Structural Chemistry (2021) 32:1365–1392
https://doi.org/10.1007/s11224-021-01793-z

REVIEW ARTICLE

Simplex representation of molecular structure as universal QSAR/QSPR tool

Victor Kuz'min[1] · Anatoly Artemenko[1] · Luidmyla Ognichenko[1] · Alexander Hromov[1] · Anna Kosinskaya[1,2] · Sergij Stelmakh[1] · Zoe L. Sessions[3] · Eugene N. Muratov[3,4]

Received: 15 January 2021 / Accepted: 7 May 2021 / Published online: 22 June 2021

Abstract

We review the development and application of the Simplex approach for the solution of various QSAR/QSPR problems. The general concept of the simplex method and its varieties are described. The advantages of utilizing this methodology, especially for the interpretation of QSAR/QSPR models, are presented in comparison to other fragmentary methods of molecular structure representation. The utility of SiRMS is demonstrated not only in the standard QSAR/QSPR applications, but also for mixtures, polymers, materials, and other complex systems. In addition to many different types of biological activity (antiviral, antimicrobial, antitumor, psychotropic, analgesic, etc.), toxicity and bioavailability, the review examines the simulation of important properties, such as water solubility, lipophilicity, as well as luminescence, and thermodynamic properties (melting and boiling temperatures, critical parameters, etc.). This review focuses on the stereochemical description of molecules within the simplex approach and details the possibilities of universal molecular stereo-analysis and stereochemical configuration description, along with stereo-isomerization mechanism and molecular fragment "topography" identification.

Keywords SiRMS approach · Stereochemistry · Chirality · QSAR/QSPR models

Introduction

Molecular modeling is a rapidly developing field of modern theoretical chemistry. There are numerous methods of molecular modeling focused on solving various problems and differing in both strategic approach and software implementation [1]. The modeling of the molecular structure is a necessary step in any QSAR/QSPR study. The descriptors used in such modeling determine the possibilities and success of solving certain QSAR/QSPR tasks. Today, a plethora of different descriptor systems (all of which depend on the models' level of (1D - nD) molecular representation) exist in the aims of accurately describing molecular structure [2]. Widely used Fragment Descriptor Systems [3] characterize each molecule by a set (ensemble) of its various fragments, as each fragment has some influence on any property in question. The advantage of such a descriptor representation is the relative ease of computation and storage of the structural information. Additionally, Fragment Descriptor Systems provide the transparent structural interpretation of their corresponding QSAR/QSPR models.

The authors of this paper have been developing and using their own approach to generate fragment descriptors for more than 25 years and present both the method and its capabilities herein as the Simplex Representation of Molecular Structure (SiRMS) method. A distinctive feature of this approach is the ability to not only to interpret QSAR/QSPR relations structurally, but also in a physical-chemical context. Moreover, the generation of unbound simplexes makes it possible to model

✉ Victor Kuz'min
theorchem@gmail.com

✉ Eugene N. Muratov
murik@email.unc.edu

Luidmyla Ognichenko
ogni@ukr.net

[1] Department of Molecular Structures and Chemoinformatics, A.V. Bogatsky Physical-Chemical Institute NAS of Ukraine, Odessa 65080, Ukraine

[2] Department of Medical Chemistry, Odessa National Medical University, Odessa 65082, Ukraine

[3] UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC 27599, USA

[4] Department of Pharmaceutical Sciences, Federal University of Paraiba, Joao Pessoa, PB 58059, Brazil

mixtures of compounds, molecular ensembles, nanoparticles, etc. Initially, SiRMS was developed not as a direct solution of "structure – properties" problems, but as a tool to describe and analyze stereochemical features of various chiral molecules. Nevertheless, in situations when the investigated property (e.g., biological activity) is connected with chirality, the correct solution to QSAR/QSPR problems is impossible to determine without an exhaustive description of the stereochemistry of the corresponding compounds. In the framework of the simplex approach, a number of fundamental problems concerning stereochemistry have been solved; in particular, the SiRMS method is able to identify any structural stereoisomers with different chirality elements. One section of this review is devoted to detailing this and other solutions of various stereochemical problems using SiRMS.

SiRMS methodology has been applied to the direct solution of QSAR/QSPR tasks for the last 20 years. In our opinion, one of the reasons this approach is so effective is the optimal size of the main fragments (simplexes). Smaller fragments (less than four vertices) are not informative enough to describe the structure of compounds. As the size of molecular fragments increases, their "occurrence" in the compounds of the training set decreases, which leads to an increase in their "uniqueness." The latter leads to a decrease in the variability of the corresponding fragment descriptors and reduces their informative value. Thus, the SiRMS descriptor system is based primarily on 4-vertice fragments (simplexes), although fragments of other sizes have been used in few specific tasks.

The main purpose of our review is to demonstrate the capabilities and effectiveness of SiRMS as it applies to a variety of QSAR/QSPR problems concerning virtual screening aimed prediction and the ensuing attempts to design novel molecules and substances with optimal properties.

Table 1 demonstrates the multitude of scientific directions in which QSAR/QSPR tasks were solved and provides references to relevant publications. The review is based only on publications of the authors, chemists-theorists. However, these publications would have suffered without the immense contributions to the successful application and development of our working from our many colleagues who specialize in the areas of chemistry, virology, pharmacology, toxicology, thermophysics, and material science, as well as other related disciplines.

In the review, we will comment on the most important and interesting publications.

The Methodology of SiRMS

SiRMS—a tool for solving fundamental stereochemical problems

Since Pasteur's pivotal discovery over 170 years ago, the concept of chirality has played a fundamental role in natural science as a whole, but especially in chemistry. The stereochemical knowledge system uses a concept such as configuration to describe the chirality of molecular structures. Although any chemist intuitively understands what the term "stereochemical configuration" means, it is difficult to provide a universal and unambiguous definition of this characteristic.

In [5] an attempt was made to formulate such a definition, as well as to understand a number of questions that arise in the analysis of the "chirality – configuration" relationship and to date they either have not been formulated, or are controversial in nature:

1.What is stereochemical configuration?

2.How to systematize the variety of chiral molecules? (The system of chiral elements of Prelog is very limited and ambiguous).

3.Is it always possible to systematize molecules into homochiral subclasses only based on their chirality?

4.Why, during the configuration of isomerization, does the enantiomer not always pass through an achiral boundary?

Table 1 SiRMS publications of the review authors

Sections review	Subsections review	References
Methodology	Stereochemical problems	[4–7]
SiRMS	Descriptors systems	[8–12]
QSAR tasks	Antiviral activity, antimicrobial activity and antitumor activity	[13–40]
	Pharmacokinetic Parameter	[25], [41–45]
	Affinity for different biological targets	[46–53]
	Different types of toxicity	[44], [54–65]
QSPR tasks	Lipophilicity and water solubility	[44], [66–72]
	Luminescent properties	[73]
	Thermodynamic properties	[74–80]
	Properties of ionic compounds and materials	[81], [82]
	Properties of nanosystems	[63], [83],[84]

The concept of chiral simplexes helped us to understand these problems. As a mathematical object, a simplex is a n-dimensional polyhedron, which is a convex shell (n+1) of points (vertexes of simplex) that do not lie in the (n-1)-dimensional plane [85]. At n = 0,1,2,3 the simplex is a point, a segment, a triangle, a tetrahedron, respectively. Chiral simplexes are not compatible with their mirror images (examples, see Fig. 1).

The simplest point object that can be chiral in the space of a corresponding dimension is the chiral simplex (ChS). In fact, the ChS is an elementary carrier of chirality. [86].

The stereoanalysis procedure we proposed—the representation of a chiral molecule as a system of simplexes (molecular multiplex)—allowed us to solve the above mentioned fundamental stereochemical problems [5 - 7].

To complete the stereoanalysis, we first obtain a spatial figure for the structure of a molecule with N atoms, four vertices, and 0-6 edges and model it with $\frac{N!}{(N-4)}$!4! simplexes, a redundant description. Then we use modified Kahn-Ingold-Prelog rules [6] to identify R,S, and achiral configurations (an example can be seen in Fig. 2). Our representation offers distinct stereoisomer representation for molecules, a distinct advantage over the classical Cahn-Ingold-Prelog (CIP) system (Fig. 2). This also allows for the differentiation of homochirality classes.

For a more detailed approach, see the original publications [4–7, 87].

The SiRMS method represents a chiral center with 5 simplexes wherein each atom is assigned a canonical number by known algorithms [87]. This representation can be used to rank the simplexes by the precedence of atoms in them (Fig. 3). This representation of single chiral center compounds can order the simplexes by their precedence and can be highly useful in determining enantiomers and their respective stereochemical configuration.

For the molecule in Fig. 2, we see a great example of the applications of simplexes. The top three ranked simplexes have the same configuration and therefore highlight common stereochemical features of the molecules. This system can be applied to any 3D structure of any molecule and so, all stereochemical peculiarities are considered.

It is well-known that the presence of chirality is a prerequisite for the existence of living matter. However, it is surprising that the molecules typical for living nature, like proteins and nucleotides, have different stereochemical configuration in their chiral centers. This introduces a hit of contradiction, given that the origin of life is due to one source of chirality. To some extent, the use of SiRMS alleviates this contradiction. As can be seen in Fig. 4, most of the simplexes have the same configuration (3 out of 5 are bolded) when comparing multiplexes describing the chiral centers of select biopolymers. This suggests that the corresponding biopolymers are largely stereochemically similar.

Furthermore, it is known that the CIP system only analyzes the environment of the chiral center, ignoring the nature and therefore inducing issues with the identification of molecular enantiomers. The proposed stereoanalysis procedure considers all atoms. As exemplified in Fig. 5, the central atom, as well as its surroundings, is crucial in determining the stereochemistry of the entire molecule.

Figure 5 also displays that with the same mutual position of the substituents, the nature of the asymmetric center (X) significantly affects the features of the stereochemical configuration. In all four examples, the stereochemistry of the molecules is different. These features can be found in the study of the stereochemistry of processes, in which the central atom of the tetrahedral chiral structure is an active participant.

Stereoanalysis can also serve as a convenient tool to evaluate stereochemical relationships (topicity) between different fragments within a molecule [88]. To assess the topicity of the atom or group pair strength, it is necessary to analyze the sequence of simplexes derived. For a molecule of n atoms the number of simplexes in which one n is included is equal to (n-1)!/(n-4)! 3!. For 5-atomic halogen-substituted methanes, each atom is included in only 4 simplexes. By the example presented in Figure 6, we see that for homotopic hydrogens the corresponding sequences of simplexes are the same. In this case all the simplexes are achiral (0), although for more complex chiral molecules identical sequences of chiral R and S simplexes will be seen. For enantiotopic atoms, the corresponding sequences are opposite, i.e. for their chiral simplexes, the configurations are necessarily different; if in one case R, then in another case S and vice versa. If the corresponding sequences also include achiral simplexes, their configuration is denoted as 0.

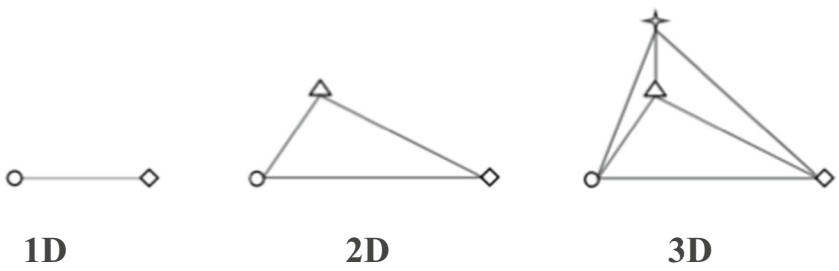


Fig. 1 Chiral simplexes of different dimensions (1D–3D)

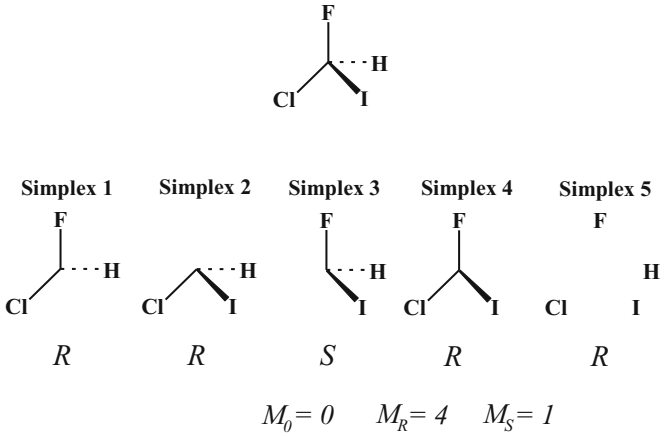
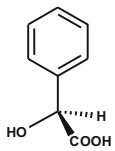
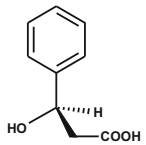


Fig. 2 An example of two compounds that are represented differently based on their stereochemical with both Cahn-Ingold-Prelog rules and simplex representation

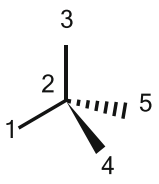
		
Cahn-Ingold-Prelog configuration	S-isomer	R-isomer
Configuration based on simplexes	RSSSS	RSSRR

For more complex chiral molecules, pairs of diastereotopic atoms in corresponding sequences will have simply different combinations of symbols R, S, 0.

Thus, the simplex sequences will simply identify different "topicity" relationships. It is important to remember that these relations characterize only the spatial (stereochemical) environment of topologically equivalent atom pairs.

In this review we do not have the opportunity to discuss the stereochemical configuration (SC) concept in detail. For each chemist, it is obvious that SC is a peculiar invariant of chiral molecules, on the basis of which it is possible to identify different stereoisomers and evaluate their stereochemical similarity (for example, subclasses of homochirality). Sequences of simplexes (R, S or 0) in the order of their seniority, discussed above, can be used as such invariants. For the simplest chiral systems, the chiral simplexes' SC reflects the duality caused by chirality (the two steric series enantiomers R and S). For multiatomic chiral molecules, the number of steric series is determined by the number of simplexes in these molecules. Thus, it is obvious that the representation of the whole variety of chiral structures by two classes (S–"left" and R–"right") is in most cases artificial and formal. Complex chiral structures usually have left and right features simultaneously. As mentioned above, simplex sequences examine all these stereochemical features. Unfortunately, such sequences are too long (n!/(n-4)!4!) and redundant in terms of stereoisomer identification. This is due to the fact that in the complete sequence, some simplexes are interdependent. Therefore, to describe the SC, it is enough to use mutually unbound simplexes, which make up shorter sequences. The corresponding procedure is described below based on a simple example already mentioned [7].

After assigning the canonical numbers, the independent simplexes are indexed based off their vertices, and are mapped off the face of the preceding simplex. An example here describes a set of N-3 independent simplexes.



RSRRR

Atom ranks in simplexes	Simplex precedence	Simplex configuration
1 2 3 4	1	R
1 2 3 5	2	S
1 2 4 5	3	R
1 3 4 5	4	R
2 3 4 5	5	R

Fig. 3 Using the stereoconfiguration of a hypothetical molecule with one chiral center (numbers are canonical numbers of atoms obtained with conventional algorithms) to rank the simplexes where the enantiomer would be SRSSS

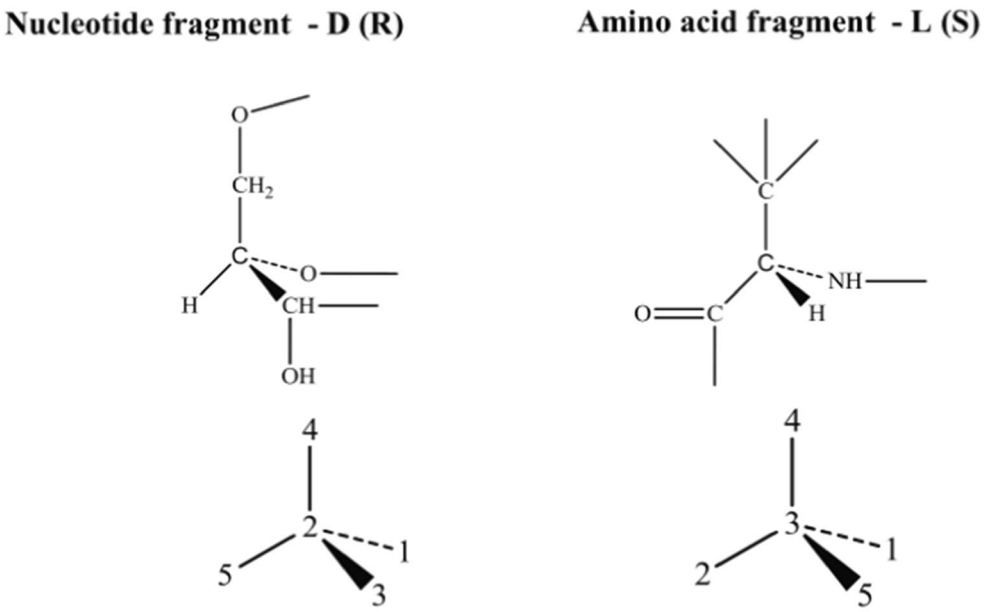


Fig. 4 Demonstration of the similarity of stereochemical configurations for biopolymers

Simplexes	Nucleotide	Amino acid
1,2,3,4	S	R
1,2,3,5	S	S
1,2,4,5	R	S
1,3,4,5	R	R
2,3,4,5	R	R

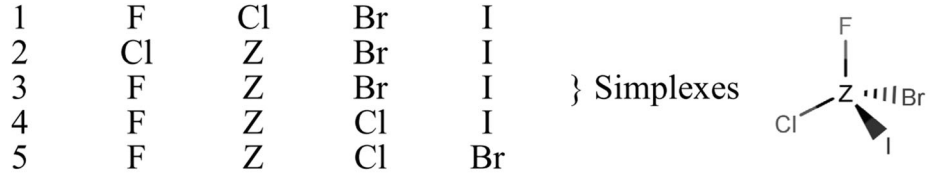


Fig. 5 Stereo-analysis of chiral molecules considering the nature of the chiral center

Z	Simplex configuration				
	1	2	3	4	5
C	*R*	*R*	*S*	*R*	*S*
Si	*R*	*R*	*R*	*S*	*R*
Ge	*R*	*S*	*R*	*R*	*S*
Pb	*R*	*S*	*R*	*S*	*R*

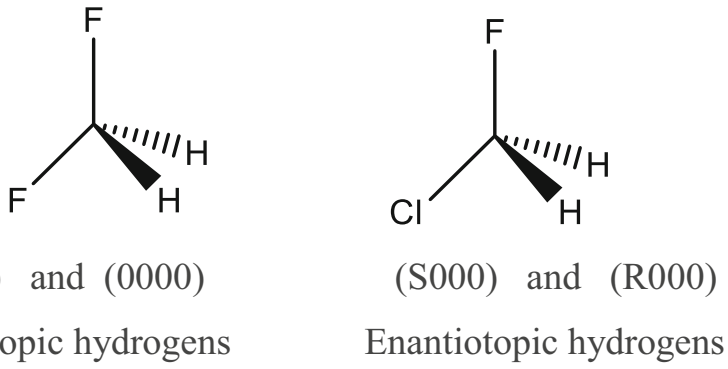
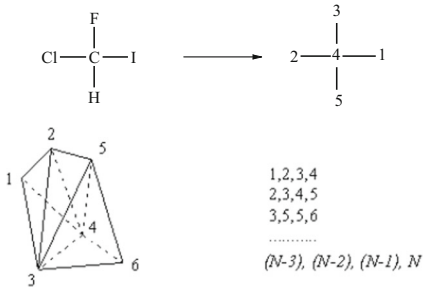
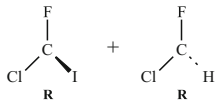


Fig. 6 Simplex sequence for homotopic and enantiotopic hydrogens in halogen-substituted methanes

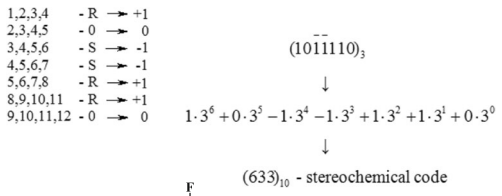


For this molecule, there are two stereochemically similar simplexes:



To identify these configurations, a stereochemical code of +1 for the R configuration, -1 for the S configuration, or 0 for achiral simplexes is assigned to each simplex so that a balanced ternary system can explicitly number and define the stereochemical configuration [89]. For convention, this number is easily converted to decimal notation. This number may be easily converted into the customary decimal notation.

