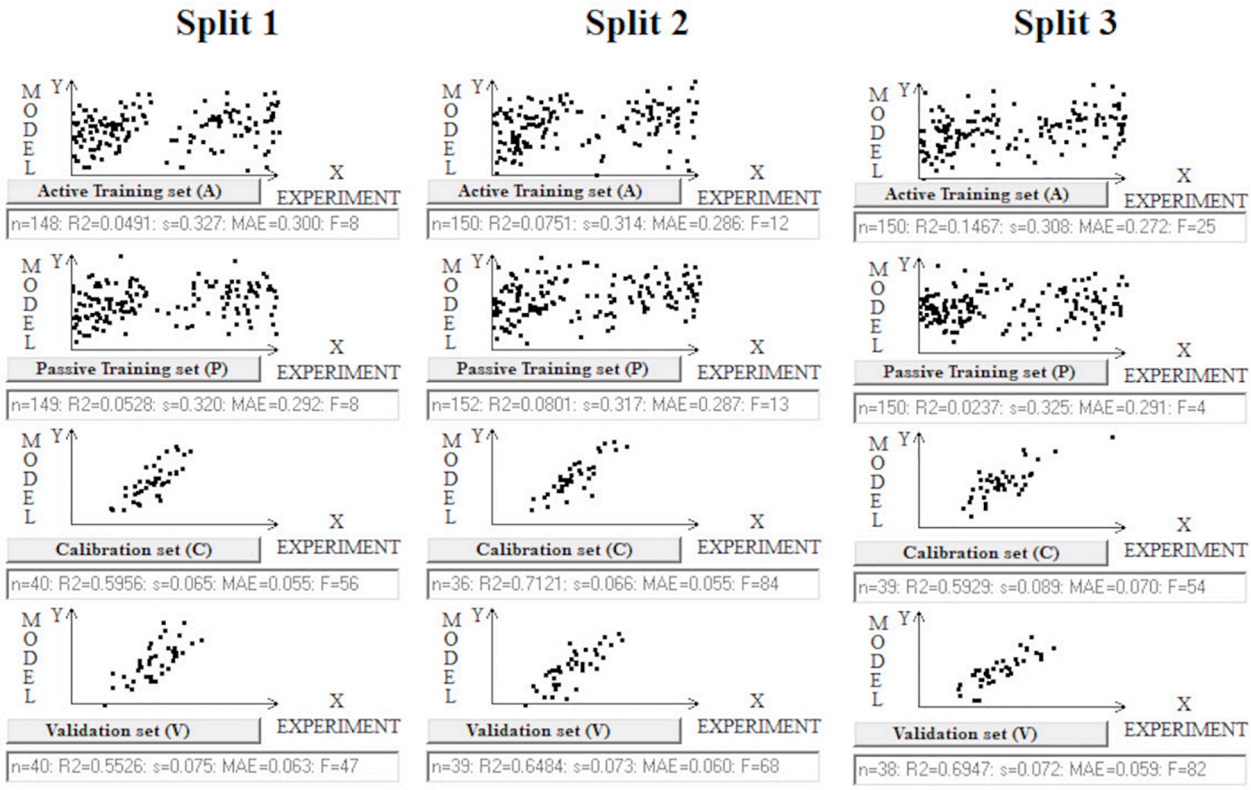
Sets	n	D	CCC	IIC	CII	Q^2	RMSE	F	Na
A	148	0.4800	0.6486	0.5576	0.7783	0.4633	1.28	135	
P	148	0.4383	0.6225	0.6112	0.8161	0.4182	1.17	114	
C	41	0.8435	0.9165	0.9184	0.9290	0.8227	0.352	210	
V	40	0.7803	-	-	-	-	0.58	-	30
A	152	0.3911	0.5623	0.4672	0.8123	0.3707	1.32	96	
P	150	0.4848	0.6118	0.5198	0.8071	0.4693	1.18	139	
C	37	0.7099	0.8425	0.8418	0.8568	0.6728	0.523	86	
V	38	0.7989	-	-	-	-	0.53	-	30
A	149	0.4017	0.5731	0.4517	0.7973	0.3791	1.23	99	
P	152	0.4886	0.5852	0.3957	0.8095	0.4708	1.29	143	
C	38	0.9009	0.9456	0.9491	0.9405	0.8851	0.256	327	
V	38	0.8946	-	-	-	-	0.47	-	29

Table 8
The statistical quality of the model for DE5.

Sets	n	D	CCC	IIC	CII	Q^2	RMSE	F	Na
A	148	0.4139	0.5855	0.4769	0.7727	0.3963	0.366	103	
P	150	0.4795	0.5888	0.5025	0.7875	0.4633	0.367	136	
C	40	0.9071	0.9497	0.9517	0.9537	0.8940	0.089	371	
V	39	0.8507	-	-	-	-	0.15	-	30
A	150	0.4644	0.6343	0.5501	0.7755	0.4485	0.393	128	
P	151	0.4028	0.6160	0.6025	0.7950	0.3817	0.338	101	
C	38	0.7510	0.8611	0.8661	0.8514	0.7278	0.183	109	
V	38	0.7856	-	-	-	-	0.23	-	27
A	148	0.5044	0.6706	0.6913	0.7760	0.4880	0.304	149	
P	148	0.4351	0.5546	0.4064	0.7690	0.4193	0.422	112	
C	41	0.8911	0.9433	0.9421	0.9242	0.8786	0.086	319	
V	40	0.8898	-	-	-	-	0.14	-	28

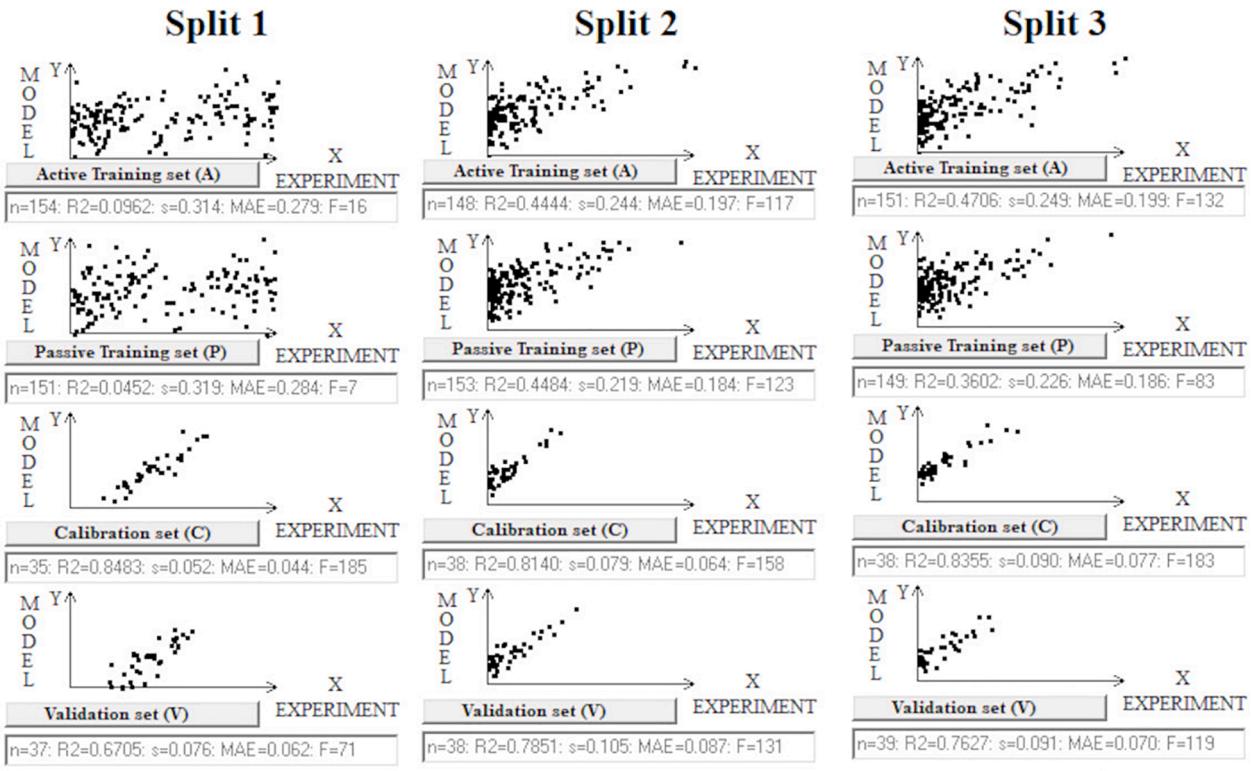
Table 9
The statistical quality of the model for DE6.

