Chemical Domain of QSAR Models from Atom-Centered Fragments

Ralph Kühne,[†] Ralf-Uwe Ebert,[†] and Gerrit Schüürmann*,[†,‡]

UFZ Department of Ecological Chemistry, Helmholtz Centre for Environmental Research, Permoserstrasse 15, 04318 Leipzig, Germany and Institute for Organic Chemistry, Technical University Bergakademie Freiberg, Leipziger Strasse 29, 09596 Freiberg, Germany

Received August 20, 2009

A methodology to characterize the chemical domain of qualitative and quantitative structure−activity relationship (QSAR) models based on the atom-centered fragment (ACF) approach is introduced. ACFs decompose the molecule into structural pieces, with each non-hydrogen atom of the molecule acting as an ACF center. ACFs vary with respect to their size in terms of the path length covered in each bonding direction starting from a given central atom and how comprehensively the neighbor atoms (including hydrogen) are described in terms of element type and bonding environment. In addition to these different levels of ACF definitions, the ACF match mode as degree of strictness of the ACF comparison between a test compound and a given ACF pool (such as from a training set) has to be specified. Analyses of the prediction statistics of three QSAR models with their training sets as well as with external test sets and associated subsets demonstrate a clear relationship between the prediction performance and the levels of ACF definition and match mode. The findings suggest that second-order ACFs combined with a borderline match mode may serve as a generic and at the same time a mechanistically sound tool to define and evaluate the chemical domain of QSAR models. Moreover, four standard categories of the ACF-based membership to a given chemical domain (outside, borderline outside, borderline inside, inside) are introduced that provide more specific information about the expected QSAR prediction performance. As such, the ACF-based characterization of the chemical domain appears to be particularly useful for QSAR applications in the context of REACH and other regulatory schemes addressing the safety evaluation of chemical compounds.

INTRODUCTION

Qualitative and quantitative structure−activity relationship (QSAR) models represent empirical rules derived from training sets of compounds. The latter are always finite and thus constrained with regard to the range of chemical structures and properties, implying corresponding limitations for the applicability domains of QSARs. In the case of QSAR predictions for compounds outside a given model domain, existing model statistics do not inform well about the expected prediction performance, and the resultant predictions may well be inaccurate or even seriously in error.

After introducing advanced techniques to characterize the descriptor domain through probabilistic analyses of the training set coverage in the *n*-dimensional descriptor space,[1,2] a comprehensive approach to address the applicability domain has been introduced.[3] In addition to considering the ranges and distributions of descriptor values, methods are needed to take into account the profile of fundamental physicochemical properties of the training set compounds as well as their structural features and mechanistic disposition. Accordingly, the applicability domain of a given QSAR model is now understood to include the physicochemical domain, the descriptor domain, the chemical domain, which is also called chemical space or structural domain, and the mechanistic domain, possibly augmented by a metabolic domain in the case of an explicit impact of biotransformation on the target value of interest.[3]

The need to characterize the model applicability domain is also reflected in the OECD guidelines for QSAR model validation.[4,5] In the European Union, the REACH (Registration, Evaluation, Authorization, and Restriction of Chemicals) Directive[6] emphasizes the inclusion of nontest methods such as QSARs and read-across for performing chemical safety evaluation, and in this context, applicability domain characterizations become crucial.

One key aspect of the model applicability is the chemical space. A method cannot be expected to yield reliable results for chemical structures significantly different from the training set compounds, keeping in mind that chemical similarity itself cannot uniquely be defined but requires reference to some concept and operational procedure. There are several case studies demonstrating the relevance of this chemical or structural domain.[7−10] Various methods for its characterization independent from the descriptor domain[11−14] or in combination with it[1] when structural features are directly applied as some of the model parameters have been suggested. Thus far, however, there is no generally accepted or even standardized approach for defining the chemical space associated with QSAR models.

This paper focuses on the chemical domain as one component of the QSAR application domain and more specifically on its characteristics that are independent of a given QSAR end point. To this end, the suitability of the atom-centered fragment (ACF)[15,16] method for characterizing

* Corresponding author phone: +49-341-235-1262; fax: +49-341-235-1785; e-mail: gerrit.schuurmann@ufz.de.
[†] UFZ Helmholtz Centre for Environmental Research.
[‡] Technical University Bergakademie Freiberg.

10.1021/ci900313u CCC: $40.75
Published on Web 11/20/2009

the chemical space of a compound set is examined. The ACF method decomposes molecules into structural fragments consisting of a central atom and bonding neighbors. An ACF is then defined through the atom type and the number and type of bonding neighbors and associated bond types. In the present context, only non-hydrogen atoms are used as central atoms, while for each ACF both non-hydrogen atoms and hydrogens may serve as neighbor atoms. If for a given central atom only directly bonded neighbors (the first neighbor shell) are taken into account, the ACF is called first-order ACF. Taking into account the first and second neighbor in each bonding direction (path length 2) leads to second-order ACFs, and accordingly *n*th-order ACFs are generated by increasing the path length in each bond direction, as far as possible, to *n*. Each atom acts as the central atom of one ACF, and its number equals the number of non-hydrogen atoms.

In the QSAR context, ACF-based structural similarity has already been demonstrated to yield significant improvements of the prediction quality through an automatized judicious correction for local model errors[16] to provide performance-optimizing model selections[17] and as a metric suitable for the semiquantitative *k*NN (*k* nearest neighbors) read-across of environmental half-lives.[18] The general usefulness of atom-centered fragments to characterize model application domains has been pointed out,[3,19] and ACFs have also been applied already to characterize the chemical domain of QSAR models.[20,21] Thus far, however, detailed guidance on the ACF setting for this kind of application and on, respectively, optimized procedures is not yet available.

While the ACF approach is straightforward in principle, practical applications require some specifications. These specifications are not obvious from theoretical considerations, and need to be optimized for the particular application in mind by test calculations.

The first question to be answered is the exact definition of the individual ACFs. This includes the atomic properties of the central atoms to be taken into account, the path length to be considered, the relevant properties of the neighbor atoms, and the properties of the bonds. Even with a fixed ACF definition, there is still ambiguity in the evaluation of the comparison of the ACF occurrences between a test compound and a model training set. Accordingly, the ACF-based chemical domain depends on both the ACF definition (first order vs second order etc.) and the ACF match mode (different degrees of strictness for the ACF comparison such as loose vs moderate) as will be shown in more detail below. It follows that both components of the ACF approach need to be considered when specifying its concrete setting for a given application such as defining chemical (structural) similarity between individual molecules or the chemical space of QSAR models.

Existing property estimation models from the literature may be utilized to explore the impact of these ACF specifications on the reliability of domain characterization. To be suitable, model candidates need to match three criteria. First, the model must be fully described and reproducible in order to run it for test compounds. Second, the training set must be fully disclosed and should be of a reasonable size. Third, a sufficient number of additional data outside the original training set are required. Preferably, these data should not have been available at the time of the model development.