For example,



For our molecule,  the stereochemical code is $(11)_3$ or $(4)_{10}$.

In Fig. 7, more complex chiral molecules are given and their stereochemical configurations are identified using the appropriate stereochemical codes (SC).

Figure 8 depicts some examples showing how the corresponding sequences of simplexes are changed for different stereoisomeric relationships.

The figure clearly shows that for conformers, the sequences of simplexes are identical, while for enantiomers they are opposite, and for diastereomers they are simply different.

As such a fundamental phenomenon, chirality manifests itself not only in the three-dimensional world, but also in spaces of other dimensions. It is obvious, for example, that the oriented segment (vector) is chiral in one-dimensional (1D) space, and the non-uniform triangle is chiral on the plane (in 2D space).

Examples of stereoisomer relations for linear molecules (conditionally 1D objects) and flat molecules (conditionally 2D objects) are given in Fig. 9.

Cis-trans isomers are actually erythro-threo isomers for two-dimensional chiral systems. One should not think that the presence of chirality in spaces with less than 3 dimensions is speculative or virtual. It is possible to create conditions for quite specific molecular systems when such chirality actually manifests itself. For example, a non-mathematical mesophase built of similarly directed rod-shaped molecules is a typical example of a 1D chiral system. Due to intermolecular interactions in the condensed phase, such extended molecules cannot be reoriented relative to each other. A similar situation can arise for asymmetric planar molecules identically oriented in Langmuir films.

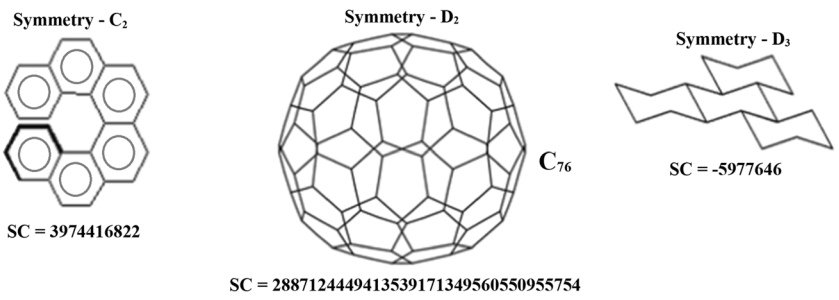


Fig. 7 Stereochemical configuration of chiral symmetrical molecules

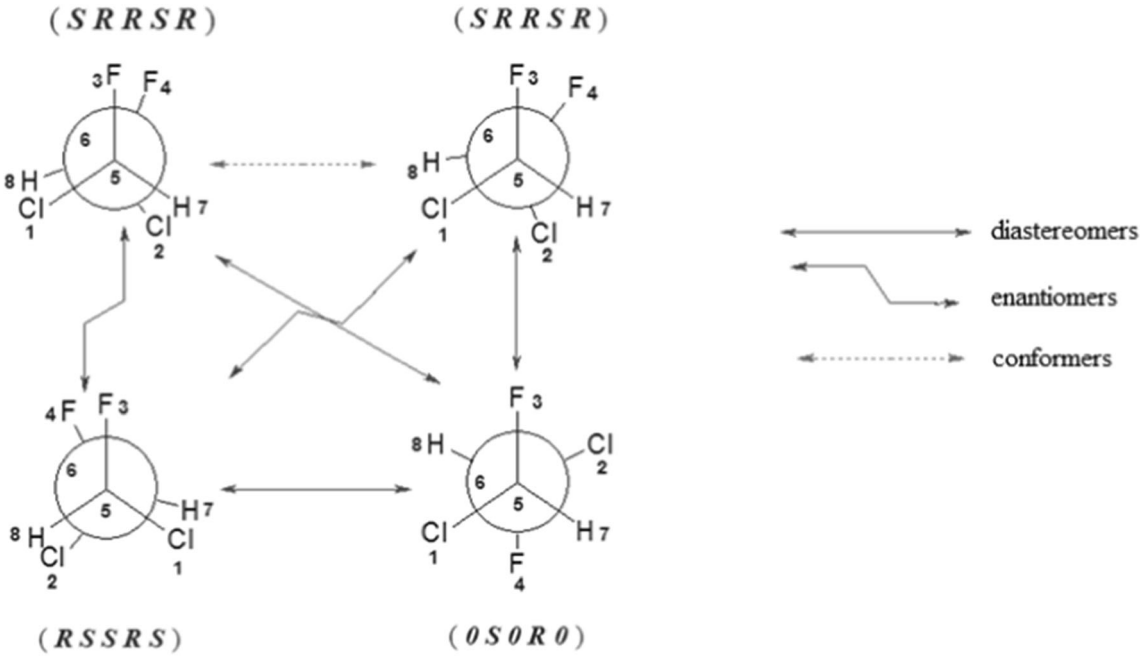
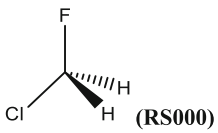


Fig. 8 Examples of describing various stereoisomeric relationships within the simplex approach

As the stereochemical section concludes, it is important to mention another fundamental result of SiRMS which follows from the analysis of the various works [4 - 7]. A new positive chirality criterion has been formed. In accordance with the symmetrical (negative) criterion for the presence of chirality, the necessary and sufficient condition is the absence of the object (molecule) mirror-rotating axes S_n. According to the positive criterion, an object (molecule) is chiral if its structure contains chiral simplexes, and if there are several of them and they have different configurations (R/S), their overall impact on chirality should not be compensated. Achiral objects (molecules) in addition to achiral simplexes, may also include chiral simplexes in their structure. However, the latter, in this case, should form a conditional mesoform, that is, compensate for each other's influence. This can be clearly seen from the example below:



(RS000)

Simplex descriptors for solving various QSAR/QSPR tasks

If we do not focus only on chiral simplexes, which are important for stereochemical problems, but instead consider all possible types of tetratomic molecular fragments, then from their totality, it is possible to generate fragment descriptors for use in various QSAR/QSPR tasks. In the framework of SiRMS, any molecule can be represented as a system of different simplexes (tetratomic fragments of fixed composition, structure, chirality and symmetry) [8 - 11]. An important and distinctive feature of our approach is that when identifying the vertices of simplexes, we use more than the labels reflecting symbols of atoms. Within SiRMS the vertices of simplexes can be characterized by weight parameters reflecting different properties of atoms such as but not limited to the partial charge, electronegativity, lipophilicity, and electronic polarizability. In these cases, the labels of the simplex vertices reflect belonging to a certain range of values of the corresponding property (see details below).

It is obvious then, that the descriptor representation of compounds depends on the level of its molecular model (1D–4D):

- 1D models reflect the formula/composition of a molecule
- 2D models incorporate structural information but only to the limited topological surface. Nonetheless, these topological models provide insights into all possible conformations and are therefore sufficient to address > 90% of existing QSAR/QSPR tasks.
- 3D-QSAR models consider the spatial shape of a molecule, but only for one conformer. These models are common but the analyzed conformer is not usually selected intentionally.
- 4D-QSAR addresses the issues for 3D-QSAR by analyzing the same information for a set of conformers as opposed to one specific conformer.

The details of how SiRMS are addressed in each dimensional model are described and depicted below (see Fig. 10).

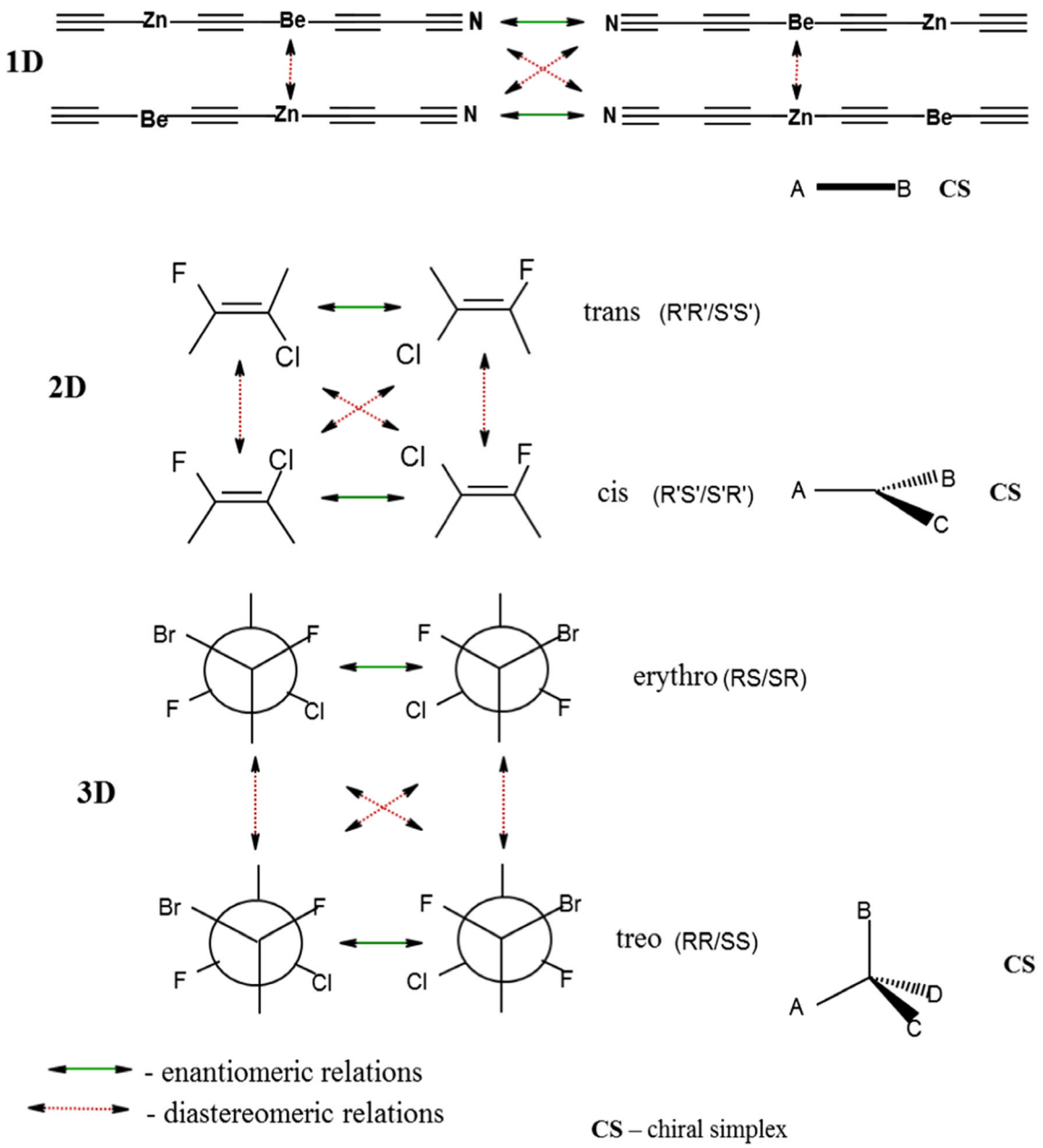


Fig. 9 1D–3D chiral molecular structures

1D models For 1D models, with the compound ($A_aB_bC_cD_dE_eF_f$ . . .), the simplex descriptor (SD) ($A_iB_jC_lD_m$) , is K = f(i)×f(j)×f(l)×f(m), where, for example, f(i) = a!/((a−i)!×i!). A quadruple is assumed for a simplex of four atoms, but smaller fragments can assume i, j, l, or m to be equal to zero as necessary.

2D models Due to their ability to consider bond nature, connectivity, and conformers, 2D models can differentiate atoms of simplexes based on an atom's individuality, partial charge, lipophilicity, atomic refraction, or ability to hydrogen bond (see Fig. 10) [90–92]. The properties with real values, such as charge or lipophilicity, are set into discrete groups and the number of groups (G) is used as a variable tuning parameter (typically G=3-7).

A critical ability of SiRMS is to be able to consider atoms by not only their nature, but also by their surroundings. To accomplish this, sundry variants are included and analyzed considering certain functions, functional groups, or identities of an atom that may not be evident from the nature alone. One great example of this is the marking of atoms that are H-bond donors or acceptors, as mentioned above.

Therefore, the SD of 2D models is fixed by the molecules composition and topology. Other structural parameters for fragment size could be used for 1D or 2D QSAR, but we have found that maintaining 1-4 atomic fragments is

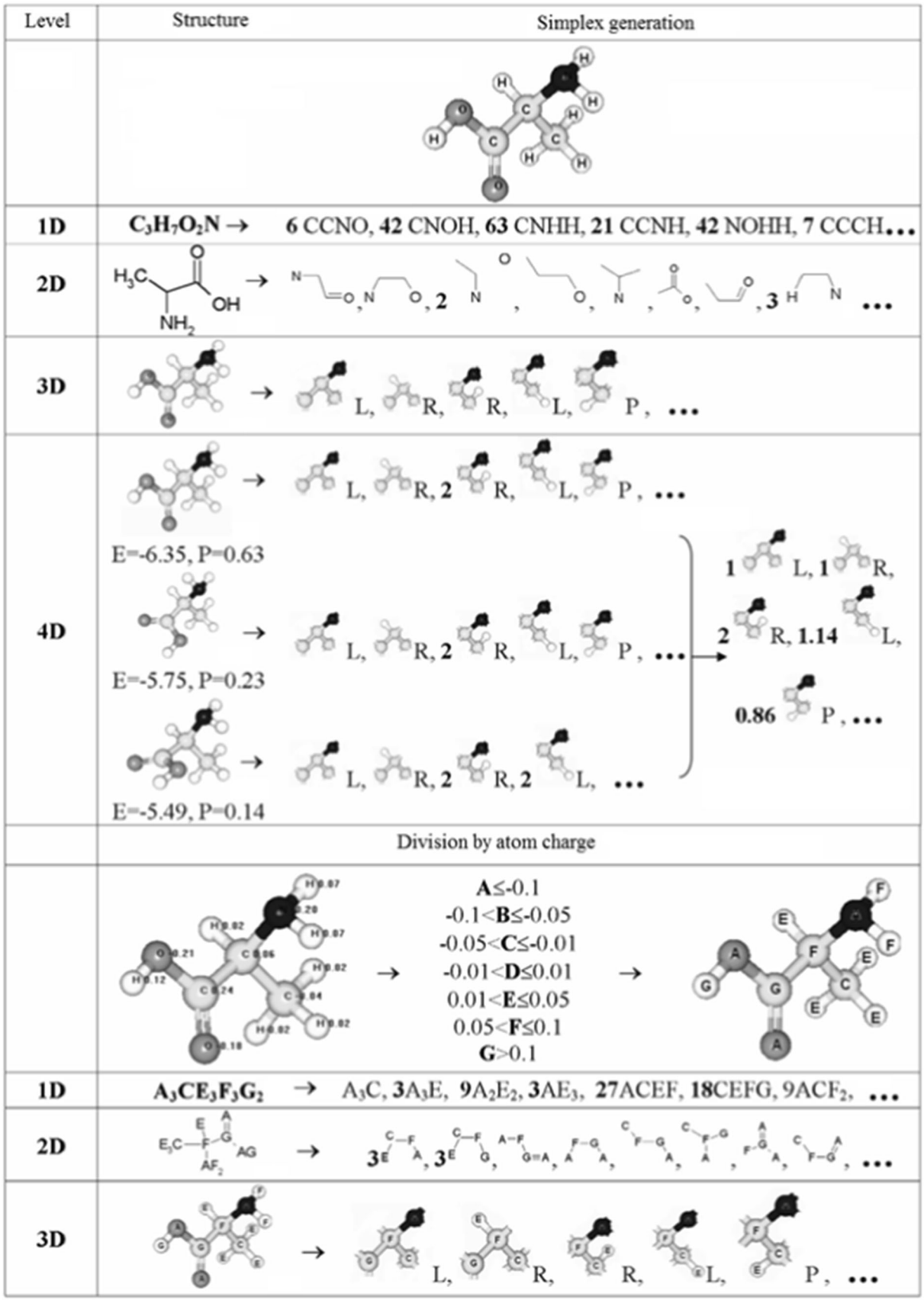
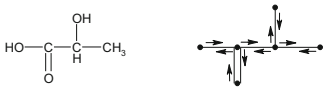


Fig. 10 Depiction of the development of simplex descriptors at varying dimensional levels

ideal to not over fit the model or decrease the predictivity and/or AD.

