Universal Approach for Structural Interpretation of QSAR/QSPR Models

Pavel G. Polishchuk,*[a] Victor E. Kuz'min,[a] Anatoly G. Artemenko,[a] and Eugene N. Muratov[a, b]

Abstract: In this paper we offer a novel approach for the structural interpretation of QSAR models. The major advantage of our developed methodology is its universality, i.e., it can be applied to any QSAR/QSPR model irrespective of chemical descriptors and machine learning methods applied. This universality was achieved by using only the information obtained from substructures of the compounds of interest to interpret model outcomes. Reliability of the offered approach was confirmed by the results of three case studies, including end-points of different types (continuous and binary classification) and nature (solubility, mutagenicity, and inhibition of Transglutaminase 2), various fragment and whole-molecule descriptors (Simplex and Dragon), and multiple modeling techniques (partial least squares, random forest, and support vector machines). We compared the global contributions of molecular fragments obtained using our methodology with known SAR rules derived experimentally. In all cases high concordance between our interpretation and results published by others was observed. Although the proposed interpretation approach could be easily extended to any type of descriptors, we would recommend using Simplex descriptors to achieve a larger variety of investigated molecular fragments. The developed approach is a good tool for interpretation of such "black box" models like random forest, neural networks, etc. Analysis of fragment global contributions and their deviation across a dataset could be useful for the identification of key fragments and structural alerts. This information could be helpful to maximize the positive influence of structural surroundings on the given fragment and to decrease the negative effects.

Keywords: QSAR · QSPR · machine learning · QSAR models interpretation

1 Introduction

While appearing more than 50 years ago,[1] the Quantitative Structure-Activity/Property Relationship (QSAR/QSPR: in this paper, the terms "QSAR" and "QSPR" are used interchangeably) field has changed dramatically in the last decade. QSAR has shifted away from the original simple and interpretable linear models developed using just a few descriptors towards complex multiparametric and, quite often, nonlinear approaches.[2] This has resulted in the popular misbelief that there is a balance between predictive ability and interpretability of models,[3] because many highly predictive models based on neural network (NN), support vector machines (SVM) and other "black box" approaches were hard or even impossible to interpret. However, interpretation is very important because it can lead to such additional benefits as (i) elicitation of structural rules and alerts; (ii) insights on the mechanism of action; (iii) structure optimization; and (iv) targeted design of new compounds. The importance of building interpretable models was demonstrated in several studies.[4–9] It was also reflected in one of the "OECD principles for the Validation, for Regulatory Purposes, of (Q)SAR Models" (http://www.oecd.org/dataoecd/33/37/37849783.pdf): "To facilitate the consideration of a QSAR model for regulatory purposes, it should be associated with … a mechanistic interpretation, if possible".

A plethora of different approaches and interpretation schemes were described previously.[3] Some models, such as decision trees,[10] are intuitively understandable and the use of interpretable descriptors will result in ready-to-use sets of structural rules and alerts. Other models need some additional operations to provide the researcher with the possibility of structural interpretation.[11] Existing approaches for interpretation of QSAR/QSPR models can be separated into model-specific and model-independent ones. Model-specific methods are more widespread. Estimation of influence of descriptors on a property value in linear models (multiple linear regression (MLR), partial least squares (PLS), etc.) can be made by considering values of regression coefficients in linear equations.[11] This approach was successfully applied in a great number of studies.[12–15]

[a] P. G. Polishchuk, V. E. Kuz'min, A. G. Artemenko, E. N. Muratov
A. V. Bogatsky Physical Chemical Institute, National Academy of Sciences of Ukraine
Lustdorfskaya Doroga 86, Odessa 65080, Ukraine
phone: +380979715161
*e-mail: pavel_polishchuk@ukr.net

[b] E. N. Muratov
Laboratory of Molecular Modeling, Division of Medicinal Chemistry and Natural Products, Eshelman School of Pharmacy, University of North Carolina
Beard Hall 301, CB#7568, Chapel Hill, NC, 27599, USA

More sophisticated methods of interpretation of PLS models based on analysis of latent variables gives a better idea about descriptors' contribution to the variation of the investigated activity.[16–17] Linear SVM models could be interpreted in the same way. Thus, Rosenbaum et al.[18] calculated the influence of descriptors on the basis of their weights in developed SVM models and their contributions in class separation. Some efforts have been made to interpret artificial neural network (ANN) models using their weights and biases.[19] Recently, a new approach for interpretation of random forest (RF) models based on the estimation of differences between mean activity values of compounds in parent and child nodes of trees caused by the corresponding descriptor has been proposed by Kuz'min et al.[20]

Estimation of descriptor importance, as originally developed for RF models by Breiman,[21] is an example of a model-independent interpretation method. It is based on the idea that the higher the importance of a given descriptor, the bigger the reduction of predictive power of the model caused by permutation of values of this descriptor. This approach can be used for interpretation of any model, as shown when Guha et al.[22] applied it to NNs. A significant disadvantage of this method is that obtained importance measures cannot indicate whether the influence of selected descriptors on the investigated property is positive or negative. Other model-independent interpretation approaches are based on the analysis of local gradients or partial derivatives of descriptors to calculate their contributions to the variation of an investigated property. These methods were used for interpretation of PLS, RF, ANN, SVM, decision tree models, and self-organized Kohonen maps.[23–26]

In all interpretation approaches described above, inverse task solution or interpretation of developed models was based on the ability to translate descriptor contributions to the molecule by color coding or by other procedures. Thus, for interpretation purposes it is necessary that the descriptors used could be directly linked back to the structure, e.g., by using fragment descriptors. If this is not possible, the estimation of their contributions will not be very helpful.

The goals of this study are: (i) to develop a universal approach for structural interpretation of any QSAR model irrespective of chemical descriptors and machine learning methods used and (ii) to validate its reliability by several case studies including a variety of end-points, activity scales, descriptors, and modeling techniques.

2 Materials and Methods

2.1 Datasets

We decided to test the applicability of our developed interpretation approach for a variety of different systems. In order to do this, we chose three datasets describing different end-points (physical-chemical properties, biological activity, and Ames mutagenicity) and expressed in different scales (continuous scale for the first two tasks and binary classification for Ames mutagenicity). The aqueous solubility dataset[27] consisted of 1033 diverse organic compounds and was the easiest for interpretation, because structural fragments increasing or decreasing the solubility of a compound are well-known. Solubility was represented as log$S$ values that were distributed in a range from $-11.62–1.58$. Ninety five structurally similar inhibitors of Transglutaminase 2 (TG2)[28] comprised the second dataset. The activity was represented as inhibition potency of compounds (pIC_{50}) in a range of 4.04–8.00. SAR rules for this dataset that were recently described by Prime et al.[28] allowed us to use it for benchmarking. The third dataset consisted of 4361 compounds with associated mutagenicity data obtained in the Ames test.[29] Activity was expressed in a binary scale and the dataset was balanced (2344 mutagens and 2017 non-mutagens). Several toxicophores and detoxifying fragments were reported earlier,[30] and thus could be compared with the results obtained by our approach. All the datasets were carefully curated according to the guidance published earlier[31] using Chemaxon Standardizer[32] and duplicates were removed.

