MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **75** (2016) 559-582

ISSN 0340 - 6253

QSAR Models for Estimating Aryl Hydrocarbon Receptor Binding Affinity of Polychlorobiphenyls, Polychlorodibenzodioxins, and Polychlorodibenzofurans

Martin Šala,[a,‡] Alexandru T. Balaban,[b,*] Marjan Veber,[a] Matevž Pompe[a,*]

[a] *University of Ljubljana, Faculty of Chemistry and Chemical Technology, Aškerčeva 5, 1000 Ljubljana, Slovenia*
[b] *Texas A&M University at Galveston, Department of Marine Sciences, Galveston, TX 77551, USA*
‡ *Current Address: National Institute of Chemistry, Hajdrihova 19, 1000 Ljubljana, Slovenia*

matevz.pompe@fkkt.uni-lj.si, balabana@tamug.edu

(Received October 22, 2015)

Abstract

In order to perform the screening of new potential pollution points and to estimate their impact on the environment, a quantitative structure-activity research (QSAR) modeling procedure was used to estimate possible toxicological effects of substances that were introduced into the environment. We have focused our study initially to known persistent toxic compounds with defined toxicological mechanisms. Such groups of studied chemicals are polychlorinated aromatic compounds. Three multiple linear regression (MLR) models were developed for the calculation of aryl hydrocarbon receptor (AhR) binding affinity of polychlorinated biphenyls (PCBs), dibenzofurans (PCDFs) and dibenzodioxins (PCDDs). In all cases models were able to explain more than 70% of the total variance. Additionally, the correlating capability of the newly developed variable distance-based topological index was tested. It showed reasonable modeling abilities and superior interpretation abilities compared to classical MLR QSAR models.

Abbreviations

AhR	aryl hydrocarbon receptor
CODESSA	comprehensive descriptors for structural and statistical analysis
DDT	1,1-(4,4'-dichlorodiphenyl)-2,2,2-trichloroethane
EC50	half maximal effective concentration (1/log EC_{50} = induction activity)
ESP	electrostatic potential
EU	European Union
LOO	leave one out
MLR	multiple linear regression
MOPAC	molecular orbital package
PCB	polychloro-biphenyl
PCDD	polychloro-dibenzodioxin
PCDF	polychloro-dibenzofuran
PNSA	partial negative surface area
PPSA	partial positive surface area
r^2	squared correlation coefficient
q^2	cross-validated squared correlation coefficient
QSAR	quantitative structure-activity relationship
QSPR	quantitative structure-property relationship
REACH	registration, evaluation and authorization of chemicals
RMS	root mean square
s	standard error
TFP	traditional food products
TMSA	total *molecular* surface area

1 Introduction

A quantitative risk assessment becomes increasingly important in the modern society and is slowly incorporated into legislation of different countries. For instance, the European Union (EU) has introduced the Registration, Evaluation and Authorization of CHemicals (REACH) program for assessment of human and environmental risk of all chemicals that are produced or imported in the amount greater than 1 ton per year. It is clear that if such a risk assessment is performed purely experimentally, it would require a huge amount of resources as well as time. Therefore, the introduced program encouraged the use of QSAR modeling and other alternatives especially for the risk assessment of chemicals that are produced or imported in smaller quantities.

In order to perform screening of new potential pollution points and to estimate their impact on environment, a QSAR modeling procedure was used to estimate possible

toxicological effects of substances that were introduced into the environment due to modification of traditional food products (TFP) production. Since we are dealing with structurally very diverse sets of compounds possessing different toxicological mechanisms, it is impossible to create just one prediction model for all possible compounds. Therefore, we have focused our study initially to known persistent toxins with defined toxicological mechanisms. Such groups of studied chemicals are the polychlorinated aromatic compounds.

Among polychlorinated aromatic compounds, the most environmentally dangerous chemicals are derivatives of polychloro-dibenzofuran (PCDF, **1**), polychloro-dibenzo-1,4-dioxin (PCDD, **2**), and polychloro-biphenyl (PCB, **3**), as shown in Fig. 1. Like the persistent chlorofluorocarbons that were extensively used until the discovery of their deleterious effect on the ozone layer, these polychloroaromatic compounds are harmful to the environment. Similarly to other polychlorinated compounds (used as efficient pesticides such as DDT, Aldrin and Dieldrin, which led to the "*Silent Spring*" due to bioaccumulation in fatty tissues of insects, birds and other higher organisms), the title compounds exert a powerful toxic effect on humans (endocrine disruptors, neurotoxic, carcinogens), and one of the polychlorodibenzo-1,4-dioxin isomers is among the most toxic organic chemicals.[1] Due to their chemical stability, these compounds make their way up the food chain, therefore they bioaccumulate in fatty tissues. Their emissions have been decreasing during last decades. This can be observed in the trend of their environmental concentration[2-7] as well as exposure measurements.[8-11] However, even as impurities of other present-day pesticides and other small sources, these polychlorinated compounds constitute a threat.[12] A comprehensive review of their different toxicological effects was published.[13]

The toxic equivalency factors (TEFs) or toxic equivalent concentrations (TEQs) released by the World Health Organization[14] indicate relative toxicities with respect to the most harmful compound, 2,3,7,8-TCDD by half-orders of magnitude. Such imprecise data discussed in literature[12] may be due to multiple mechanisms of action. In the present study, we examined more accurate data dealing with the interaction with a single receptor. An important stereochemical feature of PCBs is the coplanarity dictated by the absence of substituents in positions 2, 2',6, and 6'. It is suggested that the toxic effects are not the result of direct action of these compounds but are mediated through a specific protein complex known as the aryl hydrocarbon receptor. The structure of this complex is unknown. The

analysis of these effects is further complicated due to the chemical diversity of PCDFs, PCBs and PCDDs.

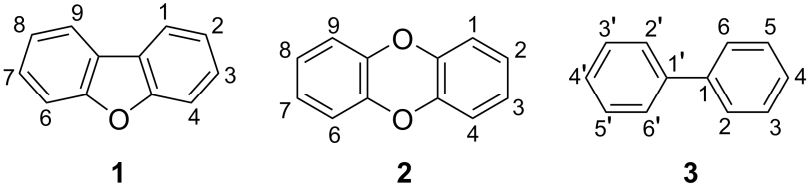


Figure 1. Numbering of carbon atoms for polychloro derivatives studied in this communication: **1** – PCDF, **2** – PCDD, **3** – PCB

The Aryl hydrocarbon Receptor (AhR) protein is a cytosolic transcription factor that is normally inactive, being attached to several co-chaperones. Upon binding to some natural ligands (e. g. bilirubin, prostaglandin G) or synthetic chemicals (such as polychloro compounds or aromatic hydrocarbons), the chaperones dissociate allowing the translocation of AhR into the nucleus leading to changes in gene transcription, and resulting in the induction of metabolizing enzymes that cause the production of metabolites which should be more easily excreted but in this case are more toxic.

Among QSAR studies about the toxicity of polyhalogenated aromatics (expressed either by the affinity for aryl hydrocarbon receptor or by the toxic equivalency factors, TEF), several papers will be mentioned with a brief characterization. TEF values assume a value of 8.444 for 2,3,7,8-TetraCDBF are larger than the original published pIC_{50} data by a factor of 1.143 (one may check if this factor is involved in data sets by looking at the values for 4-monochloroCDF: 3.00 for original AhR affinity and 3.43 for TEF).