2D information-topological models The introduction of the molecular informational field [93] allows for the superposition of a complex object, such as a molecule, over a field of its components (elements, atoms, etc.). This ideology is crucial when combined with dimensionless weight parameters and provides a framework for the influence of individual atoms on each other. The properties of each molecule can express themselves on each atom in the molecule in a quantifiable way. Given the ability to map a molecule and the respective forces within it, this is a highly useful tool especially when modeling molecular structure at the 2D level. As seen below, each vertex offers information that extends only to the edge of the graph, but that evaluates all relations between each atom.



Similar to the 2D informational potential (IP) calculation [93], the topological potential (IP) of *i*-th atom can be represented as:

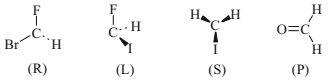
$$IP_i = w_i \cdot \sum_{j=1}^{n} \left(\frac{\sum_m lb\left(\frac{r}{2R_{ij}+1}\right)}{m} \right)$$

where m is the number of all possible paths between every atom pair, n is the number of atoms in the given molecule, R_{ij} is the number of bonds between the i-th and j-th atoms (path length), w_i is the weighed parameters describing any property (p) of the atoms, $w_i = p_i / \sum_{i=1}^{n} p_i$ (w_i=1 in the case of unweighed IP), and r is the maximal path length between atoms for the investigated set of molecules.

A central aspect of the simplex approach is the incorporation of informational field characteristics into atom differentiation. When considering the atoms nature and the topology of the molecule, evaluating scaled properties (charge, lipophilicity, refraction etc.) could prove beneficial for the understanding of atomic mutual influence.

2.5 D models In an analogous manor, stereochemical moieties could also impact biological activity. If a compound contains a chiral center on the atom X (X = C, Si, P, etc.), the special marks X^A, X^R, X^S (A—achiral X atom, R—"right" surrounding of the X atom, S—"left" surrounding of the X atom) can represent the stereochemical information. This extra information elevates the knowledge of a 2D model by adding stereochemical information. Then, X is differentiated into X^A, X^R, X^S, and the different atoms of X are analyzed in the model separately. These models are referred to as 2.5D because they include both topological (molecular graph) and stereochemical information. However, if the differentiation occurs due to some physical–chemical properties (e.g., partial charges, lipophilicity) then the atoms X^A, X^R, X^S will be leveled as in 2D models. To encompass all results, differentiated simplexes have been considered individually and in combination with those differentiated by physical–chemical properties.

3D models As mentioned above, the 3D level also considers the stereochemistry of the molecule and so simplexes can be described as right (R), left (L), symmetrical (S), and plane (P) achiral.



Modified CIP rules can be referenced in establishing stereochemical configurations [6]. The SD at this level is equal to the number of simplexes of fixed composition, topology, chirality, and symmetry.

4D models The SD of 4D-QSAR models are calculated based on the summation of the products of descriptor values for each conformer (SD_k) and the probability of the realization of the corresponding conformer (P_k) of N conformers.

$$SD = \sum_{k=1}^{N} (SD_k \cdot P_k)$$

P_k can also be defined by its energy equation [94],

$$P_k = \left\{ 1 + \sum_{i \neq k} \exp\left(\frac{-(E_i - E_k)}{RT} \right) \right\}^{-1}, \sum_k P_k = 1$$

where E_i and E_k are the energies of conformations of i and k, respectively. The energy of the conformers is assessed within a 5-7 kcal/mol energy band. The entirety of this SD accounts for the probability that any 3D conformer would actualize and so the SD can be considered with other whole- molecule spatial descriptors (e.g., characteristics of inertia ellipsoid, dipole moment).

nD models for mixtures Mixtures interactions do not occur in the same way as other interactions, as the reactivity is variable. This is also amplified by synergistic or anti-synergistic mechanisms towards a biological target [95]. Again, in this case, SiRMS can improve the ability of QSAR modeling for molecular mixtures and ensembles. One important differentiation is

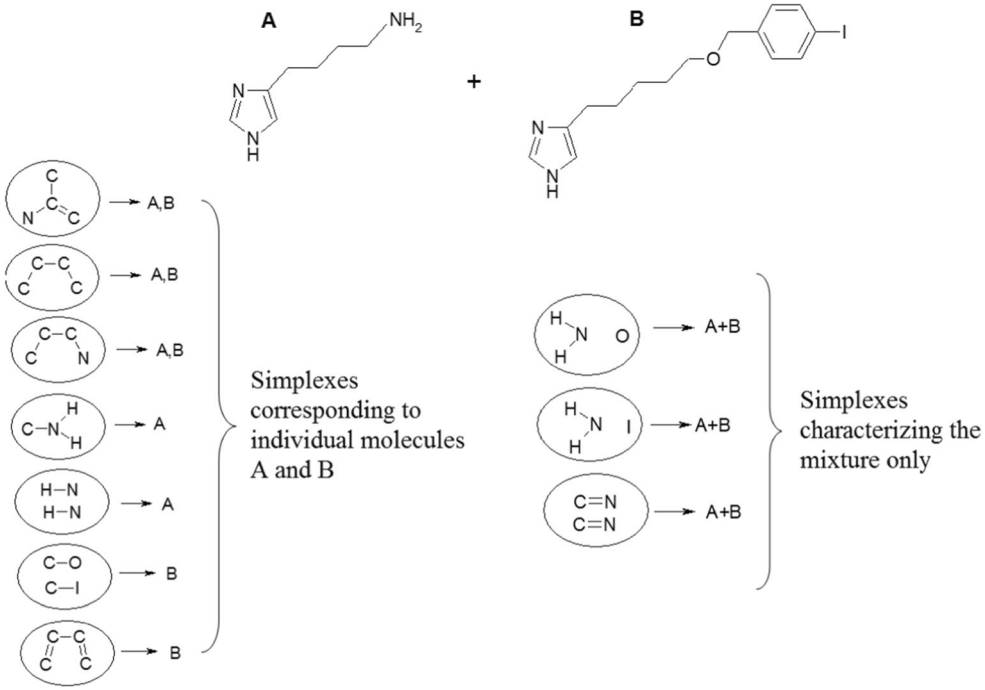


Fig. 11 Example of the structural description of the mixture

to identify what molecules parts of unbound simplexes belong to. If the part belongs to a different molecule, this provides insight into the characterization of pairs of molecules. These serve as structural descriptors for the mixture of compounds (Fig. 11) and provide for the analysis of synergism and competition as it applies to a biological target. This approach is applicable for nD-QSAR models where n = 1–4 but when individual compounds are introduced, they must be represented through the mixture of two similar molecules to maintain the descriptor system [96]. For mixtures with more than two components, one must utilize simplexes with intermolecular bonds.

QSAR models based on simplex descriptors

QSAR models of antiviral, antimicrobial, and antitumor activity

Our first work [14] that used simplex descriptors in QSAR studies of antiviral and antitumor activity was published in 2002. Based on 3D simplexes, 4D-QSAR models were built for 63 compounds, including macrocyclic pyridinophanes and their acyclic analogs, synthetic nucleosides, and a number of well-known antiviral drugs (ambenum, deiteforin, etc.). The target properties were set to study anti-influenza activity in vitro through the reproduction inhibition of the A/Hong Kong/1/68 (H3N2) and the antiviral activity of herpes simplex type1 (HSV-1) and adenovirus 5 (Ad5). The compounds tested in vitro at the National Cancer Institute (Bethesda, Maryland, USA) were investigated for anticancer activity across 60 cell lines of leukemia, CNS cancer, prostate cancer, breast cancer, melanoma, non-small cell lung cancer, colon cancer, ovarian cancer, and renal cancer and were expressed as the percent of control cell growth.

A more detailed QSAR analysis of anticancer activity is described in [18]. In all cases, the statistic characteristics for QSAR of PLS (partial least squares) models were satisfactory (R=0.92-0.97; cross-validation coefficient CVR=0.63–0.83).

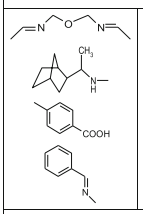
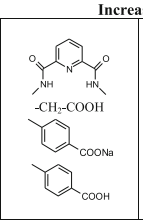
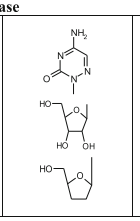
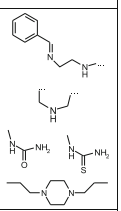
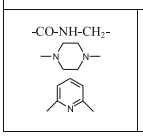
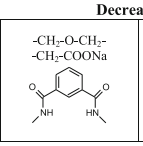
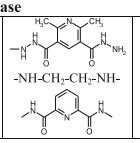
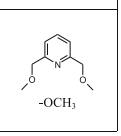
The main result of this work was that for each type of activity, fragments were identified that both increased and decreased the studied properties (see Table 2).

In [13], to evaluate the antiviral activity (focused on Influenza A/Hong Kong/1/68 H3N2) of the above compounds, a set of QSAR models with different molecular levels (2D to 4D) were constructed using the PLS method within the framework of SiRMS. The results of a comparative statistical analysis of these models are given in Table 3.

Obviously, the simplest 2D-QSAR relationships give quite acceptable results, both in terms of the adequacy of the models and their predictive ability. As will be seen from the subsequent discussion, in general, in our practice of various QSAR/QSPR studies we restrict ourselves to SiRMS descriptors of 2D molecular models.

Most of our studies of antiviral activity prior to 2010 are summarized in a review [26]. In addition to the previously mentioned anti-influenza activity and antiherpetic activity of macrocyclic pyridinophanes, in this review the antiherpetic activity of N,N′-(bis-5-nitropyrimidyl) ispirotripiperazine derivatives [19], inhibition of human rhinovirus 2 replication [20] and

Table 2 The molecular fragments which increase and decrease anticancer and antiviral activity

Anti-influenza	Anti-HSV-1	Anti-adenovirus	Anti-cancer
Increase			
			
Decrease			
			

coxsackievirus B3 replication [23] by [(biphenyloxy)propyl] isoxazole derivatives are discussed.

Another consideration herein is the research of anti-HIV activity by artificial ribonucleases. Artificial ribonucleases include compounds with the tetrapeptide Glu–X–Arg–Gly–$OC_{10}H_{21}$ and Glu–X–Lys–Gly–$OC_{10}H_{21}$ structures, where X = Gly, β-Ala, 4-aminobutanoic acid, 6-aminohexanoic acid and p-aminobenzoic acid. With the objective of inactivating viral genome RNAs, the QSAR analysis of antiviral activities of various artificial ribonucleases contributed to the molecular design of new peptide anti-HIV agents [22]. [40] completed a SAR analysis of the antiviral activities of tetrahydro-2(1H)-pyrimidinones against the fowl plague virus (FPV) and the vaccinia virus (VV).

For all the QSAR problems considered in the review [26], it was shown that the corresponding models are effective for both the virtual screening of new antiviral agents and for their molecular design. It is important to note that several of these newly designed antiviral agents have been synthesized and tested. Their experimentally determined activity, in most cases, corresponded to the predictions of QSAR models (see, for example, [20, 29].

Table 3 Statistical characteristics of the QSAR models where R^2—correlation coefficient, Q^2—cross validation correlation coefficient, R^2_{test}—correlation coefficient for test set, S_{ws}—standard error of a prediction for work set, S_{test}—standard error of a prediction for test set, A—number of PLS latent variables, N—number of descriptors in the model

Level	R^2	Q^2	R^2_{test}	S_{ws}	S_{test}	A	N
2D	0.94	0.85	0.99	0.47	0.22	2	13
4D	0.97	0.88	0.98	0.32	0.28	3	18
3D	0.98	0.95	0.98	0.30	0.33	4	14

Of the later works, [27, 30, 32, 36] deserve special attention for their discussion of the QSAR analysis of antiviral combinations against poliovirus, ebolavirus, and three enteroviruses (including poliovirus again).The QSAR studies against poliovirus alone used SiRMS mixture modeling and the PLS method to predict the antiviral effects of the binary combinations of eight picornavirus replication inhibitors in vitro. For this model, eightfold external cross validation was performed and returned CV, Q^2_{ext} = 0.67–0.93. The 2D structures were analyzed and found that fragments such as 2-(4-methoxyphenyl)-4,5-dihydrooxazole or the combination of N-hydroxybenzimidoyl and 3-methylisoxasole promoted antiviral activity. The resulting consensus model found combinations of enviroxime with pleconaril, WIN52084, and rupintrivir and the mixture of rupintrivir with disoxaril to exhibit the highest inhibition of poliovirus 1 replication [27, 30].

The QSAR models built to screen ~ 17 million compounds against ebolavirus particle entry into human cells was also based on SiRMS descriptors. Of the 102 hits selected for experimental testing, 14 compounds displayed IC_{50} values <10 μM (some having 10-fold selectivity against host cytotoxicity) and range from FDA-approved drugs to clinical candidates with non-antiviral indications to compounds with novel scaffolds and no previously known bioactivity. [36] Then, QSAR models surveying the anti-viral activity of nitrobenzonitrile derivatives against coxsackievirus B1, coxsackievirus B3 and poliovirus 1 returned a Matthew's correlation coefficient of 0.9. The results introduced the importance of nitrogen containing substituents on the 5-nitrobenzonitrile moiety for greater anti-viral activity [32].

The outbreak of a novel human coronavirus (SARS-CoV-2) has evolved into global health emergency, infecting hundreds of thousands of people worldwide. In 2020, there were many publications devoted to the search for drugs against SARS-CoV-2. In our works [33, 35, 37] dealing with SARS-CoV-2, we used QSAR models based on SiRMS descriptors. Given the 96% sequence identity and 100% active site conservation between the main protease (M^{pro}) of SARS-CoV-2 and SARS-CoV, we developed QSAR models to assess the inhibitory activity of all drugs in the DrugBank database against the SARS-CoV M^{pro}.

In our virtual screening, forty-two compounds were consensus computational hits. In subsequent experimental screenings, NCATS coincidentally tested 11 of our 42 hits in a cytopathic assay (https://opendata.ncats.nih.gov/covid19/) and found cenicriviroc, proglumenetacin, and sufugolix to be active with AC_{50} concentrations of 8.9 μM, 8.9 μM (tested again independently at 12.5 μM), and 12.6 μM respectively. These independent results endorse the abilities of QSAR modeling in the work to elicit anti-COVID-19 drug candidates.

Another undervalued approach to the battle against SARS-CoV-2 is the use of synergistic antiviral drugs. Modern AI can

be used to design drug combinations with known synergistic antiviral activities without expensive and laborious testing. One option is to use mixture specific SiRMS descriptors in QSAR models. The utilization of this technique with 38 drugs identified 281 combinations with anti-COVID-19 potential [37]. Of these, twenty binary mixtures were selected for binary experimental testing, and once the necessary infrastructure is in place twenty treble combinations will be tested.

At the end of this section, we briefly comment on QSAR studies using SiRMS descriptors of 4-thiazolidone derivatives (about 70 compounds) to assess their antimicrobial activity [24]. *Candida albicans S(I), Citrobacter freundii (II), Klebsiella pneumoniae (III), Pseudomonas aeruginosa (R (IV)and S (V)strains) and Staphylococcus aureus MSSA(VI)* were the reference organisms for our PLS QSAR models. The R^2 =0.843–0.989, Q^2 =0.679–0.864, and R^2_{test} =0.744–0.943 so the molecular fragments were analyzed based on their association to the activity. It was found that any naphthalene fragment is detrimental to activity and indole fragments are indicative of highly active compounds. Finally, the influence of a heterocyclic system evolution on the antimicrobial properties of 4-thiazolidones derivatives was also established (Fig. 12).