Sets	n	D	CCC	IIC	CII	Q^2	RMSE	F	Na
A	149	0.2321	0.3767	0.3070	0.7877	0.2100	2.34	44	
P	151	0.1879	0.3399	0.2973	0.7824	0.1673	2.34	34	
C	39	0.8260	0.9054	0.9088	0.8985	0.8078	0.368	176	
V	38	0.6889	-	-	-	-	0.78	-	29
A	148	0.2097	0.3467	0.2868	0.7735	0.1862	2.28	39	
P	149	0.1683	0.2913	0.2139	0.7895	0.1459	2.45	30	
C	39	0.7360	0.8392	0.8574	0.8538	0.6998	0.576	103	
V	41	0.7280	-	-	-	-	0.69	-	30
A	153	0.2555	0.4070	0.3538	0.7591	0.2343	2.24	52	
P	150	0.1973	0.3506	0.2294	0.7909	0.1744	2.37	36	
C	36	0.9044	0.9398	0.9507	0.9334	0.8937	0.414	321	
V	38	0.8574	-	-	-	-	0.66	-	28



	First & second	First & third	Second & third
Active training set	47%	44%	48%
Validation set	51%	38%	39%

Fig. 1. Graphical representation of models for DE1. The table below shows the percentage of identical distributions for splits 1, 2, and 3 across the active training and validation sets.



	First & second	First & third	Second & third
Active training set	39%	45%	45%
Validation set	3%	3%	33%

Fig. 2. Graphical representation of models for DE2. The table below shows the percentage of identical distributions for splits 1, 2, and 3 across the active training and validation sets.

Materials and methods

Data

Experimental data on the delivery of nanoparticles to tumours, hearts, liver, spleen, lungs, and kidneys in mice were taken from the literature [33].

The common feature between a sociological survey and the selection of correlation weights is the following:

First, from a set of possible questions, a certain subset is selected, which is then considered significant. In the case of the models considered here, such questions are the identification of the quasi-SMILES codes whose presence in the active training sample is greater than the selected threshold. By preliminary computational experiments, $T = 3$ was established as the most appropriate threshold. $T = 1$ or $T = 2$ led to instability statistical characteristics for all set considerations. In contrast, $T > 3$ led to a reduction in statistical quality for all subsets.

The survey (optimization of the correlation weights of different fragments of quasi-SMILES) is conducted only on these questions. It should be noted that the lists of questions (i.e. the list of fragments for the process of optimization correlation weights) asked most probably is varied with the different partitions of data into four subsets, namely, an active training set (40 %), a passive training set (40 %), a calibration set (10 %), and a validation set (10 %).

Secondly, one of the variants of a sociological survey is an expert survey, that is, an attempt to obtain recommendations from respondents (professional specialists) about a certain object of study. From this point of view, a survey of quasi-SMILES groups should give "recommendations" about the endpoint. That is, the model of the endpoint and possible ways of influencing the value of the endpoint.

Table 1 lists the experimental conditions (the questions) under which the endpoints considered here were measured.

Table 2 lists the parameters together with their value ranges, which are converted into the circumstances (codes of quasi-SMILES) of the corresponding experiments when the values are selected from the specified ranges.

Table 3 contains a list of the endpoints considered in this study.

These data are formatted as 377 quasi-SMILES. Six endpoints were associated with the 377 quasi-SMILES in turn. Thus, Monte Carlo experiments with six endpoints were performed (Table 3). For each endpoint, three variants of partitioning into the above-mentioned subsets (active and passive training sets, calibration set, and validation set, in the above percentages) were considered. Each partition was obtained using the Las Vegas algorithm, which sought such partitions where the best statistics were observed on the calibration set (in the hope that this would be accompanied by at least acceptable statistics on the validation set).

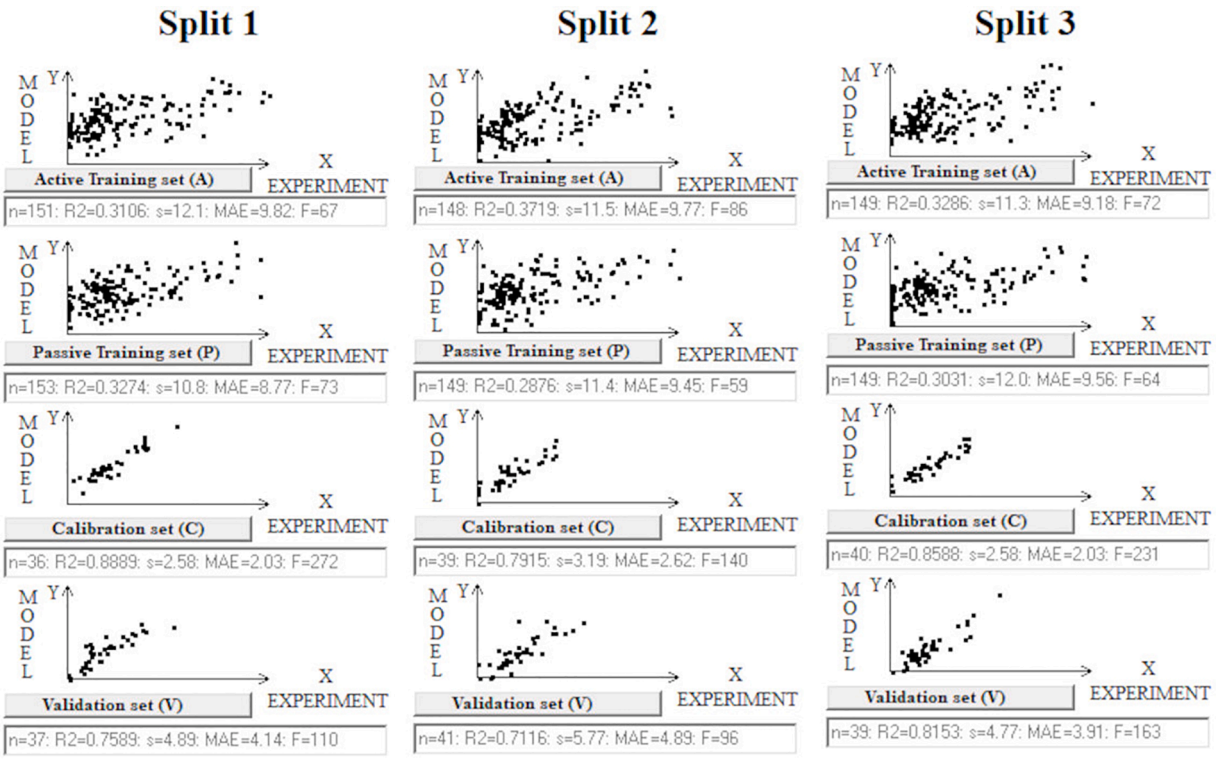
Optimal descriptor

Applying the Monte Carlo method [34], using active and passive training sets together with a calibration set, the so-called correlation weights of quasi-SMILES codes were calculated, which give the maximum value of the objective function, defined as

$$TF = D_A + D_P - |D_A - D_P| \times 0.1 + (IIC + CII) \times 0.3 \quad (1)$$

where D_A and D_P are the determination coefficients for active and passive training sets; IIC is the index of ideality of correlation [35], and CII is the correlation intensity index [36].

The IIC and CII, proposed as "efficiency enhancers" for Monte Carlo



	First & second	First & third	Second & third
Active training set	41%	46%	39%
Validation set	28%	31%	25%

Fig. 3. Graphical representation of models for DE3. The table below shows the percentage of identical distributions for splits 1, 2, and 3 across the active training and validation sets.

optimization, clearly influence the quality of the resulting models. This influence is paradoxical. Calculated based on the calibration set, the IIC and CII guide the optimization process toward improving the statistical quality of the calibration set, but to the detriment of statistical quality on the training sets. It was hypothesized, however, that good statistical performance for the calibration set would be accompanied by good (improved) statistical performance of the model for the validation set. Fortunately, this has been confirmed in many cases [35–46].