Whereas the trend of increasing cancer incidence is being counteracted by measures for restricting smoking and chlorinated insecticides, and also for reducing organics in exhaust gases, polyhalogenated aromatics are still present in the environment and are responsible for both cancer and endocrine disruption. Municipal solid waste combustion produces PCDDs and PCDFs[15-16] whereas PCBs employed in industrial processes (heat-transfer and insulating fluids in cooling systems and electrical equipment, as well as sealants, rubber, paints, plastics, printing ink, and insecticides) are still around and are accidentally released.

Previous QSAR analyses have suggested that steric, electrostatic, hydrophobic, hydrogen bonding and dispersion properties may all be important for receptor binding

affinity.[17-20] In order to take into consideration all these structural descriptor electronic eigenvalues,[21] topological charge indices,[22] quantum chemical descriptors[23] were suggested as suitable structural descriptors. The studies were not limited just to prediction of toxicological effects but were trying to model some environmental behavior as well, that is, partitioning properties,[24] sub-cooled liquid vapor pressure,[25] bioconcentration factors,[26] and even microbial degradation properties.[27]

Toxic potency differences have been discovered among species such as humans and fish.[28] Correspondences among bromo- and chloro- substituted aromatics could be accounted for.[29] Elaborated approaches using CoMFA involving alignment of ligands,[30] 3D-QSAR,[31] or molecular descriptors based on other quantum-chemical data[32-33] have been employed. The binding mechanism for AhR[34] (34) or a multidimensional approach involving two other receptors offered new insights.[35]

A thorough QSAR study by Katritzky, Karelson and their coworkers[36] used their CODESSA program for modeling *bioconcentration factors* of polychlorinated biphenyls. We report results obtained by using the CODESSA program for modeling AhR binding affinities.

Several authors analyzed the published data for AhR binding affinities for the three classes of polyhalogenated aromatics. Benfenati and coworkers[37] mentioned that in the absence of the AhR molecular structure, one has to use virtual screening, and we agree that the information provided by Song et al.[38] on the X-ray crystallographic data because it covers only the bHLH domain of AhR.[39] The pIC_{50} data used[37] were normalized to a value of 8.444 for 2,3,7,8-TetraCDBF are larger than the original published data by a factor of 1.143; three PCDF structures are duplicated, namely **54**-**56** and **60**–**62** in reference.[37] The same normalization factor was applied in ref. 18. We used log*P* values from the same reference[37] in order to examine how the Balaban index or the related variable indices behave in biparametric correlations having as the second parameter either the number of halogen atoms or the log*P* value of the polyhalogenated aromatics.

Table 1. Experimental and cross-validated AhR affinity values for polychlorobiphenyls (**3**)

No.	Structure	Experimental[a]	Cross-validated[b]	Code
1	2,4,2',4'-TetraCBP	3.89	3.93	47
2	2,3,4,4'-TetraCBP	4.55	4.87	60
3	2,3,4,5-TetraCBP	3.85	4.27	61
4	3,3',4,4'-TetraCBP	6.15	6.37	77

5	2,3,3',4,4'-PentaCBP	5.37	5.18	105
6	2,3,4,4',5-PentaCBP	5.39	5.14	144
7	2,3',4,4',5-PentaCBP	5.04	5.19	118
8	3,3',4,4',5-PentaCBP	6.89	5.90	127
9	2,3',4,4,5'-PentaCBP	4.85	4.77	123
10	2,2',4,4',5,5'-HexaCBP	4.10	4.09	153
11	2,3,3',4,4',5-HexaCBP	5.15	5.54	156
12	2,3,3',4,4',5'-HexaCBP	5.33	5.27	157
13	2,3',4,4',5,5'-HexaCBP	4.80	5.31	167
14	2,3',4,4',5',6-HexaCBP	4.00	4.36	168
15	4'-Me-2,3,4,5-TetraCBP	4.51	5.21	
16	4'-OH-2,3,4,5-TetraCBP	4.05	4.49	
17	4'-I-2,3,4,5-TetraCBP	5.82	5.68	
18	4'-F-2,3,4,5-TetraCBP	4.60	4.57	
19	4'-Br-2,3,4,5-TetraCBP	5.60	5.43	
20	4'-Et-2,3,4,5-TetraCBP	5.46	5.04	
21	4'-i-Pr-2,3,4,5-TetraCBP	5.89	5.21	
22	4'-n-Bu-2,3,4,5-TetraCBP	5.13	5.12	
23	4'-t-Bu-2,3,4,5-TetraCBP	5.17	5.42	
24	4'-Ph-2,3,4,5-TetraCBP	5.18	5.27	
25	4'-F_3C-2,3,4,5-TetraCBP	6.43	6.59	
26	4'-CN-2,3,4,5-TetraCBP	5.27	3.99	
27	4'-MeO-2,3,4,5-TetraCBP	4.80	4.33	
28	4'-Ac-2,3,4,5-TetraCBP	5.17	5.52	
29	4'-NO_2-2,3,4,5-TetraCBP	4.85	5.31	
30	4'-AcNH-2,3,4,5-TetraCBP	5.09	4.94	

[a] From refs. [42, 46]
[b] Leave-one-out cross-validation results.

Within our study, we have tested the creation of multivariate models using different structural descriptors to model aryl hydrocarbon receptor binding affinity of chlorinated biphenyls, dibenzofurans and dioxins. A special emphasis was given to the development of some simple mathematical structural representations in order to obtain structural interpretations of the prediction models. It should be stressed that the QSAR methodology applied in this study can be applied also to other classes of chemicals and it is not limited to chlorinated organic compounds (in a few cases, the present study deals also with brominated derivatives).

2 Experimental

Data sets

Aryl hydrocarbon receptor (AhR) binding affinities [log(1/EC_{50})] for 30 polychlorobiphenyls (PCBs), 20 polychlorodibenzo-1,4-dioxins (PCDDs), and 32 polychlorodibenzofurans (PCDFs) have been taken from the literature and are shown in Tables 1 to 3.[40-49] For the 209 possible PCBs the nomenclature proposed by Ballschmiter and Zell[50] (50) was completed by Schulte and Malisch,[51] then by Guitart *et al.* (52),[52] and reviewed by Mills III *et al.*;[53] this corresponding code is presented in the last column of Table 1. For PCBs derived from 2,3,4,5-TetraCBD that have a 4-substituent different from chlorine and are shown in the lower part of Table 1, no number is provided in the last column. Since the chlorine atom in abbreviations is indicated by C instead of Cl, we replace only P for 'poly' in abbreviations by Mono, Di, etc. but when bromine is present, we use Br. One should mention that halogens in mixed halogenated pollutants are occasionally designated by X in the literature.[54-55] The names used in our Tables 1 to 3 are easy to understand, using the IUPAC numbering presented in Fig, 1.

Calculation of molecular descriptors and the creation of models using CODESSA

In order to obtain molecular descriptors[56-59] (56-59) required for the creation of the models, the optimized structural co-ordinates and net atomic charges were calculated by the MOPAC software package.[60] More than 500 different topological, geometric, informational, electrostatic, electrotopological and quantum-chemical descriptors were calculated from the MOPAC output files using the CODESSA software.[36, 61] The descriptors applied in the study contain information about the connections between atoms, symmetry, shape, branching, distribution of charge, and quantum-chemical properties of the molecule.