QSAR models of various types of toxicity

A significant portion of the publications where SiRMS was used in QSAR models is devoted to surveying the toxicity of various organic compounds [44, 54–67]. Together with our American colleagues, a series of QSAR studies considered the toxicity of high-energy nitroaromatic compounds [54, 55, 57, 62]. Twenty-eight nitroaromatic conpounds were chosen to compare the non-additive effects of fragments on toxicity through SiRMS based 1D- QSAR. The LD_{50} for rats *in vivo* was used as the toxicity parameter. For all but the additive PLS QSAR models, the statistics were satisfactory (R^2 = 0.81–0.92; Q^2 = 0.64–0.83; R^2_{test} = 0.84–0.87). The success of these models and failure of the additive models speaks to the importance of the non-additive modeling. This was clealry demonstrated where the toxicity of the molecules was determined based on the relationship between the nitro group and the presence/absence of other substituents, not just the presence nitro group. For example, hydroxyl and fluorine

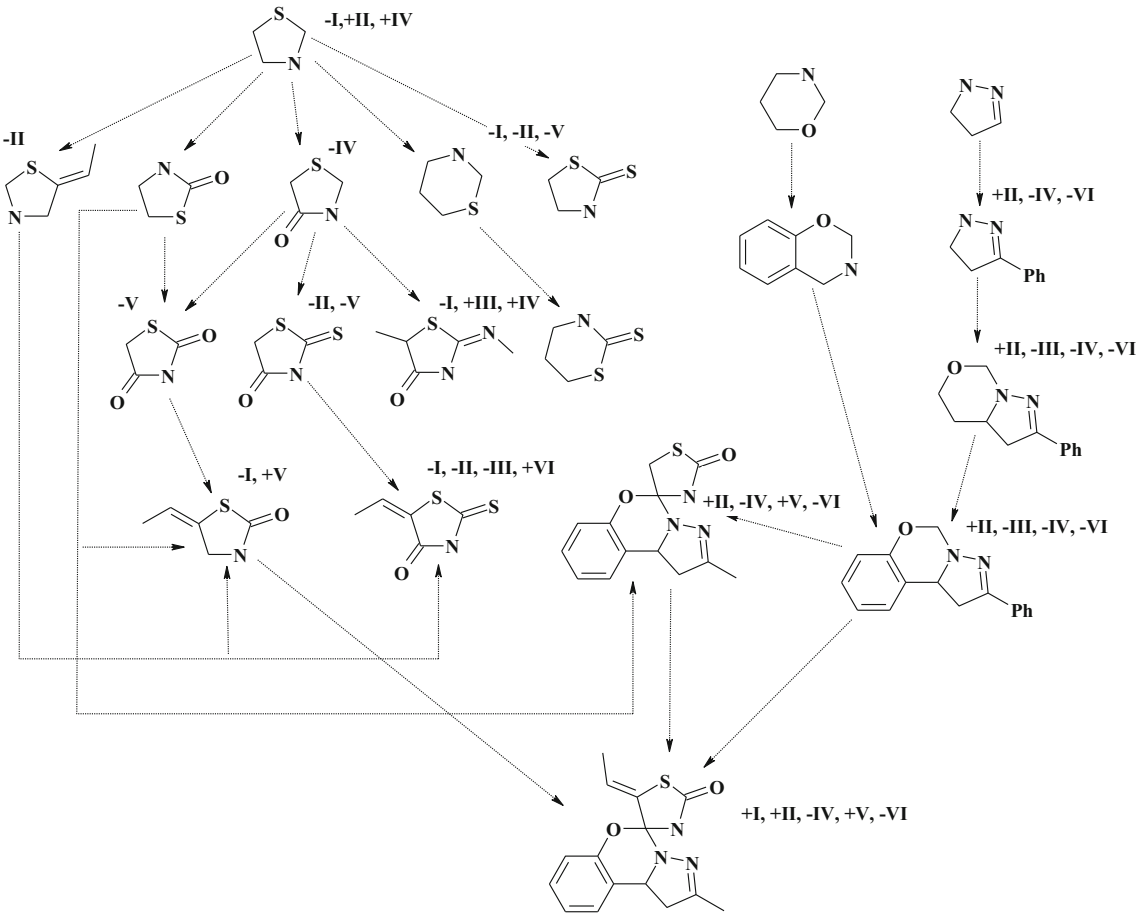


Fig. 12 The influence of a heterocyclic system evolution on antimicrobial activity ("+" indicates the strengthening of antimicrobial properties; "-" signifying the weakening of antimicrobial properties; I – VI as the investigated activities)

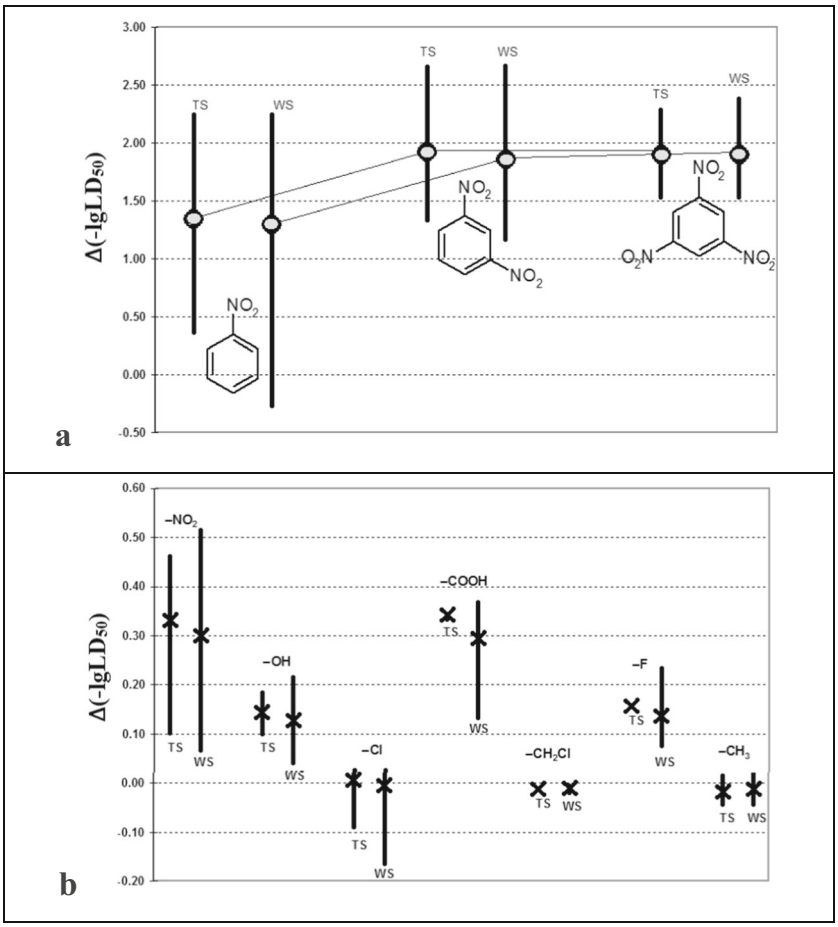


Fig. 13 Molecular fragment contributions $\Delta(-\lg LD_{50})$ to toxicity change: (a) nitroaromatic fragments; (b) substituents in benzene ring (TS: training set; WS : work set)

substituents increase toxicity and a methyl group decreases decreases toxicity, while cholrine was fairly neutral [54]. These observations were consistent with [55] in which 2D QSAR was performed and found the toxicity to depend on both substituent position and nature. More examples of fragments that impacted toxicity can be seen in Fig. 13. While

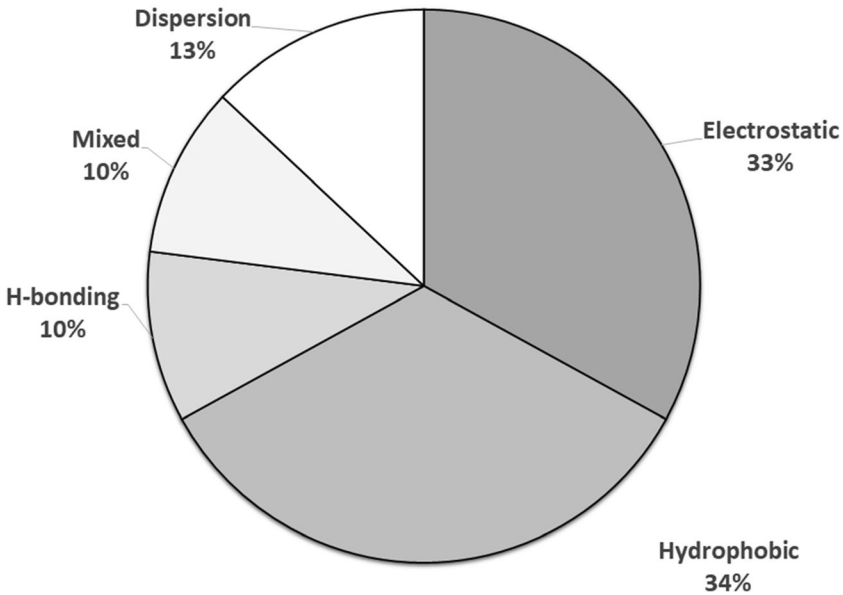


Fig. 14 Relative influences of some physical-chemical factors on the variation of toxicity estimated on the basis of consensus model

mutual influence of the substiuents does play a crucial role, the toxicity can be mediated through C-H fragments on the aromatic ring.

Toxicity was considered again in [57] using SiRMS based PLS QSAR models to analyze the 50% inhibition growth concentration, $IGC_{50,\ of}$ 95 diverse nitroaromatic against the ciliate *Tetrahymena pyriformis.* These validated models worked to classify different substituents based on their effects on toxicity, evaluate the structural descriptors of toxic compounds, and consider physical-chemical factors contributing to toxicity. As seen in Fig. 14, hydrophobic and electrostatic interactions of toxicants and their biological target are the most important factors of the interactions (see Fig. 14). Hence, it can be presumed that compound transport, which relys on lipophilicity, and the interaction of nitroaromatic compounds with their targets, which function through electrostatic, are key mechanisms in the toxicity of a nitroaromatic.

The toxicity of nitroaromatics was then computationaly examined in the context of environmental hazards. The QSAR/QSPR models built accounted for type and position of aromatic ring substituents as well as aqueous solubility, lipophilicity, Ames mutagenicity, bioavailability, blood–brain barrier penetration, aquatic toxicity on *Tetrahymena pyriformis* and acute oral toxicity on rats. Overall, nitroaromatics with electron-accepting substituents, halogens, or amino groups are the most environmentally hazardous, especially if the coumpound is hydrophobic [62].

The reproductive toxicity of various organic compounds was studied in [59]. Molecular structures were described using 2D simplex descriptors and were used with the toxicity parameter Lowest Effective Levels (LEL, mg/kg/day) leading to a miscarriage on administration by gavage. The final consensus QSAR model was adequate (R^2= 0.89), with acceptable predictive power (R^2_{test} = 0.72). The most interesting result is the identified toxiforic fragments that determine reproductive toxicity (Fig. 15).

The work in [58], featured several different computatinal techniques to predict drug hepatotoxicity in rats. The models were built using both chemical descriptors (including SiRMS descriptors) and toxicogenomics profiles. The external test set displayed a correct classification rate of 68–77% after 5-fold external cross validation and points towards the ability of models to both predict chemical factors and respond to acute treatment-induced changes in transcript levels accurately on short term assays.

Despite the common idea that QSAR models are "black boxes," [61] displays direct interpretability of the models and the meaning of structural alerts. Regardless of whether the derivation of strucural alerts were based on QSAR modeling or expert-based, experimental case studies displayed that alerts were simply hypotheses of possible toxicological effects and were not entirely trustworthy. To combat this, the authors propose a synergistic method that utilizes both structural alerts and highly validated QSAR models to accurately assess which chemicals may cause skin sensitization from repeated exposure.

To examine a chemically diverse set of compounds for skin sensitizers, a QSAR model using the Random Forest, SiRMS, and Dragon descriptor techniques was developed. The model was able to discrimate sensitizers from nonsensitizers 77–88% of the time after external validation while maintaining a broad AD, specificity of 85%, and sensitivity of 79% and has screened the Scorecard database for experimental validation [64].

The relationship and thought to be correlation between skin permeability and skin sensitization has been discreditied both experimentally and through QSAR modeling [65].

QSAR models of pharmacokinetic parameters of biologically active substances

Pharmacokinetic parameters are important characteristics of biologically active substances that describe the entry of a drug into the body, its transformation, and excretion from the body. It is obvious that any potential drug, in addition to its specific activity, must be non-toxic and have acceptable pharmacokinetic characteristics. The prediction of such characteristics and the estimation of the influence of structural parameters is an important part of QSAR modeling. A number of our works (see Table 1) are devoted to solving these problems on the basis of SiRMS. In particular, [42, 43] discuss the influence of structure on the pharmacokinetic properties of 1,4 - benzodiazepine tranquilizers.

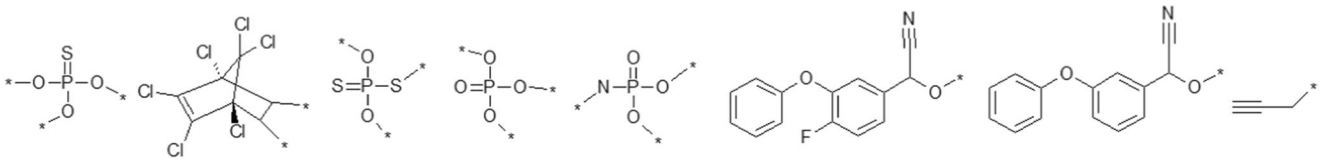


Fig. 15 Structural fragments increasing toxicity. The symbol * in structural fractures correspond to the binding site of this fragment with another part of the molecule

Originally, QSAR models were intended to approximate bioavailability, elimination half-life, clearance, and distribution volume in the human organism for the development of benzodiazepine drugs. Certain trends, such as lipophilic aromatics having high $\tau_{1/2}$ values, similar patterns in distribution volume, clearance, and refractivity, and opposite patterns between bioavailability and clearance emerged from these models. Now in modern production, drugs are classified by the Biopharmaceutics Classification System [97] based on thir water solubility and membrane permeability. This largely concerns the solubility, intestinal permeability, and dissolution rate of oral drug absorption. Like other chemical properties, QSAR tools can be used to model the properties responsible for these trends to help expedite preliminary screening of new compounds into their respective BCS classification [41]. Furthermore, using SiRMS, QSAR models can also contribute to the planning and production of compounds that would effectively permeate the blood-brain barrier (BBB). Based on [45], highly polar groups discourage the molecules ability to cross the BBB, and the presence of halogens and aromatic fragments increases this permeation.

QSAR models of the affinity of molecules (ligands) to various receptors

The biological action of a molecule requires its interaction with a biological target. One possible type of biological target is a receptor. The efficacy of most drugs depends on the affinity to the corresponding receptors. Thus, the prediction of affinity and the analysis of the structural factors determining it are important tasks of medicinal chemistry and QSAR modeling in particular. Even with the advances in mental illness treatments, anxiolytics and antidepressants remain an important field to investigate and evolve. In particular, the exploration of serotonin 5-HT_{1A} receptors works to discover ligands to help regulate anxiety, fear conditions, and depression. We utilized SiRMS methodology to test 346 ligands (Fig. 16) in an affinity QSAR model for 5-HT_{1A} receptors [46].

The relative influences (T_j) of simplex descriptors were calculated (Table 4). Some of the simplexes and corresponding structural fragments are summarized in Table 4.

See work [46] for further details, but the main trends percieved include the low affinity of 5-HT_{1A} receptors towards substituents in the para-position and polycyclic aromatic and

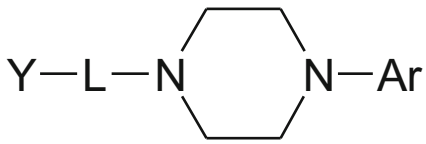
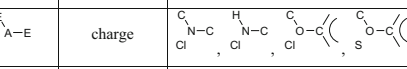
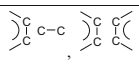
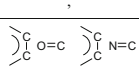
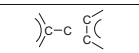


Fig. 16 The general formula of investigated compounds is, where Ar is the aromatic substituents, Y is the different cyclic substituents, and L is a carbohydrate linker $-(CH_2)_n-$

Table 4 The values of the relative influence of simplex descriptors and the corresponding ranges of their values

Simplex	Atom property	Examples of structural fragments	Relative influence (T_j)
	charge		0.31
	charge		0.94
	lipophilicity		1.50
	lipophilicity		-1.00
	atom type		-0.50 -1.50

heteroaromatic fragments, and their high affinity for p-electronodonor substituents in the ortho-position, bulky saturaed fragments, and polymethyl chains of 4 or 5 monomers. This information could prove extremely useful in the design or optimization of compounds with a desired affinity.

The high polyfunctionality of peripheral benzodiazepine receptors (PBDRs) involved in immunomodulation, cholesterol and porphyrin transport, heme and neurosteroid biosynthesis, calcium homeostasis, mitochondrial oxidation, cell proliferation, apoptosis, neurological and psychiatric disorders, raises interest in these receptor ligands.

For the quantitative analysis of the structure-affinity relationship to PBDR with the synthesized compounds (Fig. 17), a QSAR approach based on the simplex representation of the molecular structure was used [47].

It follows from the analysis of QSAR models that the presence of an amide or carboxyl group in the substituent R^1 and piperazyl and acylpiperazyl groups in position R^2 of the 1,2-dihydro-3*H*-1,4-benzadiazepine-2-one molecule leads to a decrease in affinity. The presence of a nitroaniline fragment in position R^3, bromine in position R^4, and a methoxycarbonyl group in the R^1 substituent contribute to an increase in affinity.

To antagonize the inhibition of platelet aggregation through $\alpha_{IIb}\beta_3$, [50] detailed the in silico and in vitro testing of QSAR nominated compounds. The consensus screening highlighted three hits against the closed form of the receptor. These results were validated experimentally after synthesis

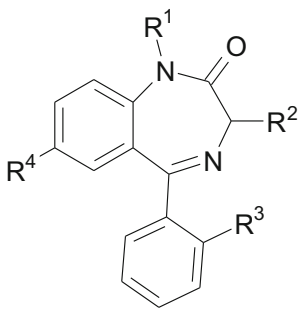
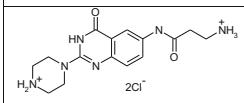
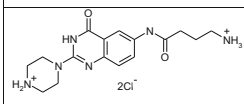
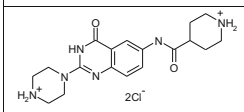


Fig. 17 1,2-dihydro-3*H*-1,4-benzadiazepine-2-one derivatives

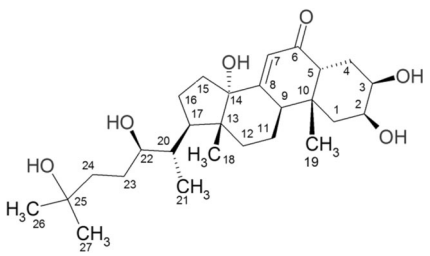
Table 5 Experimentally determined affinity of $\alpha_{IIb}\beta 3$ computationally suggested antagonists on the closed form receptor

Compound	Affinity for $\alpha_{IIb}\beta_3$, IC_{50},nM
	5.0 ± 0.8
	2.2 ± 0.3
	3.8 ± 0.4
Tirofiban	2.4 ± 0.4

and exhibited higher affinity than Tirofiban, a commercial antithrombotic (Table 5).

QSAR models in which stereochemical features of the molecules were directly taken into account (the so-called 2.5D-QSAR models, see the "Simplex descriptors for solving various QSAR/QSPR tasks" section) were developed to study the affinity of steroids to the CBG receptor (Cramer sample of 31 steroids) using a sample of 78 ecdysteroids whose affinity to the ecdysone receptor (EcR) was studied based on cell line indicators for Drosophila melanogaster BII [4]. The relative contribution of chiral simplexes in both models was 18–19%, which implies that the stereochemical features of the ligands play an essential role in the interaction of steroids with the corresponding receptors.