The optimal descriptor calculated with quasi-SMILES is calculated as follows:

$$\mathrm{DCW(T,N)} = \sum \mathrm{CW(qS_k)} \tag{2}$$

The qSk is a code of quasi-SMILES. T and N are parameters of the Monte Carlo method optimization. T is the threshold to separate qSk into active (which are involved in the simulation process) and rare (which are not involved in the simulation process). N is the number of epochs of Monte Carlo optimization. Here, $T = 3$ and $N = 15$ are used. By preliminary computational experiments, $T = 3$ was established as the most appropriate threshold. $T = 1$ or $T = 2$ led to instability statistical characteristics for all set considerations. In contrast, $T > 3$ led to a reduction in statistical quality for all subsets. In addition, preliminary computational experiments show that the number of epochs >15 does not give significant changes for statistical characteristics.

Optimizing correlation weights using the Monte Carlo method is a random process. However, different partitions of the training and validation sets will yield different predictive potentials. The Las Vegas algorithm, after a series of Monte Carlo runs, identifies the partition that is most favorable for the calibration set. Good results on the calibration set give hope for good statistical performance on the validation set as well. Thus, in this case, the Las Vegas algorithm is a random shuffling of quasi-SMILES, like a shuffle of playing cards. During this shuffling, the contents of the four aforementioned sets are changed, while the original number of quasi-SMILES in each set is preserved.

Model

The model of the endpoint is represented by an equation:

$$\mathrm{DEX} = \mathrm{C_0} + \mathrm{C_1} \times \mathrm{DCW}(3, 15) \tag{3}$$

$X = \{1, 2, 3, 4, 5, 6\}$ $\mathrm{C_0}$ and $\mathrm{C_1}$ are regression coefficients.

Applicability domain

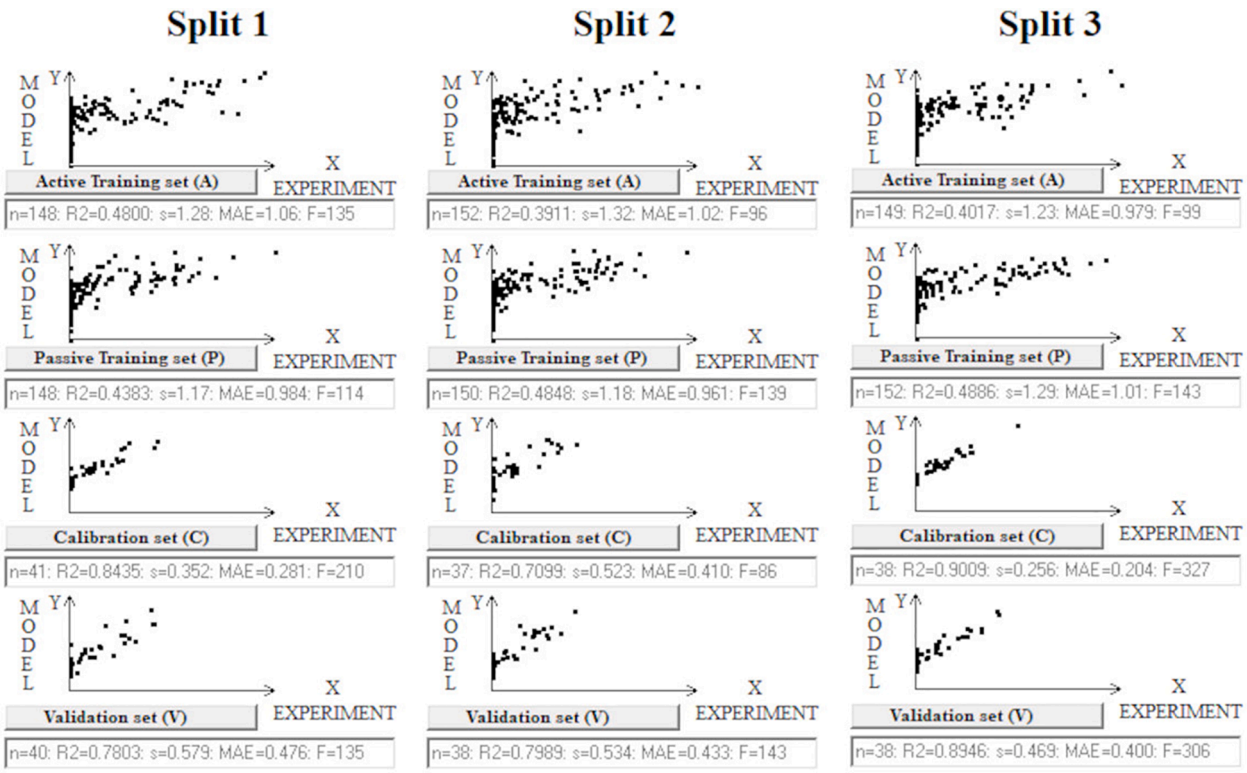
The domain of applicability for the described model is calculated via statistical defects of quasi-SMILES attributes. These defects can be calculated as:

$$\mathrm{d_k} = \frac{|\mathrm{P(qS_k)} - \mathrm{qS})|}{\mathrm{N(qS_k)} + N'(\mathrm{qS_k})} + \frac{|\mathrm{P(qS_k)} - \mathrm{P''(qS_k)}|}{\mathrm{N(qS_k)} + N''(\mathrm{qS_k})} + \frac{|\mathrm{P'(qS_k)} - \mathrm{P''(qS_k)}|}{\mathrm{N'(qS_k)} + N''(\mathrm{qS_k})} \tag{4}$$

where $\mathrm{P(qS_k)}$, $\mathrm{P'(qS_k)}$ $\mathrm{P''(qS_k)}$ are the probabilities of $\mathrm{A_k}$ in the active training set, passive training set, and calibration set, respectively; $\mathrm{N(qS_k)}$, $\mathrm{N'(qS_k)}$, and $\mathrm{N''(qS_k)}$ are frequencies of $\mathrm{qS_k}$ in the active training set, passive training set, and calibration set, respectively. The statistical SMILES-defects ($\mathrm{D_j}$) are calculated as:

$$D_j = \sum_{k=1}^{NA} d_k \tag{5}$$

where NA is the number of non-blocked quasi-SMILES codes in the SMILES.



	First & second	First & third	Second & third
Active training set	41%	42%	45%
Validation set	23%	38%	29%

Fig. 4. Graphical representation of models for DE4. The table below shows the percentage of identical distributions for splits 1, 2, and 3 across the active training and validation sets.

A quasi-SMILES falls in the domain of applicability if $Dj < 2 * \overline{D}$ (i.e. average).

Mechanistic interpretation

The developed QSAR models allow for a mechanical interpretation of the studied phenomena. Having numerical data on the correlation weights of features that take place in several runs of the Monte Carlo optimization, one can extract three categories of these features:

(i) Features that have a positive value of the correlation weight in all runs. These are promoters of endpoint increase.
(ii) Features that have a negative value of the correlation weight in all runs. These are promoters of endpoint decrease.
(iii) Features that have both negative and positive values of the correlation weight in different runs of the optimization. These are features with unclear roles (one cannot classify these features as promoters of an increase or decrease for the endpoint).

