

Review

QSPR/QSAR: State-of-Art, Weirdness, the Future

Andrey A. Toropov  and Alla P. Toropova * 

Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Science, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via Mario Negri 2, 20156 Milano, Italy; andrey.toropov@marionegri.it

* Correspondence: alla.toropova@marionegri.it; Tel.: +39-02-39014595; Fax: +39-02-39014735

Received: 11 February 2020; Accepted: 10 March 2020; Published: 12 March 2020



Abstract: Ability of quantitative structure–property/activity relationships (QSPRs/QSARs) to serve for epistemological processes in natural sciences is discussed. Some weirdness of QSPR/QSAR state-of-art is listed. There are some contradictions in the research results in this area. Sometimes, these should be classified as paradoxes or weirdness. These points are often ignored. Here, these are listed and briefly commented. In addition, hypotheses on the future evolution of the QSPR/QSAR theory and practice are suggested. In particular, the possibility of extending of the QSPR/QSAR problematic by searching for the “statistical similarity” of different endpoints is suggested and illustrated by an example for relatively “distanced each from other” endpoints, namely (i) mutagenicity, (ii) anticancer activity, and (iii) blood–brain barrier.

Keywords: QSAR evolution; multi-target QSAR; Monte Carlo method; fuzzy sets

1. Introduction

Each science meets with internal and external contradictions. Correlations many times have served as a key to the interpretation of various phenomena. Expansion of information available for analysis (e.g., search space along “traditional” substances has been extended by nanomaterials) leads to the following question: is the correlation useful or it will better to try to define a causality? [1]. Apparently, an answer to this question is deemed to be non-completed. At the first stage, Wiener had established the first correlations for physicochemical endpoints [2–4]. Later, other authors [5–9] continued the stream of similar studies. At the second stage (after almost twenty years pause), Hansh and Fujita [10] established the first correlations for biochemical endpoints.

The quantitative structure–property activity relationships (QSPRs/QSARs) are a relatively new field of natural sciences. There is a large group of aims associated with the QSPRs/QSARs technique, the main ones of these are probably the following: (i) prediction of the physicochemical behavior of various substances in industry and their further ecologic impacts [2–9]; (ii) the biochemical behavior of various substances in ecological and medicinal aspects [10]; (iii) selection of substances, which can be prospective candidates to the defined role [11,12].

Results of traditional experiments depended on properties of substances, masses, radiation; heat capacity, electronic, physicochemical, and biochemical conditions as well as porosity, Zeta potential of nanomaterials, time of exposure, irradiation, darkness, etc. Computational experiments related to QSPR/QSAR concerned with “information conditions” (available datasets) and “statistical conditions” (diversity of substances in datasets), as well as preference of the user.

Wiener has carried out the pioneer works in the field of correlation “molecular structure–macro-effect of a substance” in the 1940s [2–4]. This was the start of QSPR/QSAR history. In other words, this is the first stage of evolution of QSPR/QSAR theory and practice.

The main task of the QSPR/QSAR at this period was to establish a correlation between an endpoint and descriptor for a set of substances. Criteria of quality of those models were (i) the total number

of compounds in the available set; (ii) correlation coefficient; (iii) standard error of estimation; and (iv) the Fischer F-ratio [1–5]. In this period the family of topological indices [6–17], indices based on the mathematical theory of information [18–28], various 3D descriptors [29–32], and descriptors of quantum mechanics [33–38] were the basis of the QSPR/QSAR theory and practice.

However, the absence of the reliable statistical checking up of these models had led to intense criticism of the QSPR/QSAR research. This criticism is continuing up to now [39–42].

A set of principles was proposed for evaluating the validity of QSAR models at a conference held in Setubal, Portugal in 2002. According to the Setubal principles, QSARs should:

1. Be associated with a defined endpoint of regulatory importance;
2. Take the form of an unambiguous algorithm;
3. Have a clear domain of applicability;
4. Be associated with appropriate measures of goodness of robustness, and predictivity;
5. Have a mechanistic interpretation.

In further, these principles were renamed in OECD principles (Organization for Economic Co-operation and Development) <http://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf>

The OECD principles open the second stage of QSPR/QSAR history: “not only to establish a correlation, but to check up the predictive potential of the correlation.”

2. QSPR/QSAR: State-of-Art

There has been an improvement in the QSPR/QSAR technique during the last decade. However, some “unpleasant peculiarities” remain. The list of “main unpleasant peculiarities” of QSPR/QSAR analysis is as follows: (i) possibility of “chance correlations” [43–46]; (ii) possibility of overtraining [47]; (iii) possibility of weak reproducibility of statistical quality of an approach suggested [48,49].

A person who would like to apply a model hardly will be pleasant to the necessity to get a group of descriptors via hard-to-understand software and with the further necessity to carry out calculations with other hard-to-understand software that provides multiple linear regression analysis or the artificial neuron networks or something else. Attempts to solve problems related to the above “unpleasant peculiarities” of QSPR/QSAR have been performed. However, these attempts gave three weirdness points.

2.1. The First Weirdness of QSPR/QSAR

The distribution of available data for QSPR/QSAR analyses into the training and validation sets can be done by various manners [50,51]. The distribution has a key influence on the statistical quality of QSPR/QSAR models [52,53]. Here, one can see first weirdness in the modern QSPR/QSAR researches: the majority of the models are based solely on distribution available data into the training and validation sets.

According to many authors, a rational split into training and validation set gives better statistical results of the validation sets than models based on random splits [54]. However, the experiment confirms that there are splits successful for one approach, which are unsuccessful for another approach [55–59]. For example, three different splits (Table 1) into training and validation sets of 87 anticancer inhibitors [60] give models with different predictive abilities (Table 2).

An examination of several splits decreases the probability of “chance correlations”: solely one good correlation easily can become chance correlation, however, three (five, six, seven, ...) good correlations hardly can be “chance correlations”.

Method 1 is a one-variable model calculated with the Monte Carlo technique [61–64] for hybrid optimal descriptors, which are calculated by simplified molecular input-line entry system (SMILES) [65], together with a molecular graph [66–71]:

$$DCW(1,10) = \sum CW(EC0_k) + \sum CW(EC1_k) + \sum CW(S_j) + \sum CW(SS_j) \quad (1)$$

The $EC0_k$ is the vertex degree in hydrogen-suppressed graph (HSG); $EC1_k$ is Morgan extended connectivity [72,73] of the first order; S_j is SMILES atoms, i.e., one symbol (e.g., 'C', 'N', 'O', etc.) or a group of symbols which cannot be examined separately (e.g., 'Cl', 'Br', '@@', '%12, etc.); the SS_j are connected pairs of the SMILES atoms.

Table 1. Distribution of 87 anticancer inhibitors [60] into training and validation sets.

Split #1	Training set = 1, 4, 6, 7, 8, 9, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 25, 27, 29, 30, 31, 33, 34, 36, 37, 38, 40, 41, 42, 45, 46, 48, 51, 52, 54, 55, 56, 57, 59, 60, 61, 63, 64, 65, 67, 69, 70, 73, 74, 75, 77, 90, 94, 98, 99, 109, 112, 116, 117, 118, 120, 121, 122, 123, 124, 126, 130, 136; Validation set = 2, 3, 5, 10, 11, 22, 26, 32, 35, 39, 43, 47, 68, 71, 92, 103, 125, 143
Split #2	Training set = 1, 6, 7, 8, 9, 12, 13, 14, 15, 16, 18, 19, 23, 25, 27, 31, 33, 34, 36, 40, 41, 42, 45, 46, 48, 51, 54, 55, 56, 57, 59, 61, 63, 65, 67, 69, 73, 74, 75, 77, 98, 109, 112, 116, 117, 121, 123, 124, 130, 136, 5, 10, 11, 22, 26, 32, 39, 43, 47, 68, 71, 92, 103, 125, 143; Validation set = 4, 17, 20, 21, 29, 30, 37, 38, 52, 60, 64, 70, 90, 94, 99, 118, 120, 122, 126, 2, 3, 35
Split #3	Training set = 1, 4, 6, 7, 8, 9, 12, 13, 14, 15, 16, 17, 18, 19, 21, 23, 25, 27, 29, 31, 36, 37, 38, 40, 41, 42, 45, 48, 51, 54, 56, 57, 59, 60, 64, 65, 69, 70, 73, 74, 75, 77, 94, 98, 99, 109, 116, 118, 121, 124, 130, 136, 2, 5, 10, 11, 26, 32, 35, 39, 43, 47, 68, 71, 125, 143; Validation set = 20, 30, 33, 34, 46, 52, 55, 61, 63, 67, 90, 112, 117, 120, 122, 123, 126, 3, 22, 92, 103

Table 2. The predictive potential of different approaches observed for different splits.