MATERIAL AND METHODS

Literature Models. This study utilizes three models from the literature to optimize the ACF domain characterization. Besides the prerequisites explained in the Introduction, the model selection considered also the inclusion of both numerical and categorical models. With regard to the former, a fragment method and a model employing continuously defined descriptors have been used, and a structural alert model was taken as an example of categorical models.

The model to estimate the logarithmic molar water solubility log S_w at 25 °C by Kühne et al.[22] exploits substructure fragments together with experimental melting points. The latter is only needed for compounds that are solid at this temperature. The training set of this model covered 694 organic compounds.

The model of Abraham et al.[23] to predict the logarithmic air/water partition coefficient log K_{aw} has been selected as a test example for a quantitative model employing numerical descriptors. This model was fitted for a training set of 408 substances.

Finally, the structural alert model of von der Ohe et al.[24] allows one to predictively discriminate between narcosis-level and excess-toxic compounds with respect to *Daphnia magna*. In this context, a compound is defined as excess toxic if its actual toxicity is larger than the associated baseline narcosis toxicity by a certain threshold; the extent of excess toxicity is characterized by the respective ratio and termed T_e, toxicity enhancement. The original training set contained toxicity data for 300 compounds.

Test Sets. In all cases, external tests sets of considerably larger sizes than the respective training sets were available. For the log S_w model,[22] 1868 data for compounds not part of that training set were used. Only compounds either known to be liquid or with known experimental melting points have been considered. All experimental data were taken from the EPISUITE[25] modules WSKOWWIN[26] (S_w) and MPBPWIN[27] (melting points).

For the Abraham model to predict log K_{aw}[23] with experimental data from the EPISUITE module HENRYWIN,[28] only compounds with available experimental Abraham parameters collected in our in-house database within the software system ChemProp[29,30] have been taken into account, resulting in 569 test set compounds.

The test set for the structural alert model to identify excess-toxic compounds[24] covered 709 substances. In this model, the discrimination between narcosis level and excess toxicity was based on experimental LC_{50} data taken from the literature and baseline narcosis predictions through the associated model equation[24] and log K_{ow} (logarithmic octanol/water partition coefficient) calculated through the EPISUITE module KOWWIN.[31]

ACF Definition. Four different ACF versions have been tested. These were first-order ACFs (only neighbors directly bonded to central atom) in two versions, second-order ACFs (first and second neighbors in each bonding direction, as far as available), and third-order ACFs (first, second, and third neighbors in each bonding direction, as far as available). For first-order ACFs, two versions were included, a simple version and an extended version. Essentially, the extended first-order version already incorporates features that are required to ensure unambigu-

Table 1. Definition of Atom-Centered Fragments (ACFs)

object	feature
simple first-order ACF	
center atom	atom type: chemical element
	exact number of hydrogen atoms
	aromaticity: yes or no
neighbor atom	atom type: chemical element
	total number of neighbors including hydrogen atoms
bond	bond type: single nonaromatic, double nonaromatic, triple nonaromatic, aromatic
extended first-order ACF	
center atom	atom type: chemical element
	exact number of hydrogen atoms
	aromaticity: yes or no
	ring membership: yes or no
neighbor atom	atom type: chemical element
	exact number of hydrogen atoms
	ring membership: yes or no
	total number of neighbors including hydrogen atoms
bond	bond type: single nonaromatic, double nonaromatic, triple nonaromatic, aromatic
ring closure inside ACF (including 2nd-order neighbors outside the ACF)	to be considered explicitly
second-order and third-order ACF	
center atom	atom type: chemical element
	exact number of hydrogen atoms
	aromaticity: yes or no
	ring membership: yes or no
first-order neighbor atom	atom type: chemical element
	exact number of hydrogen atoms
	ring membership: yes or no
second-order and third-order neighbor atom	atom type: chemical element
	exact number of hydrogen atoms
	aromaticity: yes or no
	ring membership: yes or no
	total number of neighbors including hydrogen atoms
bond	bond type: single nonaromatic, double nonaromatic, triple nonaromatic, aromatic
ring closure inside ACF (including third-order or fourth-order neighbors outside the ACF, respectively)	to be considered explicitly

ousness of the results in the higher orders. The details of the ACF definitions are given in Table 1.

To characterize the chemical domain of the training set in terms of a particular ACF definition, it is not necessary to know which ACFs occur in a particular individual compound. Instead, a comprehensive list of all ACFs occurring in any of the compounds of the training sets needs to be built. For each *individual ACF* (but not, as already pointed out, for each *individual compound*), the number of compounds containing it needs to be recorded as well as the maximum number of occurrences in any individual compound. Again, it is not required to record in which compound the maximum frequencies occur.

ACF Match Mode. The strictness of comparison between the test compound ACFs and the training set ACF pool, the ACF match mode, has been varied. According to the loose ACF match mode, a compound is considered to be inside the chemical domain if all of its ACFs are found at least once in the ACF training set pool. With the moderate ACF match mode, the number of occurrence of each ACF is taken into account. Here, for a given test compound the number of occurrence of each ACF must not exceed the respective maximum number of occurrence in any of the training set compounds. In addition, still stricter variants of ACF comparison such as considering also the minimum number of occurrence of each individual ACF in the training set compounds have been tested but turned out to yield overall inferior results and thus are not reported here in detail. Finally, the results discussed below led us introduce the borderline ACF match mode that corresponds to the loose mode except that the mismatch of one ACF (one test compound ACF is missing in the training set ACF pool) is allowed for still classifying this test compound to belong to the chemical domain of the QSAR model of interest. Examples illustrating the ACF match mode with concrete chemical structures and different ACF orders are given below (see Results and Discussion).

Statistical Performance. For quantitative models (in our case log S_w and log K_{aw} models), the calibration quality was quantified through calculation of the squared correlation coefficient, r^2, and the prediction performance through the predictive squared correlation coefficient, q^2 [32]

$$r^2 = 1 - \frac{\sum_i (y_i^{fit} - y_i^{obs})^2}{\sum_i (y_i^{obs} - y^{mean})^2} \quad (1)$$

$$q^2 = 1 - \frac{\sum_i (y_i^{pred} - y_i^{obs})^2}{\sum_i (y_i^{obs} - y^{mean})^2} \quad (2)$$

In eqs 1 and 2, y_i^{obs} denotes the experimental target value of compound i and y^{mean} the mean experimental value of the data set under investigation (which is the training set mean in the case of r^2 and the prediction set mean in the case of q^2). Moreover, y_i^{fit} in eq 1 and y_i^{pred} in eq 2 denote the calibrated (regression fitted) and predicted target value, respectively. While r^2 corrects for systematic errors and