2.2 Chemical Descriptors

We chose two different sets of 2D molecular descriptors for our case studies. The simplex representation of molecular structure (SiRMS)[33–34] approach was used to generate fragment descriptors, and the set of whole-molecule parameters[35] describing the molecule as a whole along with some molecular fragments (topological, constitutional, connectivity, informational, 2D autocorrelations, molecular properties, etc,) were generated by Dragon software.[36]

2D SiRMS, or Simplex descriptors (the number of tetratomic fragments with fixed composition and topology), were generated by HiT QSAR software.[34] At the 2D level, the connectivity of atoms in a fragment, atom type, and bond nature (single, double, triple, or aromatic) have been considered. SiRMS accounts not only for the atom type, but also other atomic characteristics that may impact biological activity of molecules, e.g., partial charge, lipophilicity, refraction, and the ability of an atom to be a donor/acceptor in hydrogen-bond formation (H-bond). For atom characteristics with continuous values (i.e., charge, lipophilicity, and refraction), the division of the entire value range into discrete groups has been carried out. Firstly, the values of these properties are calculated for every atom in the molecule following Jolly-Perry algorithm of electronegativity equalization[37] for partial atom charges, Xlog$P$ scheme[38] for lipophilicity, and the atomic refraction scheme suggested by Ioffe.[39] Then, the atoms have been divided into four groups corresponding to their (i) partial charge ($A \leq -0.05 < B \leq 0 < C \leq 0.05 < D$), (ii) lipophilicity ($A \leq -0.5 < B \leq 0 < C \leq 0.5 < D$), and (iii) refraction ($A \leq 1.5 < B \leq 3 < C \leq 8 < D$). For H-bond characteristics the atoms have been divided

into three groups: *A* (acceptor of hydrogen in H-bond), *D* (donor of hydrogen in H-bond), and *I* (indifferent atom). These multiple ways to differentiate the atoms in a fragment represent the principal feature of SiRMS descriptors. Detailed description of HiT QSAR and SiRMS can be found elsewhere.[34,40]

Both bonded (all atoms within the simplex are connected by chemical bonds) and unbonded (a certain atom or atoms in the simplex could belong to different parts of a molecule and thus are not connected by chemical bonds) 2D SiRMS descriptors were used. Unbonded simplexes were critical for this study, because they could describe the molecules consisting of separate fragments.

Unlike SiRMS, Dragon software cannot calculate descriptors for molecules consisting of separate parts and having free valences. Because of this, we interpreted only the fragments with one cleaved bond (terminal groups) when using Dragon descriptors. Before calculating Dragon descriptors, an additional hydrogen was added to the atom next to a cleaved bond to pass the intrinsic filter of Dragon. Thus, the fragments interpreted on the basis of SiRMS and Dragon QSAR models were slightly different.

2.3 Modeling Techniques and Applicability Domain

The three following statistical approaches implemented in R software[41] were used for the development of QSAR models: (i) RF method from random forest package,[42]; (ii) PLS method from ppls package,[43]; and (iii) SVM method from e1071 package.[44] No variables were pre-selected prior to modeling. Five hundred trees were generated for each RF model and the number of descriptors (the second tuning parameter) was set to 500 and 1500 for Dragon and SiRMS descriptors respectively. PLS models were allowed to include up to 10 significant latent variables. Eps-regression and C-classification algorithms with RBF kernel were used for development of regression and classification SVM models. Parameters of SVM models were tuned by the grid search implemented in e1071 package.

Ten rounds of 5-fold external cross-validation were applied to each dataset. For the Ames dataset the ratio of active/inactive compounds was kept identical for every fold. For regression tasks the entire dataset was first sorted by the property values and then split into 10 consecutive equal parts. The compounds in each part were shuffled and every fifth compound was taken to the corresponding external fold. Thus, 50 models were obtained by each machine learning method for each endpoint.

Applicability domain (AD) of developed models was estimated by the minimum spanning tree approach.[40,45] The tree was built in the space of the predictions of 50 single models using Kruskal's algorithm.[46] Then average distance (d_{av}) and its root-mean-square deviation (σ) among all tree edges were calculated. Such distance characterizes the average density of molecules distributed in the considered space. If any of the external-set molecules were situated at

a distance bigger than $d_{av} + 3\sigma$ from the nearest training-set point, it was considered as being outside the AD of the given model.

2.4 Universal Approach for Structural Interpretation of QSAR Models

Contrary to existing interpretation methods described in the Introduction, we propose a clear and straightforward way to calculate the contribution of molecular fragments to the property values directly from developed QSAR models. Another advantage of the proposed interpretation approach is that it is independent from the activity scale, i.e., it can be applied to both classification and continuous endpoints.

To a certain extent, proposed methodology could be considered as somewhat similar to matched molecular pairs approach (MMPA)[47–49] in application to QSAR field. The difference between the property of two related structures, e.g., A-B and A-C, containing the common fragment A is explained by the difference in the structure of fragments B and C. To estimate the influence of B to C transformation, in MMPA one can calculate the difference between observed property values for compounds A-B and A-C or between predicted property values by corresponding QSAR models. But such comparison is valid only for compounds containing the common substructure. The same is valid for Free-Willson models, where contributions of fragments are obtained directly from the modeling equation.[50] In the case when compounds do not contain a common substructure it becomes impossible to estimate the contributions of separate molecular fragments.

In the proposed interpretation approach contribution of the molecular fragment to the investigated property is calculated as a difference between the predicted property value for the particular compound and the predicted property value for the same compound excluding the considered fragment. Because of this, in our approach it becomes possible to estimate contributions of any arbitrary atoms combinations (atoms in selected fragments can be directly linked or not) or single atoms.