Method	Split	Number of Compounds in Validation Set	Determination Coefficient for Validation Set
3D-QSAR [60]	#1	18	0.77
Method 1	#1	18	0.43
Method 2	#1	18	0.53
Method 1	#2	22	0.84
Method 1	#3	21	0.81
Method 2	#2	22	0.82
Method 2	#3	21	0.85

Method 2 is a one-variable model calculated with the Monte Carlo method for hybrid descriptors:

$$DCW(1, 10) = CW(C5) + CW(C6) + \sum CW(S_j) + \sum CW(SS_j) \quad (2)$$

The C5 and C6 are codes of molecular rings extracted from the adjacency matrix of HSG [74].

The $CW(EC0_k)$, $CW(EC1_k)$, $CW(S_j)$, $CW(SS_j)$, $CW(C5)$, and $CW(C6)$ are correlation weights of the above-listed SMILES attributes and invariants of HSG calculated with the Monte Carlo method (<http://www.insilico.eu/coral>).

The described experiment confirms successful and unsuccessful splits exist. Excellent split (Split 1) for the 3D-QSAR approach is poor for 2D approaches, i.e., models calculated by Equation (1) or Equation (2). However, (Table 2), Split 2 is excellent (at least successful) for Method 1, whereas the Split 3 is excellent (at least successful) for Method 2.

2.2. The Second Weirdness of QSPR/QSAR

The number of statistical characteristics aimed to measure the predictive potential of a model gradually increase (Table 3), despite the apparent attractiveness of a small number of criteria of the predictive potential for practical applications.

On the one hand, the diversity of different criteria of predicting potential is a tool to improve the quality of QSPR/QSAR models. On the other hand, this situation causes sometimes the uncertainty in the choice of the best model. In other words, contradictions in the recommendations of various criteria force the researcher to search for truth (i.e., the best choice) in a greater maze of possibilities.

Table 3. Statistical criteria of the predictive potential for the quantitative structure–property activity relationships (QSPR/QSAR) models.

Criterion of the Predictive Potential	Reference
$R = \frac{n \sum xy - \sum x \sum y}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}}$	[75]
$CCC = \frac{2 \sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2 + \sum (y-\bar{y})^2 + n(\bar{x}-\bar{y})^2}$	[76]
$R_0^2 = 1 - \frac{\sum (\bar{y}_i - \bar{y}_i^o)^2}{\sum (\bar{y}_i - \bar{y})^2}$ $R_0'^2 = 1 - \frac{\sum (y_i - \bar{y}_i^o)^2}{\sum (y_i - \bar{y})^2}$ $k = \frac{\bar{y}_i}{\bar{y}_i^o}$ $k' = \frac{y_i}{\bar{y}_i^o}$	[77]
$Q^2 = 1 - \frac{\sum (y_k - \bar{y}_k)^2}{\sum (y_k - \bar{y}_k)^2}$ $Q_{F1}^2 = 1 - \frac{[\sum_{i=1}^{N_{EXT}} (y_i - y_i)^2] / N_{EXT}}{[\sum_{i=1}^{N_{EXT}} (y_i - \bar{y}_{TR})^2] / N_{EXT}}$ $Q_{F2}^2 = 1 - \frac{[\sum_{i=1}^{N_{EXT}} (y_i - y_i)^2] / N_{EXT}}{[\sum_{i=1}^{N_{EXT}} (y_i - \bar{y}_{EXT})^2] / N_{EXT}}$ $Q_{F3}^2 = 1 - \frac{[\sum_{i=1}^{N_{EXT}} (y_i - y_i)^2] / N_{EXT}}{[\sum_{i=1}^{N_{TR}} (y_i - \bar{y}_{TR})^2] / N_{TR}}$	[78]
$r_m^2 = r^2 \left(1 - \sqrt{ r^2 - r_0^2 } \right)$ $\bar{R}_m^2 = \frac{R_m^2(x,y) - R_m^2(y,x)}{2}$ $\Delta R_m^2 = R_m^2(x,y) - R_m^2(y,x) $	[79]
$IIC_{CLB} = r_{CLB} \frac{\min(-MAE_{CLB}, +MAE_{CLB})}{\max(-MAE_{CLB}, +MAE_{CLB})}$ $-MAE_{CLB} = \frac{1}{-N} \sum_{k=1}^{-N} \Delta_k $ $\Delta_k 0; -N \text{ is the number of } \Delta_k < 0$ $+MAE_{CLB} = \frac{1}{+N} \sum_{k=1}^{+N} \Delta_k $ $\Delta_k 0; +N \text{ is the number of } \Delta_k \geq 0$ $\Delta_k = \text{observed}_k - \text{calculated}_k$	[80]

2.3. The Third Weirdness of QSPR/QSAR

Naturally, the contribution of the molecular structure is the key importance to an endpoint. However, any biological activity is a mathematical function of many different conditions and circumstances. In other words, toxicity or pharmaceutical effect is caused by not only molecular structure, but also physicochemical conditions (e.g., temperature, humidity) and circumstances (noise/silence, illumination/darkness). Apparently, one can disagree with the above postulation, but the majority of QSPR/QSAR has built up without taking into account something besides molecular structure.

It is to be noted, however, in some cases, the molecular structure is not informative to build up a predictive model of endpoints [81–95]. Meanwhile, the definition of a model as a mathematical function of experimental conditions (after consultations with experimentalists) is a shorter and consequently more attractive way to solve the corresponding task.

3. Discussion

The above-mentioned unpleasant peculiarities and weirdness are interacting. To avoid unpleasant peculiarities, one should build up a model without the above weirdness, namely, (i) one should study

several different splits (into training and validation sets); (ii) one should select a group of criteria of predictive potential which are agreed with each other; and (iii) one should take into account all conditions which impact corresponding endpoint (not only molecular structure). However, these actions are not enough to solve all problems.

Unfortunately, there are other problems. Fortunately, there are other solutions. The hierarchy of problems in the field of the modeling of various endpoints is not established. One group of researchers believes that the validation of a model is of key importance. Another group believes that the main result is the statistical quality of a model. A third group concentrates on mechanistic interpretation. It is curious, but non-standard tasks and solutions also exist and sometimes these are very important. Examples are below.

3.1. Multi-target QSAR Models

The limitation of almost all QSAR models is that they predict the biological activity for only one endpoint. In other words, traditional QSAR gives a model for the biological activity of drugs against only one parasite species [96], one species of a virus [97], and one type of cancer [98]. The so-called multi-target QSAR as a tool to build up models for several endpoints is suggested [96–98].

Apparently, this conception has attractive advantages, since it provides a user by extending a list of information (i.e., expected numerical data for groups of endpoints, which affect the phenomenon under consideration, e.g., therapeutic effect, inhibition, biocide potential, etc.). Nonetheless, traditional approaches serve as a basis to solve the task of building up multi-target QSARs, e.g., using multiple regression [99], partial least squares (PLS) [100], artificial neural networks (ANN) [101–103], and random forest [104].

It is to be noted, that interest to researches dedicated to multi-target QSAR in drug discovery gradually increases during the past decade, whereas interest to general QSAR in drug discovery is approximately constant. Figure 1 confirms this situation.