Table 2. Experimental and cross-validated AhR values for polyhalodibenzo-1,4-dioxins (**2**)

No.	Structure	Experimental[a]	Cross-validated[b]
1	1-MonoCD	4.00	4.52
2	2-MonoBrD	6.53	6.75
3	2,7-DiBrD	7.81	7.08
4	2,8-DiCD	5.50	5.52

5	1,2,4-TriCD	4.89	5.01
6	1,7,8-TriCD	6.66	6.33
7	2,3,7-TriCD	7.15	6.60
8	2,3,7-TriBrD	8.93	5.75
9	1,2,3,4-TetraCD	5.89	8.49
10	1,2,7,8-TetraCD	6.80	6.79
11	1,3,7,8-TetraCD	6.10	6.52
12	1,3,7,8-TetraBrD	8.70	8.22
13	2,3,7,8-TetraCD	8.00	7.23
14	1,3,7,9-TetraBrD	7.03	7.61
15	2,3,7,8-TetraBrD	8.82	9.48
16	2,3-DiBr-7,8-DiCD	8.83	9.13
17	1,2,3,7,8-PentaCD	7.10	6.61
18	1,2,4,7,8-PentaBrD	7.77	8.57
19	1,2,3,4,6,7,8-HeptaCD	6.55	6.88
20	OctaCD	5.00	5.57

[a] From ref. [21]
[b] Leave-one-out cross-validation results.

The CODESSA software was also used for the selecting the best subset of structural descriptors by minimizing errors in prediction using the MLR modeling procedure. The program is able to search for the best MLR model with the in-advance-selected number of parameters by using the following selection procedure. Initially all information-less descriptors were omitted from the study, that is, descriptors with no variation between structures, descriptors which did not cover the whole modeling space, and descriptors with the squared correlation coefficient (r^2) smaller than 0.01. Afterwards, all squared pair-wise correlation coefficients were calculated and one of the two descriptors was removed if r^2 exceeded 0.95. By using this described elimination procedure 367 descriptors were left for the stepwise selection of the best subset of structural indices.

In the next step, the remaining descriptors were sorted in the decreasing order of the squared correlation coefficient of the simple linear regression model. The best ten descriptors were selected. Afterwards, two-parameter MLR models were created by adding to each of ten pre-selected descriptors a new descriptor from the set of remaining structural indices. At the same time the descriptors within individual MLR model were pair-wise correlated and the models in which the correlation coefficient exceeded 0.8 were rejected. Finally, ten best two-parameter models showing the highest F-value were selected. The addition of new descriptors was repeated until the MLR model with the prescribed number of parameters was obtained. A new descriptor was added to the existing MLR model if it was not highly correlated with the

descriptors already included (i.e. the pair-wise correlation coefficient above 0.8) and if the resulting correlation gave F-value above $F_{old} \times n/(n + 1)$ (n = number of descriptors in the new working set). During each step the dimension of the MLR model was increased by one. As the final result of the described stepwise addition procedure, ten correlations with the pre-selected number of descriptors were obtained showing the highest squared correlation coefficient.

Application of variable topological indices in the prediction of biological toxicity

It should be mentioned that a molecular descriptor characterizing the 'topological shape' of molecules, namely the distance-based averaged molecular connectivity, is known as the Balaban index, *J*.[62] The averaged distance-based connectivity index *J* is a molecular descriptor devised so as to encode information on the "topological shape" of the molecular hydrogen-depleted graph: unlike the molecular connectivity index introduced by Randić,[63] which is based on the vertex degree and depends both on the shape and the size of the molecule, index *J* compensates the size (represented by the number n of graph vertices) by multiplication with the number q of graph edges and division by one plus the cyclomatic number $\mu = q - n + 1$:

$$J = \frac{q}{\mu + 1} \sum_{i=1}^{n} \sum_{j=1}^{i=1} (d_i \cdot d_j)^{-0.5} \qquad (1)$$

Table 3. Experimental and calculated AhR values for polychlorodibenzofurans (**1**)

No.	Structure	Exper.[a]	J	J[f c]	CV[b]	CV[c]	log*P*[d]
1	DF	3.00	2.3207	0.05447	2.86	2.54	3.71
2	2-MonoCDF	3.55	2.2977	0.05434	3.06	3.78	4.48
3	3-MonoCDF	4.38	2.2742	0.05427	4.02	4.31	4.48
4	4-MonoCDF	3.00	2.3251	0.05440	3.34	3.29	4.48
5	2,3-DiCDF	5.33	2.2884	0.05417	4.98	5.17	5.04
6	2,6-DiCDF	3.61	2.3054	0.05427	4.22	4.35	5.12
7	2,8-DiCDF	3.59	2.2832	0.05421	3.89	4.92	5.12
8	1,3,6-TriCDF	5.36	2.3367	0.05423	5.57	4.59	5.82
9	1,3,8-TriCDF	4.07	2.3158	0.05417	4.90	5.26	5.82
10	2,3,4-TriCDF	4.72	2.3406	0.05414	5.73	5.49	5.59
11	2,3,8-TriCDF	6.00	2.2763	0.05404	6.29	6.23	5.67
12	2,6,7-TriCDF	6.35	2.2959	0.05410	5.75	5.67	5.67
13	1,2,3,6-TetraCDF	6.46	2.3658	0.05414	5.82	5.31	6.23
14	1,2,3,7-TetraCDF	6.96	2.3237	0.05401	6.73	6.43	6.23
15	1,2,4,8-TetraCDF	5.00	2.3767	0.05420	5.05	4.89	6.42
16	1,3,6,8-TetraCDF	6.66	2.3432	0.05412	6.24	5.43	6.53
17	2,3,4,6-TetraCDF	6.46	2.3464	0.05408	6.31	5.87	6.23

18	2,3,4,7-TetraCDF	7.60	2.3049	0.05395	7.38	6.90	6.23
19	2,3,4,8-TetraCDF	6.70	2.3266	0.05402	6.24	6.41	6.23
20	2,3,6,8-TetraCDF	6.66	2.3049	0.05400	6.67	6.55	6.53
21	2,3,7,8-TetraCDF	7.39	2.2703	0.05388	7.45	7.61	6.23
22	1,2,3,4,8-PentaCDF	6.92	2.4000	0.05407	5.58	5.96	6.83
23	1,2,3,6,7-PentaCDF	7.17	2.3540	0.05398	7.35	6.71	6.98
24	1,2,3,7,8-PentaCDF	7.13	2.3379	0.05392	7.32	7.23	6.79
25	1,2,3,7,9-PentaCDF	6.40	2.3751	0.05405	5.55	6.18	6.93
26	1,2,4,6,8-PentaCDF	5.51	2.4014	0.05416	5.41	5.17	7.13
27	1,2,4,7,8-PentaCDF	5.89	2.3660	0.05404	6.57	6.28	6.98
28	1,3,4,7,8-PentaCDF	6.70	2.3479	0.05397	6.33	6.80	6.93
29	1,3,4,7,9-PentaCDF	4.70	2.4041	0.05416	4.81	5.34	7.13
30	1,3,6,7,8-PentaCDF	6.70	2.3544	0.05398	6.74	6.73	6.98
31	2,3,4,7,8-PentaCDF	7.82	2.3179	0.05386	7.69	7.75	6.79
32	2,3,4,7,9-PentaCDF	6.70	2.3544	0.05398	6.98	6.73	6.93
33	1,2,3,4,7,8-HexaCDF	6.64	2.3884	0.05391	7.16	7.51	7.39
34	1,2,3,6,7,8-HexaCDF	6.57	2.3810	0.05390	7.23	7.49	7.34
35	1,2,3,6,7,9-HexaCDF	5.08	2.4117	0.05402	5.90	6.59	7.53
36	2,3,4,6,7,9-HexaCDF	7.33	2.3917	0.05396	6.79	6.89	7.34

[a] From ref. [21, 32]
[b] Cross-validation results of the model with parameters presented in Table 5.
[c] Cross-validated results using the modified variable Balaban index
[d] From ref.[27]

As a consequence, when the data set indicates a systematic variation with molecular size, index *J* should be used in multiparametric (at least biparametric) correlations.

