JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

ARTICLE

pubs.acs.org/JAFC

Theoretical Molecular Descriptors Relevant to the Uptake of Persistent Organic Pollutants from Soil by Zucchini. A QSAR Study†

Barna Bordás,* Iván Bélai, and Tamás Kőmíves

Plant Protection Institute, Hungarian Academy of Sciences, P.O. Box 102, H-1525 Budapest, Hungary

ABSTRACT: The uptake of persistent organic pollutants (POPs) from soil by plants allows the development of phytoremediation protocols to rehabilitate contaminated areas. The use of diverse theoretical descriptors has been reported in the literature for developing quantitative structure−activity relationship (QSAR) models for predicting the bioconcentration factors (BCFs) of POPs in different plants. In this paper an evaluation is given on the molecular properties of POPs in terms of theoretical molecular descriptors that are relevant to the uptake and accumulation of these persistent pollutants from soil by two zucchini varieties. Statistically significant and predictive linear regression models have been developed for the BCF values of 20 polychlorinated dibenzo-*p*-dioxins/dibenzofurans and 14 polyhalogenated biphenyls in two zucchini varieties based on retrospective data. The relevant parameters have been selected from a set of 1660 DRAGON, 150 VolSurf, and 11 quantum chemical descriptors. The two most significant regression models, containing VolSurf, DRAGON GETAWAY, and quantum chemical descriptors, displayed the following statistical parameters: (eq 3) $n = 27$, $R^2 = 0.940$, $q^2 = 0.922$, SE = 0.155, $F = 392.1$; (eq 4) $n = 27$, $R^2 = 0.921$, $q^2 = 0.898$, SE = 0.161, $F = 140.4$. Predictive capabilities of the equations have been validated by using external validation sets. The QSAR models proposed might contribute to the development of viable soil remediation strategies.

KEYWORDS: BCF, bioconcentration factor, dioxin, DRAGON software, phytoremediation, persistent organic pollutants, POP, plant, QSAR, soil, SYBYL, VolSurf, zucchini

INTRODUCTION

Persistent organic pollutants (POPs)[1] are chemical substances produced by the chemical industry (xenobiotics) and generated by combustion. Many thousands of POP derivatives contaminate the environment; several of them are members of homologous chemical classes, such as polychlorinated dibenzo-*p*-doxins (PCDDs) of type I, dibenzofurans (PCDFs) of type II, and biphenyls (PCBs) of type III (Figure 1). POPs are released into the environment by many different ways. POPs are persistent in the environment and accumulate in soil, water, sediment, air, and biota.[2] Exposure to POPs elicits adverse effects on ecosystems and human health. Some of these compounds (e.g., 2378TCDD, dioxin) are extremely toxic to humans. One of the important routes of exposure of humans to POPs is the consumption of plant products contaminated by these pollutants.[3] POPs are taken up by the plants from the soil through the root system or from the air by atmospheric transport.[4]

Human industrial and agricultural activities often contaminate wide areas, and the development of methods for the removal of POPs from soils is an intense area of research. The uptake of POPs from soil by plants allows the development of phytoremediation protocols to rehabilitate contaminated areas.[5] The bioconcentration factor (BCF), defined as the ratio between the contaminant concentration in the plant tissue and the concentration in soil, can be measured to identify plant species useful for soil rehabilitation purposes.

Pharmacokinetic behavior (absorption, distribution, metabolism, and excretion; ADME) of POPs in different biological organisms have been extensively studied by employing diverse quantitative structure−property relationship (QSPR) and quantitative structure−activity relationship (QSAR) methodologies.[6] Several theoretical descriptors have been reported in the literature that were found to be useful as independent variables for modeling and predicting the BCFs of POPs mostly in aquatic target organisms. The sorption and half-life/degradation of POPs has been well studied.[7] However, the uptake and metabolic fate in plants are far less studied.