To better illustrate the proposed approach let us consider two compounds A-B and A-C with the common substructure A. Thus, if we will predict using a QSAR model the property value for the fragment A ($P^{pred}(A)$) and subtract it from the predicted property values of the compounds A-B ($P^{pred}(AB)$) and A-C ($P^{pred}(AC)$), we will be able to estimate and thus compare the contribution of the fragments B ($P_{AB}(B)$) and C ($P_{AC}(C)$) to the investigated property (Figure 1). In general, for any non-additive property, predicted property values of fragments will not be equal to their calculated contributions (i.e., $P^{pred}(A) \neq P_{AB}(A)$) (Figure 2). Although in theory for an additive property, predicted contributions of fragments could be equal to the calculated ones, they will most likely be different because of the error of prediction introduced by the QSAR model.
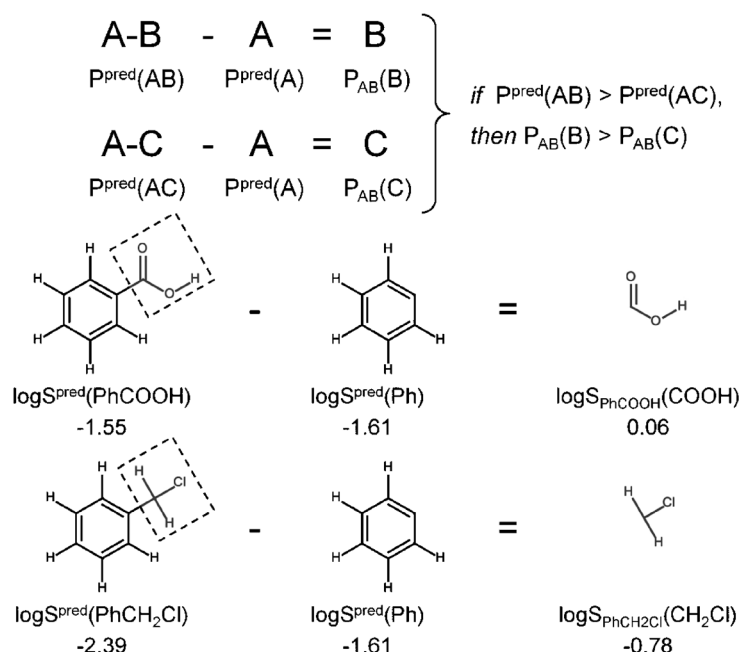
$$A\text{-}B \quad - \quad A \quad = \quad B$$
$$P^{pred}(AB) \qquad P^{pred}(A) \qquad P_{AB}(B)$$

$$A\text{-}C \quad - \quad A \quad = \quad C$$
$$P^{pred}(AC) \qquad P^{pred}(A) \qquad P_{AB}(C)$$

$$if \quad P^{pred}(AB) > P^{pred}(AC),$$
$$then \quad P_{AB}(B) > P_{AB}(C)$$

$logS^{pred}(PhCOOH)$	$logS^{pred}(Ph)$	$logS_{PhCOOH}(COOH)$
-1.55	-1.61	0.06
$logS^{pred}(PhCH_2Cl)$	$logS^{pred}(Ph)$	$logS_{PhCH2Cl}(CH_2Cl)$
-2.39	-1.61	-0.78

Figure 1. Structural interpretation of compounds consisting of two fragments.

$$A\text{-}B \quad - \quad A \quad = \quad B$$
$$P^{pred}(AB) \qquad P^{pred}(A) \qquad P_{AB}(B)$$

$$A\text{-}B \quad - \quad B \quad = \quad A$$
$$P^{pred}(AB) \qquad P^{pred}(B) \qquad P_{AB}(A)$$

$$P^{pred}(AB) \neq P_{AB}(A) + P_{AB}(B)$$

$logS^{pred}(PhCOOH)$	$logS^{pred}(Ph)$	$logS_{PhCOOH}(COOH)$
-1.55	-1.61	0.06
$logS^{pred}(PhCOOH)$	$logS^{pred}(COOH)$	$logS_{PhCOOH}(Ph)$
-1.55	-1.35	-0.20
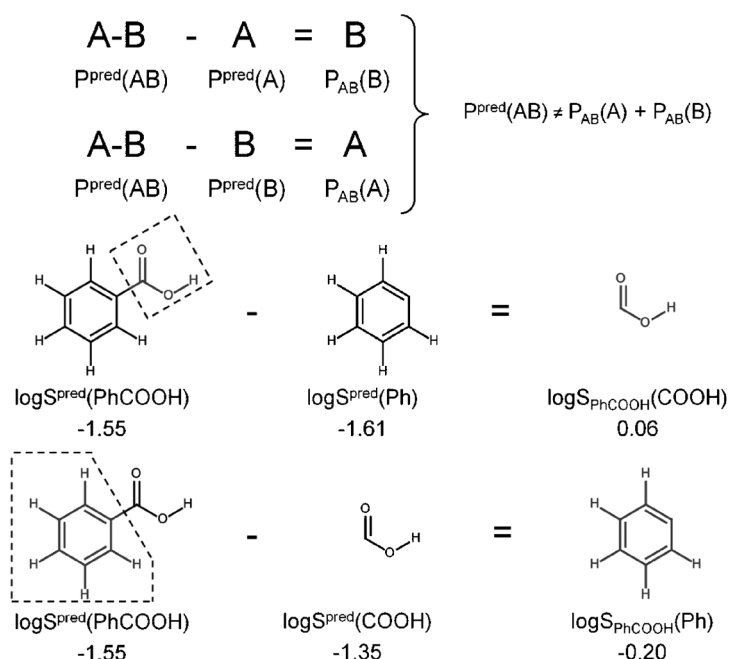
Figure 2. Non-additivity of calculated fragment contributions to property.

The developed interpretation method is not limited by two-fragment molecules, but could also be used for multi-fragment compounds, e.g., A-B-C (Figure 3). In this case to calculate the contribution of fragment B, one should predict the property value of structure A..C consisting of two separate fragments A and C, and subtract it from the predicted property value for compound A-B-C (Figure 3). This leads to one limitation of the proposed approach, because some software tools cannot correctly calculate descriptors for the multi-fragmented structures.