The stereochemical interpretation of the QSAR models performed allowed us to identify the chiral centers in steroid molecules and the changes in the stereochemical configuration that are the most critical for affinity. For example, for ecdysone receptors, changing the S-configuration at atom 22 of the steroid skeleton to R significantly decreases activity, while changing the configuration at atom 25 has almost no effect on the affinity; both enantiomers exhibit almost identical activity.



QSPR models based on simplex descriptors

As follows from Table 1, SiRMS-based descriptors have been used to solve a wide variety of QSPR challenges. In this section, we consider the lipophilicity and aqueous solubility, thermodynamic properties of substances, properties of ionic compounds, and the properties of nanoparticles.

QSPR models of lipophilicity and aqueous solubility

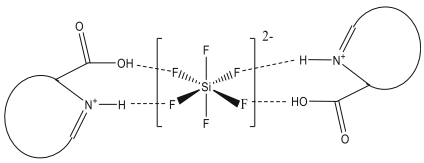
Lipophilicity, and consequentially the quantitative characteristic of lipophilicity, $LogK_{ow}$, is a crucial component in the understanding of the absorption, distribution, metabolism and elimination of many chemicals. This makes it a vital datapoint in most studies, however the experimental estimation of $LogK_{ow}$ is very costly. Hence, the determination of $LogK_{ow}$ prior to experiments is a worthy endeavor. As a result, many theoretical approaches have attempted this feat, but have done so incorrectly by assuming the $LogK_{ow}$ of molecules follows additive schemes. Therefore, [69] applies SiRMS methodology with Random Forest modeling into a 2D-QSPR of nearly 11000 organic compounds. The model was validated four times externally and was particularly strong in predicting strongly polar nitrogen containing compounds. Here it is crucial to highlight once again that the additive scheme would only account for 33% of the important parameters for these calculations.

Similarly, the aqueous solubility of organic compounds is paramount across several disciplines but is again highly costly both in time, labor, and money, not to mention difficult and dangerous. Accordingly, in [70] the authors created a SiRMS QSPR model to first predict the value of k parameter in the linear equation $lgS_w=kT+c$, where S_w is the value of solubility and T is the value of temperature and to secondly use Random Forest to create a robust and efficient model. Following cross validation and external testing, the model delivered slightly better predictive abilities compared to the quantum chemical and thermodynamically driven COSMO-RS approximation [98].

A number of our publications are devoted to more specific questions pertaining to aqueous solubility [66 - 68]. In one situation, we considered nitroaromatic compounds for military purposes, as its solubility in water poses a serious environmental threat. Particularly, in [68], PLS models were built on 135 training compounds and SiRMS methods. For the 155 tested compounds, the $R^2_{test} = 0.81$ (comparable to the ability of EPI Suite™ 4.0) and the 2D descriptors produced a well-fitted and robust QSPR model with $R^2 = 0.90$ and a $Q^2 = 0.87$.

The complex salts, ammonium hexafluorosilicates [71], proved to be interesting objects for the study of aqueous solubility. Understanding that the presence of hydrophilic groups plays a key role in H-bonding and increases the aqueous solubility of compounds, we paid special attention to the influence of hydrogen bonds in the dissolution process. The conclusions readily apply to organic compounds but are complicated when considering organic salts or ammonium compounds [99]. [71] worked to develop SiRMS QSPR models to screen for the water solubility of ammonium hexafluorosilicates and to

identify the main structural and physico-chemical factors impacting these values. The QSPR models point towards the negative influence of interionic H-bonds as well as the strength of the ammonium hexafluorosilicates ion pair from the $N^+H\cdots(SiF_6)^{2-}$ interaction.



These interpretations coincide with generally accepted physico-chemical theories surrounding the effect of ammonium cations on the water solubility of the corresponding salts along with qualitative data of previous experimental works.

QSPR models of thermodynamic properties of substances

A series of publications where SiRMS descriptors were used to build QSPR models were devoted to the thermodynamic properties of substances: the boiling temperatures, critical parameters, second virial coefficients, and adsorption parameters [74–80]. What distinguished these publications from other similar works was the demonstration of applications to mixtures of compounds (SiRMS for mixtures is described in the "Simplex descriptors for solving various QSAR/QSPR tasks" section.). In particular, [74] are devoted to QSPR modeling of boiling and condensation temperatures of two-component mixtures. For mixtures, these temperatures coincide only for compositions of azeotropes (Fig. 18).

The QSPR models were built using the 67 pure liquids and 167 mixtures from the Korean Data Base [100]. Due to the variable nature of point representation, the 167 mixtures translated to 3185 data points. The matrix managed a sparsity degree of 92.5% by incorporating only 167 mixtures, with some compounds appearing in different mixtures up to 25 times. The models were externally validated using "points out", the isolation of the boiling point temperature (T_b) predictions, then "mixtures out," the prediction of the missing T_b values within the matrix mixtures, and finally by "compounds out," the prediction T_b for mixtures formed by compounds not included in the training set. The RMSE for "points out," "mixtures out," and "compounds out" were 3.6K, 7.2K, and 10.5K respectively.

The comparison of calculated and experimental liquid-vapor equilibrium curves (Fig. 19) confirmed the satisfactory quality of the corresponding QSPR models.

Note that these models are applicable, among others, for pairs of compounds forming azeotropic mixtures (Fig. 19 c, d). In some cases, when the difference between the boiling points of individual substances was less than the prediction error, models of condensation/evaporation curves for the corresponding mixture models of condensation/evaporation were not possible.

Even compared to the COSMO-RS approach, QSPR/QSAR models have proven themselves effective for predicting any property of binary mixtures, if the mixtures' individual components were present in the modeling set.

SiRMS-based 2D-QSPR models attempting to predict the critical temperatures (Tc), volumes (Vc), and pressures (Pc) and Pitzer's acentric factors (ω) of organic compounds used 407, 382, 309, and 331 compounds, respectively, all from NIST WebBook [75, 76, 101]. Structurally diverse organic compounds were used and this resulted in high statistics for the QSPR model after 5-fold external cross validation (R^2 = 0.97–0.99, R_{5f}^2 = 0.86–0.95, predicted error T_c and V_c <3%, predicted error Pc and ω 3–10%). Conceptually, critical point

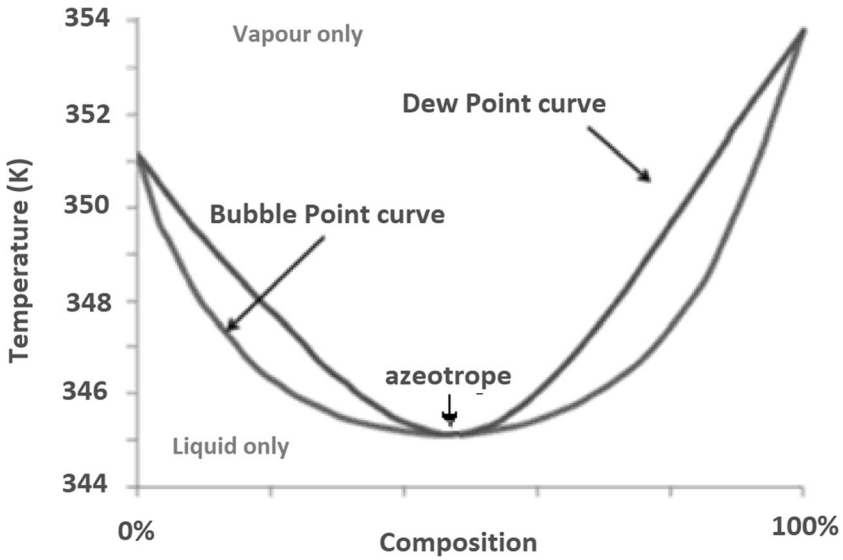


Fig. 18 Vapor-liquid equilibrium curve showing the variation of equilibrium composition of the liquid mixture with the temperature at a fixed pressure. The dew-point curve represents the temperature at which the saturated vapor starts to condense whereas the bubble-point is the temperature at which the liquid starts to boil

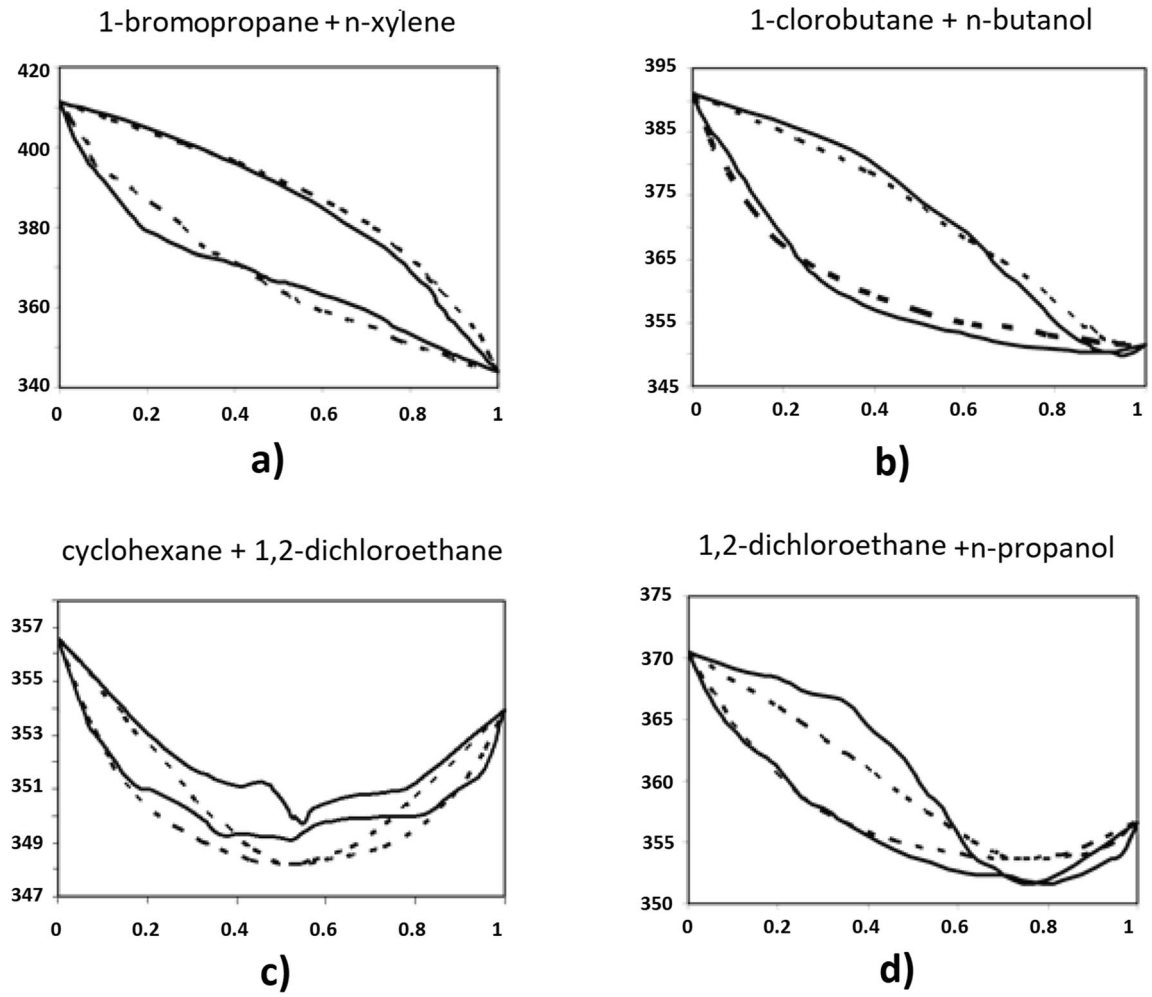


Fig. 19 Examples of experimental (dashed lines) and predicted (continues lines) liquid-vapor equilibrium curves

parameters are reliant on the energy of intermolecular interactions, and the analysis finding electrostatic and Van der Waals interactions as the primary descriptors corroborates this theory.

By combining the SiRMS methodologies for single compounds and mixtures, the "quasi-mixture" approach designates a pure compound as a mixture of two molecules and hence presents new unique mixture simplexes [76]. The QSPR models of the "quasi-mixture" simplexes display higher performance statistics and statistically significant differences in RMSE (Fig. 20).

The development of QSPR models to predict the critical properties of mixtures of organic compounds [80] has no analogues. It was possible due to the use of special SiRMS descriptors aimed at describing mixtures of compounds (see the "Simplex descriptors for solving various QSAR/QSPR tasks" section). Given 94 pure compounds and roughly 300 mixtures, the varying composition parameters resulted in ~1000 values each. The critical pressure, temperature, and volumes ranged from 20 to 100 bar, 150–800 K and from 80 to 400 cm^3/mol, respectively. Different machine learning methods were used to build the QSPR models, with the best results obtained from the RF method. Error was reported using the mean absolute percentage error (MAPE):

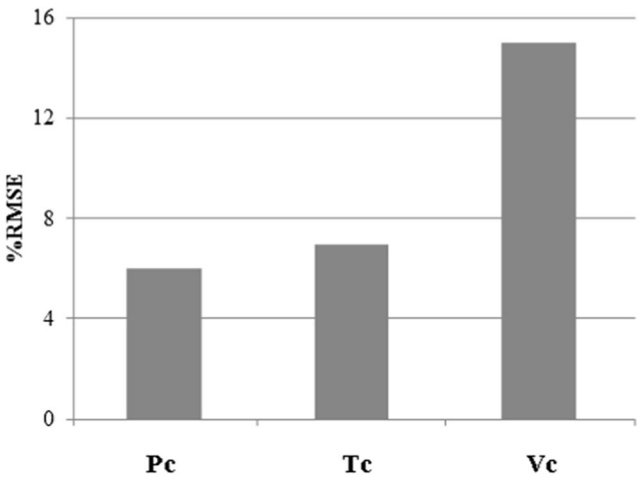


Fig. 20 Percentage increase in "quasi-mixture" models prediction accuracy relative to 'single molecule' models

$$MAPE = 1/m \cdot \sum_{i=1}^{m} \left| \left(y_i - \widehat{y}_i\right)/y_i \right| \cdot 100\%$$

Here, y_i are observed values, $\hat{y}_i$ are predicted values, and m is the number of observations. The MAPE values are more than satisfactory in determining the significance of this approach: $MAPE_{ts}$(Tc) = 6.8%, $MAPE_{ts}$(Pc) = 11.5%, $MAPE_{ts}$(Vc) = 14.6%. These numbers ascertain the ability of simplexes to predict thermodynamic properties of organic compounds at expert levels.

Considering the successful modeling and interpretability of the simplex descriptors, SiRMS methodology should be implemented into models analyzing and predicting critical properties.

Virial equations of state can be used to describe the p-v-t behavior of real gases, which are known to deviate substantially from ideal gas behavior. However, $pV_m/RT = 1 + B/V_m + C/V_m^2 + D/V_m^3 + \ldots$ has rigorous theoretical backing until extremely high pressures. All virial coefficients are temperature dependent and were established based on the real gas deviations from ideal behavior. The second virial coefficient accounts for molecular pair interactions, and therefore given the prominence of these interactions in the above theory, this is the most important coefficient. It is a calculated parameter whose experimental trials are again, expensive and time consuming. In the past, QSPR models have not been able to model temperature dependent coefficients, but given our success with these properties [70] (see above), we look to apply the simplex methods to a QSPR model for the second virial coefficient. Like any temperature dependent data, careful thought is required to ensure thoughtful, interpretable QSAR/QSPR modeling. One issue arises with the inconsistency in the temperature values and range of temperatures seen in the virial coefficients data. To solve this problem, like in [77], we used physical based methodologies that derive the following two simple but rigorous equations from the Van-der-Waals equation of state for real gases:

$$B = b - (a/RT) \quad (1)$$

$$B = b - \exp(a/RT) \quad (2)$$

a, b = f (D_1, D_2, …D_i, …), where B is the second virial coefficient, a, b are the coefficients of van-der-Waals equation, D is the descriptors, and T is the temperature. Then, two QSPR models were formed separately for parameters a and b. The second virial coefficient B was calculated using equations I or II for any given temperature. The data was taken from a comprehensive reference book [102], which covers second virial coefficients for more than 250 compounds. Given the temperature dependence, the overall number of data points is more than 4500. The "quasi-mixture" approach (see above) was used to calculate the SiRMS descriptors, while the RF method was used to develop quick models robust towards overfitting. As a result, we managed good predictive ability, with both approaches (1) and (2) being approximately equivalent (see Table 6).

The "quasi-mixture": model delivered the best consensus from the exponential equation form and is therefore used to represent variable trends in Fig. 21. Understanding **a** is representative of the repulsion between particles and **b** is the volume excluded by a mole of particles, the correlation is logical and fits into the expected physical explanation of the relationships.

For binary mixtures of compounds, the second virial coefficient has the following form: $B_{mixt} = x_1^2 B_1 + x_2^2 B_2 + 2x_1 x_2 B_{12}$, where x is the mole fraction of compounds 1 and 2, B_1 and B_2 are the second virial coefficients of pure compounds, and B_{12} - is the second virial cross-coefficient. The second virial cross-coefficient is a calculated property based solely off the mixture's component interactions and is only a measure of interactions between the two molecules. This intrinsic property opens up the opportunity to predict PVT for multicomponent mixtures as well. To our knowledge, [78] is the first attempt at a QSPR model for this coefficient. Dymond et al. [102] compilation was the source of the data for the 126 mixtures and 1211 values (each mixture selected had at least 4 values) of B_{12} at different temperatures ranging from 200-600 K. The test set comprised of compounds with less than 4 data values for a total of 102 mixtures and 188 data points at different temperatures. Given the sole focus of B_{12} on the heterogenous mixture values, the SiRMS descriptors for individual components were removed from the model. Similar to calculating the B coefficient for individual compounds (see above), two-layer QSPR models corresponding to the equation B_{12} = b - exp(a/RT) were used. The best results were obtained using the GBM (Gradient boosting machines) machine learning method for the 5-fold external cross-validation (R^2_{test} = 0.75, RMSE = 253 cm^3/mol). The external test set resulted in an R^2_{test} = 0.65 and RMSE = 224 cm^3/mol. Illustrative examples of the predictions of temperature dependences for B_{12} are shown in Fig. 22.