Table 2. log S_w Model[22] Statistics for the Training Set, External Test Set, and Associated Test Subsets Defined through Different Levels of ACF-Defined Chemical Domains, Varying Both the ACF Definition and the ACF Match Mode[a]

chemical domain coverage	n	r^2	q^2	rms	bias	mne	mpe	fraction (%) with absolute errors < 0.3	fraction (%) with absolute errors < 0.5	fraction (%) with absolute errors < 1
training set	694	0.95	0.95	0.55	±0.00	−2.62	+3.43	55.9	73.9	93.1
full external test set	1868	0.64	0.56	1.36	+0.13	−8.90	+7.29	27.8	42.2	66.0
ACF match mode										
simple first-order ACF										
loose	1237	0.73	0.71	1.14	+0.04	−8.90	+7.29	31.9	47.1	73.0
moderate	1054	0.77	0.76	1.01	+0.01	−3.90	+7.29	35.0	50.9	76.0
extended first-order ACF										
loose	752	0.80	0.79	0.97	+0.09	−2.80	+4.62	33.9	50.1	75.1
moderate	638	0.79	0.78	0.90	+0.07	−2.80	+3.16	35.7	52.8	76.8
second-order ACF										
borderline	**266**	**0.90**	**0.90**	**0.67**	**+0.03**	**−2.14**	**+1.98**	**40.2**	**58.3**	**85.3**
loose	171	0.89	0.89	0.71	+0.04	−2.14	+1.98	38.0	54.4	83.0
moderate	124	0.91	0.90	0.57	+0.13	−1.48	+1.79	46.8	62.9	91.1
third-order ACF										
loose	53	0.92	0.91	0.64	+0.09	−1.48	+1.49	37.7	52.8	84.9
moderate	28	0.92	0.92	0.51	+0.11	−1.48	+1.01	50.0	67.9	92.9

[a] n: Number of compounds. r^2: squared correlation coefficient. q^2: predictive squared correlation coefficient. rms: root-mean-square error. bias: systematic error. mne: maximum negative prediction error (underestimation). mpe: maximum positive prediction error (overestimation).

ranges between 0 (no correlation) and 1 (perfect calibration), q^2 ranges from $-\infty$ (worst case) to 1 (perfect model) with $q^2 = 0$ representing the case where the model prediction is as good as taking the experimental mean as a predictor for all values. When applying a regression model to an external test set, r^2 quantifies the model precision (trend) and q^2 the model accuracy (absolute value agreement).

To further characterize the quantitative model performance, the following additional parameters have been used: root-mean-square error (rms), systematic error (bias), maximum negative error (mne, largest underestimation), maximum positive error (mpe, largest overestimation), absolute error mean, error range, absolute error mean of top-10%, absolute error mean of top-30%, and absolute error mean of top-50%.

For the structural alert model to discriminate between narcosis level and excess toxicity as a binary classification scheme, the following contingency table statistics have been used

$$\text{sensitivity}(+) = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{predictivity}(+) = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

$$\text{sensitivity}(-) = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (5)$$

$$\text{predictivity}(-) = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (6)$$

$$\text{concordance} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} = \frac{\text{TP} + \text{TN}}{n} \quad (7)$$

In eqs 3−7, TP represents the number of true positive compounds (in our case positive = excess toxic), TN the number of true negative compounds (in our case negative = not excess toxic = narcosis level), and FP and FN the numbers of false positive (actually narcosis level but predicted to be excess toxic) and false negative (actually excess toxic but predicted to be narcosis level) compounds; n in eq 7 denotes the total number of compounds (in our case training set, $n = 300$ and test set, $n = 709$). Sensitivity(+) quantifies the fraction of excess-toxic compounds recognized by the model and sensitivity(−) the fraction of narcosis-level compounds correctly recognized (recognition power). By contrast, predictivity(+) quantifies the fraction of compounds predicted to be excess toxic that are actually excess toxic and predictivity(−) the fraction of compounds predicted to be narcosis level that are actually narcosis level (prediction power). Finally, the concordance quantifies the fraction of compounds classified correctly by the model and as such corresponds to q^2 of quantitative models.

Computer Implementation. Except for the KOWWIN calculation, all model estimations as well as the ACF domain investigations were carried out within the in-house software system ChemProp.[29,30]

RESULTS AND DISCUSSION

Threshold for the Chemical Domain Coverage. The three models have been applied to their original training sets as well as to the above-described external test sets. For each test set, combination of the four different ACF definitions (first-order simple, first-order extended, second-order, third-order) and the level of ACF comparison (loose, moderate) led to correspondingly different subsets that represented different levels of membership to its chemical domain. The major statistical results are summarized in Tables 2−6.

Taking the log S_w prediction model[22] as an example, the first results line in Table 2 refers to the original training set of 694 compounds with a calibration r^2 (and q^2) of 0.95, a root-mean-square error (rms) of 0.55 log units, and maximum negative and positive prediction errors (mne, mpe) of −2.62 and 3.43 log units. Here, the fractions of compounds with prediction errors below 0.3, 0.5, and 1 log unit are 55.9%, 73.9%, and 93.1%, respectively.

Application to the full external set of 1868 compounds (second entry line in Table 2) yields substantial decreases in both r^2 and q^2 (now 0.56), a much larger rms (1.36), maximum under- and overestimations of log S_w by almost 9 and 7 orders of magnitude, respectively, and a total of

Table 3. log K_{aw} Model[23] Statistics for the Training Set, External Test Set, and Associated Test Subsets Defined through Different Levels of ACF-Defined Chemical Domains, Varying Both the ACF Definition and the ACF Match Mode[a]

chemical domain coverage	n	r^2	q^2	rms	bias	mne	mpe	fraction (%) with absolute errors < 0.3	fraction (%) with absolute errors < 0.5	fraction (%) with absolute errors < 1
training set	408	0.99	0.99	0.16	±0.00	−1.02	+0.40	93.9	99.3	99.8
full external set	569	0.92	0.90	0.83	−0.27	−6.66	+3.42	44.1	62.4	87.2
ACF match mode										
simple first-order ACF										
loose	370	0.90	0.87	0.69	−0.28	−3.80	+2.54	40.3	59.2	88.6
moderate	234	0.91	0.90	0.52	−0.19	−2.12	+1.22	49.6	69.2	95.7
extended first-order ACF										
loose	236	0.92	0.90	0.63	−0.21	−1.86	+2.54	40.3	61.0	89.8
moderate	138	0.92	0.91	0.47	−0.13	−1.19	+1.22	50.0	71.0	97.1
second-order ACF										
borderline	**72**	**0.98**	**0.98**	**0.37**	**+0.09**	**−1.19**	**+0.84**	**68.1**	**84.7**	**98.6**
loose	41	0.99	0.98	0.38	+0.19	−0.30	+0.84	65.9	82.9	100
moderate	28	0.97	0.97	0.35	+0.15	−0.30	+0.84	67.9	89.3	100
third-order ACF										
loose	11	0.98	0.96	0.41	+0.29	−0.21	+0.74	54.5	72.7	100
moderate	5	0.99	0.89	0.40	+0.25	−0.21	+0.54	40.0	80.0	100

[a] n: Number of compounds. r^2: squared correlation coefficient. q^2: predictive squared correlation coefficient.[32] rms: root-mean-square error. bias: systematic error. mne: maximum negative prediction error (underestimation). mpe: maximum positive prediction error (overestimation).