Calculated contributions of the same fragment situated in different molecules could be different. The issue of context-dependence is inherent for all interpretation methods due to the naturally observed mutual influence of molecular fragments in chemical compounds. Fragment contribution to an activity as calculated from a single compound
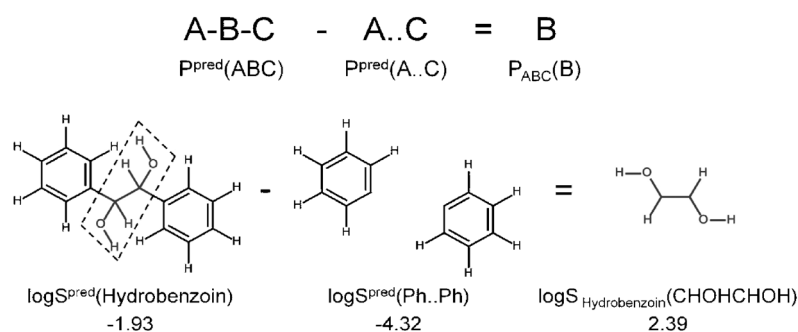
$$\underset{\underset{P^{pred}(ABC)}{}}{A\text{-}B\text{-}C} \quad - \quad \underset{\underset{P^{pred}(A..C)}{}}{A..C} \quad = \quad \underset{\underset{P_{ABC}(B)}{}}{B}$$

logS^pred(Hydrobenzoin)   logS^pred(Ph..Ph)   logS _Hydrobenzoin(CHOHCHOH)
        -1.93   -4.32   2.39

Figure 3. Structural interpretation of compounds consisting of three fragments.

gives a local interpretation of the fragment's influence on a property. To obtain a more broadly applicable (global) contribution of the fragment, one should average the values obtained from all compounds of the dataset containing this fragment. Thus, we could estimate the influence of the structural surroundings on the fragment's influence on an investigated property by considering the variation of fragment contribution throughout the whole dataset, which can be represented by standard deviation, variance, lower and upper quartiles, etc. Obviously, the greater such variation the higher the influence of structural surroundings. In addition, low variability of fragments contributions may indicate that the property is more additive by nature. Then, one can assume that using our approach it is possible to quantitatively estimate an additivity of a property. However, the validation of this assumption is out of scope of this paper and will be conducted as a separate study.

The interpretation workflow consisted of the following steps: (i) removal of the chosen molecular fragment from applicable molecule; (ii) processing of the remaining part(s) of the molecule and descriptors calculation for them; (iii) prediction (taking into account AD) of activity value for the remaining part(s) from step (ii); (iv) estimation of fragment contribution as the difference between predicted activity of the molecule and predicted values of the fragments obtained in step (iii) (Equation 1); (v) repetition of steps (i)-(iv) for all the molecules containing chosen fragment; and (vi) averaging of obtained fragment contributions (Equation 2). The obtained average value represents the global contribution of the fragment to the target property.

$$P_{AB_{(i)}}(A) = P^{pred}(AB_{(i)}) - P^{pred}(B) \tag{1}$$

$$P(A) = \frac{1}{n}\sum_{l=1}^{n} P_{AB_{(i)}}(A) \tag{2}$$

where $P_{AB_{(i)}}(A)$ is the local contribution of fragment A to the property value of the compound AB_i, $P^{pred}(AB_{(i)})$ is the predicted activity value of the compound $AB_{(i)}$ containing the fragment A, $P^{pred}(B)$ is the predicted activity for the molecu-

lar part B_i of the compound $AB_{(i)}$ remaining after removal of the fragment A, $P(A)$ is the global contribution of the fragment A to the property, and n is the number of compounds of the dataset containing fragment A.

The global contributions of the fragments could be calculated for both classification and regression tasks. In both cases the contributions will be continuous values that can be ranked in order to identify the most influential fragments responsible for increases or decreases of the investigated activity. Alternatively, global contributions can be calculated as a median value of all fragment contributions, not as an average one.

Molecular fragments for each investigated dataset were prepared differently. For the TG2 dataset they were drawn manually according to the SAR rules observed by Prime et al.[28] For the Ames dataset the fragments were represented as SMARTS according to toxicophores and detoxifying groups reported earlier.[30] The structures of the compounds from the solubility dataset were split into fragments using the Chemaxon fragment tool[51] according to the following simple rules: (i) bonds between two rings were cleaved; (ii) bonds between ring and chain were cleaved; and (iii) if after the bond cleavage one of the obtained fragments had only one heavy atom then such a split was ignored.

Removal of generated fragments from the molecules was performed with an in-house Python script using OpenEye OEChem library,[52] and 2D Dragon and SiRMS descriptors were generated for the remaining part(s) of molecules. Consensus prediction of their activity was then calculated separately for every modeling technique by averaging 50 single outputs of regression models (RF, PLS, or SVM) or by choosing the major class from results of 50 classification models (RF or SVM). Predicted values for the fragments that were outside of AD were discarded and then the global contributions of fragments were calculated according to Equation 2. Finally, we compared the global contributions obtained from different models using Pearson correlation coefficient (*R*) and root-mean-squared deviation (*RMSD*).

3 Results and Discussion

3.1 QSPR Analysis of Solubility Dataset

2D Simplex and Dragon descriptors were calculated for 1033 compounds of the solubility dataset (11335 and 1331 descriptors, respectively). Statistical characteristics of developed models are presented in Table 1. In total, 472 unique molecular fragments were automatically generated for this dataset by the Chemaxon fragment tool.[51] The global contributions of fragments occurring in the dataset at least 10 times are shown in Figure 4 and Figure 5. The global contributions of molecular fragments obtained by our approach correspond to the known SAR trends of solubility. For instance, the following fragments enhance the solubility: $-COOH$, $-OCH_3$, $-C(O)NH_2$, $-C(O)CH_3$, $-CH_2OH$, $-N(CH_3)_2$, $-SO_2NH_2$, etc., while ethyl and nitro groups slightly decrease the solubility. Phenyl-containing fragments significantly

reduce the solubility. We observed that the higher the degree of halogenation of the phenyl group, the stronger the negative influence of the fragment on the solubility (phenyl → 4-chlorophenyl → 3,4-dichlorophenyl) (see Figure 4 and Figure 5).

Table 1. Statistical characteristics of developed models for continuous datasets (average results of 10 runs of 5-fold external cross-validation).