Contrary to other states, the QSPR models used 2D descriptors but do not require additional experimental data.

[79] works to characterize the interactions of the surface groups of A-300 aerosil with more than 40 different benzo-, dibenzo-, and aliphatic crown ethers. The QSPR models analyzed the Henry (K_H) and Langmuir (K_L) constants in

Table 6 Statistical characteristics of the Random Forest consensus models

Equation	R^2_{ws}	$RMSE_{ws}$	R^2_{ts}	$RMSE_{ts}$
Linear form	0.98	20	0.79	190±21
Exponential form	0.99	18	0.81	185±25

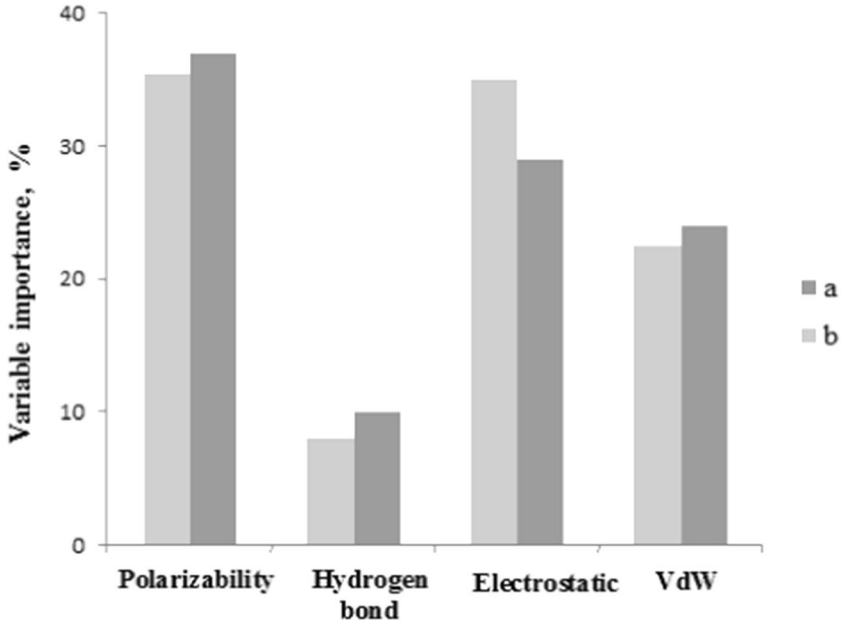


Fig. 21 Relative variable importance

addition to properties or fragments seen to impact the surface group formation. The models were validated with five-fold cross validation and the 2D-PLS models displayed a sataisfactory R^2 = 0.86–0.94, Q^2 = 0.82–0.92 and R^2_{test} = 0.65–0.88. The best 2D-QSPR models using SiRMS descriptors were for K_H. The analysis concluded that electron polarizability (33%) and electrostatics (29%) are the most influential on the Henry constant, and once again this concurred with the accepted general knowledge of polar molecule interactions with aerosil surfaces.

QSPR models of the luminescent properties of complex compounds

In [73], the QSPR analysis of the luminescence properties of complexes of Eu(III) and Tb(III) ions with 2-oxo-4-

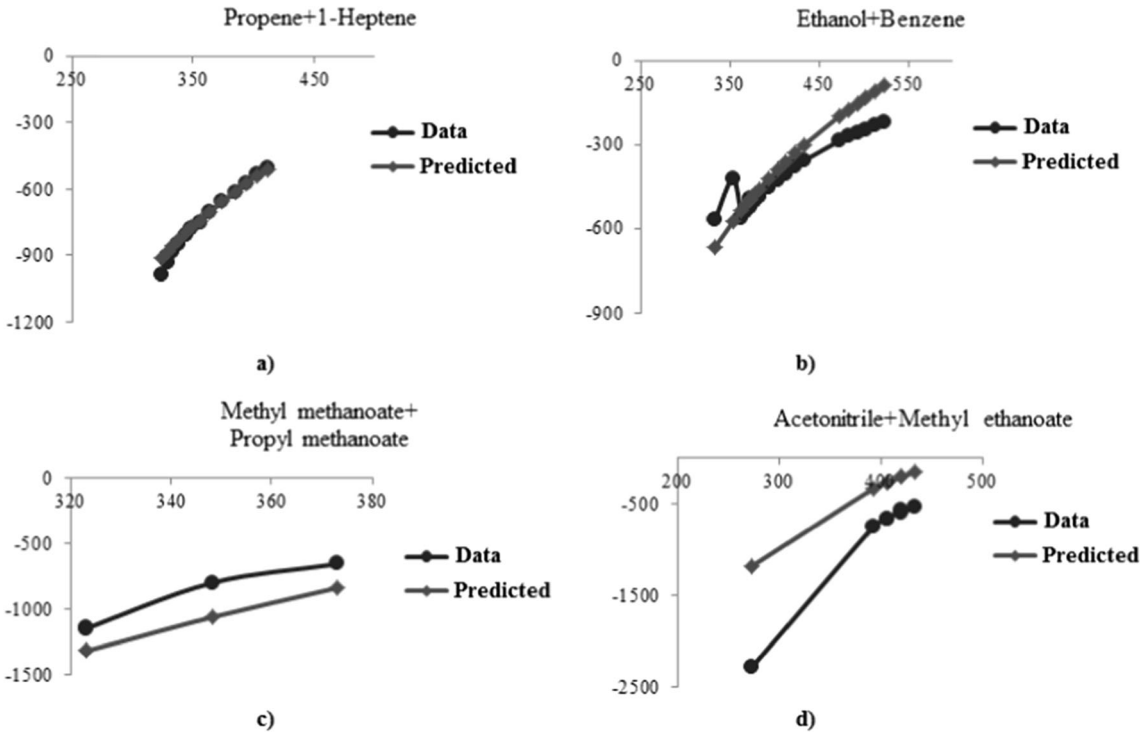


Fig. 22 Examples of temperature curves for B_{12} prediction

hydroxyquinoline-3-carboxylic acid amides was detailed. In these works, the information-topological version of SiRMS descriptors was used. For these tasks, it proved to be much more efficient than the standard 2D SiRMS descriptors. The properties under study were lifetime и quantum yield luminescence of the above complexes, with a total of 42 compounds being studied. All models were built by the PLS method and the five-fold procedure was used to evaluate the predictive ability of the models. The R^2_{test} = 0.92–0.97, so the models were used for virtual screening of new promising compounds. Structural interpretation of the QSPR models showed that the most promising ligands for luminescence where those containing unsubstituted cyclohexane or benzene rings as a fragment "A" (Fig. 23). Unbranched alkyls and furfuryl fragment are the most promising as the "B" fragment (C_2 - C_6). Alkylsubstituted 1,3,4-thiadiazoles and picolines are beyond competition for complexes of Eu(III) and Tb(III) ions as the "D" fragment.

As a result of this work, a terbium (III) complex with one of the best model predicted ligands has been used as an analytical form for the highly sensitive luminescent determination of terbium in high-purity lanthanum, yttrium, and gadolinium oxides.

QSPR models of the properties of ionic inorganic compounds

Even though cheminformatics approaches are frequently used in the study of organic compounds, there are almost no publications devoted to QSPR models of inorganic compounds. Objectively, typical molecular descriptor schemes rarely apply to inorganics. A few reasons for this include the significantly smaller variety of elements in organic compounds as opposed to inorganic compounds and the molecular diversity of organic compounds. Interestingly, aside from coordination complexes, isomerism is not as prevalent for many crystalline inorganics, and therefore the term "molecule" is rather conditional.

Overall, QSPR approaches are uncommon in the study of inorganic compounds. However, the information provided from them is undeniably valuable, especially given the current limited development.

QSPR models were developed to predict the melting points (MP) and refractive indices (RI) of various inorganic compounds in [81]. These data points are essential for the development of new optical materials. The authors point out that the language of structural formulas, which is the basis for the calculation of 2D descriptors of organic molecules, is often not suitable for the description of inorganic compounds. A typical example of such a situation is shown in Fig. 24.

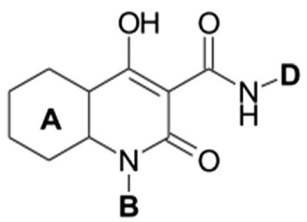


Fig. 23 Structure of investigated ligands

Despite the allowance of different structural formulas due to valence, inorganic crystals do not typically conform to formulas (Fig. 24). Thus, given that information on the spatial structure (3D) of inorganic compounds is not always available, and 2D structures are not correct, 1D descriptors were used to build appropriate QSPR models (see the "Simplex descriptors for solving various QSAR/QSPR tasks" section). In fact, the number of different combinations of atoms (twos, threes, fours, etc.) included in the gross formulation of an inorganic compound was calculated. The estimation of weight parameters characterizing atoms took into account the specificity of inorganic compounds, so as the key atomic characteristics were used, including the group number, oxidation level, nuclear charge, belonging to s-, p-, d-, f- elements, and the electronegativity. Information on melting points and refractive indices of various inorganic compounds was collected from reference books [103, 104]. In total, about 400 compounds were studied and 13 QSPR models were built using the RF method. The predictive ability of these models, evaluated by the "out-of-bag" (oob) procedure, was quite satisfactory with the R^2_{oob} = 0.66–0.88. The mean relative error of these predictions was 6–15%, and our models demonstrated that even simple 1D-QSPR models can both screen important properties non-experimentally, as well as provide meaning and direct interpretations. These interpretations suggest the relevance of electrostatic factors on the considered properties and while this may seem obvious due to the ionic nature of inorganic compounds, it only validates the interpretation of QSPR models. It should also be noted that these QSPR models are more practical for preliminary nonexperimental screening of inorganics compared to quantum-chemical based models. QSPR models have also been built to consider qualitative and quantitative values of superconducting critical temperature and geometrical features helping/hindering criticality [82].

QSPR modeling of nanoparticle properties

Some of the most complex objects for QSPR modeling are nanoparticles. Herein, it is necessary to distinguish two types of nanoparticles :

- large nanoscale individual molecules (e.g., fullerenes, nanotubes, etc.)
- aggregates or agglomerates of molecules (atoms) forming nanoscale particles.

Obviously, the approaches to modeling different types of nanoparticles must be different. In the first case, known

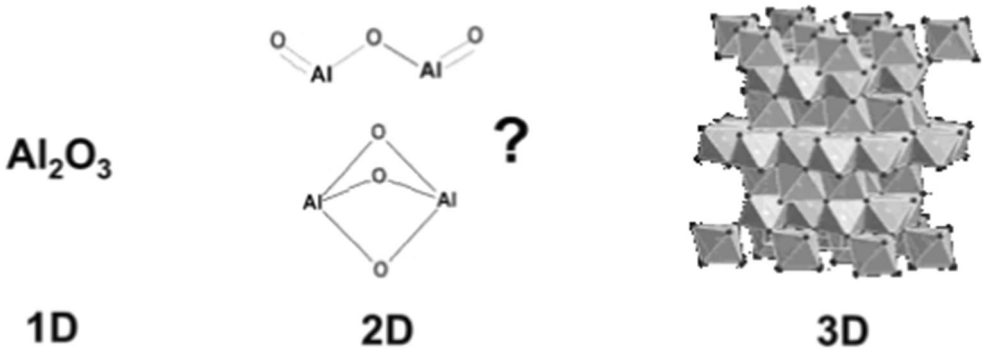


Fig. 24 The formula-based issue in the structural modeling of inorganic compounds

descriptor systems can be used. Although, in the second case, it is necessary to know both the information about the molecules composing the nanoparticle as well as the parameters of the nanoparticle, such as the size, surface area, shape, etc. of the integral object. Solving QSPR problems of the first type, the researcher nevertheless faces the problem of atom differentiation in carbon skeletons of fullerenes or nanotubes. The information-topological 2D SiRMS descriptors developed by us (see above the "QSAR models of various types of toxicity" section) successfully solve this problem. This can be demonstrated by the results of [84], where the QSPR model for the solubility of 27 fullerene (C60 and C70) derivatives in chlorobenzene was developed.

The developed PLS model is characterized by good statistical characteristics as for the training set R^2 = 0.939 and RMSE = 0.120, for the validation set Q^2 = 0.904 and RMSE = 0.141, for the test set R^2 = 0.873 and RMSE = 0.146, and lastly with scrambling, the R^2 = 0.026 and the Q^2 = 0.031. Interpretation of the QSPR model shows that when varying the aromatic fragment solubility decreases in the series: furan > benzene > thiophene. The greater number of lipophilic fragments (-C-C=) also promotes better solubility in chlorobenzene.

The results indicate that the SiRMS informational descriptors are sufficient to encode and describe the variation of the experimental solubility of fullerene.

QSPR models for type II nanoparticles (nanoaggregates) are discussed in [63, 83]. In these studies, in vitro cytotoxicity data (EC_{50} and LC_{50}) of metal oxide nanoparticles (ZnO, CuO, V_2O_3, Y_2O_3, Bi_2O_3, In_2O_3, Sb_2O_3, Al_2O_3, Fe_2O_3, SiO_2, ZrO_2, SnO_2, TiO_2, CoO, NiO, Cr_2O_3, La_2O_3) against *Escherichia coli* bacteria and the human keratinocyte cell line HaCaT was considered. 1D SiRMS descriptors were used to describe the chemical nature of the nanoparticles, similar to those for ionic inorganic compounds. We developed the «liquid drop model» (LDM) to characterize these nanoparticles [83]. The LDM represents each nanoparticle as a spherical drop so that elementary particles (molecules) can be densely packed and the mass density can be calculated. It is important to note that this model assumes the minimum radius of interactions between the molecules in the cluster is the Wigner-Seitz radius.

Using the HaCaT and *E. Coli* cell lines, we developed two nano-QSAR models. The HaCaT model displayed an R^2 = 0.83, Q^2_{cv} = 0.71, R^2_{ext} = 0.91 and RMSE = 0.12, while the *E. coli* model showed R^2 = 0.93, Q^2_{cv} = 0.90, R^2_{ext} = 0.97, and RMSE = 0.12.

These results suggest the combinatorial 1D and size-dependent descriptors are capable of producing meaningful nano-QSAR models as it applied to metal oxide cytotoxicity on HaCaT and *E. coli*. The weighted cross product of descriptors revealed that while both size-dependent parameters and the chemical nature of metal ions are important to cytotoxicity, the magnitude of the charge of the metal ion is the most important.

Conclusions

In conclusion, the authors are pleased to note that the SiRMS approach is quite popular among our colleagues who are solving various QSAR/QSPR problems. A list of some works known to us is presented in the final Table 7.

To summarize, the simplex representation of the molecular structure is a sufficiently versatile and flexible tool for solving a variety of structural problems from detailed stereochemical analysis to QSAR/QSPR. The multiplicity of simplex descriptors based on well-understood physical-chemical principles allows for not only predictive modeling, but also detailed structural and physical-chemical interpretations of these models. The list of objects to which SiRMS can be applied is also very broad, ranging from simple inorganic compounds to complex organic molecular and supramolecular systems, including nanoparticles. Thus, SiRMS was successfully used for wide variety (all major types of bioactivities and toxicities, phys-chem properties, etc.) of 1-4D QSAR/QSPR tasks described in this review. Moreover, we have pioneered the development of both SiRMS-based descriptors for chemical mixtures [137] and strategies for robust validation of QSAR models for mixtures [137, 138]. These approaches were successfully applied to the modeling of mixtures of organic solvents [74], drug delivery systems

Table 7 QSAR/QSPR works of external authors, which use SiRMS descriptors

Tasks	Investigated properties	References
Comparison of efficiency descriptors		[105]
QSAR	Antiviral activity	[106]
	Antitumor activity	[107–110]
	Respiratory sensitizers, radiosensitizers	[111, 112]
	Affinity to different biological targets	[113–115]
	Different types of toxicity	[116–127]
QSPR	Equilibrium liquid-liquid	[128]
	Optical rotations	[129]
	Optoelectronic properties	[130, 131]
	Reaction ability	[132]
	Flash points of binary mixtures, auto-ignition temperatures of binary liquid mixtures	[133–135]
	Properties of the nanosystems	[136]

[139], inorganic materials [140], and drug-drug interactions [141]. Importantly, we have addressed the very difficult task of predicting the synergistic effects in drug mixtures [27]. Advances of the Simplex approach related to modeling of mixtures and interpretation of QSAR models were highlighted in two highly cited perspectives of QSAR field [142, 143].

Acknowledgements A large number of authors have participated in the publications discussed in this review. Unfortunately, we are not able to include all of them in the list of co-authors of this review. We express our acknowledgment to all these colleagues for the productive collaboration, which allowed us to show the broad possibilities of the simplex approach and contributed to its development.