Table 4. Additional log S_w Model[22] Statistics for the Training Set, External Test Set, and Associated Test Subsets Defined through Different Levels of ACF-Defined Chemical Domains, Varying Both the ACF Definition and the ACF Match Mode[a]

chemical domain coverage	n	absolute error mean	error range	absolute top-10% error mean	absolute top-30% error mean	absolute top-50% error mean
training set	694	0.38	6.05	1.30	0.84	0.64
full external test set	1868	0.94	16.19	3.19	2.08	1.59
ACF match mode						
simple first-order ACF						
loose	1237	0.78	16.19	2.67	1.72	1.31
moderate	1054	0.70	11.19	2.38	1.55	1.19
extended first-order ACF						
loose	752	0.70	7.42	2.24	1.52	1.17
moderate	638	0.66	5.96	2.08	1.43	1.10
second-order ACF						
borderline	**266**	**0.51**	**4.12**	**1.47**	**1.06**	**0.84**
loose	171	0.55	4.12	1.53	1.12	0.89
moderate	124	0.44	3.27	1.25	0.88	0.72
third-order ACF						
loose	53	0.51	2.97	1.28	0.99	0.82
moderate	28	0.38	2.49	1.08	0.80	0.63

[a] n: Number of compounds. Error range equals mpe − mne from Table 2. Absolute top-10% (30%, 50%) error mean is the mean of the absolute prediction errors of the compounds belonging to the top-10% (30%, 50%) error range with regard to both over- or underestimations of log S_w.

34.0% (100 − 66.0) with prediction errors ≥ 1 log unit. Note, however, that this external model application is without any domain consideration and thus ignores the limitations imposed by the chemical space of the training set used for the model derivation.

Restriction of the test set compounds to those that meet the loose ACF match mode applied to simple first-order ACFs reduces its number by ca. 1/3 to 1237 compounds, and application of the moderate match mode eliminates further 183 compounds. The reduction in the application range is accompanied by a significant improvement in the prediction performance, with q^2 increasing from 0.56 over 0.71 to 0.76, and rms decreasing from 1.36 over 1.14 to 1.01. Correspondingly, the fraction of compounds with absolute prediction errors below 0.5 log units increases from 42.2% over 47.1% to 50.9% (second last column of Table 2).

Introduction of the extended first-order ACFs results in substantial further reductions in the application range to about 40% of the external test set, again accompanied by a systematic increase in prediction performance according to all statistical parameters listed in Table 2. With second-order ACFs, the loose and moderate match modes yield a particularly large reduction in the application range to around 10% of the external test set. At the same time, the prediction performance comes close to the one with the original training set.

When defining the chemical domain through third-order ACFs, however, most statistical parameters indicate an

Table 5. Additional log K_{aw} Model[23] Statistics for the Training Set, External Test Set, and Associated Test Subsets Defined through Different Levels of ACF-Defined Chemical Domains, Varying Both the ACF Definition and the ACF Match Mode[a]

chemical domain coverage	*n*	absolute error mean	error range	top-10% error means	top-30% error means	top-50% error means
training set	408	0.12	1.42	0.36	0.25	0.20
full external test set	569	0.53	10.08	1.90	1.18	0.90
ACF match mode						
			simple first-order ACF			
loose	370	0.51	6.34	1.52	1.05	0.84
moderate	234	0.40	3.33	1.10	0.82	0.65
			extended first-order ACF			
loose	236	0.48	4.40	1.37	0.98	0.79
moderate	138	0.37	2.41	0.94	0.74	0.60
			second-order ACF			
borderline	**72**	**0.28**	**2.03**	**0.86**	**0.58**	**0.45**
loose	41	0.30	1.14	0.81	0.59	0.46
moderate	28	0.28	1.14	0.74	0.53	0.42
			third-order ACF			
loose	11	0.33	0.95	0.74	0.63	0.51
moderate	5	0.34	0.75		0.48	0.46

[a] *n*: Number of compounds. Error range equals mpe − mne from Table 2. Absolute top-10% (30%, 50%) error mean is the mean of the absolute prediction errors of the compounds belonging to the top-10% (30%, 50%) error range with regard to both over- or underestimations of log K_{aw}.

Table 6. Structural Alert Model[24] Statistics for the Training Set, External Test Set, and Associated Test Subsets Defined through Different Levels of ACF-Defined Chemical Domains, Varying Both the ACF Definition and the ACF Match Mode[a]

chemical domain coverage	*n*	TN	FP	FN	TP	sensitivity (−)	predictivity (−)	sensitivity (+)	predictivity (+)	concordance
training set	300	215	7	14	64	0.97	0.94	0.82	0.90	0.93
full external test set	709	548	34	77	50	0.94	0.88	0.39	0.60	0.84
ACF match mode										
					simple first-order ACF					
loose	397	328	16	27	26	0.95	0.92	0.49	0.62	0.89
moderate	307	255	14	17	21	0.95	0.94	0.55	0.60	0.90
					extended first-order ACF					
loose	193	160	8	12	13	0.95	0.93	0.52	0.62	0.90
moderate	150	128	7	5	10	0.95	0.96	0.67	0.59	0.92
					second-order ACF					
borderline	**79**	**66**	**2**	**5**	**6**	**0.97**	**0.93**	**0.55**	**0.75**	**0.91**
loose	51	43	0	5	3	1.00	0.90	0.38	1.00	0.90
moderate	30	26	0	2	2	1.00	0.93	0.50	1.00	0.93
					3rd-order ACF					
loose	0									
moderate	0									

[a] *n*: Number of compounds. TN = number of true negatives = number of true narcosis-level compounds. TP = number of true positives = number of true excess-toxic compounds. FN = number of false negatives = number of false narcosis-level compounds. TP = number of false positives = number of false excess-toxic compounds. Sensitivity(−) = fraction of narcosis-level compounds recognized by the model (eq 5). Predictivity(−) = fraction of compounds predicted to belong to the narcosis level that actually belong to this toxicity range (eq 6). Sensitivity(+) = fraction of excess-toxic compounds recognized by the model (eq 3). Predictivity(+) = fraction of compounds predicted to exert excess toxicity that actually do so (eq 5). Concordance = fraction of compounds classified correctly (eq 7).

inferior prediction performance as compared with the second-order ACF variants. Apart from that, the remaining application domain is only 2−3%, suggesting that this ACF level is of little practical use in the context of chemical domain definitions.