As a side remark, in a paper by H. X. Zhang and coworkers[64] on a QSAR study of PCDD, PCDF, and PCB using a heuristic method and support vector machine, the heuristic method for selecting CODESSA-parameters found that the Balaban index was the first and most powerful descriptor.

Variable topological indices have been introduced in early nineties[65-67] as an alternative way to account for characterization of heteroatoms in the molecule. The individual contributions of different atoms were introduced by replacing zero diagonal entries in the adjacency matrix with the variable weights whose values are optimized during modeling procedure. Despite initial promising results and reported high quality regressions (68-70),[68-70] it was soon obvious that only adjacency-matrix-related indices cannot be used in all cases because they somehow postulate intramolecular interaction/influences only for adjacency between atoms in the molecules. To account for longer-range interactions, several classes of distance related indices were introduced from the universal distance connectivity indices .[71-72] Then variable distance connectivity index to pure variable distance indices based solely on distance matrix. However, especially for the latest two classes, very limited numbers of applications were reported.[73-74]

The aim of this report is to apply the variable distance connectivity index and the newly-developed distance-weighted variable connectivity index for modeling the toxicity of polychlorobenzofurans, where presumably long-range intramolecular interactions play very important role. However, an even more important goal of this study was the structural interpretation of the obtained models.

The distance-weighted variable connectivity index was recently introduced for the modeling of pK_a values of organic and halogenated organic acids.[75] It was defined as a topological distance-weighted variable connectivity index of the complete graph (equation 1):

$$ {}^{1}\chi^{f} = \sum_{i=1}^{k}\sum_{j=i+1}^{k} d_{i,j}{}^{-\lambda} \cdot {}^{1}\chi_{i,j}^{f} = \sum_{i=1}^{n}\sum_{j=i+1}^{k} d_{i,j}{}^{-\lambda} \prod_{i=1}^{2} (\delta_i^{f})_j^{-0.5} \qquad (2) $$

where $d_{i,j}$ represents the topological distance between vertices i and j, n is the number of vertices in the molecule, λ is the distance weighting exponent, $\chi_{i,j}^{f}$ is the partial contribution of the vertex pair i,j, and δ_i^{f} is the row sum of the i-th vertex in the augmented adjacency matrix. The index is calculated for the complete graph representing the hydrogen-depleted compound.

On the other hand, the variable distance-based index (or variable Balaban index) is a topological-distance-related index and is calculated from the modified distance matrix,[74] where individual atom contributions are introduced into the diagonal of the distance matrix. Afterwards the variable index is calculated in the same way as with the modified Balaban index,[62] by using equation 2:

$$ J^{f} = \sum_{i=1}^{n}\sum_{j=1}^{i=1} (d_i \cdot d_j)^{-0.5} \qquad (3) $$

where d_j and d_i are row sums of the modified distance matrix, and where diagonal zero elements were substituted by the weights for the individual types of heteroatoms. In equation 3, we leave out the normalizing factor containing the number of edges and the cyclomatic number, in order to allow the size of the graph to influence the value of the descriptor. Earlier, it was shown that the advantage of distance-based topological indices consisted in the fact that local vertex invariants such as distance sums could easily allow including information on the nature of heteroatoms and bond multiplicity.[76-78] For the Balaban index, one could choose between parameters characterizing the relative electronegativity or covalent radius of the heteroatom with respect to those of the carbon atom.

Testing of the prediction capabilities of the MLR models

The leave-one-out (LOO) cross-validation procedure was used for the evaluation of the prediction capabilities of MLR models during the stepwise selection of structural descriptors. The MLR model with the best cross-validation results was chosen for further evaluation. Since the number of data points in each data set was relatively low, the same cross-validation procedure was used also for the final evaluation of the created models.

3 Results and Discussion

Most data points were found for AhR binding affinities and this is why the most detailed studies were performed on this data set. Separate QSPR models were developed for the prediction of binding affinities for the three sets of polychlorinated compounds **1**, **2**, and **3**. A chemical structure was encoded by 367 different informational, topological, geometric, electrostatic, electrotopological and quantum-chemical indices calculated by the CODESSA software. Afterwards an optimal *m*-parameter MLR model with up to 10 descriptors was selected based on the best cross-validation capabilities obtained by the leave-one-out cross-validation procedure. The influence of the number of selected parameters of the MLR model on its cross-validation capabilities was evaluated. The squared correlation coefficients r^2 and q^2 as well as standard error(s) were calculated for each *m*-dimensional MLR model. As an example, the dependence of the values of these parameters on the number *m* of descriptors for modeling the toxicity of polychloro-biphenyls is shown in Fig. 2.

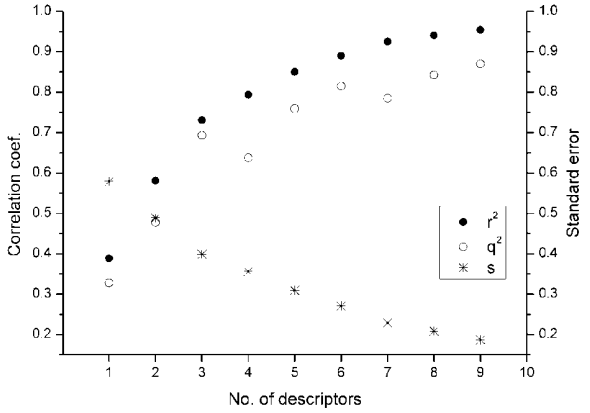
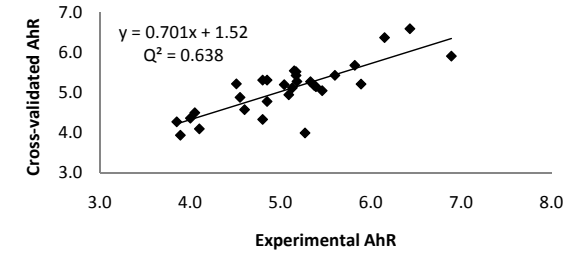
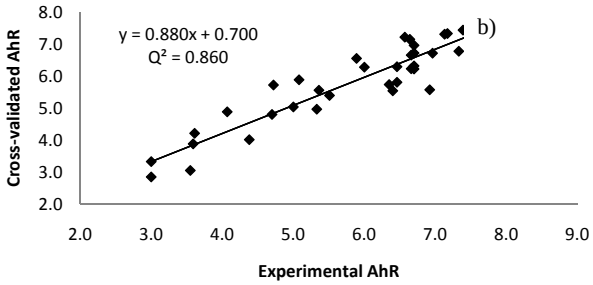