Code availability Not applicable

Author contribution Optional

Data availability Not applicable

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Holtje H-D, Sippl W, Rognan D, Folkers G (2009) Molecular modeling 3-rd ed. Wiley-VCH Weinheim
2. Todeschini R, Consonni V (2009) Handbook of Molecular Descriptors, 2-nd ed. Wiley-VCH Weinheim
3. Baskin II, Madzhidov TI, Varnek A (2015) Introduction to Chemoinformatics. Part 3. "Structure - properties" modeling. Kazan university Kazan (In Russian)
4. Polischuk P, Mokshina E, Kosinskaja A, Muats A, Kulinsky M, Tinkov O, Ognichenko L, Khristova T, Artemenko A, Kuz'min V (2017) Structural, physico-chemical and stereochemical interpretation of QSAR models based on simplex representation of molecular structure. In "Advances in QSAR modeling: Applications in Pharmaceutical, Chemical, Food, Agricultural and Environmental Sciences" Ed. Kunal Roy Springer: 107-148
5. Kuz'min VE (1994) The structure of chiral molecules. Analysis of the concept of configuration and mechanisms of stereoisomerization. Russian Journal of Physical Chemistry 63:936–941
6. Kuz'min VE (1995) Homo- and heterochirality of dissymmetrical tetrahedra (chiral simplices). Stereochemical tunneling. Journal of Structural Chemistry 36(5):794–797
7. Kuz'min VE, Chelombitko VA, Yudanova IV, Stelmakh IB, Rublev IS (1998) Stereochemical analysis by simplex representation of molecules. Journal of Structural Chemistry 39(3):452–456
8. Kuz'min VE, Artemenko AG, Muratov EN, Polischuk PG, Ognichenko LN, Liahovsky AV, Hromov AI, Varlamova EV (2010) Virtual screening and molecular design based on hierarchical QSAR technology. Challenges and Advances in Computational Chemistry and Physics. T. Puzyn, J. Leszczynski and M. Cronin 8:127-176
9. Kuz'min VE, Artemenko AG, Polischuk PG et al (2005) Hierarchic system of QSAR models (1D-4D) on the base of simplex representation of molecular structure. J Mol Model 11:457–467
10. Kuz'min VE, Artemenko AG, Muratov EN (2008) Hierarchical QSAR technology on the base of simplex representation of molecular structure. J Comp Aid Mol Des 22:403–421
11. Ognichenko LN, Kuz'min VE, Artemenko AG (2009) New structural descriptors of molecules on the basis of symbiosis of the informational field model and simplex representation of molecular structure. QSAR & Comb Sci 28(9):939–945
12. Alves VM, Golbraikh A, Capuzzi SJ, Liu K, Lam WI, Korn DR, Pozefsky D, Andrade CH, Muratov EN, Tropsha A (2018) Multi-descriptor read across (mudra): a simple and transparent approach for developing accurate quantitative structure-activity relationship models. J Chem Inf Model 58:1214–1223
13. Kuz'min VE, Artemenko AG, Muratov EN, Ognichenko LN, Hromov AI, Liahovskij AV, Polischuk PG (2008) The Hierarchic Informational Technology for QSAR Investigations: Molecular Design of Antiviral Compounds. In: National Institute of Allergy and Infectious Diseases, NIH Frontiers in Antiviral Research VST Georgiev, Humana Press Inc, Totowa NJ 1 163-178
14. Kuz'min VE, Artemenko AG, Lozitsky VP et al (2002) The analysis of structure-anticancer and antiviral activity relationships for macrocyclic pyridinophanes and their analogues on the basis of 4D QSAR models (simplex representation of molecular structure). Acta Biochim Polon 49:157–168
15. Muratov EN, Artemenko AG, Kuz'min VE et al (2005) Investigation of anti-influenza activity using hierarchic QSAR technology on the base of simplex representation of molecular structure. Antiviral Research 65:A62–A63
16. Kuz'min VE, Artemenko AG, Muratov EN et al (2005) The hierarchical QSAR technology for effective virtual screening and molecular design of the promising antiviral compounds. Antiviral Research 65:A70–A71
17. Artemenko AG, Kuz'min VE, Muratov EN et al (2005) Investigation of antiherpetic activity using hierarchic QSAR technology on the base of simplex representation of molecular structure. Antiviral Research 65:A77
18. Kuz'min VE, Artemenko AG, Lozitska RN, Fedtchouk AS, Lozitsky VP, Muratov EN, Mescheriakov AK (2005) Investigation of anticancer activity by means of 4D QSAR based on simplex representation of molecular structure. SAR and QSAR in Env Res 16(3):219–230
19. Artemenko AG, Muratov EN, Kuz'min VE et al (2007) Identification of individual structural fragments of N,N-(bis-5-nitropyrimidyl)dispirotripiperazine derivatives for cytotoxicity

and antiherpetic activity allows the prediction of new highly active compounds. J Antimicrob Chemother 60:68–77
20. Kuz'min VE, Artemenko AG, Muratov EN et al (2007) Quantitative structure–activity relationship studies of [(biphenyloxy)propyl]isoxazole derivatives – human rhinovirus 2 replication inhibitors. J Med Chem 50:4205–4213
21. Kuz'min V, Artemenko A, Muratov E, Varlamova E, Makarov V, Riabova O, Wutzler P, Schmidtke M (2008) QSAR analysis of cytotoxicity in Hela cells. Antiviral Research 78:A43
22. Artemenko A, Kuz'min V, MuratovE KM, Lozitsky V, Fedchuk A, Gridina T, Koroleva L, Silnikov V (2008) QSAR analysis of influence of artifical ribonucleases structure on their anti-influenza activity. Antiviral Research 78:A53
23. Muratov E, Kuz'min V, Artemenko A, Varlamova E, Makarov V, Riabova O, Wutzler P, Schmidtke M (2008) HiT QSAR analysis of anti-coxsackie virus B3 activity of [(biphenyloxy)propyl]isoxazole derivatives. Antiviral Research 78:A60–A61
24. Artemenko AG, Muratov EN, Atamanyuk DV, Kuz'min VE, Khromov AI, Kutsyk RV, Lesyk RB (2009) QSAR analysis of antimicrobial activity of 4-thiazolidone derivatives. QSAR Comb Sci 28:194–205
25. Artemenko AG, Muratov EN, Kuz'min VE, Kulinskiy M, Borisuk I, NYA G (2009) HiT QSAR study of antivirals' bioavailability. Antiviral Research 82:A56
26. Muratov EN, Artemenko AG, Varlamova EV, Polischuk PG, Lozitsky VP, Fedchuk AS, Lozitska RN, Gridina TL, Koroleva LS, Sil'nikov VN, Galabov AS, Makarov VA, Riabova OB, Wutzler P, Schmidtke M, Kuz'min VE (2010) Per aspera ad astra: application of Simplex QSAR approach in antiviral research. Future Medicinal Chemistry 2:1205–1226
27. Muratov E, Varlamova E, Kuz'min V, Artemenko A, Nikolaeva-Glomb L, Galabov A (2010) QSAR analysis of poliovirus inhibition by dual combinations of antivirals. Antiviral Research 86:A62
28. Muratov E, Varlamova E, Artemenko A, Kuz'min V, Anfimov P, Zarubaev V, Saraev V, Kiselev O (2011) QSAR analysis of anti-influenza (A/H1N1) activity of azoloadamantanes. Antiviral Research 90:A74
29. Muratov EN, Varlamova EV, Artemenko AG, Khristova T, Kuz'min VE, Makarov VA, Riabova OB, Wutzler P, Schmidtke M (2011) QSAR analysis of [(biphenyloxy)propyl] isoxazoles: agents against coxakievirus B3. Future Med Chem 3(1):31–43
30. Muratov EN, Varlamova EV, Artemenko AG, Polishchuk PG, Nikolaeva-Glomb L, Galabov AS, Kuz'min VE (2013) QSAR analysis of poliovirus inhibition by dual combinations of antivirals. Struct Chem 53:1665–1679
31. Gridina TL, Fedchuk AS, Basok SS, Artemenko AG, Ognichenko LN, Shitikova LI, Lutsyuk AF, Gruzevskii AA, Kuz'min VE (2019) The effect of the structure of derivatives of nitrogen-containing heterocycles on their anti-influenza activity. Chem heterocycle compounds 55(4/5):455–462
32. Nikolova I, Slavchev I, Ravutsov M et al (2019) Anti-entero viral activity of new MDL-860 analogues: Synthesis, in vitro/in vivo studies and QSAR analysis. Bioorg Chem 85:487–497
33. Alves VM, Bobrowski T, Melo-Filho CC, Korn D, Auerbach S, Schmitt C, Muratov EN, Tropsha A (2020) QSAR modeling of SARS-CoV M^{pro} inhibitors identifies Sufugolix, Cenicriviroc, Proglumetacin and other drugs as candidates for repurposing against SARS-CoV-2. Mol Inf https://doi.org/10.1002/minf.202000113
34. Thompson CG, Sedykh A, Nicol MR, Muratov E, Fourches D, Tropsha A, Kashuba ADM (2014) Short communication: cheminformatics analysis to identify predictors of antiviral drug penetration into the female genital tract. AIDS Research and Human Retroviruses 30(11):1058–1064. https://doi.org/10.1089/aid.2013.0254
35. Bobrowski T, Alves V, Melo-Filho CC, Korn D, Auerbach SS, Schmitt C, Muratov E, Tropsha A (2020) Computational models identify several FDA approved or experimental drugs as putative agents against SARS-CoV-2. Chem Rhiv. https://doi.org/10.26434/chemrxiv.12153594.v1
36. Capuzzi SJ, Sun W, Muratov EN, Martínez-Romero C, He S, Zhu W, Li H, Tawa G, Fisher EG, Xu M, Shinn P, Qiu X, García-Sastre A, Zheng W, Tropsha A (2018) Computer-aided discovery and characterization of novel ebola virus inhibitors. J Med Chem 61:3582–3594
37. Muratov E, Zakharov A (2020) Viribus unitis: drug combinations as a treatment against COVID-19. Chem Rhiv. https://doi.org/10.26434/chemrxiv.12143355.v1
38. Bobrowski T, Chen L, Eastman RT, Itkin Z, Shinn P, Chen CZ, Guo H, Zheng W, Michael S, Simeonov A, Hall MD, Zakharov AV, Muratov EN (2021) Synergistic and antagonistic drug combinations against SARS-CoV-2. Molecular Therapy 29(2):873–885. https://doi.org/10.1016/ymthe.2020.12.016
39. Soares Rodrigues GC, Maia MS, Silva Cavalcanti AB, Costa Barros RP, Scotti L, Cespedes CL, Muratov EN, Scotti MT (2021) Computer-assisted discovery of compounds with insecticidal activity against Musca domestica and Mythimna separate. Food and Chemical Toxicology. https://doi.org/10.1016/j.fct.2020.111899
40. Kuz'min VE, Muratov EN, Artemenko AG, Sidzhakova D, Galabov AS (2009) Antiviral activity of tetrahydro-2(1H)-pyrimidinones and related compounds: classification SAR study. Antiviral Research 82:A61
41. Golovenko NYA, Borisyuk IYU, Kulinskiy MA, Polishchuk PG, Muratov EN and Kuz'min VE (2014) Quantitative structure-property relationship analysis of drugs' pharmacokinetics within the framework of biopharmaceutics classification system using simplex representation of molecular structure. In: Application of Computational Techniques in Pharmacy and Medicine. L Gorb, V Kuz'min, E Muratov Springer Dordrecht Hiedelberg New York London 461-499
42. Artemenko AG, Polishchuk PG, Borysyuk IY, Muratov EN, Kuz'min VE, NYA G (2007) Prediction of the half-life of 1,4-benzdiazepine derivatives based on a combination of simplexes. Medical chemistry 9(3):10–17
43. Artemenko AG, Kuz'min VE, Muratov EN, Polishchuk PG, Borisyuk IY, NYA G (2009) Influence of the structure of substituted benzodiazepines on their pharmacokinetic properties. Pharm Chem J 43(8):27–35
44. Kolumbin OG, Ognichenko LN, Artemenko AG, Polischuk PG, Kulinskiy MA, Muratov EN, Kuz'min VE, Bobeica VA (2013) Nonexperimental screening of the water solubility, lipophilicity, bioavailability, mutagenicity and toxicity of various pesticides with QSAR models aid. Chem J Moldova 8(1):95–100
45. Polishchuk PG, Kosinskaya AP, Larionov VB, Ognichenko LN, Kuz'min VE, NYA G (2017) Ranked series of molecular fragments defining neuroavailability of drugs. Pharm Chem J 51(1):35–38
46. Kuz'min VE, Polischuk PG, Artemenko AG, Makan SY, Andronati SA (2008) Quantitative structure-affinity relationship of 5-HT_{1A}receptor ligands by the classification tree method. SAR & QSAR in Envir Res 19:213–244
47. Burenkova NA, Pavlovsky VI, Oleinich IA, Boyko IA, Makan SY, Artemenko AG, Kuz'min VE (2009) Synthesis and selectivity of 1-methoxycarbonyl-methyl-3-arylamino-7-bromo-5-phenyl-1,2-dihydro-3H-1,4-benzodi-azepin-2-ones binding for CNS benzodiazepine receptors. Ukrainica Bioorganica Acta 1:8–15

48. Krysko AA, Samoylenko GV, Polishchuk PG, Andronati SA, Kabanova TA, Khristova TM, Kuz'min VE, Kabanov VM, Krysko OL, Varnek AA, Grygorash RY (2011) RGD mimetics containing phthalimidine fragment, novel ligands of fibrinogen receptor. Bioorg & Med Chem Lett 21:5971–5974
49. Krysko AA, Samoylenko GV, Polishchuk PG, Fonari MS, Kravtsov VC, Andronati SA, Kabanova TA, Lipkowski J, Khristova TM, Kuz'min VE, Kabanov VM, Krysko OL, Varnek AA (2013) Synthesis, biological evaluation, X-ray molecular structure and molecular docking studies of RGD mimetics containing 6-amino-2,3-dihydroisoindolin-1-one fragment as ligands of integrin αIIbβ3. Bioorg& Med Chem 21:4646–4661
50. Polishchuk PG, Samoylenko GV, Khristova TM, Krysko OL, Kabanova TA, KabanovVM KAY, Klimchuk O, Langer T, Andronati SA, Kuz'min VE, Krysko AA, Varnek A (2015) Design, virtual screening, and synthesis of antagonists of αIIbβ3 as antiplatelet agents. J Med Chem 58:7681–7694
51. Yilmaz H, Sizochenko N, Rasulev B, Toropov A, Ya G, Kuz'min V, Leszczynska D, Leszczynski J (2015) Amino substituted nitrogen heterocycle ureas as kinase insert domain containing receptor (KDR) inhibitors: Performance of structure-activity relationship approaches. J food and drug analysis 23:168–175
52. Klimenko K, Lyakhov S, Shibinskaya M, Karpenko A, Marcou G, Horvath D, Zenkova M, Goncharova E, Amirkhanov R, Krysko A, Andronati S, Levandovskiy I, Polishchuk P, Kuz'min V, Varnek A (2017) Virtual screening, synthesis and biological evaluation of DNA intercalating antiviral agents. Bioorg & Med Chem Letters 27:3915–3919
53. Fourches D, Muratov E, Ding F, Dokholyan NV, Tropsha A (2013) Predicting binding affinity of CSAR ligands using both structure based and ligand-based approaches. J Chem Inf Model 53:1915–1922
54. Kuz'min VE, Muratov EN, Artemenko AG, Gorb L, Qasim M, Leszczynski J (2008) The effect of nitroaromatics' composition on their toxicity in vivo: novel, efficient nonadditive 1D QSAR analysis. Chemosphere 72(9):1373–1380
55. Kuz'min VE, Muratov EN, Artemenko AG, Gorb L, Qasim M, Leszczynski J (2008) The effects of characteristics of substituents on toxicity of the nitroaromatics: HiT QSAR study. J Comput Aided Mol Design 22:747–759
56. Polishchuk PG, Muratov EN, Artemenko AG, Kolumbin OG, Muratov NN, Kuz'min VE (2009) Application of random forest approach to QSAR prediction of aquatic toxicity. J Chem Inf Model 49:2481–2488
57. Artemenko AG, Muratov EN, Kuzmin VE, Muratov NN, Varlamova EV, Kuz'mina AV, Gorb LG, Golius A, Hill FC, Leszczynski J, Tropsha A (2011) QSAR analysis of nitroaromatics' toxicity in Tetrahymena pyriformis: structural factors and possible modes of action. SAR QSAR Env Res 22(5-6):575–601
58. Low Y, Uehara T, Minowa Y, Yamada H, Ya o, Urushidani T, Sedykh A, Muratov E, Kuz'min V, Fourches D, Zhu H, Rusyn I, Tropsha A (2011) Predicting druginduced hepatotoxicity using QSAR and toxicogenomics approaches. Chemical Research in Toxicology 24:1251–1262
59. Tin'kov OV, Muratov EN, Artemenko AG, Kuz'min VE (2013) Investigation and prediction of reproductive toxicity of organic compounds of various classes using 2D simplex representation of their molecular structures. Pharm Chem J 47(8):30–36
60. Tin'kov OV, Polishchuk PG, Artemenko AG, Kuz'min VE (2015) QSAR investigation of acute toxicity of organic acids and their derivatives upon intraperitoneal injection in mice. Pharm Chem J 49(2):34–40
61. Alves V, Eugene M, Capuzzi S, Politi R, Yen Low Y, Braga RC, Zakharov AV, Sedykh A, Mokshyna E, Farag S, Andrade CH, Kuz'min VE, Fourches D, Tropsha A (2016) Alarms about structural alerts. Green Chem 18:4348–4360
62. Tinkov OV, Ognichenko LN, Kuz'min VE, GorbLG (2016) Computational assessment of environmental hazards of nitroaromatic compounds: influence of the type and position of aromatic ring substituents on toxicity. Struct Chem 27(1):191–198
63. Kuz'min VE, Ognichenko LN, Sizochenko N (2019) Combining features of metal oxide nanoparticles: nano-QSAR for cytotoxicity. Int J QSPR 4(1):28–40
64. Alves VM, Muratov E, Fourches D, Strickland J, Kleinstreuer N, Andrade CH, Tropsha A (2014) Predicting chemically-induced skin reactions. Part I: QSAR models of skin sensitization and their application to identify potentially hazardous compounds. Toxicology and Applied Pharmacology 284(2):262–272
65. Alves VM, Muratov E, Fourches D, Strickland J, Kleinstreuer N, Andrade CH, Tropsha A (2015) Predicting chemically-induced skin reactions. Part II: QSAR models of skin permeability and the relationships between skin permeability and skin sensitization. Toxicology and Applied Pharmacology 284(2):273–280
66. Kholod YA, Muratov EN, Gorb LG, Hill FC, Artemenko AG, Kuz'min VE, Qasim M, Leszczynski J (2009) Application of quantum chemical approximations to environmental problems: prediction of water solubility for nitro compounds. Environ Sci Technol 43(24):9208–9215
67. Kovdienko NA, Polishchuk PG, Muratov EN, Artemenko AG, Kuz'min VE, Gorb L, Hill F, Leszczynski J (2010) Application of random forest and multiple linear regression techniques to QSPR prediction of an aqueous solubility for military compounds. Mol Inf 29:394–406
68. Muratov EN, Kuz'min VE, Artemenko AG, Kovdienko NA, Gorb L, Hill F, Leszczynski J (2010) New QSPR equations for prediction of aqueous solubility for military compounds. Chemosphere 79:887–890
69. Ognichenko LN, Kuz'min VE, Gorb L, Hill F (2012) QSPR prediction of lipophilicity for organic compounds using random forest technique on the basis of simplex representation of molecular structure. Mol Inf 31:273–280
70. Klimenko K, Kuz'min V, Ognichenko L (2016) Novel enhanced applications of QSPR models: temperature dependence of aqueous solubility. J Comput Chem 37:2045–2051
71. Gelmboldt V, Ognichenko L, Shyshkin I, Kuz'min V (2020) QSPR models for water solubility of ammonium hexafluorosilicates: analysis of the effects of hydrogen bonds. Struct Chem. https://doi.org/10.1007/s11224-020-01652-3
72. Alves VM, Hwang D, Muratov E, Sokolsky-Papkov M, Varlamova E, Vinod N, Lim C, Andrade CH, Tropsha A, Kabanov A (2019) Cheminformatics-driven discovery of polymeric micelle formulations for poorly soluble drugs. Sci Adv 5:eaav9784
73. Leonenko II, Yegorova AV, Ognichenko LN, Liahovsky AV, Aleksandrova DI, Ukrainets IV, Kuz'min VE, Antonovich VP (2011) QSPR analysis of the luminescent characteristics of Eu(III) and Tb(III) complexes with 2-oxo-4-hydroxyquinoline-3-carboxylic acid amides. Methods and Objects of Chem Analysis 6(1):38–50
74. Oprisiu I, Varlamova E, Muratov E, Marcou G, Polishchuk P, Kuz'min V, Varnek A (2012) QSPR approach to predict nonadditive properties of mixtures. Application to bubble point temperatures of binary mixtures of liquids. Mol Inf 31:491–502
75. Mokshina EG, Kuz'min VE, Nedostup VI (2014) QSPR modeling of critical parameters of organic compounds belonging to different classes in terms of the simplex representation of molecular structure. Russ J Organic Chem 50(3):314–321
76. Mokshyna E, Nedostup VI, Polishchuk PG, Kuz'min VE (2014) Quasi-mixture descriptors for QSPR analysis of molecular macroscopic properties. The critical properties of organic compounds. Mol Inf 33(10):647–654