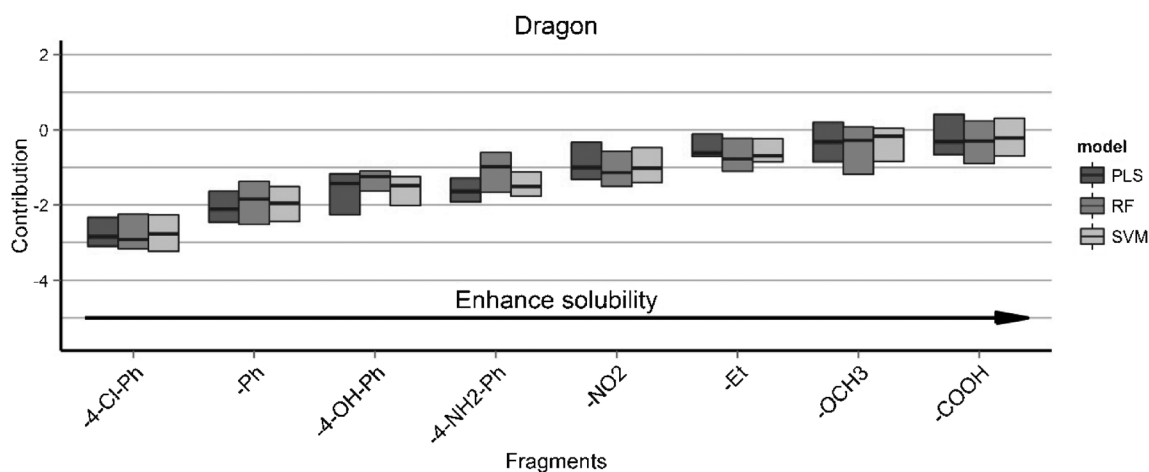| Endpoint	Model	SiRMS		Dragon	
		R_{CV}^2	RMSE	R_{CV}^2	RMSE
Solubility (logS)	PLS	0.84	0.82	0.91	0.60
	RF	0.88	0.71	0.91	0.62
	SVM	0.87	0.72	0.92	0.59
TG2 inhibition (IC_{50})	PLS	0.70	0.67	0.65	0.72
	RF	0.74	0.62	0.64	0.74
	SVM	0.70	0.67	0.68	0.70

Figure 4. Global contributions of fragments on aqueous solubility based on Dragon descriptors. Median values with upper (75%) and lower (25%) quartiles are shown by the boxplots.
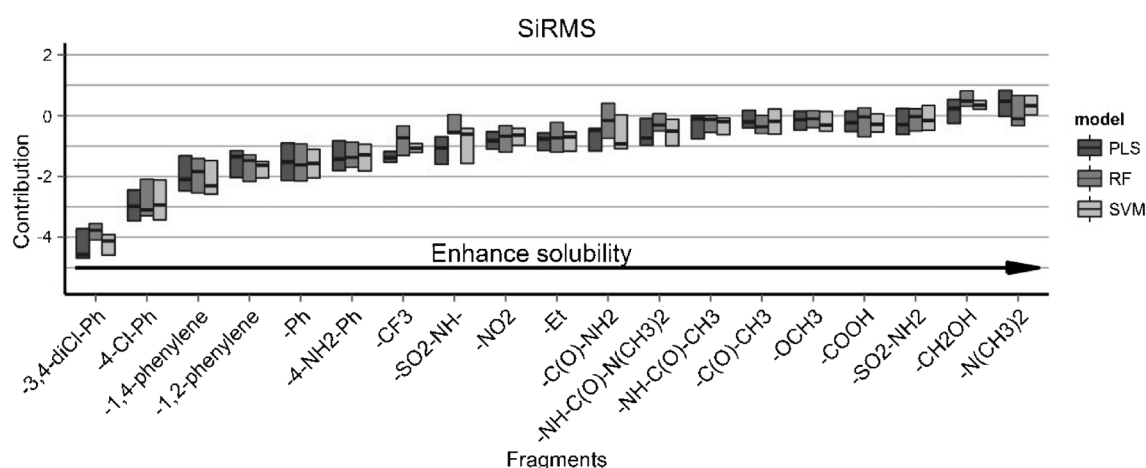

Figure 5. Global contributions of fragments on aqueous solubility based on SiRMS descriptors. Median values with upper (75%) and lower (25%) quartiles are shown by the boxplots.

The global contributions of such fragments as −C(O)CH₃, −OCH₃, −CH₂OH, −CF₃, etc. were consistent, i.e., these fragments were resistant to the influence of structural surroundings. Per contra, the global contributions of phenyl, 1,4-phenylene groups, etc., have large deviations that indicate that these fragments are sensitive to the influence of their structural neighborhood.

Reasonably high correlation ($R = 0.86–0.99$) between global fragment contributions obtained from different QSPR models was observed (Figure S1, see Supporting Information). The concordance between the contributions calculated from the same descriptors but from different modeling techniques was higher than the concordance between the contributions obtained from the same machine learning approach but different descriptors. This agrees with the observation that descriptors are more important for model building than modeling technique.[53] It should be also noted that the deviations of fragment contributions estimated by different modeling techniques are almost the same, thus the proposed interpretation approach could be considered as a robust one.

We chose the solubility dataset to demonstrate the non-additive nature of our approach. In other words, calculated fragment contributions obtained as a result of modeling of a non-additive property were also non-additive. In the example of benzoic acid, one can see that the calculated contributions of separate phenyl and carboxylic acid residues (−0.20 and 0.06, see Figure 2) were very different from predicted ones (−1.61 and −1.35, see Figure 2). The sum of calculated contributions (−0.14) is not equal to the sum of predicted property values (−2.96) because in our approach, the influence of structural surroundings was considered for the calculated contributions. The same observation could be made for any other compounds of the solubility dataset. It is worth noting that calculated contributions strongly depend on the modeling set, and they should not be considered as absolute values but rather used for the comparison of the influence of different structural fragments within a given dataset.

3.2 QSAR Analysis of Inhibitors of Transglutaminase 2 (TG2)

For 95 TG2 inhibitors, 4590 2D Simplex and 1017 2D Dragon descriptors were calculated. Statistical characteristics of developed models are presented in Table 1. Twenty four fragments reflecting the SAR rules published earlier[28] were considered. Global contributions of different molecular fragments are shown in Figures 6 and 7.