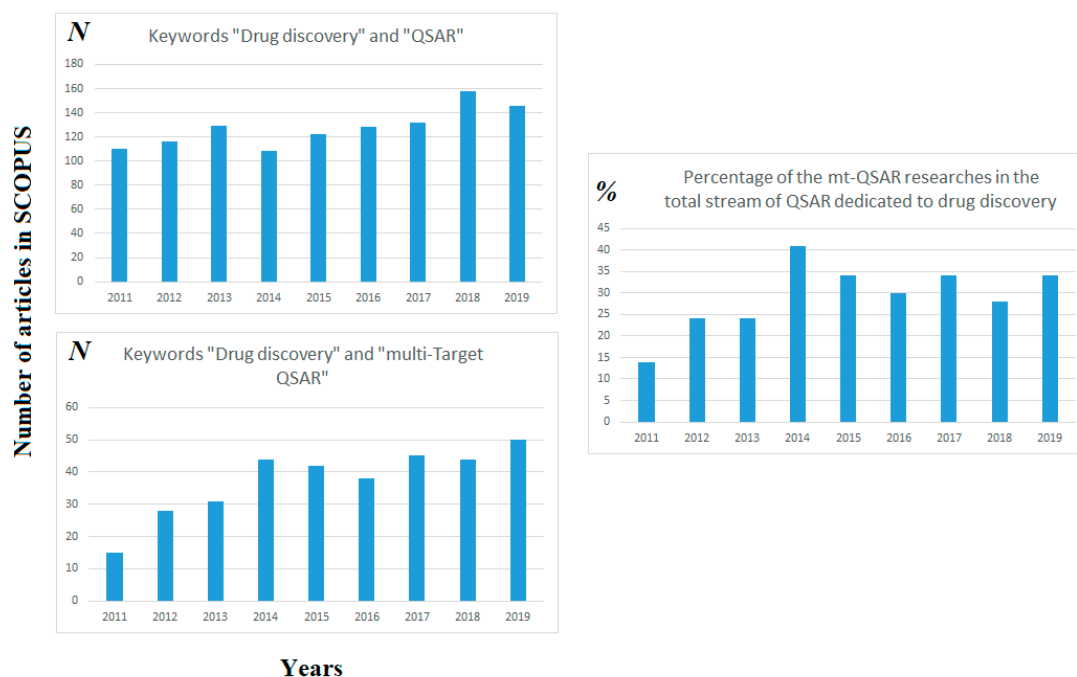


Figure 1. Comparison of frequencies of using quantitative structure–activity relationships (QSAR) and multi-target QSAR (mt-QSAR) in drug discovery researches.

3.2. Similarity of Endpoints

As noted in the previous section, the simultaneous examination of two endpoints is an attractive way in the QSPR/QSAR analysis. In addition to multi-target QSAR, the similarity of endpoints may be a heuristic tool of control of the biochemical knowledge [105–107]. Similarity/dissimilarity of endpoints can be expressed via correlation weights of molecular features extracted from SMILES [105]. In principle, the spectrum of physicochemical conditions with a clear impact on biochemical endpoints (toxicity, therapeutic potential) able to provide hints to establish similarity (dissimilarity) for two endpoints relevant to drug discovery, toxicity, risk assessment, and others.

The similarity of endpoints was analyzed in the literature [105]. The task is to extract molecular features involved in the modeling process, which play an analogous role for corresponding models of (i) mutagenicity; (ii) blood–brain barrier; and (iii) anticancer activity.

3.2.1. Mutagenicity

The endpoint for QSAR analysis is the mutagenic potential. The mutagenic potential in *Salmonella typhimurium* TA98+S9 microsomal repair is represented by the natural logarithm of R, where R is the number of revertants per nanomole (lnR).

3.2.2. Anticancer Activity

The endpoint considered here is IC50 which represents the concentration of the agent necessary to reduce cell viability by 50% against Murine P388 Leukemia (in vitro cytotoxic activity). The endpoint is expressed on a logarithmic scale (pIC50).

3.2.3. Blood–Brain Barrier (BBB)

The database for BBB permeation ($n = 291$) is taken from the literature [105].

QSAR models for the above-listed endpoints are based on the following descriptor:

$$DCW(T^*, N^*) = CW(BOND) + CW(NOSP) + CW(HALO) + CW(PAIR) + \sum CW(S_k) + \sum CW(SS_k) \quad (3)$$

Here, the global SMILES attributes are the following *BOND*, *NOSP*, *HALO*, and *PAIR*. The S_k and SS_k are local SMILES attributes. Tables 4 and 5 contain comments on these attributes. The $CW(BOND)$, $CW(NOSP)$, $CW(HALO)$, $CW(PAIR)$, $CW(S_k)$, and $CW(SS_k)$ are correlation weights of the above-listed attributes.

Table 4. Simplified molecular input-line entry system (SMILES) attributes applied to build up a model.

SMILES Attribute	Comments
S_k	One symbol or two symbols which cannot be examined separately in SMILES, e.g., Cl, Br, etc.
SS_k	A combination of two connected S_k
BOND	Descriptor reflects the presence in SMILES of the following symbols: '@', '=', and '#' (i.e. presence of different bonds)
NOSP	Descriptor reflects the presence of the following chemical elements nitrogen (i.e., symbol 'N'), oxygen (i.e., symbols 'O'), Sulfur (i.e., symbol 'S'), and phosphorus (i.e., symbol 'P')
HALO	Descriptor reflects the presence of fluorine (i.e., symbol 'F'), chlorine (i.e., symbols 'Cl'), bromine (i.e., symbols 'Br'), and iodine (i.e., 'I')
PAIR	Descriptor reflects simultaneous the presence of pair of the above elements (i.e. details related to BOND, NOSP, and HALO, without any details about their places in molecular structure)

Table 5. Generalized representation of above SMILES attributes for Clc1cc(Cl)ccc1C(O)=O.

ID	Attribute	1	2	3	4	5	6	7	8	9	10	11	12	
1	S_k	C	l	
		c	
		l	
		c	
		c	
		(.
		C	l
		*
		c
		c
		c
		l
		C
		(.
		O
		(.
		=
		O
		2	SS_k	c	.	.	.	C	l
c	.			.	.	l	
c	.			.	.	l	
c	.			.	.	c	
c	.			.	.	(.	
C	l			.	.	(.	
C	l			.	.	(.	
c	.			.	.	(.	
c	.			.	.	c	
c	.			.	.	c	
c	.			.	.	l	
c	.			.	.	l	
C	.			.	.	l	
O	.			.	.	(.	
O	.			.	.	(.	
=	.			.	.	(.	
=	.			.	.	=	
3	BOND	B	O	N	D	1	0	0	0	0	0	0		
4	NOSP	N	O	S	P	0	1	0	0	0	0	0		
5	HALO	H	A	L	O	0	1	0	0	0	0	0		
6	PAIR	+	+	+	+	C	l	.	.	O	=	=	=	
		+	+	+	+	C	l	.	.	B	2	=	=	
		+	+	+	+	O	.	.	.	B	2	=	=	

^{*)} Used only "(" (not ')'; ^{**)} Symbols in SS_k are placed according to ASCII code, in order to avoid situation wrong interpretations AB and BA as non-equivalent features.

The scheme of estimation of similarity and dissimilarity for the above-mentioned endpoints demonstrated by Table 6 is adapted from [105].

Table 6. Definition of similarities to models for mutagenicity, anticancer activity and blood–brain barrier (BBB). Here, model-1, denoted m1; model-2, denoted m2. The “m1.1” means first run of optimization for endpoint 1. Each plus denotes a promoter of an increase for endpoints (#1 or #2). Each minus denotes a promoter for a decrease for endpoints (#1 or #2).