Interestingly, the change from the moderate extended first-order scheme to the loose second-order scheme is accompanied by both a particularly large decrease in the ACF-defined application range and a significant increase in prediction quality. This finding prompted us to investigate the second-order ACF mismatch situations between test set compounds and the training set pool in more detail. As a result, the borderline ACF match mode as still more loose ACF comparison mode was introduced. In this new borderline mode, test compounds with one missing second-order ACF are still classified as belonging to the chemical domain or more precisely to the border of the chemical domain. As can be seen from the highlighted entry line in Table 2, this borderline mode applied to second-order ACFs yields an overall good compromise between the prediction quality (q^2 = 0.90 vs 0.95 training, rms = 0.67 vs 0.55 training, outliers about 2 log units) and the application range (n = 266). While the latter covers still only 14% of the total test set, it is significantly larger than when employing the loose match mode (n = 171 = 9%).

For the Abraham model to predict log K_{aw} employing experimental input parameters,[23] similar results with regard to the dependence of both prediction quality and application

range are obtained (Table 3). When taking into account all 569 test set compounds and thus ignoring chemical domain issues, q^2 decreases from 0.99 (training) to 0.90 and rms increases from 0.16 (training) to 0.83. Increasingly strict consideration of the chemical domain through the above-described ACF schemes yields a generally increasing prediction quality accompanied by a significant decrease in the application range. Again, the most striking difference is between the moderate extended first-order scheme and the loose second-order scheme, and again the borderline second-order mode (highlighted in Table 3) yields essentially the prediction quality of the (more strict) loose mode but for a significantly larger subset (72 vs 41 compounds). As before, third-order ACFs appear to have no practical value for specifying the chemical domain coverage, resulting in very small application ranges (1−2% of the test set).

Tables 4 and 5 provide additional error statistics to further characterize the impact of the ACF level variation on the resultant chemical domain and the associated prediction performance. For the two quantitative models (log S_w, log K_{aw}), the absolute error mean, the error range, and the absolute top-10%, top-30%, and top-50% error means generally decrease with increasing ACF order and strictness of matching criteria. For the log S_w model,[22] ignoring its chemical domain results in almost one log unit as rms for the full external test set of 1868 compounds, which goes down to rms = 0.51 for the borderline match mode applied to second-order ACFs for 266 compounds. In this context, we would like to emphasize that the selection of these 266 test set compounds is solely driven by the ACF definition and its match mode, without taking into account the model performance. As such, the ACF approach provides an intelligent and automatized means to identify those compounds for which a good or at least reasonable QSAR model performance can be expected, acknowledging that there may still be individual outliers as with all statistically derived models.

The significant increase in prediction performance when confining the test set to compounds meeting the borderline second-order ACF criterion is also reflected by the error range (16 log units vs 4 log units; see Table 4), the absolute top-10% error mean (3.19 vs 1.47), and the corresponding absolute error means of the compounds belonging to the top-30% and top-50% error range with regard to both over- and underestimations of log S_w (2.08 vs 1.06, and 1.59 vs 0.84).

Table 5 summarizes the additional error statistics for the log K_{aw} model.[23] Again, the absolute error mean goes down by ca. 50% (from 0.53 to 0.28) when confining the test set compounds to the chemical domain defined through the borderline second-order ACF mode, the error range is reduced by eight log units, and the absolute top-10%, top-30%, and top-50% error means by a factor of 2.

The structural alert model to discriminate narcosis level from excess toxicity of organic compounds toward *Daphnia magna*[24] serves as test example for analyzing the practical usefulness of ACF-based chemical domain definitions of categorical models. As can be seen from Table 4, the original training set of 300 compounds yielded a good predictivity for both narcosis-level and excess-toxic compounds (0.94 vs 0.90), a somewhat larger difference for the respective sensitivities (0.97 vs 0.80). In absolute numbers, there were 7 false positives (FP) as compared to 215 true negatives (TN, narcosis level) and 14 false negatives (FN) as compared to 64 true positives (TP, excess toxic). As such, the model appeared particularly useful to identify narcosis-level compounds, for which simple regression relationships based on log K_{ow} are available to provide even quantitative estimates of their toxicity.

As expected, application to the external set of 709 compounds yields significantly inferior statistics. Consideration of increasingly strict chemical domain definitions results in increasingly improved concordance, although the trends of the individual sensitivities and predictivities show some variations. In particular, the recognition and prediction of excess-toxic compounds is less successful than the ones of narcosis-level compounds. Again, the borderline second-order ACF mode represents a good compromise between the moderate extended first-order and loose second-order modes, yielding predictivities for narcosis-level and excess-toxic compounds of 0.93 (vs 0.94 training) and 0.75 (vs 0.90 training), respectively. For the latter, however, it should be noted that the absolute number of 11 excess-toxic compounds is already rather small and probably too small to draw general conclusions about the associated predictivity and sensitivity.

Overall, the present analysis suggests that the ACF approach is a very valuable and powerful tool for characterizing, in an automatized way, the chemical domain of quantitative and qualitative structure−activity relationships. The distinct dependence of the model performance on the degree of strictness of the ACF-based domain definition demonstrates the sound mechanistic basis of this approach. According to the present findings, the borderline second-order ACF mode is recommended as general-purpose method for identifying compounds that belong to the chemical (structural) domain of a given QSAR model.

General ACF Chemical Domain Characteristics. Interestingly, the ACF methodology as applied here does not make use of any information about the target property. The only information taken into account is whether or not a given compound is sufficiently similar (according to the ACF match mode selected) in terms of its ACF composition to those compounds for which the model has been demonstrated to generally work fine (which is usually the case for the training set of a given model). As such, the ACF setting introduced here provides a procedure to extract and specify that part of the structural characteristics that appears to be generally relevant for the chemical domain of QSAR models.

Note further that in the QSAR model examples analyzed here the chemical domain was simply defined through the ACFs generated from the respective QSAR training set. In general, however, the chemical domain may be larger than the one resulting from the training set through inclusion of ACFs from additional compounds for which a sufficiently good QSAR performance was already demonstrated (such as a previously analyzed external set used for model validation).