We suggest that once a suitable plant for soil remediation has been found, QSAR modeling of the ADME behavior of pollutant(s) in the given soil−plant system is a useful approach to identify those structural analogues or chemical types that will be preferentially taken up by the plant. These predictions can be carried out in silico before any wet experiments are undertaken. The QSAR models developed can be used for filtering databases of xenobiotics and identifying those POPs that might be preferentially taken up by zucchini. The presence of highly accumulating POPs in a soil, identified by a QSAR model, might alert zucchini growers to possible contamination of their product by these pollutants.

Some zucchini (*Cucurbita pepo* subspecies *pepo*) cultivars accumulate relatively higher levels of POPs, including PCDDs, PCDFs, and PCBs, than other plant species, rendering them

†Part of the Symposium on Pesticide Toxicology in Honor of Professor John Casida.

Special Issue: Casida Symposium

Received: June 17, 2010
Accepted: December 2, 2010
Revised: December 1, 2010
Published: January 18, 2011

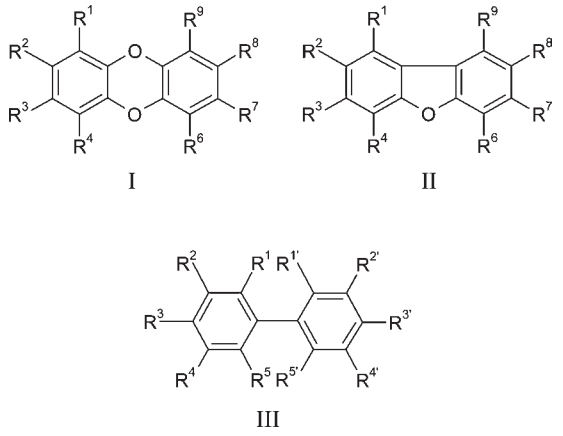

Figure 1. Structures of polychlorinated dioxins (PCDDs, I), dibenzofurans (PCDFs, II), and biphenyls (PCBs, III).

valuable as possible candidates for soil remediation purposes. Inui et al.[8] have studied the differential uptake and reported the experimental BCF values of a set of 20 PCDDs/PCDFs and 14 PCBs (Figure 1) in two zucchini cultivars (*C. pepo* subsp. *pepo* cv. `Black Beauty' (BB) and cv. `Gold Rush' (GR)) as well as in a non-zucchini cultivar squash (*C. pepo* subsp. *ovifera* cv. `Patty Green' (PG)).

According to Inui et al.[8] there are four critical stages for the uptake, translocation, and accumulation of hydrophobic compounds from soil in plants: (i) desorption of hydrophobic compounds from soil particles into soil pore-water, (ii) absorption into roots, (iii) translocation into aerial parts, and (iv) metabolic stability in plants. It has been suggested by Inui et al.[8] that the release of POPs from soil particles is the key step for uptake into roots. The mechanisms underlying the difference in accumulation capabilities of diverse plant species is not understood.

These authors found that (i) the BCFs for PCDD and PCDF congeners negatively correlated with the log*P* values in all three subspecies; (ii) the BCFs for PCBs in BB and GR did not correlate with log*P*; (iii) in contrast, the BCFs for PCBs in PG significantly correlated with log*P* (log*P* denotes the partition coefficient of the substance in 1-octanol/water system). The authors concluded that the species with high accumulating capabilities had unknown, unique mechanisms for the uptake of PCBs, whereas PCDDs and PCDFs were taken up on the basis of their physicochemical properties.

In this study, we have performed a retrospective QSAR study using the BCF values of 20 PCDDs and PCDFs as well as 14 PCBs in two zucchini varieties based on their BCF values reported by Inui et al.[8] (Table 1). The objective of this study was the identification of theoretical quantitative descriptors that can be used for modeling and predicting the BCFs of POPs of types I, II, and III in BB and GR (BCF_{BBGR}).