Results

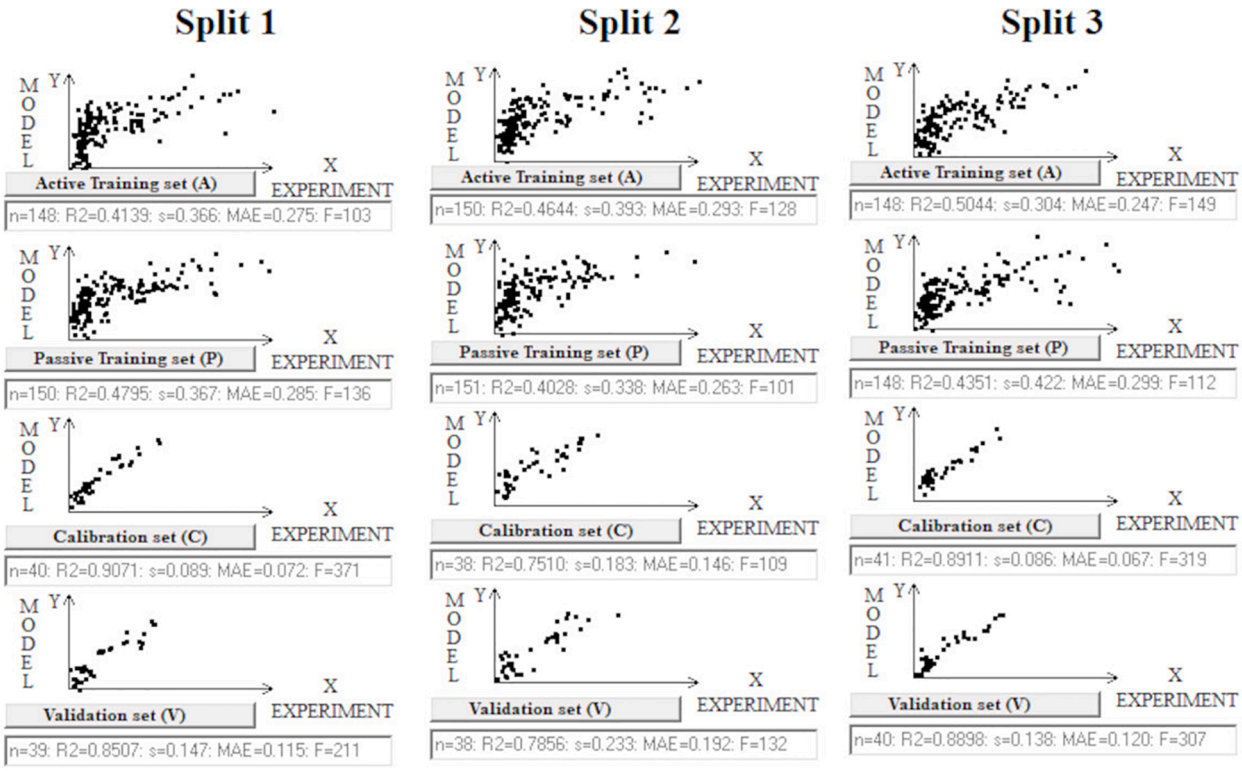
The statistical quality of models for the different organs is represented in Tables 4–9 by frequencies (n) of quasi-SMILES in four sets: active training set (A), passive training set (P), calibration set (C), and validation set (V); the statistical parameters are: determination coefficients (D), concordance correlation coefficients (CCC), index of ideality of correlation (IIC), correlation intensity index (CII), leave-one-out cross-validated R^2, root mean square error (RMSE), Fischer F-ratio (F), and number of active qSk (Na). Figs. 1–6 provide the graphical representation of the performance of the models for the different organs, as codified in Table 3. The supplementary materials section contains technical details of models for DE1, DE2, …, DE6 (all results relate to the corresponding Split 1 for DE1, DE2, …, DE6, respectively).

It should be noted that the percentage of identical quasi-SMILES distributions in the models under consideration is quite high (sometimes close to 50 %). This means that the models obtained using the Las Vegas algorithm are not completely chaotic. The Las Vegas algorithm grants some quasi-SMILES the privilege of being included in the training set, and for some quasi-SMILES, conversely, grants the statistical privilege of being included in the calibration set.

Tables 10–15 provide the correlation weights (CW) associated with the most representative parameters identified by the models, for the different organs. NA, NP, and NC are frequencies of corresponding SMILES attributes in the active training set, passive training set, and calibration set, respectively. As was said earlier in the section on Mechanistic interpretation, if the values CW(X) are positive (for all runs), then this is the ability to increase promoters of the endpoint; if the values are negative, then this is the ability to decrease promoters of the endpoint.

Discussion

The results obtained are somewhat tricky since they relate to six phenomena (endpoints) at once. However, to make them more convenient for analysis, they are all presented in a single format. There are three logical stages of the computational experiment under consideration, which are discussed below in three sections.



	First & second	First & third	Second & third
Active training set	42%	43%	37%
Validation set	20%	35%	20%

Fig. 5. Graphical representation of models for DE5. The table below shows the percentage of identical distributions for splits 1, 2, and 3 across the active training and validation sets.

Chaos

From Figs. 1–6, it is evident that the superposition of points in the "experiment – calculation" coordinates for active and passive training sets is poorly ordered. Despite this, the superpositions of points in the "experiment-calculation" coordinates for the calibration and validation sets demonstrate correlations. The reason that correlations on the calibration and validation sets occur is due to the use of the Las Vegas algorithm to select the splits most favorable for the calibration set [32,37]. IIC and CII also give some improvement in the predictive potential [35, 36]. It was assumed that good statistics on the calibration set should be accompanied by good correlation on the validation set. Indeed, this is what happened in the computer experiments described. This can be verified by comparing the results for the calibration and validation sets. The validation sets contain data related to nanoparticles not used in the phase of the model building, while the results on the calibration sets correspond to the values obtained on the developed models in their last phase, when all parameters are optimized.

The number of parameters being optimized, if we assume $T = 1$, is so large that the processes under consideration become not so random but simply chaotic. Nevertheless, if $T = 3$ is chosen, the number of parameters to be optimized is reduced to approximately 30. In this case, a stochastic process is observed, ending with reproducible results (Table 4–9).

Thus, it becomes clear that conducting a "sociological" survey with a large number of complex questions cannot give acceptable results. It is necessary to rely on "popular questions that concern all participants". Speaking in terms of quasi-SMILES, it is necessary to rely on experimental circumstances that unite quasi-SMILES groups, and not on experimental conditions that are rare and strongly individualize survey participants.

Gambling

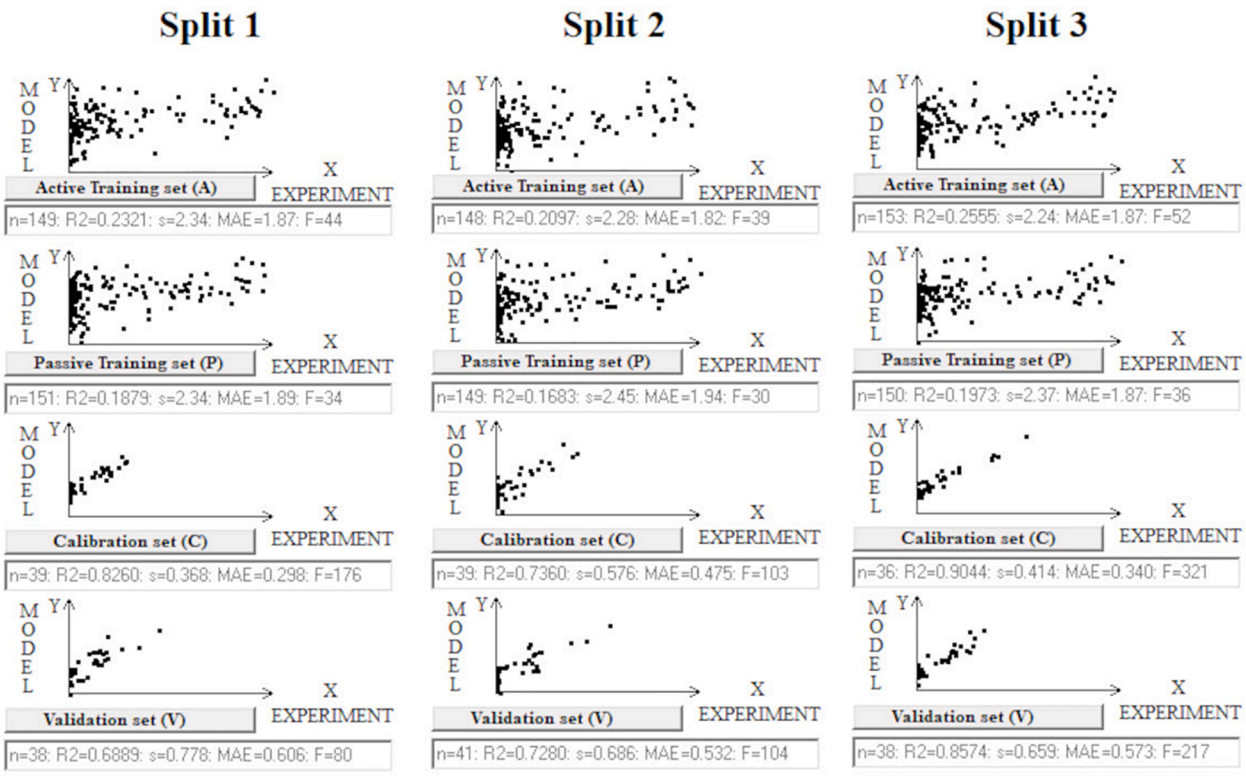
In order to obtain an informative model capable of suggesting how to control the behavior of nanoparticles carrying certain molecular agents in the organs of a living organism, it is sufficient to start with a mechanistic interpretation of models in terms of promoters of growth or reduction of the corresponding endpoints (Tables 10, 11, 12, 13, 14, 15).