Figure 2. Dependence of r^2, q^2 and *s* on the number of descriptors in MLR model (PCBs)

a)





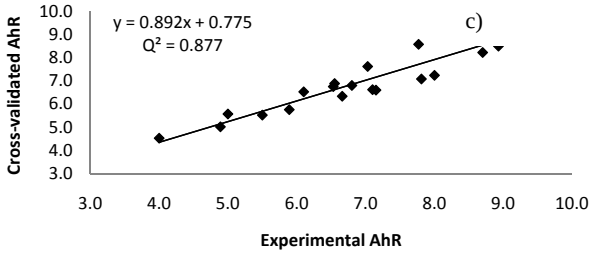


Figure 3. Experimental vs. calculated activities: a - polychloro-biphenyls, b - polychloro-dibenzofurans and c - polyhalo-dibenzodioxins

In all cases, r^2 for the retrieved values increases with the number of parameters of the MLR model. On the other hand, the parameter for the predicted values (q^2) increases markedly only at the beginning and then more slowly with fluctuations. Such a result is

expected because with the addition of each new descriptor to the model we are improving the structural description of the chemical compound, that is, we are introducing structural features that contain some information about modeled property. This goes only up to the point where new descriptors cannot improve the structural representation any more. The r^2 for the retrieved values still increases but the cross-validation ability of the models decreases. The model is over-fitted and the new structural indices are introducing noise into the model.

In the case of polychloro-biphenyls, three-parameter MLR models were selected for the calculating AhR binding affinities. Using a similar approach, four-parameter and three-parameter MLR models were selected for the modeling AhR binding affinities for chlorinated dibenzofurans and dioxins, respectively. The selected descriptors and the regression coefficients of all three models are listed in Tables 4. Interestingly, there is no common descriptor for these three polyhalo-aromatics.

All final models were cross-validated in order to obtain estimates for the prediction ability of the developed models. The experimental vs. cross-validated results are shown in Figure 3 and Tables 1 to 3. The corresponding q^2 values for chlorinated biphenyls, dibenzofurans and dibenzodioxins were 0.638, 0.860 and 0.877, respectively, as seen in Table 5. This Table also contains data for the correlation factor r^2, standard error (s), root mean square error (RMS), and RMS of cross-validation procedures (RMS_{cv}). The most homogeneous data set was for chlorinated dibenzodioxins. On first sight, the models for toxicity were different between the three chemical classes. However, a closer analysis of the obtained models showed that intramolecular distances between chlorine atoms as well as distribution of the charges play an important role in the determining the biological activity. This is why we have tested some topological indices that are able to encode such structural features.

Two types of variable topological indices (the variable distance connectivity index and the variable Balaban distance index) were tested using the polychloro-dibenzofuran data set. Both indices are able to encode information about heteroatoms and distant intramolecular interactions. However, the first index failed to encode structural features that are important for the particular modeling property. During the optimization procedure, the Simplex algorithm did not converge to the better solution but was oscillating around the value that was obtained by random weights. On the other hand, a strong convergence was noticed in the case of the

variable Balaban index. In Fig. 4 we show the experimental vs. calculated results for the variable Balaban index.

Table 4. Regression coefficients and selected structural descriptors for predicting AhR receptor binding affinities of polychloro-biphenyls, polychloro-dibenzofurans and polychloro-dibenzodioxins by the CODESSA program

Regression coefficient	Standard error of regression coefficient	Structural descriptor
Polychloro-biphenyls (**3**)		
-445.1	84.9	Intercept
1.189	0.21	Minimal electron-nucleus attraction for a C–Cl bond
-222.8	47.8	Fractional PPSA (PPSA-3/TMSA) [Zefirov's PC]
-0.0893	0.022	Minimal electron-nucleus attraction for a C–C bond
Polychloro-dibenzofurans (**1**)		
40.2	6.0	Intercept
47.2	4.1	Topographic electronic index (all bonds) [Zefirov's PC]
-27.0	3.0	Balaban index
-170	64	Minimal nucleophilic reactivity index for a Cl atom
85	35	Maximal 1-electron reactivity index for an O atom
Polychloro-dibenzodioxins (**2**)		
11.0	1.4	Intercept
40.8	3.7	Minimal net atomic charge for a C atom
0.172	0.037	ESP-PNSA-3 Atomic charge weighted PNSA [Semi-MO PC]
-471.2	135.9	Maximum nucleophilic reaction index for an O atom

The following equation was obtained for modeling the AhR of PCDFs:

$$\log \frac{1}{EC_{50}} = -8273 J^{f} + 453.3 \qquad (4)$$

where weights for C, O, and Cl atoms were 208.6, 31126, and 7317.6, respectively. As it can be seen, the much higher weight of chlorine indicates its decisive role in toxicity; the even

higher weight for oxygen is not essential because it does not vary among the PCDFs. Since this index is bond-additive, we can analyze which structural feature of the molecule is predominant in determining the toxicity: the equation 4 contains a negative slope for J^f, therefore the lower the value of J^f, the higher the toxicity. From equation 3 one can see that small bond contributions are associated with chlorine atoms situated at larger distances from the central furan ring in PCDFs, which is consistent with experiments, and may be useful for a better understanding of the interaction between the agonist and the AhR protein.

Table 5. Statistical parameters for MLR and variable descriptor models

	PCBs[a]	PCDDs[a]	PCDFs[a]	PCDFs[b]
r^2	0.73	0.92	0.89	0.80
q^2	0.64	0.88	0.86	0.76
s	0.40	0.43	0.48	0.62
RMS	0.37	0.39	0.45	0.60
RMS_{cv}	0.44	0.48	0.51	0.66

[a] MLR procedures
[b] Variable descriptor model

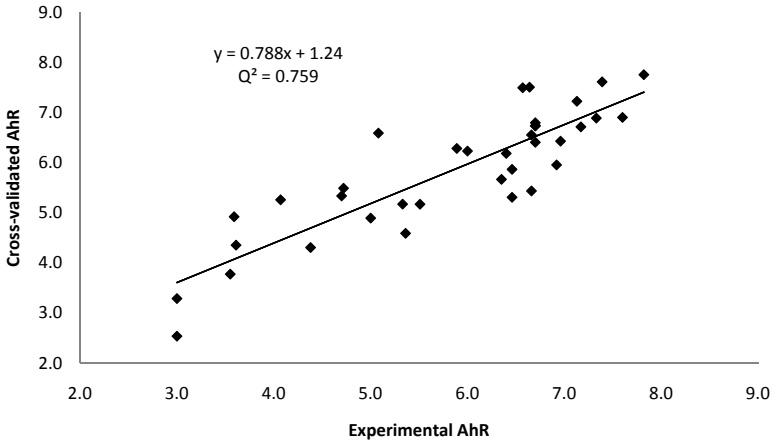


Figure 4. Calculated *vs.* experimental biological activities using the variable Balaban index for polychloro-dibenzofurans **1.**

Although the result of MLR models using electrostatic and quantum-chemical descriptors appears to be better than when using topological models, the advantage of the latter is the structural interpretation. The analysis of the optimal weights showed that the position of chlorine atoms at the highest distance between them in the molecule is the most important structural factor for the modeling toxicity in the case of polychloro-dibenzofurans.