COMPUTATIONAL METHODS

Structures of the 34 POPs for the calculation of DRAGON and VolSurf descriptors were built manually in the SYBYL v8.0 suite of molecular modeling programs (SYBYL 8.0; Tripos Inc., St. Louis, MO). In the next step molecular mechanics geometry optimization was performed using the Tripos force field with Gasteiger–Hückel charges and the conjugate gradient method, with a gradient of 0.01 kcal mol^{-1} $Å^{-1}$ as termination criterion. Full geometry optimizations and the calculation of the semiempirical quantum chemical descriptors were performed with the AM1 method included in MOPAC of SYBYL using the "precise" option. SYBYL runs under the Irix 6.5 operating system implemented on a Silicon Graphics Octane2 R12000 workstation.

Table 1. Bioconcentration Factors (BCF_{BBGR}) of the POPs Studied and Their Negative Logarithm Values ($pBCF_{BBGR}$)

compound	BCF_{BBGR}[a]	$pBCF_{BBGR}$
dioxins		
368TCDD	6.57×10^{-2}	1.183
1379TCDD	3.16×10^{-2}	1.500
2378TCDD	1.39×10^{-2}	1.858
12378PCDD	2.05×10^{-2}	1.688
123478HCDD	7.82×10^{-3}	2.107
123678HCDD	9.36×10^{-3}	2.029
123789HCDD	9.91×10^{-3}	2.004
1234678HCDD	2.23×10^{-3}	2.652
OCDD	2.16×10^{-3}	2.666
dibenzofurans		
1278TCDF	1.40×10^{-2}	1.853
2378TCDF	3.98×10^{-2}	1.400
12378PCDF	1.86×10^{-2}	1.732
23478PCDF	1.52×10^{-2}	1.818
123478HCDF	6.43×10^{-3}	2.192
123678HCDF	6.47×10^{-3}	2.189
123789HCDF	2.18×10^{-3}	2.661
234678HCDF	5.56×10^{-3}	2.255
1234678HCDF	2.04×10^{-3}	2.691
1234789HCDF	1.01×10^{-3}	2.998
OCDF	2.09×10^{-4}	3.680
biphenyls		
33′44′TCB	1.35×10^{-2}	1.870
344′5TCB	1.55×10^{-2}	1.810
33′44′5PCB	2.23×10^{-2}	1.651
33′44′55′HCB	8.24×10^{-3}	2.084
233′44′PCB	1.03×10^{-1}	0.986
2344′5PCB	7.17×10^{-2}	1.145
23′44′5PCB	1.21×10^{-1}	0.918
2′344′5PCB	8.91×10^{-2}	1.050
233′44′5HCB	4.76×10^{-2}	1.322
233′44′5′HCB	4.76×10^{-2}	1.322
23′44′55′HCB	1.76×10^{-2}	1.754
233′44′55′HCB	1.09×10^{-2}	1.961
22′33′44′5HCB	2.66×10^{-2}	1.575
22′344′55′HCB	2.95×10^{-2}	1.530

[a] Bioconcentration factors (BCF_{BBGR}) reported by Inui et al.[8]

The DRAGON descriptors were calculated by means of the software package DRAGON v5.4, available on the Web,[9] providing a diverse set of 1660 molecular descriptors.[10,11]

The DRAGON GETAWAY descriptors[12−14] are based on the representation of molecular geometry in terms of an influence matrix (H-GETAWAY) (eq 1) or influence−distance matrix (R-GETAWAY). The molecular influence matrix (**H**) is defined as

$$\mathbf{H} = \mathbf{M} \cdot (\mathbf{M}^{T} \cdot \mathbf{M})^{-1} \cdot \mathbf{M}^{T} \quad (1)$$

where **M** is the molecular matrix constituted by the centered Cartesian coordinates and the superscript T indicates the transposed matrix.

The Moriguchi (Mlog*P*)[15] and the Ghose–Crippen (Alog*P*)[16] octanol–water partition coefficients, derived by atom-based prediction algorithms, were among the DRAGON descriptors.