	Attributes, SA _k	m1.1	m1.2	m1.3	m2.1	m2.2	m2.3
Mutagenicity (#1) vs. Anticancer Activity (#2)							
1	1.....	+	+	+	+	+	+
2	c...2.....	+	+	+	+	+	+
3	c...(.....	+	+	+	+	+	+
4	3.....	+	+	+	+	+	+
5	C.....	+	+	+	+	+	+
6	1...(.....	+	+	+	+	+	+
7	C...1.....	+	+	+	+	+	+
8	C...3.....	+	+	+	+	+	+
9	Cl...(.....	+	+	+	+	+	+
10	Cl.....	+	+	+	+	+	+
1	c.....	+	+	+	-	-	-
2	O.....	+	+	+	-	-	-
3	O...(.....	+	+	+	-	-	-
4	N...(.....	-	-	-	+	+	+
5	++++N—O===	-	-	-	+	+	+
6	NOSP11000000	-	-	-	+	+	+
7	C...(.....	-	-	-	+	+	+
8	C...C.....	-	-	-	+	+	+
Mutagenicity (#1) vs. BBB (#2)							
1	1.....	+	+	+	+	+	+
2	BOND00000000	+	+	+	+	+	+
3	HALO00000000	+	+	+	+	+	+
4	NOSP10000000	+	+	+	+	+	+
5	1...(.....	+	+	+	+	+	+
6	++++CL—N===	+	+	+	+	+	+
7	-.....	+	+	+	+	+	+
8	=...(.....	+	+	+	+	+	+
9	C...1.....	+	+	+	+	+	+
10	BOND10000000	+	+	+	+	+	+
11	Cl...(.....	+	+	+	+	+	+
12	Cl.....	+	+	+	+	+	+
13	N...+.....	+	+	+	+	+	+
14	N.....	-	-	-	-	-	-
1	O.....	+	+	+	-	-	-
2	O...(.....	+	+	+	-	-	-
3	N...1.....	+	+	+	-	-	-
4	[...+.....	+	+	+	-	-	-
5	NOSP11000000	-	-	-	+	+	+
6	C...(.....	-	-	-	+	+	+
7	C...C.....	-	-	-	+	+	+
BBB (#1) vs. anticancer activity (#2)							
1	C...C.....	+	+	+	+	+	+
2	C...(.....	+	+	+	+	+	+
3	1.....	+	+	+	+	+	+
4	C...1.....	+	+	+	+	+	+
5	C...=.....	+	+	+	+	+	+
6	++++N—B2==	+	+	+	+	+	+
7	C...2.....	+	+	+	+	+	+
8	NOSP11000000	+	+	+	+	+	+
9	1...(.....	+	+	+	+	+	+
10	O...C.....	+	+	+	+	+	+

Table 6. Cont.

	Attributes, SA _k	m1.1	m1.2	m1.3	m2.1	m2.2	m2.3
Mutagenicity (#1) vs. Anticancer Activity (#2)							
11	2...(.....	+	+	+	+	+	+
12	4.....	+	+	+	+	+	+
13	Cl.....	+	+	+	+	+	+
14	Cl...(.....	+	+	+	+	+	+
15	++++S—B2==	+	+	+	+	+	+
16	HALO1000000	+	+	+	+	+	+
17	++++F—B2==	+	+	+	+	+	+
18	++++F—N===	+	+	+	+	+	+
19	HALO1000000	+	+	+	+	+	+
20	N...4.....	+	+	+	+	+	+
21	++++CL—S===	+	+	+	+	+	+
22	(.....	–	–	–	–	–	–
23	O.....	–	–	–	–	–	–
24	O...(.....	–	–	–	–	–	–
25	5.....	–	–	–	–	–	–
26	C...5.....	–	–	–	–	–	–
1	++++Cl—B2==	+	+	+	–	–	–
2	F...(.....	+	+	+	–	–	–
3	++++F—Cl==	+	+	+	–	–	–
4	++++O—B2==	–	–	–	+	+	+
5	2.....	–	–	–	+	+	+
6	=...2.....	–	–	–	+	+	+
7	3...(.....	–	–	–	+	+	+
8	++++O—S===	–	–	–	+	+	+

Table 7 contains numerical measures of similarity and dissimilarity of the corresponding endpoints (Table 6).

Table 7. The matrix of similarity for examining endpoints.

Similarity			
	Mutagenicity	Anticancer Activity	Blood–Brain Barrier
Mutagenicity	41	10	14
Anticancer activity	10	61	26
Blood–brain barrier	14	26	92
Dissimilarity			
Mutagenicity	11	8	7
Anticancer activity	8	24	8
Blood–brain barrier	7	8	52

The similarity of endpoints defined according to suggested scheme can become the beginning of “a next generation” of the QSPR/QSAR evolution.

3.3. Gender-Oriented QSAR Models

Usually, the categorization of eco-toxic effects is related to a different animal (fishes, birds, and insects). However, in addition, at least for animals, categorization related to sex also may be useful from practical and theoretical points of view. QSAR models of carcinogenicity separately for male and female rats can have wide applications for both the agriculture and theoretical biochemistry [69]. The matrix to recognize the difference of corresponding models build up by scheme analogous to the above-mentioned scheme applied for Table 6. The promoters of the increase of pTA50 have been examined separately in the cases of male and female rats. In both cases, symbol ‘1’ means the stable

positive correlation weight, whereas symbol ‘0’ means the stable negative correlation weight. Table 8 contains the results of the above-mentioned computational experiments.

Table 8. Promoters for increase carcinogenicity in male rats (MR) and female rats (FR).

Promoters of Carcinogenicity Increase	Male Rats, MR									Total MR	Female Rats, FR									Total FR
	Split 1			Split 2			Split 3				Split 1			Split 2			Split 3			
	1	2	3	1	2	3	1	2	3		1	2	3	1	2	3	1	2	3	
Molecular features extracted from SMILES																				
1...(.....	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	1	0	1	7
2...(.....	1	1	1	0	0	0	1	1	1	6	0	0	0	0	0	0	0	0	0	0
2...1.....	1	1	0	0	0	1	1	1	1	5	0	0	0	0	0	0	0	0	0	0
C...1.....	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	2
C...2.....	0	1	0	1	1	0	0	1	0	4	0	0	0	0	0	0	0	0	0	0
N...=.....	1	0	0	0	1	1	1	0	0	4	0	0	0	0	0	0	0	0	0	0
N...1.....	0	0	0	1	1	1	0	0	0	3	0	0	0	0	0	0	0	0	0	0
HALO00000000	1	1	1	1	1	1	1	1	1	9	0	0	0	0	0	0	0	0	0	0
BOND00000000	1	0	1	0	0	0	1	0	1	4	1	1	1	1	1	1	1	1	1	9
BOND10000000	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	0	0	0	5
BOND10100000	1	1	1	1	1	1	1	1	1	9	0	0	0	0	0	0	0	0	0	0
Molecular features (invariants) extracted from molecular graph*																				
C5.....0...	0	0	0	1	0	1	0	0	0	2	1	0	0	1	1	1	0	0	0	4
C6.....0...	1	1	1	1	1	1	1	1	1	9	1	1	1	1	1	1	1	1	0	8
NNC-C...101.	1	1	1	1	1	1	1	1	1	9	0	0	0	0	0	0	1	1	0	2
NNC-C...110.	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	5
NNC-C...211.	1	1	1	1	1	1	1	1	1	9	1	1	1	1	1	1	1	1	1	9
NNC-C...303.	1	1	1	1	0	0	1	1	1	7	0	0	0	0	0	0	0	0	0	0
NNC-C...321	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	7
NNC-O...101	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	2
Summation										63										50

*) Detailed description for C5 and C6 represented in [74]; detailed description for NNC-Y ... xxx represented in [80].

Examples of molecular features acted in a different manner for male rats and female rats are the following (i) BOND1010000; (ii) HALO00000000; (iii) NNC-C ... 303.; and (iv) NNC-C ... 321. This information can be useful, e.g., for developers of corresponding biocides. Algorithms able to generate gender-oriented models may have wider applications (e.g., drug design).