The three MLR models that were developed for the calculation of aryl hydrocarbon receptor (AhR) binding affinity of chlorinated biphenyls, dibenzofurans and dibenzodioxins were able to explain more than 70% of the total variance. The best prediction models were obtained in the case of polychloro-dibenzodioxins and the worst in the case of biphenyls. The structural interpretation of the obtained models shows that the distribution of charge and the distant intramolecular interactions play important roles for modeling the toxicity. The newly developed variable Balaban's distance-based topological index showed reasonable modeling abilities and superior interpretation abilities when compared to classical MLR QSAR models. It should be mentioned that Sabljic et al.[79-80] and Jäntschi et al.[81-83] have also published reports correlating the toxicity of polychlorinated aromatics with topological molecular descriptors.

As a side remark to data for PCBs, the logP values listed in Table 3 show weak correlations ($r^2 = 0.51$) with experimental AhR values and with the *J* and *J'* indices ($r^2 = 0.64$)

Acknowledgment: Financial support is gratefully acknowledged from the Ministry of Science and Higher Education (grants Grant P1-0153), Slovenian Human Resources Development and Scholarship Ad-Futura (grant No. 1000-07-780002) and EU Commission (TRUEFOOD project). TRUEFOOD - "Traditional United Europe Food" is an Integrated Project financed by the European Commission under the 6th Framework Programme for RTD Contract FOOD-CT-2006-016264. The information in this document reflects only the author's views and the Community is not liable for any use that may be made of the information contained therein.

References

[1] A. Beteringhe, A. T. Balaban, QSAR for toxicities of polychlorodibenzofurans, polychlorodibenzo-1,4-dioxins, and polychlorobiphenyls, *Arkivoc* (2004) 163-182.

[2] P. Sun, I. Basu, R. A. Hites, Temporal trends of polychlorinated biphenyls in precipitation and air at Chicago, *Environ. Sci. Technol.* **40** (2006) 1178-1183.

[3] J. Fu, T. Wang, P. Wang, G. Qu, Y. Wang, Q. Zhang, A. Zhang, G. Jiang, Temporal trends (2005-2009) of PCDD/Fs, PCBs, PBDEs in rice hulls from an e-waste dismantling area after stricter environmental regulations, *Chemosphere* **88** (2012) 330-335.

[4] A. Katsoyiannis, R. Gioia, A. J. Sweetman, K. C. Jones, Continuous monitoring of PCDD/Fs in the UK atmosphere: 1991-2008, *Environ. Sci. Technol.* **44** (2010) 5735-5740.

[5] Y. Li, W. Thanh, P. Wang, L. Ding, X. Li, Y. Wang, Q. Zhang, A. Li, G. Jiang, Reduction of atmospheric polychlorinated dibenzo-p-dioxins and dibenzofurans (PCDD/Fs) during the 2008 Beijing olympic games, *Environ. Sci. Technol.* **45** (2011) 3304-3309.

[6] E. Niemirycz, D. Jankowska, Concentrations and profiles of PCDD/Fs in sediments of major Polish rivers and the Gdansk basin - Baltic sea, *Chemosphere* **85** (2011) 525-532.

[7] M. Venier, J. Ferrario, R. A. Hites, Polychlorinated dibenzo-p-dioxins and dibenzofurans in the atmosphere around the Great lakes, *Environ. Sci. Technol.* **43** (2009) 1036-1041.

[8] L. L. Aylward, S. M. Hays, Temporal trends in human TCDD body burden: Decreases over three decades and implications for exposure levels, *J. Expo. Anal. Env. Epid.* **12** (2002) 319-328.

[9] R. A. Hites, Dioxins: An overview and history, *Environ. Sci. Technol.* **45** (2011) 16-20.

[10] G. Perello, J. Gomez-Catalan, V. Castell, J. M. Llobet, J. L. Domingo, Assessment of the temporal trend of the dietary exposure to PCDD/Fs and PCBs in Catalonia, over Spain: Health risks, *Food Chem. Toxicol.* **50** (2012) 399-408.

[11] H. T. O'Brien, R. Blanchet, D. Gagne, J. Lauziere, C. Vezina, E. Vaissiere, P. Ayotte, S. Dery, Exposure to toxic metals and persistent organic pollutants in inuit children attending childcare centers in Nunavik, Canada, *Environ. Sci. Technol.* **46** (2012) 4614-4623.

[12] E. Holt, R. Weber, G. Stevenson, C. Gaus, Polychlorinated dibenzo-p-dioxins and dibenzofurans (PCDD/Fs) impurities in pesticides: A neglected source of contemporary relevance, *Environ. Sci. Technol.* **44** (2010) 5409-5415.

[13] S. H. Safe, Comparative toxicology and mechanism of action of polychlorinated dibenzo-p-dioxins and dibenzofurans, *Annu. Rev. Pharmacol.* **26** (1986) 371-399.

[14] M. Van den Berg, L. S. Birnbaum, M. Denison, M. De Vito, W. Farland, M. Feeley, H. Fiedler, H. Hakansson, A. Hanberg, L. Haws, M. Rose, S. Safe, D. Schrenk, C. Tohyama, A. Tritscher, J. Tuomisto, M. Tysklind, N. Walker, R. E. Peterson, The 2005 World Health Organization reevaluation of human and mammalian toxic equivalency factors for dioxins and dioxin-like compounds, *Toxicol. Sci.* **93** (2006) 223-241.

[15] S. Jansson, H. Antti, S. Marklund, M. Tysklind, Multivariate Relationships between molecular descriptors and isomer distribution patterns of PCDD/Fs formed during MSW combustion, *Environ. Sci. Technol.* **43** (2009) 7032-7038.

[16] J. Y. Ryu, J. A. Mulholland, J. E. Dunn, F. Iino, B. K. Gullett, Potential role of chlorination on pathways in PCDD/F formation in a municipal waste incinerator, *Environ. Sci. Technol.* **38** (2004) 5112-5119.

[17] J. D. McKinney, T. Darden, M. A. Lyerly, L. G. Pedersen, PCB and related compound binding to the ah receptor(s) theoretical-model based on molecular-parameters and molecular mechanics, *Quant Struct-Act Rel* **4** (1985) 166-172.

[18] J. S. Murray, P. Evans, P. Politzer, A comparative-analysis of the electrostatic potentials of some structural analogs of 2,3,7,8-tetrachlorodibenzo-p-dioxin and of related aromatic systems, *Int. J. Quantum Chem.* **37** (1990) 271-289.

[19] M. Wagener, J. Sadowski, J. Gasteiger, Autocorrelation of molecular-surface properties for modeling corticosteroid-binding globulin and cytosolic ah receptor activity by neural networks, *J. Am. Chem. Soc.* **117** (1995) 7769-7775.

[20] C. L. Waller, J. D. McKinney, 3-dimensional quantitative structure-activity-relationships of dioxins and dioxin-like compounds - model validation and ah receptor characterization, *Chem. Res. Toxicol.* **8** (1995) 847-858.

[21] K. Tuppurainen, J. Ruuskanen, Electronic eigenvalue (EEVA): a new QSAR/QSPR descriptor for electronic substituent effects based on molecular orbital energies. A QSAR approach to the Ah receptor binding affinity of polychlorinated biphenyls (PCBs), dibenzo-p-dioxins (PCDDs) and dibenzofurans (PCDFs), *Chemosphere* **41** (2000) 843-848.

[22] J. M. Luco, J. Galvez, R. Garcia-Domenech, J. V. de Julian-Ortiz, Structural invariants for the prediction of relative toxicities of polychloro dibenzo-p-dioxins and dibenzofurans, *Mol. Divers.* **8** (2004) 331-342.