Reasonably high correlation ($R = 0.83–0.98$) between global contributions obtained from different QSAR models was observed (Figure S2, see Supporting Information). However, much more important information could be derived from the comparison of the contributions obtained by our approach with known SAR rules. For instance, Prime et al.[28] showed that the stronger the electron withdrawing substituents that are situated in position 4 of the phenyl ring (Scaffold I), the higher the inhibitory activity. The same trend was observed by us. Thus, according to the calculated fragment contributions, *p*-nitrophenyl group increased the inhibition potency and *p*-methoxyphenyl decreased it (Figure 6). Numerical values of global contributions varied across the models, but the observed trend remained the same irrespective of descriptors and modeling technique. As can be seen in Figure 7, any changes in the structure of the acryl residue lead to decrease or loss of inhibitory activity. Thus, our conclusion about the importance of the acryl residue for high inhibition activity served as another example of perfect concordance between our interpretation and existing SAR rules.[28]

3.3 QSAR Analysis of Ames Mutagenicity Dataset

For the Ames dataset, 19,612 2D Simplex and 1601 2D Dragon descriptors were calculated. Statistical characteristics of developed models are presented in Table 2. Known

Table 2. Statistical characteristics of developed models for Ames datasets (average results of 10 runs of 5-fold external cross-validation).

| Descriptors	Algorithm	Accuracy	Balanced Accuracy
SiRMS	RF	0.818	0.817
	SVM	0.800	0.800
Dragon	RF	0.817	0.816
	SVM	0.794	0.793

toxicophores and detoxifying fragments (Figure S3, see Supporting Information) described by Kazius et al.[30] were represented as SMARTS strings. In addition, in some cases (i.e., epoxide, aziridine, triazene, aromatic hydroxylamine, K and bay-region polycyclic aryl) all hydrogen atoms attached to the core structure were considered as a part of the fragment.
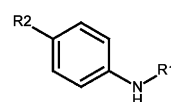
Scaffold I
R1 = various acyl groups(preferably acryl group);
R2 = NO₂, F, Br, CF₃, CH₃, OCH₃.
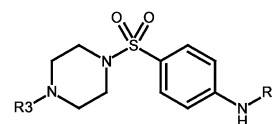
Scaffold II
R1 = various acyl groups (preferably acryl group and its derivatives);
R3 = various acyl groups (preferably Boc, Cbz and its derivatives), substituted phenyl and pyridyl groups.

Our analysis of the global contributions of fragments (Figure 8) showed that polycyclic aromatic groups, nitrosamine, alkyl nitrite, and triazene fragments strongly increased the mutagenicity. These conclusions completely
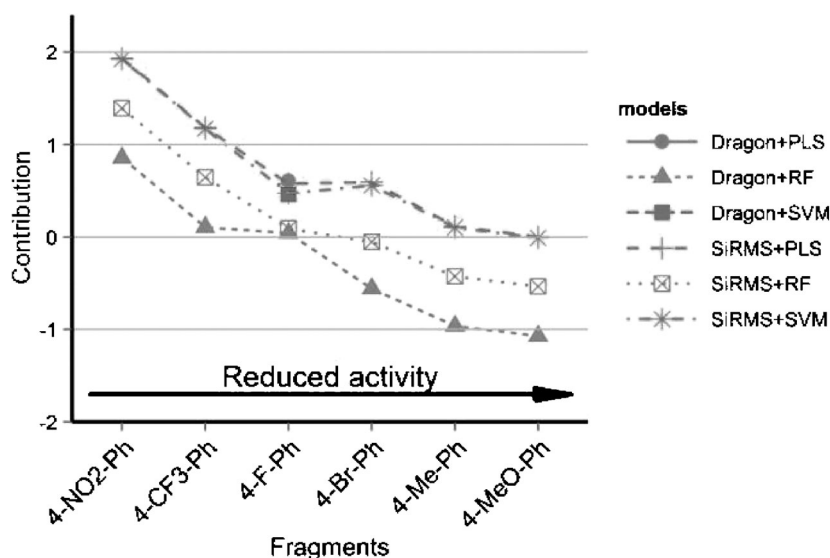
Figure 6. Global contributions of different molecular fragments (phenyl ring with attached substituent R2) of Scaffold I compounds on TG2 inhibition estimated by PLS, RF, and SVM models. For Dragon-PLS and Dragon-SVM models, only the predictions of 4-fluorophenyl group fell inside AD of corresponding models.
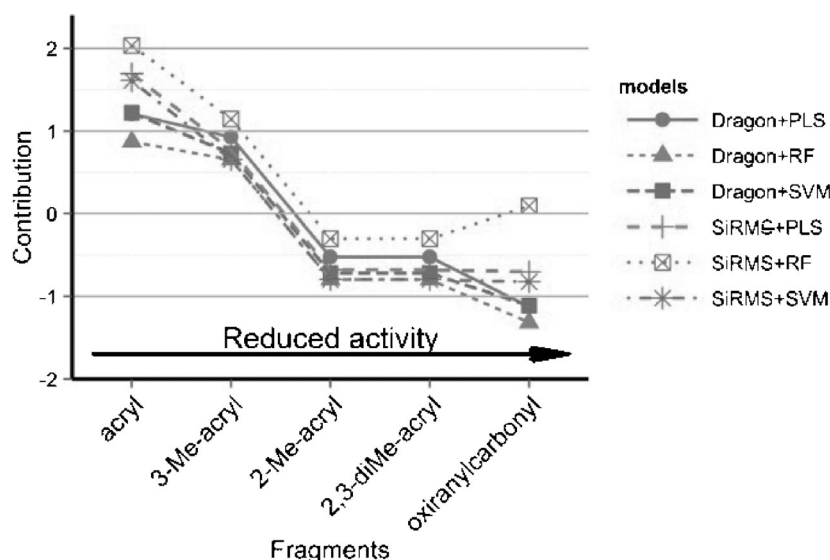

Figure 7. Global contributions of different molecular fragments R1 of Scaffold I and II compounds on TG2 inhibition estimated by PLS, RF and SVM models.

corresponded to already published findings.[30,54–55] Metabolites of polycyclic aromatic compounds can form DNA adducts and produce toxic effects.[54] A nitrosamine group, after enzymatic hydroxylation of the neighboring α-carbon and prior to the formation and cleavage of the carbon-diazonium bond, transforms in an electrophilic carbocation[55] which also causes toxic effects. However, according to our analysis, the contribution to mutagenicity of such a well-known toxicophore as an aromatic nitro group[30] is pretty low. This could be explained by two reasons. First, the contributions were always calculated separately for each frag-

ment. Thus, these are local contributions of single fragments. Often, compounds can simultaneously contain several mutagenic fragments, e.g., two or more nitro groups, or nitro group and polycyclic aromatic system. Then, removing one of such groups from the mutagenic compound will not affect the predicted mutagenicity of the remaining part of the molecule, which will still contain another nitro group. Then, the contribution of excluded nitro group will be equal to 0 because, according to QSAR predictions, mutagenicity of the whole compound = mutagenicity of remaining part of the molecule = 1; and the contribution of
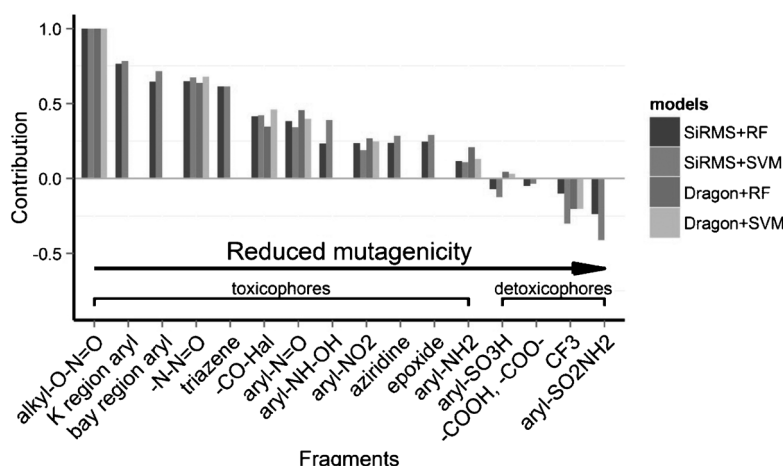
Figure 8. Global contributions of different molecular fragments on mutagenicity estimated by RF and SVM models.