3.4. The Simplicity or the Efficiency: Which is Better?

QSAR should be assessed as a surrogate of a real experiment. QSAR aimed to measure an endpoint value. However, to expect adequate prediction physicochemical and biochemical behavior of an arbitrary substance by means of the QSPR/QSAR-model is naive.

Despite the above-mentioned thesis, QSPR/QSAR has become an integral part of modern science as a tool to detect “fuzzy tendencies” in the behavior of groups of substances. This fact logically echoes the theory of fuzzy sets [108]. This is not surprising, as fuzzy set theory has success in solving some problems of QSPR/QSAR analysis [109–111].

One can extract two components in the total big variety of QSAR studies: (i) “extensive” studies and (ii) “intensive” studies. The aim of “extensive” studies is the integration of the results of applying current approaches to solve practical tasks. The aim of “intensive” studies is attempting to develop new conceptions of the QSPR/QSAR analysis. Naturally, a small part of the results of the “intensive” studies gradually become a tool of robust “extensive” studies.

Nowadays, multi-target QSAR is a part of “intensive” studies [96–104]. The development of criteria of the predictive potential of models (Table 3) also is a part of the “intensive” studies. Maybe, search for the similarity of endpoints [105–107], also, will become part of “intensive” QSPR/QSAR researches.

4. Conclusions

The evolution of the field of QSPR/QSAR has two components: intensive and extensive. The intensive component is responsible for developing the quality and epistemology potential of various QSPR/QSAR approaches. The multi-target QSAR is the perspective field of the evolution of the QSAR theory and practices. Other perspective components of the “intensive” evolution of the QSPR/QSAR are (i) applying fuzzy set theory; (ii) developing statistical methods to detect similarity of biochemical endpoints; and (iii) extending “input data” for QSPR/QSAR by means of taking into account experimental conditions and circumstances also can be a component of intensive evolution of the QSPR/QSAR.

Author Contributions: Authors have done equivalent contributions to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors are grateful for the contribution of the project LIFE-VERMEER contract (LIFE16 ENV/ES/000167) for financial support.

Conflicts of Interest: No potential conflicts of interest were reported by the authors.

References

1. Sizochenko, N.; Gajewicz, A.; Leszczynski, J.; Puzyn, T. Causation or only correlation? Application of causal inference graphs for evaluating causality in nano-QSAR models. *Nanoscale* **2016**, *8*, 7203–7208. [[CrossRef](#)] [[PubMed](#)]
2. Wiener, H. Correlation of Heats of Isomerization, and Differences in Heats of Vaporization of Isomers, Among the Paraffin Hydrocarbons. *J. Am. Chem. Soc.* **1947**, *69*, 2636–2638. [[CrossRef](#)]
3. Wiener, H. Relation of the Physical Properties of the Isomeric Alkanes to Molecular Structure. Surface Tension, Specific Dispersion, and Critical Solution Temperature in Aniline. *J. Phys. Chem.* **1948**, *52*, 1082–1089. [[CrossRef](#)] [[PubMed](#)]
4. Wiener, H. Vapor Pressure-Temperature Relationships Among the Branched Paraffin Hydrocarbons. *J. Phys. Chem.* **1948**, *52*, 425–430. [[CrossRef](#)] [[PubMed](#)]
5. Hosoya, H. Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339. [[CrossRef](#)]
6. Katritzky, A.R.; Gordeeva, E.V. Traditional Topological Indexes vs Electronic, Geometrical, and Combined Molecular Descriptors in QSAR/QSPR Research. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 835. [[CrossRef](#)]
7. Balaban, T.; Mills, D.; Ivanciuc, O.; Basak, S.C. Reverse Wiener Indices. *Croat. Chem. Acta.* **2000**, *73*, 923.
8. Ivanciuc, O.; Ivanciuc, T.; Klein, D.J.; Seitz, W.A.; Balaban, A.T. Wiener Index Extension by Counting Even/Odd Graph Distances. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 536–549. [[CrossRef](#)]
9. Torrens, F. Valence Topological Charge-Transfer Indices for Dipole Moments. *Molecules* **2003**, *8*, 169–185. [[CrossRef](#)]
10. Hansch, C.; Fujita, T. ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616–1626. [[CrossRef](#)]
11. Balaban, A. Topological and Stereochemical Molecular Descriptors for Databases Useful in QSAR, Similarity/Dissimilarity and Drug Design. *SAR QSAR Environ. Res.* **1998**, *8*, 1–21. [[CrossRef](#)]
12. Estrada, E. On the Topological Sub-Structural Molecular Design (TOSS-MODE) in QSPR/QSAR and Drug Design Research. *SAR QSAR Environ. Res.* **2000**, *11*, 55–73. [[CrossRef](#)] [[PubMed](#)]
13. Randić, M. Topological indices. In *Encyclopedia of Computational Chemistry*, 1st ed.; John Wiley & Sons: Hoboken, NJ, USA, 1998; Volume 5, pp. 3018–3032. ISBN 978-0471965886.
14. Randić, M. Generalized Molecular Descriptors. *J. Math. Chem.* **1991**, *7*, 155–168. [[CrossRef](#)]
15. Estrada, E.; Uriarte, E. Recent Advances on the Role of Topological Indices in Drug Discovery Research. *Curr. Med. Chem.* **2001**, *8*, 1699–1714. [[CrossRef](#)] [[PubMed](#)]
16. Basak, S.C.; Balaban, A.T.; Grunwald, G.D.; Gute, B.D. Topological Indices: Their Nature and Mutual Relatedness. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 891–898. [[CrossRef](#)] [[PubMed](#)]