[23] G. Zheng, M. Xiao, X. H. Lu, QSAR study on the Ah receptor-binding affinities of polyhalogenated dibenzo-p-dioxins using net atomic-charge descriptors and a radial basis neural network, *Anal. Bioanal. Chem.* **383** (2005) 810-816.

[24] J. Dai, C. Sun, S. Han, L. Wang, QSAR for polychlorinated organic compounds (PCOCs). I. Prediction of partition properties for PCOCs using quantum chemical parameters, *B. Environ. Contam. Tox.* **62** (1999) 530-538.

[25] X. Zeng, Z. Wang, Z. Ge, H. Liu, Quantitative structure-property relationships for predicting subcooled liquid vapor pressure (P-L) of 209 polychlorinated diphenyl ethers (PCDEs) by DFT and the position of Cl substitution (PCS) methods, *Atmos. Environ.* **41** (2007) 3590-3603.

[26] E. Papa, J. C. Dearden, P. Gramatica, Linear QSAR regression models for the prediction of bioconcentration factors by physicochemical properties and structural theoretical molecular descriptors, *Chemosphere* **67** (2007) 351-358.

[27] V. Hornak, S. Balaz, K. J. Schaper, J. K. Seydel, Multiple binding modes in 3D-QSAR: Microbial degradation of polychlorinated biphenyls, *Quant. Struct. Act. Rel.* **17** (1998) 427-436.

[28] C. C. Abnet, R. L. Tanguay, W. Heideman, R. E. Peterson, Transactivation activity of human, zebrafish, and rainbow trout aryl hydrocarbon receptors expressed in COS-7 cells: Greater insight into species differences in toxic potency of polychlorinated dibenzo-p-dioxin, dibenzofuran, and biphenyl congeners, *Toxicol. Appl. Pharm.* **159** (1999) 41-51.

[29] G. Luthe, J. A. Jacobus, L. W. Robertson, Receptor interactions by polybrominated diphenyl ethers versus polychlorinated biphenyls: A theoretical structure-activity assessment, *Environ. Toxicol. Phar.* **25** (2008) 202-210.

[30] V. Lukacova, S. Balaz, Multimode ligand binding in receptor site modeling: Implementation in CoMFA, *J. Chem. Inf. Comp. Sci.* **43** (2003) 2093-2105.

[31] S. S. So, M. Karplus, Three-dimensional quantitative structure-activity relationships from molecular similarity matrices and genetic neural networks. 2. Applications, *J. Med. Chem.* **40** (1997) 4360-4371.

[32] S. Arulmozhiraja, M. Morita, Structure-activity relationships for the toxicity of polychlorinated dibenzofurans: Approach through density functional theory-based descriptors, *Chem. Res. Toxicol.* **17** (2004) 348-356.

[33] J. Diao, Y. Li, S. Shi, Y. Sun, Y. Sun, QSAR models for predicting toxicity of polychlorinated dibenzo-p-dioxins and dibenzofurans using quantum chemical descriptors, *B. Environ. Contam. Tox.* **85** (2010) 109-115.

[34] P. I. Petkov, J. C. Rowlands, R. Budinsky, B. Zhao, M. S. Denison, O. Mekenyan, Mechanism-based common reactivity pattern (COREPA) modelling of aryl hydrocarbon receptor binding affinity, *SAR QSAR Environ. Res.* **21** (2010) 187-214.

[35] M. A. Lill, M. Dobler, A. Vedani, In silico prediction of receptor-mediated environmental toxic phenomena - Application to endocrine disruption, *SAR QSAR Environ. Res.* **16** (2005) 149-169.

[36] A. R. Katritzky, K. Kasemets, S. Slavov, M. Radzvilovits, K. Tamm, M. Karelson, Estimating the toxicities of organic chemicals in activated sludge process, *Water Res.* **44** (2010) 2451-2460.

[37] E. Lo Piparo, K. Koehler, A. Chana, E. Benfenati, Virtual screening for aryl hydrocarbon receptor binding prediction, *J. Med. Chem.* **49** (2006) 5702-5709.

[38] J. J. Song, M. K. Kim, E. Hur, H. Park, S. H. Eom, Crystallization and preliminary X-ray crystallographic analysis of bHLH domain of aryl hydrocarbon receptor nucleus translocator, *Protein Peptide Lett.* **8** (2001) 419-422.

[39] H. I. Swanson, C. A. Bradfield, The Ah-receptor - genetics, structure and function, *Pharmacogenetics* **3** (1993) 213-230.

[40] S. Bandiera, S. Safe, A. B. Okey, Binding of polychlorinated-biphenyls classified as either phenobarbitone-type, 3-methylcholanthrene-type or mixed-type inducers to cytosolic ah receptor, *Chem-Biol Interact* **39** (1982) 259-277.

[41] S. Bandiera, T. Sawyer, M. Romkes, B. Zmudzka, L. Safe, G. Mason, B. Keys, S. Safe, Polychlorinated dibenzofurans (PCDFS) - effects of structure on binding to the 2,3,7,8-TCDD cytosolic receptor protein, AHH induction and toxicity, *Toxicology* **32** (1984) 131-144.

[42] S. Bandiera, T. W. Sawyer, M. A. Campbell, T. Fujita, S. Safe, Competitive-binding to the cytosolic 2,3,7,8-tetrachlorodibenzo-para-dioxin receptor - effects of structure on the affinities of substituted halogenated biphenyls - a QSAR analysis, *Biochem. Pharmacol.* **32** (1983) 3803-3813.

[43] S. Safe, S. Bandiera, T. Sawyer, L. Robertson, L. Safe, A. Parkinson, P. E. Thomas, D. E. Ryan, L. M. Reik, W. Levin, M. A. Denomme, T. Fujita, PCBS - structure-function-relationships and mechanism of action, *Environ. Health. Persp.* **60** (1985) 47-56.

[44] G. Mason, K. Farrell, B. Keys, J. Piskorskapliszczynska, L. Safe, S. Safe, polychlorinated dibenzo-para-dioxins - quantitative invitro and invivo structure-activity-relationships, *Toxicology* **41** (1986) 21-31.

[45] G. Mason, T. Sawyer, B. Keys, S. Bandiera, M. Romkes, J. Piskorskapliszczynska, B. Zmudzka, S. Safe, Polychlorinated dibenzofurans (PCDFS) - correlation between invivo and invitro structure-activity-relationships, *Toxicology* **37** (1985) 1-12.

[46] A. Ashek, C. Lee, H. Park, S. J. Cho, 3D QSAR studies of dioxins and dioxin-like compounds using CoMFA and CoMSIA, *Chemosphere* **65** (2006) 521-529.

[47] O. G. Mekenyan, G. D. Veith, D. J. Call, G. T. Ankley, A QSAR evaluation of Ah receptor binding of halogenated aromatic xenobiotics, *Environ. Health. Persp.* **104** (1996) 1302-1310.

[48] S. Safe, Polychlorinated-biphenyls (PCBS), dibenzo-para-dioxins (PCDDS), dibenzofurans (PCDFS), and related-compounds - environmental and mechanistic considerations which support the development of toxic equivalency factors (TEFS), *CRC Cr. Rev. Toxicol.* **21** (1990) 51-88.