Table 16 contains a map of the synergetic and antagonistic effects of experimental conditions. There are more synergetic effects (upper triangle of the square matrix, total number of synergetic effects 136) than antagonistic effects (lower triangle, total number of antagonistic effects 65). In addition, the synergetic effect is provided by the selected experimental conditions (shape, material used, nature of tumors), whereas antagonistic effects are provided to a greater extent by the organs to which the effects of nanoparticles are directed.

For example, according to Table 16, for the penetration of nanoparticles into DE1 (tumor) and DE2 (heart), synergists are [Activ], [Silic], [Spher], [ONM], [Rod], [Dendr], [Sarco], [XO], [Prost], [XH], [Pancr], [Cervi]. Whereas the antagonist is only one [Brain].

Simplicity

The approach under consideration, implementing a kind of survey on a quasi-SMILES set, turned out to be quite effective, since the statistical quality of the models on the validation sets is quite good (Table 17). However, it is obvious that the prospects for increasing the strength of information of such an approach are quite possible with the use of continuous translation data in discrete data, as described in the



	First & second	First & third	Second & third
Active training set	43%	46%	49%
Validation set	38%	42%	28%

Fig. 6. Graphical representation of models for DE6. The table below shows the percentage of identical distributions for splits 1, 2, and 3 across the active training and validation sets.

Table 10
Mechanistic interpretation for the DE1 model.

No.	qS_k	CWs in run 1	CWs in run 2	CWs in run 3	NA	NP	NC	d_k (Eq.4)
1	[Activ]	0.7184	0.9873	0.5353	74	75	15	0.0016
2	[Silic]	0.5131	0.5624	0.5684	56	69	16	0.0012
3	[Spher]	0.4681	0.4000	0.3420	54	58	15	0.0004
4	[Plate]	0.3845	0.2422	0.3598	50	38	13	0.0016
5	[ONM]	0.3696	0.3338	0.4345	45	50	9	0.0021
6	[Rod]	1.2992	1.5073	1.2006	36	35	9	0.0005
7	[Dendr]	0.5611	1.2753	0.8607	28	27	3	0.0039
8	[XO]	0.7569	1.2990	0.2974	28	28	13	0.0040
9	[Sarco]	1.2848	1.4101	0.9940	27	22	10	0.0035
10	[XH]	0.1227	0.8023	0.1478	26	23	4	0.0029
11	[Brain]	0.3461	0.6122	0.1050	20	30	8	0.0023
12	[Prost]	0.5727	1.1497	0.4008	17	16	3	0.0022
13	[Pancr]	1.8885	2.7970	1.8544	12	11	0	1.0000
14	[Liver]	1.0706	1.3901	0.7196	11	17	2	0.0043
15	[Cervi]	1.4660	1.1153	1.0704	6	7	3	0.0043
1	[INM]	−0.5303	−0.0558	−0.0870	103	99	31	0.0009
2	[Iron]	−0.4724	−0.4164	−0.1969	19	9	9	0.0089
3	[Other]	−0.3133	−0.2526	−0.1932	9	23	7	0.0059
4	[Lung]	−0.4242	−0.4723	−0.4978	4	1	1	0.0068

literature [47]. It ensures a reasonable division of the surveyed objects in accordance with their sizes, zeta potential, and concentration, into a smaller number of equivalence groups (i.e., by intervals, and not by individual values). Most likely, the indicated measures will significantly improve the models without complicating them.

In accordance with the principle that "QSAR is a random event," if several splits into training and validation sets are considered, the predictive potential most probably will be quite different [48]. These models can be successful, but unfortunately, these models can also be unsuccessful.

In this study, the Las Vegas algorithm, in conjunction with the correlation ideality index, is used to select only "successful" models. It is shown that there may be multiple "successful" models. If enough "successful" models are collected for each quasi-SMILES, the probability of

Table 11
Mechanistic interpretation for the DE2 model.

No.	qS_k	CWs in run 1	CWs in run 2	CWs in run 3	NA	NP	NC	d_k (Eq.4)
1	[Activ]	0.1808	0.4393	1.2327	72	79	17	0.0007
2	[Silic]	0.5023	0.4637	0.5329	65	65	17	0.0009
3	[Spher]	0.5408	0.1702	0.2527	55	56	16	0.0016
4	[ONM]	0.3247	0.4860	0.9448	48	47	8	0.0016
5	[Rod]	1.0143	0.7170	1.4512	39	33	9	0.0010
6	[XH]	0.1521	0.1211	1.1335	30	20	3	0.0041
7	[Dendr]	0.5737	0.6764	0.6729	28	26	4	0.0023
8	[Sarco]	0.4732	0.6812	0.9352	27	28	6	0.0005
9	[XO]	0.4780	0.2589	1.4292	25	36	7	0.0022
10	[Prost]	0.2078	0.5207	0.2149	18	16	4	0.0006
11	[Liver]	0.3780	0.5668	0.5341	12	17	4	0.0022
12	[Polym]	0.1939	0.1616	0.2764	10	19	4	0.0037
13	[Lipos]	0.3204	0.2304	0.3930	8	2	0	1.0000
14	[Pancr]	1.1730	1.4586	2.6499	8	13	1	0.0052
15	[Cervi]	0.8422	0.4100	0.3316	6	7	3	0.0058
1	[Brain]	−0.3965	−0.2316	−0.3432	36	16	4	0.0046
2	[Other]	−0.0018	−0.0200	−0.3347	19	19	6	0.0022
3	[IronÄ]	−0.3135	−0.0309	−0.5412	16	12	6	0.0054
4	[Breas]	−0.6079	−0.1467	−0.6090	10	14	5	0.0054
5	[Skin]	−0.0775	−0.1549	−0.3340	6	10	4	0.0075
6	[S0,99]	−0.3194	−0.0166	−0.2971	4	0	0	1.0000

Table 12
Mechanistic interpretation for the DE3 model.

No.	qS_k	CWs in run 1	CWs in run 2	CWs in run 3	NA	NP	NC	d_k (Eq.4)
1	[INM]	0.3879	1.0255	1.3602	109	91	29	0.0018
2	[Passi]	0.3471	0.9028	0.7684	83	78	17	0.0009
3	[AO]	0.2791	0.5166	1.0450	76	83	13	0.0021
4	[Rod]	1.3715	1.7311	2.0954	45	33	6	0.0031
5	[ONM]	0.0087	0.5720	0.9495	42	62	7	0.0038
6	[Gold]	0.1816	1.1345	1.0508	30	23	4	0.0031
7	[AH]	0.0334	0.5031	0.9561	27	22	9	0.0037
8	[XH]	0.7040	1.2590	1.0718	23	16	6	0.0028
9	[Dendr]	0.8219	0.9764	1.7171	20	39	4	0.0046
10	[Iron]	0.0200	0.3085	0.2140	19	14	1	0.0058
11	[Prost]	0.9139	1.1454	0.8487	19	17	3	0.0022
12	[Breas]	1.2821	1.4978	1.4483	18	10	4	0.0034
13	[Gliom]	1.0740	1.0698	1.3643	14	19	2	0.0039
14	[Pancr]	1.6182	2.3363	2.0294	9	10	3	0.0022
15	[Cervi]	0.7696	0.4836	0.9240	7	10	0	1.0000
1	[Spher]	−0.0294	−0.1471	−0.1359	51	56	18	0.0026
2	[Plate]	−0.3174	−0.3456	−0.2545	40	48	10	0.0010
3	[XO]	−0.4927	−0.2573	−0.1220	25	32	8	0.0017
4	[Brain]	−0.4221	−0.4192	−0.6775	22	24	2	0.0042
5	[Other]	−0.2859	−0.1182	−0.1692	15	22	6	0.0031
6	[Liver]	−0.2831	−0.0182	−0.2188	9	15	5	0.0055
7	[Ovary]	−0.3669	−0.2595	−0.4090	5	3	4	0.0153

being in the calibration set can be estimated. High values of the above probability for quasi-SMILES are indicators of the informativeness of such quasi-SMILES for the approach considered. The proposed approach allows us to take into account experimental conditions, and it also allows us to evaluate the probabilities of the influence of these conditions (compare the statistical significance of these conditions).