17. Patel, H.; Cronin, M.T.D. A Novel Index for the Description of Molecular Linearity. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1228–1236. [CrossRef]
18. Ivanciuc, O.; Ivanciuc, T.; Cabrol Bass, D.; Balaban, A.T. Evaluation in Quantitative StructureProperty Relationship Models of Structural Descriptors Derived from Information Theory Operators. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 631–643. [CrossRef]
19. Randić, M. Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *69*, 6609–6615.
20. Plavšić, D.; Nikolić, S.; Trinajstić, N.; Mihalić, Z. On the Harary Index for the Characterization of Chemical Graphs. *J. Math. Chem.* **1993**, *12*, 235–250. [CrossRef]
21. Estrada, E. Spectral Moment of Edge Adjacency Matrix in Molecular Graphs.1. Definition and Application to the Prediction of Physical Properties of Alkanes. *J. Chem. Inf. Comp. Sci.* **1996**, *36*, 846–849. [CrossRef]
22. Estrada, E. Spectral Moment of Edge Adjacency Matrix in Molecular Graphs. 2. Molecules Containing Heteroatom and QSAR Applications. *J. Chem. Inf. Comp. Sci.* **1997**, *37*, 320–328. [CrossRef]
23. Estrada, E. Spectral Moment of Edge Adjacency Matrix in Molecular Graphs 3. Molecules Containing Cycles. *J. Chem. Inf. Comp. Sci.* **1998**, *38*, 123–127. [CrossRef]
24. Mihalic, Z.; Trinajstić, N. A Graph-Theoretical Approach to Structure-Property Relationships. *J. Chem. Educ.* **1992**, *69*, 701–712. [CrossRef]
25. Diudea, M.V. *QSPR/QSAR Studies by Molecular Descriptors*; Diudea, M.V., Ed.; Nova Science, Huntington: New York, NY, USA, 2001; ISBN 1560728590.
26. Marino, D.J.G.; Peruzzo, P.J.; Castro, E.A.; Toropov, A.A. QSAR Carcinogenic Study of Methylated Polycyclic Aromatic Hydrocarbons Based on Topological Descriptors Derived from Distance Matrices and Correlation Weights of Local Graph Invariants. *Internet Electron. J. Mol. Des.* **2002**, *1*, 115–133. Available online: <http://www.biochempress.com> (accessed on 31 March 2002).
27. Ivanciuc, O. QSAR Comparative Study of Wiener Descriptors for Weighted Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1412–1422. [CrossRef]
28. Broach, J.R.; Thorner, J. High-Throughput Screening for Drug Discovery. *Nature* **1996**, *384*, 14–16.
29. Raichurkar, A.V.; Shah, U.A.; Kulkarni, V.M. 3D-QSAR of novel Phosphodiesterase-4 inhibitors by genetic function approximation. *Med. Chem.* **2011**, *7*, 543–552. [CrossRef]
30. Gaurav, A.; Yadav, M.R.; Giridhar, R.; Gautam, V.; Singh, R. 3D-QSAR studies of 4-quinolone derivatives as high-affinity ligands at the benzodiazepine site of brain GABAA receptors. *Med. Chem. Res.* **2011**, *20*, 192–199. [CrossRef]
31. Sinha, N.; Sen, S. Predicting hERG activities of compounds from their 3D structures: Development and evaluation of a global descriptors based QSAR model. *Eur. J. Med. Chem.* **2011**, *46*, 618–630. [CrossRef]
32. Singh, H.P.; Chaturvedi, A.P.; Sharma, C.S. 3D-QSAR and insilico study: Modeling parameters for designing new selective 12-LO enzymes inhibitors. *Int. J. Pharm. Tech. Res.* **2011**, *3*, 231–236.
33. Clare, B.W.; Supuran, C.T. Carbonic anhydrase inhibitors. Part 61. Quantum chemical QSAR of a group of benzenedisulfonamides. *Eur. J. Med. Chem.* **1999**, *34*, 463–474. [CrossRef]
34. Hameed, A.J.; Ibrahim, M.; El Haes, H. Computational notes on structural, electronic and QSAR properties of [C60]fulleropyrrolidine-1-carbodithioic acid 2; 3 and 4-substituted-benzyl esters. *J. Mol. Struct. THEOCHEM* **2007**, *809*, 131–136. [CrossRef]
35. Tekiner-Gulbas, B.; Temiz-Arpaci, O.; Oksuzoglu, E.; Eroglu, H.; Yildiz, I.; Diril, N.; Aki-Sener, E.; Yalcin, I. QSAR of genotoxic active benzazoles. *SAR QSAR Environ. Res.* **2007**, *18*, 251–263. [CrossRef] [PubMed]
36. Panda, P.; Samanta, S.; Alam, S.M.; Basu, S.; Jha, T. QSAR for analogs of 1,5-N,N'-disubstituted-2-(substituted benzenesulphonyl) glutamamides as antitumor agents. *Internet Elect. J. Mol. Des.* **2007**, *6*, 280–301.
37. Eroglu, E.; Türkmen, H. A DFT-based quantum theoretic QSAR study of aromatic and heterocyclic sulfonamides as carbonic anhydrase inhibitors against isozyme, CA-II. *J. Mol. Graph. Model.* **2007**, *26*, 701–708. [CrossRef] [PubMed]
38. Pasha, F.A.; Cho, S.J.; Beg, Y.; Tripathi, Y.B. Quantum chemical QSAR study of flavones and their radical-scavenging activity. *Med. Chem. Res.* **2007**, *16*, 408–417. [CrossRef]
39. Doweiko, A.M. Is QSAR relevant to drug discovery? *IDrugs* **2008**, *11*, 894–899.
40. Doweiko, A.M. QSAR: Dead or alive? *J. Comput. Aid. Mol. Des.* **2008**, *22*, 81–89. [CrossRef]
41. Maggiora, G.M. On outliers and activity cliffs - Why QSAR often disappoints? *J. Chem. Inf. Model.* **2006**, *46*, 1535. [CrossRef]
42. Johnson, S.R. The trouble with QSAR (or how I learned to stop worrying and embrace fallacy). *J. Chem. Inf. Model.* **2008**, *48*, 25–26. [CrossRef]

43. Topliss, J.G.; Costello, R.J. Chance Correlations in Structure-Activity Studies Using Multiple Regression Analysis. *J. Med. Chem.* **1972**, *15*, 1066–1068. [[CrossRef](#)] [[PubMed](#)]
44. Clark, M.; Cramer, R.D., III. The Probability of Chance Correlation Using Partial Least Squares (PLS). *Quant. Struct. Act. Rel.* **1993**, *12*, 137–145. [[CrossRef](#)]
45. Baumann, K. Chance correlation in variable subset regression: Influence of the objective function, the selection mechanism, and ensemble averaging. *QSAR Comb. Sci.* **2005**, *24*, 1033–1046. [[CrossRef](#)]
46. Toropov, A.A.; Toropova, A.P. QSAR as a random event: Criteria of predictive potential for a chance model. *Struct. Chem.* **2019**, *30*, 1677–1683. [[CrossRef](#)]
47. Tetko, I.V.; Sushko, I.; Pandey, A.K.; Zhu, H.; Tropsha, A.; Papa, E.; Öberg, T.; Todeschini, R.; Fourches, D.; Varnek, A. Critical assessment of QSAR models of environmental toxicity against tetrahymena pyriformis: Focusing on applicability domain and overfitting by variable selection. *J. Chem. Inf. Model.* **2008**, *48*, 1733–1746. [[CrossRef](#)] [[PubMed](#)]
48. Roy, P.P.; Kovarich, S.; Gramatica, P. QSAR model reproducibility and applicability: A case study of rate constants of hydroxyl radical reaction models applied to polybrominated diphenyl ethers and (benzo-)triazoles. *J. Comput. Chem.* **2011**, *32*, 2386–2396. [[CrossRef](#)] [[PubMed](#)]
49. Patel, M.; Chilton, M.L.; Sartini, A.; Gibson, L.; Barber, C.; Covey-Crump, L.; Przybylak, K.R.; Cronin, M.T.D.; Madden, J.C. Assessment and Reproducibility of Quantitative Structure-Activity Relationship Models by the Nonexpert. *J. Chem. Inf. Model.* **2018**, *58*, 673–682. [[CrossRef](#)] [[PubMed](#)]
50. Roy, P.P.; Leonard, J.T.; Roy, K. Exploring the impact of size of training sets for the development of predictive QSAR models. *Chemometr. Intell. Lab. Syst.* **2008**, *90*, 31–42. [[CrossRef](#)]
51. Roy, K.; Mitra, I.; Ojha, P.K.; Kar, S.; Das, R.N.; Kabir, H. Introduction of rm2 (rank) metric incorporating rank-order predictions as an additional tool for validation of QSAR/QSPR models. *Chemometr. Intell. Lab. Syst.* **2012**, *118*, 200–210. [[CrossRef](#)]
52. Masand, V.H.; Mahajan, D.T.; Nazeruddin, G.M.; Hadda, T.B.; Rastija, V.; Alfeefy, A.M. Effect of information leakage and method of splitting (rational and random) on external predictive ability and behavior of different statistical parameters of QSAR model. *Med. Chem. Res.* **2015**, *24*, 1241–1264. [[CrossRef](#)]
53. Ghaemian, P.; Shayanfar, A. Quantitative structure activity relationship (QSAR) of methylated polyphenol derivatives as permeability glycoprotein (P-gp) inhibitors: A comparison of different training and test set selection methods. *Letts. Drug Des. Discov.* **2017**, *14*, 999–1007. [[CrossRef](#)]
54. Martin, T.M.; Harten, P.; Young, D.M.; Muratov, E.N.; Golbraikh, A.; Zhu, H.; Tropsha, A. Does rational selection of training and test sets improve the outcome of QSAR modeling? *J. Chem. Inf. Model.* **2012**, *52*, 2570–2578. [[CrossRef](#)] [[PubMed](#)]
55. Toropov, A.A.; Toropova, A.P.; Puzyn, T.; Benfenati, E.; Gini, G.; Leszczynska, D.; Leszczynski, J. QSAR as a random event: Modeling of nanoparticles uptake in PaCa2 cancer cells. *Chemosphere* **2013**, *92*, 31–37. [[CrossRef](#)] [[PubMed](#)]
56. Veselinović, J.B.; Nikolić, G.M.; Trutić, N.V.; Živković, J.V.; Veselinović, A.M. Monte Carlo QSAR models for predicting organophosphate inhibition of acetylcholinesterase. *SAR QSAR Environ. Res.* **2015**, *26*, 449–460. [[CrossRef](#)] [[PubMed](#)]
57. Toropova, A.P.; Toropov, A.A.; Benfenati, E.; Leszczynska, D.; Leszczynski, J. QSAR model as a random event: A case of rat toxicity. *Bioorg. Med. Chem.* **2015**, *23*, 1223–1230. [[CrossRef](#)]
58. Toropova, A.P.; Toropov, A.A.; Veselinović, J.B.; Veselinović, A.M. QSAR as a random event: A case of NOAEL. *Environ. Sci. Pollut. Res.* **2015**, *22*, 8264–8271. [[CrossRef](#)] [[PubMed](#)]
59. Toropova, M.A.; Raska, I., Jr.; Toropova, A.P.; Raskova, M. CORAL software: Analysis of impacts of pharmaceutical agents upon metabolism via the optimal descriptors. *Curr. Drug Metab.* **2017**, *18*, 500–510. [[CrossRef](#)]
60. Alam, S.; Khan, F. 3D-QSAR, Docking, ADME/Tox studies on Flavone analogs reveal anticancer activity through Tankyrase inhibition. *Sci. Rep.* **2019**, *9*, 5414. [[CrossRef](#)] [[PubMed](#)]
61. Jain, S.; Amin, S.A.; Adhikari, N.; Jha, T.; Gayen, S. Good and bad molecular fingerprints for human rhinovirus 3C protease inhibition: Identification, validation, and application in designing of new inhibitors through Monte Carlo-based QSAR study. *J. Biomol. Struct. Dyn.* **2020**, *38*, 66–77. [[CrossRef](#)]
62. 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 Environ. Res.* **2019**, *30*, 525–541. [[CrossRef](#)]