[49] S. Safe, S. Bandiera, T. Sawyer, B. Zmudzka, G. Mason, M. Romkes, M. A. Denomme, J. Sparling, A. B. Okey, T. Fujita, effects of structure on binding to the 2,3,7,8-TCDD receptor protein and AHH induction - halogenated biphenyls, *Environ. Health. Persp.* **61** (1985) 21-33.

[50] K. Ballschmiter, M. Zell, Analysis of polychlorinated-biphenyls (pcb) by glass-capillary gas-chromatography - composition of technical aroclor-pcb and clophen-pcb mixtures, *Fresen. Z. Anal. Chem.* **302** (1980) 20-31.

[51] E. Schulte, R. Malisch, Calculation of the real PCB content in environmental-samples .1. Investigation of the composition of 2 technical PCB mixtures, *Fresen. Z. Anal. Chem.* **314** (1983) 545-551.

[52] R. Guitart, P. Puig, J. Gomezcatalan, Requirement for a standardized nomenclature criterium for PCBS - computer-assisted assignment of correct congener denomination and numbering, *Chemosphere* **27** (1993) 1451-1459.

[53] S. A. Mills, D. I. Thal, J. Barney, A summary of the 209 PCB congener nomenclature, *Chemosphere* **68** (2007) 1603-1612.

[54] A. R. Fernandes, D. Mortimer, R. J. Wall, D. R. Bell, M. Rose, M. Carr, S. Panton, F. Smith, Mixed halogenated dioxins/furans (PXDD/Fs) and biphenyls (PXBs) in food: Occurrence and toxic equivalent exposure using specific relative potencies, *Environ. Int.* **73** (2014) 104-110.

[55] R. J. Wall, A. Fernandes, M. Rose, D. R. Bell, I. R. Mellor, Characterisation of chlorinated, brominated and mixed halogenated dioxins, furans and biphenyls as potent and as partial agonists of the Aryl hydrocarbon receptor, *Environ. Int.* **76** (2015) 49-56.

[56] A. T. Balaban, T. S. Balaban, Correlations using topological indexes based on real graph invariants, *J. Chim. Phys. PCB* **89** (1992) 1735-1745.

[57] J. Devillers, A. T. Balaban (Eds), *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach, 1999.

[58] M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley, New York, 2000.

[59] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH: Weinheim, 2002; *Molecular Descriptors for Chemoinformatics*, 2002nd Revised and Enlarged Edition, 2009.

[60] J. J. P. Stewart, Special issue - MOPAC - a semiempirical molecular-orbital program, *J. Comput. Aid. Mol. Des.* **4** (1990) 1-45.

[61] A. R. Katritzky, V. S. Lobanov, M. Karelson, CODESSA training manual Gainesville, 1995.

[62] A. T. Balaban, Highly discriminating distance-based topological index, *Chem. Phys. Lett.* **89** (1982) 399-404.

[63] M. Randić, Characterization of molecular branching, *J. Am. Chem. Soc.* **97** (1975) 6609-6615.

[64] F. Luan, W. P. Ma, X. Y. Zhang, H. X. Zhang, M. C. Liu, Z. D. Hu, B. T. Fan, QSAR study of polychlorinated dibenzodioxins, dibenzofurans, and biphenyls using the heuristic method and support vector machine, *QSAR Comb. Sci.* **25** (2006) 46-55.

[65] M. Pompe, M. Veber, M. Randić, A. T. Balaban, Using variable and fixed topological indices for the prediction of reaction rate constants of volatile unsaturated hydrocarbons with OH radicals, *Molecules* **9** (2004) 1160-1176.

[66] M. Randić, On computation of optimal parameters for multivariate-analysis of structure-property relationship, *J. Comput. Chem.* **12** (1991) 970-980.

[67] M. Randić, Novel graph theoretical approach to heteroatoms in quantitative structure - activity relationships, *Chemometr. Intell. Lab.* **10** (1991) 213-227.

[68] M. Randić, High quality structure-property regressions. Boiling points of smaller alkanes, *New J. Chem.* **24** (2000) 165-171.

[69] M. Randić, S. C. Basak, Construction of high-quality structure-property-activity regressions: The boiling points of sulfides, *J. Chem. Inf. Comp. Sci.* **40** (2000) 899-905.

[70] M. Randić, M. Pompe, D. Mills, S. C. Basak, Variable connectivity index as a tool for modeling structure-property relationships, *Molecules* **9** (2004) 1177-1193.

[71] A. R. Matamala, E. Estrada, Simplex optimization of generalized topological index (GTI-simplex): A unified approach to optimize QSPR models, *J. Phys. Chem. A* **109** (2005) 9890-9895.

[72] A. R. Matamala, E. Estrada, Generalised topological indices: Optimisation methodology and physico-chemical interpretation, *Chem. Phys. Lett.* **410** (2005) 343-347.

[73] O. Ivanciuc, Design of topological indices. Part 23 - Structural descriptors derived from the distance-path matrix of vertex- and edge-weighted molecular graphs, *Rev. Roum. Chim.* **46** (2001) 543-552.

[74] M. Randić, M. Pompe, The variable molecular descriptors based on distance related matrices, *J. Chem. Inf. Comp. Sci.* **41** (2001) 575-581.

[75] A. Ščavničar, A. T. Balaban, M. Pompe, Application of variable anti-connectivity index to active sites. Modelling pK (a) values of aliphatic monocarboxylic acids, *SAR QSAR Environ. Res.* **24** (2013) 553-563.

[76] O. Ivanciuc, T. Ivanciuc, A. T. Balaban, Design of topological indices. Part 10. Parameters based on electronegativity and covalent radius for the computation of molecular graph descriptors for heteroatom-containing molecules, *J. Chem. Inf. Comput. Sci.* **38** (1998) 395-401.

[77] A. T. Balaban, Chemical graphs. 48. Topological index *J* for heteroatom-containing 6molecules taking into account periodicities of element properties, *MATCH Commun. Math. Comput. Chem.* **21** (1986) 115-122.

[78] O. Ivanciuc, A. T. Balaban, Design of topological indices. Part 5. Precision and error in computing graph theoretic invariants for molecules containing heteroatoms and multiple bonds, *MATCH Commun. Math. Comput. Chem.* **30** (1994) 117-139.

[79] A. Sabljić, H. Güsten, J. Hermens, A. Opperhuizen, Modeling octanol water partition-coefficients by molecular topology - chlorinated benzenes and biphenyls, *Environ. Sci. Technol.* **27** (1993) 1394-1402.

[80] A. Sabljić, M. Protić, Molecular connectivity - a novel method for prediction of bioconcentration factor of hazardous chemicals, *Chem. Biol. Interact.* **42** (1982) 301-310.

[81] L. Jäntschi, S. D. Bolboacă, Molecular descriptors family on structure activity relationships 6. Octanol-water partition coefficient of polychlorinated biphenyls, *Leonardo El. J. Pract. Technol.* **8** (2006) 71-86.

[82] L. Jäntschi, S. D. Bolboacă, M. V. Diudea, Chromatographic retention times of polychlorinated biphenyls: from structural information to property characterization, *Int. J. Mol. Sci.* **8** (2007) 1125-1157.

[83] L. Jäntschi, S. D. Bolboacă, R. E. Sestraş, Meta-heuristics on quantitative structure-activity relationships: study on polychlorinated biphenyls, *J. Mol. Model.* **16** (2010) 377-386.