Conclusions

The approach based on the quasi-SMILES representation of the experimental conditions, together with data on architecture nanoparticles tested and considered in this work, provides good models of nanoparticle delivery efficiency to the mice's organs (heart, liver, lung, spleen, kidney, as well as tumors). Thus, it is confirmed that the ability to use the described approach (Monte Carlo optimization of responses to questionnaires) can be used to develop models of nanoparticle biochemical behavior under different experimental conditions.

Data availability statement

Data are available in the article or its supplementary materials. Supplementary materials contain: Table S1 includes quasi-SMILES and numerical data on delivery efficiency in tumor (DE1); Table S2 includes quasi-SMILES and numerical data on delivery efficiency in heart (DE2); Table S3 includes quasi-SMILES and numerical data on delivery efficiency in liver (DE3); Table S4 includes quasi-SMILES and numerical data on delivery efficiency in lung (DE4); Table S5 includes quasi-SMILES and numerical data on delivery efficiency in spleen (DE5); Table S6 includes quasi-SMILES and numerical data on delivery efficiency in kidney (DE6).

Author contributions

The authors contributed equally to this work.

Table 13
Mechanistic interpretation for the DE4 model.

No.	qS_k	CWs in run 1	CWs in run 2	CWs in run 3	NA	NP	NC	d_k (Eq.4)
1	[INM]	1.3344	0.0359	0.8041	107	108	24	0.0012
2	[AO]	0.6567	0.5912	0.1512	69	67	21	0.0008
3	[Silic]	0.2388	0.7950	0.1528	68	64	13	0.0020
4	[Spher]	0.6511	0.8957	0.2769	55	59	17	0.0007
5	[Rod]	1.5647	1.4129	0.6308	40	39	4	0.0042
6	[Brain]	0.1809	0.1889	0.4940	27	18	7	0.0023
7	[XH]	0.1894	0.4889	0.0373	26	18	7	0.0021
8	[Gold]	1.4700	1.9859	0.9370	25	23	6	0.0008
9	[Dendr]	0.4107	0.2733	0.1163	23	18	10	0.0048
10	[Other]	0.1953	0.0521	0.3159	21	9	6	0.0048
11	[Sarco]	1.3094	1.3574	1.0726	21	37	5	0.0041
12	[Iron]	0.6381	1.4973	0.3032	14	21	5	0.0024
13	[Prost]	2.0409	1.9891	1.7090	14	24	2	0.0057
14	[Cervi]	1.2028	0.8203	0.7200	8	7	0	1.0000
15	[S1,05]	0.1174	0.2592	0.1790	4	1	0	1.0000
1	[Plate]	−0.6633	−0.3134	−0.3973	35	45	16	0.0032
2	[XO]	−0.4921	−0.1262	−0.3063	30	33	5	0.0030
3	[AH]	−0.3581	−0.1537	−0.1552	23	30	8	0.0016
4	[Gliom]	−1.4817	−0.7209	−0.4120	17	16	2	0.0038
5	[Polym]	−0.4595	−0.3635	−0.2969	12	17	5	0.0024
6	[Skin]	−1.2831	−0.8260	−0.4437	11	6	4	0.0054
7	[Lipos]	−0.4672	−0.4882	−0.1107	4	5	1	0.0019

Table 14
Mechanistic interpretation for the DE5 model.

No.	qS_k	CWs in run 1	CWs in run 2	CWs in run 3	NA	NP	NC	d_k (Eq.4)
1	[Activ]	1.0255	0.9317	1.3045	70	84	14	0.0025
2	[AO]	0.8204	1.6329	1.5764	63	81	21	0.0014
3	[Spher]	0.4633	0.8754	1.0623	61	56	9	0.0030
4	[Plate]	0.5486	0.9151	0.8285	37	50	16	0.0029
5	[Sarco]	1.1806	1.4067	1.5344	31	26	5	0.0027
6	[AH]	0.0011	0.4982	0.6940	28	25	5	0.0022
7	[Dendr]	1.5743	1.8590	2.3967	27	27	4	0.0028
8	[Prost]	0.5489	1.0148	0.4839	18	17	3	0.0025
9	[Iron]	1.0809	0.4365	0.7740	16	19	2	0.0041
10	[Pancr]	0.4500	0.9432	0.5139	13	7	1	0.0060
11	[Breas]	1.0003	1.1571	1.4343	12	14	3	0.0013
1	[XO]	−0.3870	−0.0433	−0.0925	30	29	10	0.0016
2	[Other]	−0.5833	−0.5816	−0.3788	19	16	6	0.0021
3	[Gliom]	−0.4420	−0.4323	−0.2354	12	12	7	0.0061
4	[Polym]	−0.5099	−0.8598	−0.4340	11	18	2	0.0045
5	[Liver]	−0.5837	−0.6695	−0.6540	9	14	3	0.0025
6	[Cervi]	−0.1566	−0.1431	−0.2334	4	8	4	0.0091
7	[Lipos]	−0.4177	−0.4720	−0.6532	4	3	5	0.0175

Table 15
Mechanistic interpretation for the DE6 model.

No.	qS_k	CWs in run 1	CWs in run 2	CWs in run 3	NA	NP	NC	d_k (Eq.4)
1	[AO]	0.5717	0.4612	0.6811	71	72	22	0.0011
2	[Silic]	0.2196	0.2516	0.4383	60	62	14	0.0008
3	[ONM]	0.2821	0.8712	1.1056	47	47	14	0.0009
4	[Plate]	1.2342	1.8473	1.2093	47	55	3	0.0055
5	[Rod]	1.1947	1.4802	1.2672	40	39	2	0.0054
6	[XO]	1.1263	1.4452	1.0121	33	30	7	0.0012
7	[Sarco]	1.8839	1.3537	1.3157	32	27	6	0.0019
8	[Iron]	0.7919	0.8173	0.8787	20	19	2	0.0040
9	[Liver]	0.4911	0.6830	0.9483	16	11	2	0.0039
10	[Breas]	1.0283	1.2165	0.6099	15	14	2	0.0032
11	[Pancr]	1.1668	1.1240	0.9394	9	8	3	0.0024
12	[Cervi]	0.3072	0.6830	0.3919	8	6	2	0.0017
13	[S1,65]	0.8315	0.2280	1.4006	4	0	0	1.0000
1	[Spher]	−0.2520	−0.1698	−0.2220	52	49	24	0.0047
2	[AH]	−0.4902	−0.2614	−0.4741	23	23	7	0.0010
3	[Other]	−0.7908	−0.4222	−0.5665	13	13	13	0.0127
4	[Prost]	−0.5644	−0.3835	−0.8972	10	19	6	0.0050
5	[Skin]	−0.4586	−0.3049	−0.3899	10	8	2	0.0016
6	[Ovary]	−0.3428	−0.5525	−0.4094	4	4	2	0.0050

Table 16
Lists of common experimental circumstances that act synergistically or, on the contrary, antagonistically on the endpoints under consideration.