77. Mokshyna EG, Polishchuk PG, Nedostup VI, Kuz'min VE (2015) Predictive QSPR modelling for the second virial coefficient of the pure organic compounds. Mol Inf 34:53–59
78. Mokshyna E, Polishchuk P, Nedostup V, Kuz'min V (2016) QSPR-modeling for the second virial cross-coefficients of binary organic mixtures. Int J QSPR 1(2):73–86
79. Voloshina NS, Ognichenko LN, Kuz'min VE, Pluzhnik-Gladyr SM, Kamalov GL (2015) Structural factors of the interaction of crown ethers with the aerosil surface. Protection of Metals and Physical Chemistry of Surfaces 51(1):93–105
80. Mokshyna EG, Polishchuk PG, Nedostup VI, Kuz'min VE (2016) QSPR modeling of critical properties of organic binary mixtures. Russ J Org Chem 52(1):5–10
81. Kuz'min VE, Ognichenko LN, Zinchenko VF (2020) QSPR models for prediction of the melting points and refractive indexes for inorganic substances - components of the optical film-forming materials. Int J QSPR 5(1):1–21
82. Isayev O, Fourches D, Muratov EN, Oses C, Rasch KM, Tropsha A, Curtarolo S (2014) Materials cartography: representing and mining material space using structural and electronic fingerprints. Chem Mater 27(3):735–743
83. Sizochenko N, Rasulev B, Gajewicz A, Kuzmin VE, Puzyn T, Leszczynski J (2014) From basic physics to mechanisms of toxicity: liquid drop approach applied to develop predictive classification models for toxicity of metal oxide nanoparticles. Nanoscale 6(22):13986–13993
84. Sizochenko N, Kuz'min V, Ognichenko L (2016) Introduction of simplex-informational descriptors for QSPR analysis of fullerene derivatives. J Math Chem 54(3):698–706
85. Zalgaller VA (1984) Simplex. Mathematical encyclopedia Vol. 4, Ch. ed. IM Vinogradov, M. Soviet encyclopedia (In Russian)
86. Wirth K, Dreiding AS (2007) Kants Hand, Chiralität und konvexe Polytope. Elemente der Mathematik 62(1):8–29. https://doi.org/10.4171/EM/50
87. Weininger D, Weininger A, Weininger JL (1989) SMILES. 2. Algorithm for generation of unique SMILES notation. J Chem Inf and Comp Sci 29(2):97–101. https://doi.org/10.1021/ci00062a008
88. Mislow K, Raban M (1967) Stereoisomeric relations of groups in Molecules. Top Stereochem, eds. Alinger NL Eliel EL 1
89. Glusker M, Hogan DM, Vass P (2005) The ternary calculating machine of Thomas Fowler. IEEE Annals of the History of Computing 27(3):4–22. https://doi.org/10.1109/MAHC.2005.49
90. Jolly WL, Perry WB (1973) Estimation of atomic charges by an electronegativity equalization procedure calibration with core binding energies. J Am Chem Soc 95:5442–5450
91. Wang R, Fu Y, Lai L (1997) A new atom-additive method for calculating partition coefficients. J Chem Inf Comp Sci 37:615–621
92. Londolt-Bornstein (1923) Physikalisch-chemische Tabellen 5 Auflage Band II Berlin
93. Kuz'min V, Ognichenko L, Artemenko A (2001) Modeling of the informational field of molecules. J Mol Model 7:278–285
94. Burkert U, Allinger N (1982) Molecular mechanics. ACS Publication, Washington DC 430
95. Hodges G, Roberts DW, Marshall SJ et al (2006) Defining the toxic mode of action of ester sulphonates using the joint toxicity of mixtures. Chemosphere 64:17–25
96. Kuz'min VE, Muratov EN, Artemenko AG et al (2009) Consensus QSAR modeling of phosphor containing hiral AChE inhibitors. J Comp Aid Mol Des 28:664–677
97. FDA (1999) Draft guidance for industry: bioavailability and bioequivalence studies for orally administered drug products-general considerations, US Department of Health, Food and Drug administration, Center for Drug Evaluation and Research BP August
98. Klamt A (1995) Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. J Phys Chem 99(7):2224–2235
99. Gelmboldt VO, Kravtsov VC, Fonari MS (2019) Ammonium hexafluoridosilicates: Synthesis, structures, properties, applications. J Fluorine Chem 221:91–102
100. Kang JW, Yoo KP, Kim HY et al (2001) Development and current status of the Korea Thermophysical Properties Databank (KDB). Int J Thermophysics 22:487–494
101. NIST WebBook: http://webbook.nist.gov/chemistry
102. Dymond J, Marsh K, Wilhoit R, Wong K (2002) Virial Coefficients of Pure Gases. Numerical Data and Functional Relationships in Science and Technology, Landolt-Bornstein
103. Nikolsky BP (1971) The chemist's handbook. Khimiya (In Russian)
104. Binnewies M, Milke E (2002) Thermochemical data of elements and compounds. Weinheim: Wiley-VCH Verlag GmbH. https://doi.org/10.1002/9783527618347
105. Adilova F, Davronov R, Rasulev B (2019) Comparison of the effectiveness of molecular descriptors in modeling the «structure-activity» relationship. Problems of Computational and Applied Mathematics 4(22):5–11
106. Ghosh K, Amin SA, Gayen S, Jha T (2020) Chemical-informatics approach to COVID-19 drug discovery: Exploration of important fragments and data mining based prediction of some hits from natural origins as main protease (Mpro) inhibitors. J Mol Struct 1224:129026. https://doi.org/10.1016/j.molstruc.2020.129026
107. Tinkov OV, Polishchuk PG, Khachatryan DS, Kolotaev AV, Balaev AN, Osipov VN, Grigorev VY (2019) Quantitative analysis of "structure – anticancer activity" and rational molecular design of bi-functional VEGFR-2/HDAC-inhibitors. Computer Research and Modeling 11(5):911–930
108. Ghosh K, Bhardwaj B, Amin SA, Jha T, Gayen S (2020) Identification of structural fingerprints for ABCG2 inhibition by using Monte Carlo optimization, Bayesian classification, and structural and physicochemical interpretation (SPCI) analysis. SAR and QSAR in Environmental Research. https://doi.org/10.1080/1062936X.2020.1771769
109. Amin SkA, Ghosh K, Mondal D, Jha T, Gayen S (2020) Exploring indole derivatives as myeloid cell leukaemia-1 (Mcl-1) inhibitors with multi-QSAR approach: a novel hope in anticancer drug discovery. The Royal Society of Chemistry and the Centre National de la Recherche Scientifique. https://doi.org/10.1039/d0nj03863f
110. Sidorov P, Naulaerts S, Ariey-Bonnet J, Pasquier E, Ballester PJ (2019) Predicting synergism of cancer drug combinations using NCI-Almanac data. Front Chem 7:509. https://doi.org/10.3389/fchem.2019.00509
111. Cui X, Yang R, Li S,·Liu J, Wu Q, Li X (2020) Modeling and insights into molecular basis of low molecular weight respiratory sensitizers. Molecular Diversity https://doi.org/10.1007/s11030-020-10069-3
112. De P, Bhattacharyya D, Roy K (2020) Exploration of nitroimidazoles as radiosensitizers: application of multilayered feature selection approach in QSAR modeling. Struct Chem. https://doi.org/10.1007/s11224-019-01481-z
113. Anju CP, Subhramanian S, Sizochenko N, Melge AR, Leszczynski J, Mohan CG (2018) Multiple e-Pharmacophore modeling to identify a single molecule that could target both streptomycin and paromomycin binding sites for 30S ribosomal subunit inhibition. Journal of Biomolecular Structure and Dynamics. https://doi.org/10.1080/07391102.2018.1462731
114. Chauhan S, Kumar A (2018) Consensus QSAR modeling of SIRT1 activators using simplex representation of molecular structure. SAR and QSAR in Environmental Research. https://doi.org/10.1080/1062936X.2018.1426626

115. Klimenko K (2019) In silico identification of endogenous and exogenous agonists of Estrogenrelated receptor α. Computational Toxicology 10:105–112
116. Tinkov O, Polishchuk P, Grigorev V, Yu P (2020) The cross-interpretation of QSAR toxicological models. Springer Nature Switzerland AG, Eds. Z. Cai et al: ISBRA 2020. LNBI 12304: 262–273. https://doi.org/10.1007/978-3-030-57821-3_23
117. Tinkov OV, Grigorev VY, Razdolsky AN, Grigoryeva LD, Dearden JC (2020) Effect of the structural factors of organic compounds on the acute toxicity toward Daphnia magna. SAR and QSAR in Environ Res. https://doi.org/10.1080/1062936X.2020.1791250
118. Tinkov O, Polishchuk P, Matveieva M, Grigorev V, Grigoreva L, Yu P (2020) The influence of structural patterns on acute aquatic toxicity of organic compounds. Mol Inf. https://doi.org/10.1002/minf.202000209
119. Capuzzi SJ, Politi R, Isayev O, Farag S, Tropsha A (2016) QSAR Modeling of Tox21 Challenge stress response and nuclear receptor signaling toxicity assays. Front Environ Sci 4:3. https://doi.org/10.3389/fenvs.2016.00003
120. Khan K, Kar S, Sanderson H, Roy K, Leszczynski J (2018) Ecotoxicological assessment of pharmaceuticals using computational toxicology approaches: QSTR and interspecies QTTR modeling. MOL2NET 4. https://doi.org/10.3390/mol2net-04-xxxx
121. Gooch A, Sizochenko N, Rasulev B, Gorb L, Leszczynski J (2017) In vivo toxicity of nitroaromatics: a comprehensive QSAR study. Environ Toxicol Chem. https://doi.org/10.1002/etc.3761
122. Jillella GK, Khan K, Roy K (2020) Application of QSARs in identification of mutagenicity mechanisms of nitro and amino aromatic compounds against Salmonella typhimurium species. Toxicology in Vitro 65:104768
123. Khan K, Kar S, Sanderson H, Roy K, Leszczynski J (2018) Ecotoxicological modeling, ranking and prioritization of pharmaceuticals using QSTR and i-QSTTR approaches: application of 2D and fragment based descriptors. Mol Inf 37:1800078. https://doi.org/10.1002/minf.201800078
124. Khan K, Baderna D, Cappelli C, Toma C, Lombardo A, Roy K, Benfenati E (2019) Ecotoxicological QSAR modeling of organic compounds against fish: Application of fragment based descriptors in feature analysis. Aquatic Toxicology 212:162–174. https://doi.org/10.1016/j.aquatox.2019.05.011
125. Moon H, Cong M (2016) Predictive models of cytotoxicity as mediated by exposure to chemicals or drugs. SAR and QSAR in Environmental Research 27(6):455–468
126. Eduati F, Mangravite LM, Wang T, Tang H, Bare JC et al (2015) Prediction of human population responses to toxic compounds by a collaborative competition. Nat Biotechnol 33(9):933–940
127. Sosnin S, Karlov D, Tetko IV, Fedorov MV (2019) Comparative study of multitask toxicity modeling on a broad chemical space. J Chem Inf Model 59(3):1062–1072. https://doi.org/10.1021/acs.jcim.8b00685
128. Klimenko KO, Inês JM, Esperança JMSS, Rebelo LPN et al (2020) QSPR modeling of liquid-liquid equilibria in two-phase systems of water and ionic liquid. Mol Inf 39:2000001. https://doi.org/10.1002/minf.202000001
129. Kapusta K, Sizochenko N, Karabulut S, Okovytyy S, Voronkov E, Leszczynski J (2018) QSPR modeling of optical rotation of amino acids using specific quantum chemical descriptors. Journal of Molecular Modeling 24:59. https://doi.org/10.1007/s00894-018-3593-z
130. Kar S, Sizochenko N, Ahmed L, Batista VS, Leszczynski J (2016) Quantitative structure-property relationship model leading to virtual screening of fullerene derivatives: exploring structural attributes critical for photoconversion efficiency of polymer solar cell acceptors. Nano Energy https://doi.org/10.1016/j.nanoen.2016.06.011
131. Roy JK, Supratik Kar S, Leszczynski J (2019) Optoelectronic properties of C60 and C70 fullerene derivatives: designing and evaluating novel candidates for efficient P3HT polymer solar cells. Materials 12:2282. https://doi.org/10.3390/ma12142282
132. Polishchuk P, Madzhidov T, Gimadiev T, Bodrov A, Nugmanov R, Varnek A (2017) Structure–reactivity modeling using mixture-based representation of chemical reactions. J Comput Aided Mol Des. https://doi.org/10.1007/s10822-017-0044-3
133. Cao W, Pan Y, Yi L, Jiang J (2020) A novel method for predicting the flash points of binary mixtures from molecular structures. Safety Science 126:104680. https://doi.org/10.1016/j.ssci.2020.104680
134. Shen S, Pan Y, Ji X, Yu N, Jiang J (2019) Prediction of the auto-ignition temperatures of binary miscible liquid mixtures from molecular structures. Int J Mol Sci 20:2084. https://doi.org/10.3390/ijms20092084
135. Yao J, Qi R, Pan Y, He H, Fan Y, Jiang J, Jiang J (2020) Prediction of the flash points of binary biodiesel mixtures from molecular structures. Journal of Loss Prevention in the Process Industries 65:104137. https://doi.org/10.1016/j.jlp.2020
136. Ojha PK, Kar S, Roy K, Leszczynski J (2018) Toward comprehension of multiple human cells uptake of engineered nano metal oxides: quantitative inter cell line uptake specificity (QICLUS) modeling. Nanotoxicology. https://doi.org/10.1080/17435390.2018.1529836
137. Muratov EN, Varlamova EV, Artemenko AG, Polishchuk PG, Kuz'min VE (2012) Existing and developing approaches for QSAR analysis of mixtures. Mol Inform 31(3-4):202–221
138. Muratov EN, Tropsha A, Varlamova EV, Kuzmin VE, Artemenko AG, Muratov NN, Mileyko S, Fourches D (2014) "Everything Out" Validation Approach for QSAR Models of Chemical Mixtures. J Clin Pharm 1(1):1005
139. Alves VM, Hwang D, Muratov E, Sokolsky-Papkov M, Varlamova E, Vinod N, Lim C, Andrade CH, Tropsha A, Kabanov A (2019) Cheminformatics-driven discovery of polymeric micelle formulations for poorly soluble drugs. Sci Adv 5(6):eaav9784. https://doi.org/10.1126/sciadv.aav9784
140. Isayev O, Oses C, Toher C, Gossett E, Curtarolo S, Tropsha A (2017) Universal fragment descriptors for predicting properties of inorganic crystals. Nat Commun 8:15679
141. Zakharov AV, Varlamova EV, Lagunin AA, Dmitriev AV, Muratov EN, Fourches D, Kuz'min VE, Poroikov VV, Tropsha A, Nicklaus MC (2016) QSAR modeling and prediction of drug-drug interactions. Mol Pharm 13(2):545–556
142. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, Dearden J, Gramatica P, Martin YC, Todeschini R, Consonni V, Kuz'min VE, Cramer R, Benigni R, Yang C, Rathman J, Terfloth L, Gasteiger J, Richard A, Tropsha A (2014) QSAR modeling: where have you been? Where are you going to? J Med Chem 57(12):4977–5010
143. Muratov EN, Bajorath J, Sheridan RP, Tetko IV, Filimonov D, Poroikov V, Oprea TI, Baskin II, Varnek A, Roitberg A, Isayev O, Curtalolo S, Fourches D, Cohen Y, Aspuru-Guzik A, Winkler DA, Agrafiotis D, Cherkasov A, Tropsha A (2020) QSAR without borders. Chem Soc Rev 49:3525–3564