nitro group $= 1 - 1 = 0$. Obviously, the presence of such compounds will decrease the relative contributions of such mutagenic fragments as nitro group. However, obtained positive value of the nitro group contribution demonstrates its certain mutagenic potency. The second reason was given by Kazius et al.,[30] who indicated that detoxifying groups, such as carboxylic acid or esters, arylsulfonic acid, trifluoromethyl or arylsulfonylamide fragments, can neutralize the influence of the toxic fragments, in particular NO_2. As follows from our analysis, the presence of these detoxifiers in the compounds decreases their toxic effect (Figure 8). Moreover, we have calculated the "mutagenicity ratio" for nitro group $=$ number of mutagens containing nitro group/total number of NO_2-containing compounds $= 530/626 = 0.82$. This value is close to nitro group "mutagenicity ratio" $= 0.89$ obtained by Kazius et al.[30] Overall, similarly to the first two datasets, good concordance between fragment contributions was observed, irrespective of the descriptors and modeling methods used.

Concluding this section, we especially want to note that the models developed in this paper were used as proof-of-concept case studies and were not aimed for very comprehensive investigation of solubility, mutagenicity, or inhibition of Transglutaminase 2. Good concordance between our interpretation results and experimentally derived SAR rules previously published was observed for all QSAR models irrespective of the endpoint nature, activity scale, molecular descriptors, or modeling technique.

4 Conclusions

A universal approach for structural interpretation of QSAR models was developed. We have proved that our methodology is capable of retrieving reliable SAR rules or non-additive global contributions of molecular fragments to investigated activity. Our developed interpretation approach is applicable to any QSAR models irrespective of endpoint type (continuous or classification) and nature (physical-chemical property or biological activity), descriptors, and machine learning methods. We observed that estimated fragment contributions are more sensitive to chemical descriptors than to modeling technique. Correct choice of descriptors is even more important because some descriptors cannot describe structures consisting of several unbounded fragments or lacking hydrogens. However, poor descriptor choice can only reduce the number of considered molecular fragments but will not affect the quality of interpretation.

We validated the reliability of our developed approach using several case studies (including a variety of endpoints, activity scales, descriptors, and modeling techniques) by comparing the global contributions of molecular fragments obtained using our methodology with known SAR rules derived experimentally. In all cases, high concordance between our interpretation and results published by others was observed. Only SiRMS and Dragon descriptors were used in the current study, but the proposed interpretation approach could be easily extended to other types of descriptors. However, we would recommend SiRMS descriptors to achieve a larger variety of fragments for which the contributions can be calculated.

The proposed approach is a good tool for interpretation of such "black box" models like random forest, neural networks, etc. In addition, it could be useful for the identification of key fragments (structural alerts) that can be used in the development of expert systems. The most important fragments can be chosen for the design of novel compounds. Analysis of global contributions and their deviation could help to maximize the positive influence of structural surroundings on the given fragment and to decrease the negative effects.

The developed approach was implemented as a web-based software tool which will be available soon at http://qsar4u.com.

References

[1] C. Hansch, P. P. Maloney, T. Fujita, R. M. Muir, *Nature* **1962**, *194*, 178–180.

[2] A. Tropsha, *Mol. Inf.* **2010**, *29*, 476–488.

[3] R. Guha, *J. Comput.-Aided Mol. Des.* **2008**, *22*, 857–871.

[4] A. G. Artemenko, E. N. Muratov, V. E. Kuz'min, N. A. Kovdienko, A. I. Hromov, V. A. Makarov, O. B. Riabova, P. Wutzler, M. Schmidtke, *J. Antimicrob. Chemother.* **2007**, *60*, 68–77.

[5] V. E. Kuz'min, A. G. Artemenko, R. N. Lozitska, A. S. Fedtchouk, V. P. Lozitsky, E. N. Muratov, A. K. Mescheriakov, *SAR QSAR Environ. Res.* **2005**, *16*, 219–230.

[6] V. E. Kuz'min, A. G. Artemenko, E. N. Muratov, I. L. Volineckaya, V. A. Makarov, O. B. Riabova, P. Wutzler, M. Schmidtke, *J. Med. Chem.* **2007**, *50*, 4205–4213.

[7] A. Lagunin, A. Zakharov, D. Filimonov, V. Poroikov, *SAR QSAR Environ. Res.* **2007**, *18*, 285–298.

[8] G. V. Kokurkina, M. D. Dutov, S. A. Shevelev, S. V. Popkov, A. V. Zakharov, V. V. Poroikov, *Eur. J. Med. Chem.* **2011**, *46*, 4374–4382.

[9] A. Lagunin, A. Zakharov, D. Filimonov, V. Poroikov *Mol. Inf.* **2011**, *30*, 241–250.

[10] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, *Classification and Regression Trees*, Wadsworth, Belmont, CA, **1984**, p 368.

[11] S. Wold, A. Ruhe, H. Wold, W. J. Dunn III, *SIAM J. Sci. Stat. Comp.* **1984**, *5*, 735–743.

[12] V. E. Kuz'min, P. G. Polischuk, A. G. Artemenko, S. Y. Makan, S. A. Andronati *SAR QSAR Environ. Res.* **2008**, *19*, 213–244.

[13] R. Guha, P. C. Jurs *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1440–1449.

[14] S. B. Larsen, F. S. Jorgensen, L. Olsen, *J. Chem. Inf. Model.* **2008**, *48*, 233–241.

[15] V. Kuz'min, E. Muratov, A. Artemenko, L. Gorb, M. Qasim, J. Leszczynski, *J. Comput.-Aided Mol. Des.* **2008**, *22*, 747–759.

[16] S. Wold, M. Sjöström, L. Eriksson, *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130.