	DE1	DE2	DE3	DE4	DE5	DE6
DE1		[Activ] [Silic] [Spher] [ONM] [Rod] [Dendr] [Sarco] [XO] [Prost] [XH] [Pancr] [Cervi]	[Chervi] [Other] [ONM] [Rod] [Dendri] [XH] [Prost] [Pancr]	[Silic] [Spher] [Rod] [Dendr] [Sarco] [XH] [Brain] [Prost] [Cervi]	[Activ] [Spher] [Plate] [Dendr] [Sarco] [Prost] [Pancr]	[Plate] [ONM] [Rod] [XO] [Sarco] [Pancr] [Liver] [Cervi] [Other]
DE2	[Brain]		[ONM] [Rod] [XH] [Dendr] [Prost] [Pancr] [Cervi] [Brain] [Other]	[Silic] [Spher] [Rod] [XH] [Dendr] [Sarco] [Prost] [Cervil] [Skin]	[Activ] [Spher] [Dendr] [Sarco] [Prost] [Pancr]	[Silic] [ONM] [Rod] [Sarco] [XO] [Liver] [Pancr] [Cervi] [Other] [Iron] [Breas]
DE3	[Spher] [XO] [Brain] [Plate] [Liver] [INM] [Iron]	[Spher] [XO] [Liver] [Breas]		[INM] [AO] [Rod] [Gold] [XH] [Dendr] [Iron] [Prost] [Cervi] [Plate] [XO]	[AO] [AH] [Dendr] [Iron] [Prost] [Breas] [Pancr] [XO] [Other] [Liver]	[AO] [Rod] [ONM] [Iron] [Breas] [Pancr] [Cervi] [Spred] [Other] [Ovary]
DE4	[Plate] [XO] [Iron]	[XO] [Polym] [Lipos] [Brain] [Other] [Iron]	[AH] [Gliom] [Spher] [Brain] [Other]		[AO] [Spher] [Dendr] [Sarco] [Iron] [Prost] [XO] [Gliom] [Polym] [Lipos]	[AO] [Silic] [Rod] [Sarco] [Iron] [Cervi] [AH] [Skin]
DE5	[XO] [Liver] [Cervi] [Iron]	[XO] [Liver] [Polym] [Lipos] [Cervil] [Iron] [Breas]	[Gliom] [Cervi] [Spher] [Plate]	[Other] [Cervi] [Plate] [AH]		[AO] [Plate] [Sarco] [Iron] [Pancr] [Breas] [Other]
DE6	[Silic] [Spher] [Prost] [Iron]	[Spred] [Prost]	[AH] [Prost] [Plate] [XO] [Liver]	[Spher] [Other] [Prost] [Plate] [XO]	[Spher] [AH] [Prost] [Liver]	

CRediT authorship contribution statement

Alla P. Toropova: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Andrey A. Toropov:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ivan Raska Jr:** Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Maria Raskova:** Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Emilio Benfenati:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Conceptualization. **Danuta Leszczynska:** Writing – review & editing, Writing – original draft, Visualization, Supervision. **Jerzy Leszczynski:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision.

Table 17
Comparison models obtained here with models suggested in the literature [15].

Endpoint	In this work		In [15]	
	D_{train}	$D_{validation}$	D_{train}	$D_{validation}$
DE1	0.15	0.69	0.73	0.41
DE2	0.44	0.78	0.78	0.42
DE3	0.33	0.81	0.77	0.45
DE4	0.40	0.89	0.90	0.79
DE5	0.50	0.90	0.90	0.87
DE6	0.25	0.86	0.87	0.83

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.insi.2025.100105.

References

[1] Wicki A, Witzigmann D, Balasubramanian V, Huwyler J. Nanomedicine in cancer therapy: challenges, opportunities, and clinical applications. J Control Release 2015;200:138–57. https://doi.org/10.1016/j.jconrel.2014.12.030.

[2] Sun Q, Zhou Z, Qiu N, Shen Y. Rational design of cancer nanomedicine: nanoproperty integration and synchronization. Adv Mater 2017;29(14):1606628. https://doi.org/10.1002/adma.20160662.

[3] Asgharzadeh F, Binabaj MM, Fanoudi S, Cho WC, Yang Y-J, Azarian M, Ardestani MS, Nasiri N, Farani MR, Huh YS. Nanomedicine strategies utilizing lipid-based nanoparticles for liver cancer therapy: exploring signaling pathways and therapeutic modalities. Adv Pharm Bull 2024;14(3):513–23. https://doi.org/10.34172/apb.2024.061.

[4] Hussain S. Nanomedicine for treatment of lung cancer. Adv Exp Med Biol 2016: 890137–47. https://doi.org/10.1007/978-3-319-24932-2_8.

[5] Hauser PV, Chang H-M, Yanagawa N, Nanotechnology Hamon M. nanomed kidney Appl Sci (Basel) 2021;11(16):7187. https://doi.org/10.3390/app11167187.

[6] Concu R, Kleandrova VV, Speck-Planche A, Cordeiro MNDS. Probing the toxicity of nanoparticles: a unified in silico machine learning model based on perturbation theory. Nanotoxicology 2017;11(7):891–906. https://doi.org/10.1080/17435390.2017.1379567.

[7] Speck-Planche A, Kleandrova VV, Feng LF, Cordeiro MNDS. Computational modeling in nanomedicine: prediction of multiple antibacterial profiles of nanoparticles using a quantitative structure-activity relationship perturbation model. Nanomedicine 2015;10(2):193–204. https://doi.org/10.2217/nnm.14.96.

[8] Feng LF, Kleandrova VV, González-Díaz H, Ruso JM, Melo AS, Speck-Planche A, Cordeiro MNDS. Computer-aided nanotoxicology: assessing cytotoxicity of nanoparticles under diverse experimental conditions by using a novel QSTR-perturbation approach. Nanoscale 2014;6(18):10623–30. https://doi.org/10.1039/c4nr01285b.

[9] Urista DV, Carrué DB, Otero I, Arrasate S, VF Quevedo-Tumailli, Gestal M, González-Díaz H, Munteanu CR. Prediction of antimalarial drug-decorated nanoparticle delivery systems with random forest models. Biol (Basel) 2020;9(8): 198. https://doi.org/10.3390/biology9080198.

[10] Santana RR, Zuluaga Gallego RO, Gañán-Rojo PS, Arrasate SOE. González-Díaz H. Designing nanoparticle release systems for drug-vitamin cancer co-therapy with multiplicative perturbation-theory machine learning (PTML) models. Nanoscale 2019;11(45):21811–23. https://doi.org/10.1039/c9nr05070a.

[11] Moradi Kashkooli F, Soltani M, Souri M, Meaney C, Kohandel M. Nexus between in silico and in vivo models to enhance clinical translation of nanomedicine. Nano Today 2021;36:101057. https://doi.org/10.1016/j.nantod.2020.101057.

[12] Paknia F, Mohabatkar H, Ahmadi-Zeidabadi M, Zarrabi A. The convergence of in silico approach and nanomedicine for efficient cancer treatment; in vitro investigations on curcumin loaded multifunctional graphene oxide nanocomposite structure. J Drug Deliv Sci Tech 2022;71:103302. https://doi.org/10.1016/j.jddst.2022.103302.

[13] Stillman NR, Kovacevic M, Balaz I, Hauert S. In silico modelling of cancer nanomedicine, across scales and transport barriers. npj Comput Mater 2020;6(1): 92. https://doi.org/10.1038/s41524-020-00366-8.

[14] Mascheroni P, Schrefler BA. In silico models for nanomedicine: recent developments. Curr Med Chem 2018;25(34):4192–207. https://doi.org/10.2174/0929867324666170417120725.

[15] Fontana F, Gelain F. Modeling of supramolecular biopolymers: leading the in silico revolution of tissue engineering and nanomedicine. Nanotechnol Rev 2022;11(1): 2965–96. https://doi.org/10.1515/ntrev-2022-0455.

[16] Weininger D. SMILES, a chemical language and information system: 1: introduction to methodology and encoding rules. J Chem Inf Comput Sci 1988;28 (1):31–6. https://doi.org/10.1021/ci00057a005.

[17] Kumar P, Kumar ACORAL. QSAR models of CB1 cannabinoid receptor inhibitors based on local and global SMILES attributes with the index of ideality of correlation and the correlation contradiction index. Chemom Intell Lab Syst 2020; 200:103982. https://doi.org/10.1016/j.chemolab.2020.103982.

[18] Ghaedi A. Predicting the cytotoxicity of ionic liquids using QSAR model based on SMILES optimal descriptors. J Mol Liq 2015;208:269–79. https://doi.org/10.1016/j.molliq.2015.04.049.

[19] Kumar P, Kumar A, Sindhu J. In silico design of diacylglycerol acyltransferase-1 (DGAT1) inhibitors based on SMILES descriptors using Monte-Carlo method. SAR QSAR Env Res 2019;30(8):525–41. https://doi.org/10.1080/1062936X.2019.1629998.

[20] Costa AS, Martins JPA, de Melo EB. SMILES-based 2D-QSAR and similarity search for identification of potential new scaffolds for development of SARS-CoV-2 MPRO inhibitors. Struct Chem 2022;33(5):1691–706. https://doi.org/10.1007/s11224-022-02008-9.