Moreover, the chemical domain could also be defined as to not even contain all training set compounds. The latter may result from elimination of those training set compounds that do not meet a certain performance criterion (e.g., omitting those training set compounds where a categorical model is systematically in error or where a quantitative model yields statistically outlying prediction errors). Finally, one could fine tune the chemical domain definition still further

Table 7. Log S_w Model[22] and log K_{aw} Model[23] Statistics According to the Categories of Membership with Respect to the Chemical Domain[a]

chemical domain membership category	n	r^2	q^2	rms	bias	mne	mpe	fraction (%) with errors < 0.3	fraction (%) with errors < 0.5	fraction (%) with errors < 1
log S_w model[22]										
outside	1194	0.57	0.43	1.55	+0.16	−8.90	+7.29	23.9	36.9	60.6
borderline outside	408	0.74	0.71	1.03	+0.08	−2.80	+3.16	31.1	47.1	69.4
borderline inside	142	0.89	0.89	0.75	−0.05	−2.14	+1.98	34.5	54.2	80.3
inside	124	0.91	0.90	0.57	+0.13	−1.48	+1.79	46.8	62.9	91.1
log K_{aw} model[23]										
outside	414	0.91	0.88	0.93	−0.33	−6.66	+3.42	42.0	59.4	83.3
borderline outside	83	0.89	0.86	0.54	−0.26	−1.11	+1.22	33.7	57.8	96.4
borderline inside	44	0.98	0.98	0.39	+0.05	−1.19	+0.81	68.2	81.8	97.7
inside	28	0.97	0.97	0.35	+0.15	−0.30	+0.84	67.9	89.3	100

[a] n: Number of compounds. r^2: squared correlation coefficient. q^2: predictive squared correlation coefficient.[32] rms: root-mean-square error. bias: systematic error. mne: maximum negative prediction error (underestimation). mpe: maximum positive prediction error (overestimation).

through elimination of those ACFs that are untypical (such that each of them occurs in only one compound) for the ACF pool derived from the training set and possibly additional well-behaving compounds.

A further issue is the fact that the ACF approach is confined to the chemical domain and does not consider explicitly other components of the QSAR application domain such as the physicochemical domain, the descriptor domain, the mechanistic domain, and the metabolic domain (if applicable).[3] From this viewpoint, the present results demonstrate the importance and power of suitably addressing the chemical domain that alone provides already pertinent information about the expected QSAR performance.

At the same time, taking into account other domain components is expected to still increase the level of confidence associated with QSAR predictions. Note, however, that ACFs are able to indirectly capture mechanistic aspects to a certain degree. An example of the latter is the structural discrimination between aldehydes and ketones that can already be achieved with first-order ACFs. The latter thus could be used in principle to reflect associated differences in toxicological reaction pathways, keeping in mind that aldehydes are much more ready to undergo Schiff base formation than ketones, provided the respective ACF difference is weighted sufficiently in the respective ACF match mode. Accordingly, the ACF methodology provides also opportunities to include end-point-specific mechanistic information, which may be the subject of future investigations.

Categories of Chemical Domain Membership. Beyond providing a single threshold, the chemical domain assessment for a given compound may yield a graded answer instead of the inside−outside alternative, acknowledging the partially fuzzy character of this issue. Higher degrees of chemical domain membership generally relate to a higher degree of confidence in QSAR predictions.

To this end, we propose characterizing the degree of membership of a test compound to the chemical domain of a given QSAR model in terms of four categories, taking the borderline second-order ACF mode that had turned out as optimum general-purpose scheme as a major reference. Compounds not matching this borderline criterion because of at least two second-order ACFs outside the ACF training set pool are evaluated further, now employing extended first-order ACFs with the moderate match mode (which takes into account the maximum number of occurrence of each ACF observed for the training set compounds). Those compounds that meet this criterion are classified as borderline outside the chemical domain, and the remaining compounds with at least two second-order ACFs outside the training set pool are classified as outside the chemical domain.

Compounds meeting the borderline second-order ACF criterion may or may not contain one second-order ACF outside the training set ACF pool. Apart from that, they may contain second-order ACFs with frequencies above the respective maximum numbers of individual occurrences observed for the training set compounds. Compounds with one second-order ACF outside the training set pool or with at least one second-order ACF with a frequency above the training set compound maximum (or with both) are classified as borderline inside the chemical domain. The remaining compounds have only second-order ACFs inside the training set pool and, in addition, meet the moderate match criterion for second-order ACFs. These latter compounds where all ACFs do not occur more often than in any of the training set compounds are classified as inside the chemical domain.

In Table 7 the external test sets for both log S_w ($n = 1868$, see Table 2) and log K_{aw} ($n = 569$, see Table 3) are subdivided further according to the just introduced four membership categories in terms of outside, borderline outside, borderline inside, and inside the chemical domain and are characterized with respect to the associated prediction performances. As before, increasing the strictness of the chemical domain criterion is accompanied by a decrease of the application range, which reflects the trade off between the goals of a high prediction quality and a broad model applicability.

With regard to the log S_w model[22] (upper part of Table 7), a clear relationship between the increasing degree of belonging of the compounds to the chemical domain and an increasing prediction performance is observed. Taking q^2 and rms as examples, the former increases from 0.43 (outside) over 0.74 (borderline outside) and 0.89 (borderline inside) to 0.90 (inside), and the latter decreases from 1.55 (outside) over 1.03 (borderline outside) and 0.75 (borderline inside) to 0.57 (inside). Corresponding trends are observed for the maximum over- and underestimations (mpe, mne) as well as for the compound fractions with absolute errors below 0.3, 0.5, and 1.0 log units, respectively. Note further the significant increase in prediction quality when going from the borderline outside category to the borderline inside category. It confirms the particular suitability of the borderline second-order ACF mode introduced above to discrimi-

nate between compounds inside and outside the chemical domain (see Tables 2−6).

For the log K_{aw} prediction model[23] (lower part of Table 7) the overall trends are similar but with slight individual variations, probably mainly due to the much smaller subset sizes. Again, the borderline inside category yields a significantly larger q^2 and significantly smaller rms than the borderline outside category ($q^2 = 0.86$ vs 0.98; rms = 0.39 vs 0.54), and the fraction of compounds with absolute prediction errors below 0.5 log units is much larger for the two inside categories than for the two outside categories (>80% vs < 60%).

Because of the much smaller sizes of the test subsets for the structural alert model[24] (see Table 6), a respective analysis is not useful. Nevertheless and in view of the findings summarized in Table 6, we assume that the general trends observed for the quantitative models (Table 7) will hold correspondingly for categorical models.

Examples Illustrating the ACF Match Mode. Scheme 1 contains three chemical compounds where a given ACF match or mismatch situation is illustrated in terms of the relevant ACFs. These examples were taken from the log K_{aw} data set, employing the respective ACF domain of the 408 training set compounds[23] (see Materials and Methods above) as chemical domain.

As can be seen from the scheme, diphenylmethane contains four different extended first-order ACFs. Three of them (those that occur eight times, two times, and two times in the molecule) are also found in training set compounds, while the one ACF containing the aliphatic methylene carbon as the central atom does not belong to the training set ACFs. Following the categories of domain membership introduced in the last subsection, this compound is classified as being outside the chemical domain of the log K_{aw} QSAR model.

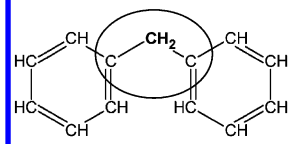
The second example concerns 1-chloro-2-methylpropane. Here, all four extended first-order ACFs are found in training set compounds. However, among the four second-order ACFs there are two that differ from all second-order training set ACFs. Hence, the compound is classified as borderline outside.