[17] D. T. Stanton, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1423–1433.

[18] L. Rosenbaum, G. Hinselmann, A. Jahn, A. Zell, *J. Cheminf.* **2011**, *3*, 11.

[19] R. Guha, D. T. Stanton, P. C. Jurs, *J. Chem. Inf. Model.* **2005**, *45*, 1109–1121.

[20] V. E. Kuz'min, P. G. Polishchuk, A. G. Artemenko, S. A. Andronati, *Mol. Inf.* **2011**, *30*, 593–603.

[21] L. Breiman, *Mach. Learn.* **2001**, *45*, 5–32.

[22] R. Guha, P. C. Jurs, *J. Chem. Inf. Model.* **2005**, *45*, 800–806.

[23] L. Carlsson, E. A. Helgee, S. Boyer, *J. Chem. Inf. Model.* **2009**, *49*, 2551–2558.

[24] K. Hasegawa, K. Migita, K. Funatsu, in *Knowledge-Oriented Applications in Data Mining* (Ed: K. Funatsu), InTech, New York, **2011**, pp. 167–182.

[25] I. I. Baskin, A. O. Ait, N. M. Halberstam, V. A. Palyulin, N. S. Zefirov, *SAR QSAR Environ. Res.* **2002**, *13*, 35–41.

[26] G. Marcou, D. Horvath, V. Solov'ev, A. Arrault, P. Vayer, A. Varnek, *Mol. Inf.* **2012**, *31*, 639–642.

[27] J. Huuskonen, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 773–777.

[28] M. E. Prime, O. A. Andersen, J. J. Barker, M. A. Brooks, R. K. Y. Cheng, I. Toogood-Johnson, S. M. Courtney, F. A. Brookfield, C. J. Yarnold, R. W. Marston, P. D. Johnson, S. F. Johnsen, J. J. Palfrey, D. Vaidya, S. Erfan, O. Ichihara, B. Felicetti, S. Palan, A. Pedret-Dunn, S. Schaertl, I. Sternberger, A. Ebneth, A. Scheel, D. Winkler, L. Toledo-Sherman, M. Beconi, D. Macdonald, I. Muñoz-Sanjuan, C. Dominguez, J. Wityak, *J. Med. Chem.* **2012**, *55*, 1021–1046.

[29] K. Hansen, S. Mika, T. Schroeter, A. Sutter, A. ter Laak, T. Steger-Hartmann, N. Heinrich, K.-R. Müller, *J. Chem. Inf. Model.* **2009**, *49*, 2077–2081.

[30] J. Kazius, R. McGuire, R. Bursi, *J. Med. Chem.* **2005**, *48*, 312–320.

[31] D. Fourches, E. Muratov, A. Tropsha, *J. Chem. Inf. Model.* **2010**, *50*, 1189–1204.

[32] *Standartizer*, 5.4, Chemaxon, Budapest, Hungary.

[33] V. E. Kuz'min, A. G. Artemenko, P. G. Polischuk, E. N. Muratov, A. I. Khromov, A. V. Liahovskiy, S. A. Andronati, S. Y. Makan, *J. Mol. Model.* **2005**, *11*, 457–467.

[34] V. Kuz'min, A. Artemenko, E. Muratov, *J. Comput.-Aided Mol. Des.* **2008**, *22*, 403–421.

[35] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, **2000**, p. 667.

[36] *Dragon*, 5.5, Talete srl, Milano, Italy.

[37] W. L. Jolly, W. B. Perry, *J. Am. Chem. Soc.* **1973**, *95*, 5442–5450.

[38] R. Wang, Y. Fu, L. Lai, *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 615–621.

[39] B. V. Ioffe, *Chemistry Refractometric Methods*, 3rd ed., Himiya, Leningrad, **1983**.

[40] E. N. Muratov, A. G. Artemenko, E. V. Varlamova, P. G. Polischuk, V. P. Lozitsky, A. S. Fedtchouk, R. L. Lozitska, T. L. Gridina, L. S. Koroleva, V. N. Sil'nikov, A. S. Galabov, V. A. Makarov, O. B. Riabova, P. Wutzler, M. Schmidtke, V. E. Kuz'min, *Future Med. Chem.* **2010**, *2*, 1205–1226.

[41] *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, **2012**.

[42] A. Liaw, M. Wiener, *R News* **2002**, *2(3)*, 18–22.

[43] N. Kraemer, A.-L. Boulesteix, *ppls: Penalized Partial Least Squares*, R package version 1.05; http://CRAN.R-project.org/package=ppls.

[44] E. Dimitriadou, K. Hornik, F. Leisch, D. Meyer, A. Weingessel, *Misc Functions of the Department of Statistics (e1071)*, TU Wien, R package version 1.6; http://CRAN.R-project.org/package=e1071.

[45] P. G. Polischuk, E. N. Muratov, A. G. Artemenko, O. G. Kolumbin, N. N. Muratov, V. E. Kuz'min, *J. Chem. Inf. Model.* **2009**, *49*, 2481–2488.

[46] J. B. Kruskal, *Proc. Am. Math. Soc.* **1956**, *7*, 48–50.

[47] A. G. Leach, H. D. Jones, D. A. Cosgrove, P. W. Kenny, L. Ruston, P. MacFaul, J. M. Wood, N. Colclough, B. Law, *J. Med. Chem.* **2006**, *49*, 6672–6682.

[48] E. Griffen, A. G. Leach, G. R. Robb, D. J. Warner, *J. Med. Chem.* **2011**, *54*, 7739–7750.

[49] G. Papadatos, M. Alkarouri, V. J. Gillet, P. Willett, V. Kadirkamanathan, C. N. Luscombe, G. Bravi, N. J. Richmond, S. D. Pickett,

J. Hussain, J. M. Pritchard, A. W. J. Cooper, S. J. F. Macdonald, *J. Chem. Inf. Model.* **2010**, *50*, 1872–1886.

[50] S. M. Free, J. W. Wilson, *J. Med. Chem.* **1964**, *7*, 395–399.

[51] *Fragmenter*, 5.4, Chemaxon, Budapest, Hungary.

[52] *OEChem*, 1.7.4, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, **2010**.

[53] S. S. Young, F. Yuan, M. Zhu, *Mol. Inf.* **2012**, *31*, 707–710.

[54] B. Singer, J. T. Kusmierek, *Annu. Rev. Biochem.* **1982**, *51*, 655–691.

[55] L. Fishbein, *Potential Industrial Carcinogens and Mutagens*, Elsevier, Amsterdam, The Netherlands, **1979**.