[21] Li Q, Ding X, Si H, Gao H. QSAR model based on SMILES of inhibitory rate of 2, 3-diarylpropenoic acids on AKR1C3. Chemom Intell Lab Syst 2014;139:132–8. https://doi.org/10.1016/j.chemolab.2014.09.013.

[22] Lotfi S, Ahmadi S, Zohrabi P. QSAR modeling of toxicities of ionic liquids toward Staphylococcus aureus using SMILES and graph invariants. Struct Chem 2020;31 (6):2257–70. https://doi.org/10.1007/s11224-020-01568-y.

[23] Ouabane M, Zaki K, Tabti K, Alaqarbeh M, Sbai A, Sekkate C, Bouachrine M, Lakhlifi T. Molecular toxicity of nitrobenzene derivatives to tetrahymena pyriformis based on SMILES descriptors using Monte Carlo, docking, and MD simulations. Comput Biol Med 2024;169:107880. https://doi.org/10.1016/j.compbiomed.2023.107880.

[24] 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): 183–90. https://doi.org/10.1021/acs.chemrestox.7b00303.

[25] He Y, Liu F, Min W, Liu G, Wu Y, Wang Y, Yan X, Yan B. De novo design of biocompatible nanomaterials using quasi-SMILES and recurrent neural networks. ACS Appl Mater Interfaces 2024;16(48):66367–76. https://doi.org/10.1021/acsami.4c15600.

[26] Ahmadi S, Aghabeygi S, Farahmandjou M, Azimi N. The predictive model for band gap prediction of metal oxide nanoparticles based on quasi-SMILES. Struct Chem 2021;32(5):1893–905. https://doi.org/10.1007/s11224-021-01748-4.

[27] Ahmadi S, Ketabi S, Qomi M. CO2 uptake prediction of metal-organic frameworks using quasi-SMILES and Monte Carlo optimization. New J Chem 2022;46(18): 8827–37. https://doi.org/10.1039/d2nj00596d.

[28] Kumar P, Kumar A, Sindhu J, Lal S. Quasi-SMILES as a basis for the development of QSPR models to predict the CO2 capture capacity of deep eutectic solvents using correlation intensity index and consensus modelling. Fuel 2023;345:128237. https://doi.org/10.1016/j.fuel.2023.128237.

[29] Lemoine CP, Madadi-Sanjani O, Petersen C, Chardot C, de Ville de Goyet J, Superina R. Pediatric liver and transplant surgery: results of an international survey and expert consensus recommendations. J Clin Med 2023;12(9):3229. https://doi.org/10.3390/jcm12093229.

[30] Di Iorgi N, Morana G, Cappa M, D'Incerti L, Garrè ML, Grossi A, Iughetti L, Matarazzo P, Parpagnoli M, Pozzobon G, Salerno M, Sardi I, Wasniewska MG, Zucchini S, Rossi A, Maghnie M. Expert opinion on the management of growth hormone deficiency in brain tumor survivors: results from an Italian survey. Front Endocrinol 2022;13:920482. https://doi.org/10.3389/fendo.2022.920482.

[31] Fastner S, Shen N, Hartman RI, Chu EY, Kim CC, Kirkwood JM, Grossman D. Prognostic gene expression profile testing to inform use of adjuvant therapy: a survey of melanoma experts. Cancer Med 2023;12(24):22103–8. https://doi.org/10.1002/cam4.6819.

[32] Toropova AP, Toropov AA. The coefficient of conformism of a correlative prediction (CCCP): building up reliable nano-QSPRs/QSARs for endpoints of nanoparticles in different experimental conditions encoded via quasi-SMILES. Sci Total Env 2024;927:172119. https://doi.org/10.1016/j.scitotenv.2024.172119.

[33] Mi K, Chou W-C, Chen Q, Yuan L, Kamineni VN, Kuchimanchi Y, He C, Monteiro-Riviere NA, Riviere JE, Lin Z. Predicting tissue distribution and tumor delivery of nanoparticles in mice using machine learning models. J Control Release 2024;374: 219–29. https://doi.org/10.1016/j.jconrel.2024.08.015.

[34] Toropova AP, Toropov AA. Hybrid optimal descriptors as a tool to predict skin sensitization in accordance to OECD principles. Toxicol Lett 2017;275:57–66. https://doi.org/10.1016/j.toxlet.2017.03.023.

[35] Toropova AP, Toropov AA. The index of ideality of correlation: a criterion of predictability of QSAR models for skin permeability? Sci Total Env 2017;586: 466–72. https://doi.org/10.1016/j.scitotenv.2017.01.198.

[36] Toropov AA, Toropova AP. Correlation intensity index: building up models for mutagenicity of silver nanoparticles. Sci Total Env 2020;737:139720. https://doi.org/10.1016/j.scitotenv.2020.139720.

[37] Veselinović AM, Toropova AP, Toropov AA, Roncaglioni A, Benfenati E. Las Vegas algorithm in the prediction of intrinsic solubility of drug-like compounds. J Mol Graph Model 2025;137:109004. https://doi.org/10.1016/j.jmgm.2025.109004.

[38] Valizadeh N, Ahmadi S, Lotfi S, Ketabi S. Integrating QSAR modeling, ADMET screening, molecular docking, and molecular dynamics simulations to identify potential MCF-7 inhibitors. Comput Biol Med 2025;197(PtA):110972. https://doi.org/10.1016/j.compbiomed.2025.110972.

[39] Toropova AP, Toropov AA, Raska I, Raskova M, Carbö-Dorca R. The prediction of the retention time of pesticide based on the Monte Carlo method with the use of the

vector of the ideality of correlation and correlation weights of local symmetry fragments. J Math Chem 2024;62:2373–87. https://doi.org/10.1007/s10910-023-01517-0.

[40] Bhawna Kumar S, Kumar P, Kumar A. Correlation intensity index-index of ideality of correlation: a hyphenated target function for furtherance of MAO-B inhibitory activity assessment. Comput Biol Chem 2024;108:107975. https://doi.org/10.1016/j.compbiolchem.2023.107975.

[41] Ahmadi S, Lotfi S, Hamzehali H, Kumar P. A simple and reliable QSPR model for prediction of chromatography retention indices of volatile organic compounds in peppers. RSC Adv 2024;14:3186–201. https://doi.org/10.1039/d3ra07960k.

[42] Toropov AA, Toropova AP, Leszczynska D, Leszczynski J. Development of self-consistency models of anticancer activity of nanoparticles under different experimental conditions using Quasi-SMILES approach. Nanomaterials 2023;13(12):1852. https://doi.org/10.3390/nano13121852.

[43] Bagri K, Kapoor A, Kumar P, Kumar A. Hybrid descriptors–conjoint indices: a case study on imidazole-thiourea containing glutaminyl cyclase inhibitors for design of novel anti-Alzheimer's candidates. SAR QSAR Env Res 2023;34(5):361–81. https://doi.org/10.1080/1062936X.2023.2212175.

[44] Toropova AP, Toropov AA, Fjodorova N. Quasi-SMILES for predicting toxicity of nano-mixtures to Daphnia Magna. NanoImpact 2022;28:100427. https://doi.org/10.1016/j.impact.2022.100427.

[45] Toropov AA, Sizochenko N, Toropova AP, Leszczyńska D, Leszczynski J. Advancement of predictive modeling of zeta potentials (ζ) in metal oxide nanoparticles with correlation intensity index (CII). J Mol Liq 2020;317:113929. https://doi.org/10.1016/j.molliq.2020.113929.

[46] Toropova AP, Toropov AA. Fullerenes C60 and C70: a model for solubility by applying the correlation intensity index. Fuller Nanotub Car N 2020;28(11):900–6. https://doi.org/10.1080/1536383X.2020.1779705.

[47] Toropova AP, Toropov AA. Nano-QSAR in cell biology: model of cell viability as a mathematical function of available eclectic data. J Theor Biol 2017;416:113–8. https://doi.org/10.1016/j.jtbi.2017.01.012.

[48] Toropov AA, Toropova AP. QSAR as a random event: criteria of predictive potential for a chance model. Struct Chem 2019;30:1677–83. https://doi.org/10.1007/s11224-019-01361-6.