63. Ahmadi, S.; Mardinia, F.; Azimi, N.; Qomi, M.; Balali, E. Prediction of chalcone derivative cytotoxicity activity against MCF-7 human breast cancer cell by Monte Carlo method. *J. Mol. Struct.* **2019**, *1181*, 305–311. [[CrossRef](#)]
64. Bhargava, S.; Patel, T.; Gaikwad, R.; Patil, U.K.; Gayen, S. Identification of structural requirements and prediction of inhibitory activity of natural flavonoids against Zika virus through molecular docking and Monte Carlo based QSAR Simulation. *Nat. Prod. Res.* **2019**, *33*, 851–857. [[CrossRef](#)] [[PubMed](#)]
65. Weininger, D. SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36. [[CrossRef](#)]
66. Kumar, A.; Chauhan, S. Monte Carlo method based QSAR modelling of natural lipase inhibitors using hybrid optimal descriptors. *SAR QSAR Environ. Res.* **2017**, *28*, 179–197. [[CrossRef](#)]
67. Bhargava, S.; Adhikari, N.; Amin, S.A.; Das, K.; Gayen, S.; Jha, T. Hydroxyethylamine derivatives as HIV-1 protease inhibitors: A predictive QSAR modelling study based on Monte Carlo optimization. *SAR QSAR Environ. Res.* **2017**, *28*, 973–990. [[CrossRef](#)]
68. Kumar, P.; Kumar, A. Monte Carlo Method Based QSAR Studies of Mer Kinase Inhibitors in Compliance with OECD Principles. *Drug Res.* **2018**, *68*, 189–195. [[CrossRef](#)] [[PubMed](#)]
69. Toropova, A.P.; Toropov, A.A. CORAL: QSAR models for carcinogenicity of organic compounds for male and female rats. *Comput. Biol. Chem.* **2018**, *72*, 26–32. [[CrossRef](#)]
70. Kumar, P.; Kumar, A.; Sindhu, J.; Lal, S. QSAR Models for Nitrogen Containing Monophosphonate and Bisphosphonate Derivatives as Human Farnesyl Pyrophosphate Synthase Inhibitors Based on Monte Carlo Method. *Drug Res.* **2019**, *69*, 159–167. [[CrossRef](#)]
71. Manisha; Chauhan, S.; Kumar, P.; Kumar, A. Development of prediction model for fructose- 1,6-bisphosphatase inhibitors using the Monte Carlo method. *SAR QSAR Environ. Res.* **2019**, *30*, 145–159. [[CrossRef](#)]
72. Toropov, A.A.; Toropova, A.P.; Ismailov, T.T.; Voropaeva, N.L.; Ruban, I.N. Extended molecular connectivity: Prediction of boiling points of alkanes. *J. Struct. Chem.* **1997**, *38*, 965–969. [[CrossRef](#)]
73. Toropov, A.A.; Toropova, A.P. QSAR modeling of toxicity on optimization of correlation weights of Morgan extended connectivity. *J. Mol. Struct.: THEOCHEM* **2002**, *578*, 129–134. [[CrossRef](#)]
74. Toropov, A.A.; Toropova, A.P.; Marzo, M.; Benfenati, E. Use of the index of ideality of correlation to improve aquatic solubility model. *J. Mol. Graph. Model.* **2020**, *96*, 107525. [[CrossRef](#)] [[PubMed](#)]
75. Pearson, K. Notes on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.* **1895**, *58*, 240–242.
76. Lin, L.I.-K. Assay validation using the concordance correlation coefficient. *Biometrics* **1992**, *48*, 599–604. [[CrossRef](#)]
77. Golbraikh, A.; Tropsha, A. Beware of q^2 ! *J. Mol. Graph. Model.* **2002**, *20*, 269–276. [[CrossRef](#)]
78. Consonni, V.; Ballabio, D.; Todeschini, R. Comments on the definition of the Q^2 parameter for QSAR validation. *J. Chem. Inf. Model.* **2009**, *49*, 1669–1678. [[CrossRef](#)]
79. Mitra, I.; Roy, P.P.; Kar, S.; Ojha, P.K.; Roy, K. On further application of r_m^2 as a metric for validation of QSAR models. *J. Chemometr.* **2010**, *24*, 22–33. [[CrossRef](#)]
80. Toropova, A.P.; Toropov, A.A. The index of ideality of correlation: A criterion of predictability of QSAR models for skin permeability? *Sci. Total Environ.* **2017**, *586*, 466–472. [[CrossRef](#)]
81. Toropov, A.A.; Toropova, A.P. Quasi-SMILES and nano-QFAR: United model for mutagenicity of fullerene and MWCNT under different conditions. *Chemosphere* **2015**, *139*, 18–22. [[CrossRef](#)]
82. Toropova, A.P.; Toropov, A.A.; Benfenati, E. A quasi-QSPR modelling for the photocatalytic decolourization rate constants and cellular viability (CV%) of nanoparticles by CORAL. *SAR QSAR Environ. Res.* **2015**, *26*, 29–40. [[CrossRef](#)]
83. Toropova, A.P.; Toropov, A.A.; Rallo, R.; Leszczynska, D.; Leszczynski, J. Optimal descriptor as a translator of eclectic data into prediction of cytotoxicity for metal oxide nanoparticles under different conditions. *Ecotoxicol. Environ. Saf.* **2015**, *112*, 39–45. [[CrossRef](#)] [[PubMed](#)]
84. Toropova, A.P.; Toropov, A.A. Mutagenicity: QSAR -quasi-QSAR -nano-QSAR. *Mini-Rev. Med. Chem.* **2015**, *15*, 608–621. [[CrossRef](#)] [[PubMed](#)]
85. Toropova, A.P.; Toropov, A.A.; Benfenati, E.; Korenstein, R.; Leszczynska, D.; Leszczynski, J. Optimal nano-descriptors as translators of eclectic data into prediction of the cell membrane damage by means of nano metal-oxides. *Environ. Sci. Pollut. Res.* **2015**, *22*, 745–757. [[CrossRef](#)] [[PubMed](#)]