With regard to 2-hexanol as the third and last example in Scheme 1, all extended first-order ACFs belong to the training set domain (not shown explicitly), and the same holds for six of the seven different second-order ACFs. Because there is precisely one second-order ACF not belonging to any of the training set compounds, the compound is classified as borderline inside. This latter ACF is shown both as a linear line notation and as an enclosed part of the graphical representation of the chemical structure and has as central atom C (bonded to two H atoms and to two carbon atoms) at position three of the molecule.

Scheme 1. Examples for the ACF Domain Match and Mismatch with Reference to the Chemical Domain Derived from the log K_{aw} Training Set of 408 Compounds[23a]

Outside: Missing ACF(s) even in extended first order

Example: Diphenylmethane



Extended 1st order ACF's occurring in training set:

8 · **C$_{aro}$H**(~C$_{aro}$H~)~C$_{aro}$H~

2 · **C$_{aro}$H**(~C$_{aro}$H~)~C$_{aro}$~

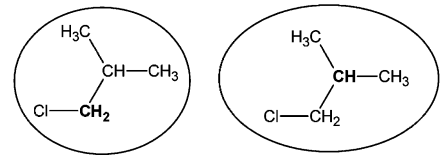
2 · **C$_{aro}$**(~C$_{aro}$H~)(~C$_{aro}$H~)–C$_{no\ ring}$H$_2$–

Extended 1st order ACF not occurring in training set:

1 · **C$_{no\ ring}$H$_2$**(–C$_{aro}$~)–C$_{aro}$~

Borderline outside: No missing ACF in extended first order
More than one ACF missing in 2nd order

Example: 1-Chloro-2-methylpropane



2nd order ACF's occurring in training set:

1 · **Cl**–CH$_2$–C$_{no\ ring}$H<

2 · **CH$_3$**–CH(–CH$_3$)–CH$_2$–

2nd order ACF's not occurring in training set:

1 · **CH$_2$**(–Cl)–CH(–CH$_3$)–CH$_3$

1 · **CH**(–CH$_3$)(–CH$_3$)–CH$_2$–Cl

All extended 1st order ACF's represented in training set:

1 · **Cl**–CH$_2$–

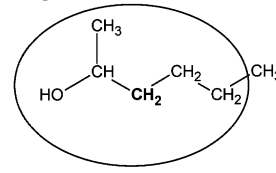
1 · **CH$_2$**(–Cl)–C$_{no\ ring}$H<

1 · **CH**(–CH$_3$)(–CH$_3$)–CH$_2$–

2 · **CH$_3$**–C$_{no\ ring}$H<

Borderline inside: No missing ACF in extended first order
Exactly one ACF missing in 2nd order

Example: 2-Hexanol



2nd order ACF's occurring in training set:

1 · **CH$_3$**–CH$_2$–CH$_2$–

1 · **CH$_2$**(–CH$_3$)–CH$_2$–CH$_2$–

1 · **CH$_2$**(–CH$_2$–CH$_3$)–CH$_2$–C$_{no\ ring}$H<

1 · **CH**(–CH$_3$)(–OH)–CH$_2$–CH$_2$–

1 · **CH$_3$**–CH(–OH)–CH$_2$–

1 · **OH**–CH(–CH$_3$)–CH$_2$–

2nd order ACF not occurring in training set:

1 · **CH$_2$**(–CH$_2$–CH$_2$–)–CH(–OH)–CH$_3$

[a] In the graphical representations, the crucial ACFs missing in the chemical domain are marked through enclosures.

CONCLUSIONS

The ACF methodology is a powerful tool to characterize the chemical domain of qualitative and quantitative structure−activity relationship (QSAR) models. For concrete applications, both the ACF definition (such as first order vs second order) and the ACF match mode have to be specified to yield transparent and reproducible results. Our findings suggest that second-order ACFs combined with a borderline match mode, allowing one second-order ACF mismatch for compounds still belonging to the chemical domain, may serve as general-purpose method for defining and evaluating the chemical domain in the context of QSAR model predictions. Because the derived ACF setting is end point independent, the accordingly defined chemical domain can also be understood as representing that part of the structural characteristics that appears to be generally relevant for the chemical domain of QSAR models. As such, the ACF approach provides an intelligent and automatized means to

identify those compounds for which a good or at least reasonable QSAR model performance can be expected. In this context, four standard categories of the ACF-based chemical domain membership in terms of outside, borderline outside, borderline inside, and inside the chemical domain have been derived that inform further about the expected prediction performance of a given QSAR model for a compound of interest. The derived methodology offers a generic and mechanistically sound way to address the chemical domain of QSAR models and as such is expected to facilitate the use of in silico approaches as nontest methods in the context of the REACH regulation for industrial chemicals. At the same time, one should keep in mind that additional consideration of other components of the QSAR applicability domain such as the physicochemical domain, descriptor domain, mechanistic domain, and metabolic domain (if applicable) offers room for still improved levels of confidence in predictive model applications.

ACKNOWLEDGMENT

The authors thank Barbara Wagner and Daniel Exner for programming work. Financial support was provided by the European Commission through the projects CAESAR (Contract No. 022674), 2-FUN (Contract No. 036976), and OSIRIS (Contract No. 037017), which is gratefully acknowledged.

REFERENCES AND NOTES

(1) Mekenyan, O.; Dimitrov, S.; Serafimova, R.; Thompson, E.; Kotov, S.; Dimitrova, N.; Walker, J. D. Identification of the Structural Requirements for Mutagenicity by Incorporating Molecular Flexibility and Metabolic Activation of Chemicals I: TA100 Model. *Chem. Res. Toxicol.* **2004**, *17*, 753–766.

(2) Jaworska, J.; Nikolova-Jeliazkova, N.; Aldenberg, T. QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review. *ATLA* **2005**, *33*, 445–459.

(3) Dimitrov, S.; Dimitrova, G.; Pavlov, T.; Dimitrova, N.; Patlewicz, G.; Niemela, J.; Mekenyan, O. A Stepwise Approach for Defining the Applicability Domain of SAR and QSAR Models. *J. Chem. Inf. Model.* **2005**, *45*, 839–849.

(4) Organisation for Economic Co-operation and Development Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. 2007, OECD Document ENV/JM/MONO(2007)2.

(5) Netzeva, T. I.; Worth, A. P.; Aldenberg, T.; Benjamin, I.; Cronin, M. T. D.; Gramatica, P.; Jaworska, J. S.; Kahn, S.; Klopman, G.; Marchant, C. A.; Myatt, G.; Nikolova-Jeliazkova, N.; Patlewicz, G. Y.; Perkins, R.; Roberts, D. W.; Schultz, T. W.; Stanton, D. T.; van de Sandt, J. J. M.; Tong, W.; Veith, G.; Yang, C. Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships-The Report and Recommendations of ECVAM Workshop 52. *ATLA* **2005**, *33*, 155–173.

(6) European Commission Directive of the European Parliament and of the Council: Amending Council Directive 67/548/EEC in Order to Adapt It to Regulation (EC) of the European Parliament and of the Council Concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals. 2003/0256/COD, 2003/0257/COD (Oct 29, 2003). 2003.

(7) Tetko, I. V.; Poda, G. I.; Ostermann, C.; Mannhold, R. Accurate in Silico Log *P* Predictions: One Can't Embrace the Unembraceable. *QSAR Comb. Sci.* **2009**, *28*, 845–849.

(8) Weaver, S.; Gleeson, M. P. The Importance of the Domain of Applicability in QSAR Modeling. *J. Mol. Graphics Modell.* **2008**, *26*, 1315–1326.

(9) Tetko, I. V.; Bruneau, P.; Mewes, H. W.; Rohrer, D. C.; Poda, G. I. Can We Estimate the Accuracy of ADME-Tox Predictions. *Drug Discovery Today* **2006**, *11*, 700–707.

(10) Netzeva, T. I.; Saliner, A. G.; Worth, A. P. Comparison of the Applicability Domain of a Quantitative Structure-Activity Relationship for Estrogenicity With a Large Chemical Inventory. *Environ. Toxicol. Chem.* **2006**, *25*, 1223–1230.

(11) Dragos, H.; Gilles, M.; Alexandre, V. Predicting the Predictability: A Unified Approach to the Applicability Domain Problem of QSAR Models. *J. Chem. Inf. Model.* **2009**, *49*, 1762–1776.

(12) Wang, Y.; Geppert, H.; Bajorath, J. Shannon Entropy-Based Fingerprint Similarity Search Strategy. *J. Chem. Inf. Model.* **2009**, *49*, 1687–1691.

(13) Sprous, D. G. Fingerprint-Based Clustering Applied to Define a QSAR Model Use Radius. *J. Mol. Graphics Modell.* **2008**, *27*, 225–232.

(14) Casalegno, M.; Sello, G.; Benfenati, E. Definition and Detection of Outliers in Chemical Space. *J. Chem. Inf. Model.* **2008**, *48*, 1592–1601.

(15) Adamson, G. W.; Lynch, M. F.; Town, W. G. Analysis of Structural Characteristics of Chemical Compounds in a Large Computer-Based File. Part II. Atom-Centred Fragments. *J. Chem. Soc. (C)* **1971**, *1971*, 3702–3706.

(16) Kühne, R.; Kleint, F.; Ebert, R.-U.; Schüürmann, G. In *Software Development in Chemistry 10*; Gasteiger, J., Ed.; PROserv Springer Produktionsgesellschaft: Berlin, Germany, 1996; pp 125−134.

(17) Kühne, R.; Ebert, R.-U.; Schüürmann, G. Model Selection Based on Structural Similarity-Method Description and Application to Water Solubility Prediction. *J. Chem. Inf. Model.* **2006**, *46*, 636–641.

(18) Kühne, R.; Ebert, R.-U.; Schüürmann, G. Estimation of Compartmental Half-Lives of Organic Compounds - Structural Similarity *Versus* EPI-Suite. *QSAR Comb. Sci.* **2007**, *26*, 542–549.

(19) Kühne, R.; Ebert, R.-U.; Schüürmann, G. Prediction of the Temperature Dependency of Henry's Law Constant From Chemical Structure. *Environ. Sci. Technol.* **2005**, *39*, 6705–6711.

(20) Dimitrov, S. D.; Low, L. K.; Patlewicz, G. Y.; Kern, P. S.; Dimitrova, G. D.; Comber, M. H. I.; Phillips, R. D.; Niemela, J.; Bailey, P. T.; Mekenyan, O. G. Skin Sensitization: Modeling Based on Skin Metabolism Simulation and Formation of Protein Conjugates. *Int. J. Toxicol.* **2005**, *24*, 189–204.

(21) Patlewicz, G.; Dimitrov, S. D.; Low, L. K.; Kern, P. S.; Dimitrova, G. D.; Comber, M. I. H.; Aptula, A. O.; Phillips, R. D.; Niemelä, J.; Madsen, C.; Wedebye, E. B.; Roberts, D. W.; Bailey, P. T.; Mekenyan, O. G. TIMES-SS−A Promising Tool for the Assessment of Skin Sensitization Hazard. A Characterization With Respect to the OECD Validation Principles for (Q)SARs and an External Evaluation for Predictivity. *Regul. Toxicol. Pharmacol.* **2007**, *48*, 225–239.

(22) Kühne, R.; Ebert, R.-U.; Kleint, F.; Schmidt, G.; Schüürmann, G. Group Contribution Methods to Estimate Water Solubility of Organic Chemicals. *Chemosphere* **1995**, *30*, 2061–2077.

(23) Abraham, M. H.; Andonian-Haftvan, J.; Whiting, G. S.; Leo, A.; Taft, R. S. Hydrogen Bonding. Part 34. The Factors That Influence the Solubility of Gases and Vapors in Water at 298 K, and a New Method for Its Determination. *J. Chem. Soc., Perkin Trans. 2* **1994**, 1777–1791.

(24) Von der Ohe, P. C.; Kühne, R.; Ebert, R.-U.; Altenburger, R.; Liess, M.; Schüürmann, G. Structural Alerts - a New Classification Model to Discriminate Excess Toxicity From Narcotic Effect Levels of Organic Compounds in the Acute Daphnid Assay. *Chem. Res. Toxicol.* **2005**, *18*, 536–555.

(25) U.S. Environmental Protection Agency EPISUITE 4.0, 2008.

(26) U.S. Environmental Protection Agency WSKOWWIN 1.41a, 2008.

(27) U.S. Environmental Protection Agency MPBPWIN 1.43, 2008.

(28) U.S. Environmental Protection Agency HENRYWIN 3.20, 2008.

(29) Schüürmann, G.; Kühne, R.; Kleint, F.; Ebert, R.-U.; Rothenbacher, C.; Herth, P. In *Quantitative Structure-Activity Relationships in Environmental Sciences-VII*; Chen, F., Schüürmann, G., Eds.; SETAC Press: Pensacola, FL, 1997; pp 93−114.,

(30) Schüürmann, G.; Ebert, R.-U.; Nendza, M.; Dearden, J. C.; Paschke, A.; Kühne, R. In *Risk assessment of chemicals: An Introduction*, 2nd ed.; Van Leeuwen, K., Vermeire, T., Eds.; Springer: Dordrecht, The Netherlands, 2007; pp 375−426.

(31) U.S. Environmental Protection Agency KOWWIN 1.67a, 2008.

(32) Schüürmann, G.; Ebert, R.-U.; Chen, J.; Wang, B.; Kühne, R. External Validation and Prediction Employing the Predictive Squared Correlation Coefficient−Test Set Activity Mean vs Training Set Activity Mean. *J. Chem. Inf. Model.* **2008**, *48*, 2140–2145.

CI900313U