86. Toropov, A.A.; Achary, P.G.R.; Toropova, A.P. Quasi-SMILES and nano-QFPR: The predictive model for zeta potentials of metal oxide nanoparticles. *Chem. Phys. Lett.* **2016**, *660*, 107–110. [[CrossRef](#)]
87. Toropov, A.A.; Toropova, A.P.; Begum, S.; Achary, P.G.R. Towards predicting the solubility of CO₂ and N₂ in different polymers using a quasi-SMILES based QSPR approach. *SAR QSAR Environ. Res.* **2016**, *27*, 293–301. [[CrossRef](#)]
88. Toropova, A.P.; Toropov, A.A.; Manganeli, S.; Leone, C.; Baderna, D.; Benfenati, E.; Fanelli, R. Quasi-SMILES as a tool to utilize eclectic data for predicting the behavior of nanomaterials. *NanoImpact* **2016**, *1*, 60–64. [[CrossRef](#)]
89. Toropova, A.P.; Toropov, A.A.; Veselinović, A.M.; Veselinović, J.B.; Benfenati, E.; Leszczynska, D.; Leszczynski, J. Nano-QSAR: Model of mutagenicity of fullerene as a mathematical function of different conditions. *Ecotoxicol. Environ. Saf.* **2016**, *124*, 32–36. [[CrossRef](#)]
90. Toropova, A.P.; Toropov, A.A. Nano-QSAR in cell biology: Model of cell viability as a mathematical function of available eclectic data. *J. Theor. Biol.* **2017**, *416*, 113–118. [[CrossRef](#)]
91. Toropova, A.P.; Toropov, A.A.; Leszczynska, D.; Leszczynski, J. CORAL and Nano-QFAR: Quantitative feature–Activity relationships (QFAR) for bioavailability of nanoparticles (ZnO, CuO, Co₃O₄, and TiO₂). *Ecotoxicol. Environ. Saf.* **2017**, *139*, 404–407. [[CrossRef](#)]
92. Trinh, T.X.; 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*, 183–190. [[CrossRef](#)]
93. Toropov, A.A.; Sizochenko, N.; Toropova, A.P.; Leszczynski, J. Towards the development of global nano-quantitative structure–property relationship models: Zeta potentials of metal oxide nanoparticles. *Nanomaterials* **2018**, *8*, 243. [[CrossRef](#)] [[PubMed](#)]
94. Choi, J.-S.; Trinh, T.X.; Yoon, T.-H.; Kim, J.; Byun, H.-G. Quasi-QSAR for predicting the cell viability of human lung and skin cells exposed to different metal oxide nanomaterials. *Chemosphere* **2019**, *217*, 243–249. [[CrossRef](#)] [[PubMed](#)]
95. Ahmadi, S. Mathematical modeling of cytotoxicity of metal oxide nanoparticles using the index of ideality correlation criteria. *Chemosphere* **2020**, *242*, 125192. [[CrossRef](#)] [[PubMed](#)]
96. Prado-Prado, F.J.; de la Vega, O.M.; Uriarte, E.; Ubeira, F.M.; Chou, K.-C.; González-Díaz, H. Unified QSAR approach to antimicrobials. 4. Multi-target QSAR modeling and comparative multi-distance study of the giant components of antiviral drug-drug complex networks. *Bioorg. Med. Chem.* **2009**, *17*, 569–575. [[CrossRef](#)]
97. Prado-Prado, F.J.; García-Mera, X.; González-Díaz, H. Multi-target spectral moment QSAR versus ANN for antiparasitic drugs against different parasite species. *Bioorg. Med. Chem.* **2010**, *18*, 2225–2231. [[CrossRef](#)]
98. Speck-Planche, A.; Kleandrova, V.V.; Luan, F.; Cordeiro, M.N.D.S. Rational drug design for anti-cancer chemotherapy: Multi-target QSAR models for the in silico discovery of anti-colorectal cancer agents. *Bioorg. Med. Chem.* **2012**, *20*, 4848–4855. [[CrossRef](#)]
99. Liu, Q.; Che, D.; Huang, Q.; Cao, Z.; Zhu, R. Multi-target QSAR study in the analysis and design of HIV-1 inhibitors. *Chin. J. Chem.* **2010**, *28*, 1587–1592. [[CrossRef](#)]
100. Nikolic, K.; Filipic, S.; Agbaba, D. Multi-target QSAR and docking study of steroids binding to corticosteroid-binding globulin and sex hormone-binding globulin. *Curr. Comput. Aided. Drug Des.* **2012**, *8*, 296–306. [[CrossRef](#)]
101. Garcia-Domenech, R.; Zanni, R.; Galvez-Llompart, M.; Galvez, J. Predicting antiprotozoal activity of benzyl phenyl ether diamine derivatives through QSAR multi-target and molecular topology. *Mol. Divers.* **2015**, *19*, 357–366. [[CrossRef](#)]
102. Abdolmaleki, A.; Ghasemi, J.B.; Ghasemi, F. Computer aided drug design for multi-target drug design: SAR /QSAR, molecular docking and pharmacophore methods. *Curr. Drug Targets* **2017**, *18*, 556–575. [[CrossRef](#)]
103. Speck-Planche, A.; Scotti, M.T. BET bromodomain inhibitors: Fragment-based in silico design using multi-target QSAR models. *Mol. Divers.* **2019**, *23*, 555–572. [[CrossRef](#)]
104. Halder, A.K.; Cordeiro, M.N.D.S. Development of multi-target chemometric models for the inhibition of class I PI3K enzyme isoforms: A case study using QSAR-Co tool. *Int. J. Mol. Sci.* **2019**, *20*, 4191. [[CrossRef](#)]
105. Toropov, A.A.; Toropova, A.P.; Benfenati, E.; Salmona, M. Mutagenicity, anticancer activity and blood–brain barrier: Similarity and dissimilarity of molecular alerts. *Toxicol. Mech. Methods* **2018**, *28*, 321–327. [[CrossRef](#)]

106. Toropova, A.P.; Toropov, A.A.; Begum, S.; Achary, P.G.R. Blood brain barrier and Alzheimer's disease: Similarity and dissimilarity of molecular alerts. *Curr. Neuropharmacol.* **2018**, *16*, 769–785. [[CrossRef](#)]
107. Toropova, A.P.; Toropov, A.A.; Veselinović, A.M.; Veselinović, J.B.; Leszczynska, D.; Leszczynski, J. Quasi-SMILES as a Novel Tool for Prediction of Nanomaterials' Endpoints. In *Multi-Scale Approaches in Drug Discovery: From Empirical Knowledge to In silico Experiments and Back*, 1st ed.; Speck-Planche, A., Ed.; Elsevier Science: Amsterdam, The Netherlands, 2017; pp. 191–221.
108. Zadeh, L.A. Fuzzy sets. *Inform. Control* **1965**, *8*, 338–353. [[CrossRef](#)]
109. Kumar, M.; Thurow, K.; Stoll, N.; Stoll, R. Robust fuzzy mappings for QSAR studies. *Eur. J. Med. Chem.* **2007**, *42*, 675–685. [[CrossRef](#)]
110. Pérez-Garrido, A.; Girón-Rodríguez, F.; Bueno-Crespo, A.; Soto, J.; Pérez-Sánchez, H.; Helguera, A.M. Fuzzy clustering as rational partition method for QSAR. *Chemometr. Intell. Lab. Syst.* **2017**, *166*, 1–6. [[CrossRef](#)]
111. Abdolmaleki, A.; Ghasemi, J.B. Inhibition activity prediction for a dataset of candidates' drug by combining fuzzy logic with MLR/ANN QSAR models. *Chem. Biol. Drug Des.* **2019**, *93*, 1139–1157. [[CrossRef](#)]

Sample Availability: Samples of the compounds are not available from the authors.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).