Table 2. Descriptors Showing $r^2 \geq 0.65$ Correlation with BCF_{BBGR}

type	descriptor	definition	r^2 with $pBCF_{BBGR}$
DRAGON descriptors			
connectivity indices[19]	IVDE	mean information content on the vertex degree equality	0.778
3D-MoRSE descriptors[20]	Mor03m	3D-MoRSE − signal 03/weighted by atomic masses	0.688
	Mor03v	3D-MoRSE − signal 03/weighted by atomic van der Waals volumes	0.684
WHIM descriptors[21]	L2u	second component size directional WHIM index/unweighted	0.712
	L2v	second component size directional WHIM index/weighted by atomic van der Waals volumes	0.689
	L2e	second component size directional WHIM index/weighted by atomic Sanderson electronegativities	0.749
	L2p	second component size directional WHIM index/weighted by atomic polarizabilities	0.670
GETAWAY descriptors			
	H4v	H autocorrelation of lag 4/weighted by atomic van der Waals volumes	0.660
	H5v	H autocorrelation of lag 5/weighted by atomic van der Waals volumes	0.645
	H6v	H autocorrelation of lag 6/weighted by atomic van der Waals volumes	0.684
	H4p	H autocorrelation of lag 4/weighted by atomic polarizabilities	0.725
	H5p	H autocorrelation of lag 5/weighted by atomic polarizabilities	0.661
	H6p	H autocorrelation of lag 6/weighted by atomic polarizabilities	0.683
VolSurf descriptors			
hydrophobic probes	D4DRY	calculated at −0.8 kcal/mol energy level	0.685
	D6DRY	calculated at −1.2 kcal/mol energy level	0.704
	D7DRY	calculated at −1.4 kcal/mol energy level	0.693
	D8DRY	calculated at −1.6 kcal/mol energy level	0.676
quantum chemical descriptor			
	LUMO energy	lowest unoccupied molecular orbital energy	0.695

The VolSurf descriptors[17] were calculated using the VolSurf program module of SYBYL. VolSurf is a computational program to generate 2D molecular descriptors from 3D molecular interaction energy grid maps. The basic idea of VolSurf is to compress the information present in 3D maps into a few 2D numerical descriptors that are simple to understand and to interpret. VolSurf descriptors are specifically designed for the in silico optimization of ADME properties governing the absorption, distribution, metabolism, and excretion of chemical substances in humans. These descriptors quantitatively characterize the size, shape, polarity, and hydrophobicity of molecules as well as the balance between them. All of the available GRID probes have been used in the calculation of the VolSurf descriptors. These were the water probe (OH2); the hydrophobic probe (DRY); the amphipatic probe (BOTH); the carbonyl oxygen atom probe (O); anionic sp2 carboxy oxygen atom probe (O::); neutral flat NH (e.g., amide) probe (N1); sp2 nitrogen with one lone pair probe (N:=); sp3 amine NH3 cation probe and (N3+); and sp2 phenolate oxygen atom probe (O^-). Grid spacing was set to 0.5 Å.

Forward stepwise regression analysis of the data set was performed using STATISTICA v6.1 software (StatSoft, Inc., 2003) running on a PC (Intel(R) Core(TM)2 Duo, 2.33 GHz CPU; Microsoft Windows XP operating system).

Partial least-squares (PLS) analysis of latent variables was carried out using the PLS module in SYBYL v8.0. PLS regression analysis is based on linear combination reducing a large number of original descriptors to a small number of orthogonal factors (latent variables) providing the optimal linear model in terms of predictivity. Predictive capability of the model is quantified by cross-validation using the leave-one-out (LOO) cross-validation technique (q^2). The number of accepted latent variables was based on the first local maximum of q^2.

Variable selection from the set of 1660 DRAGON descriptors has been carried out as follows: variables with zero variance and the ones that showed $r^2 \leq 0.3$ coefficient of determination with pBCF have been omitted, yielding a subset of 467 descriptors. For the set of 150 VolSurf descriptors and the 11 quantum chemical descriptors, calculated using SYBYL MOPAC, the cutoff value of the coefficient of determination with pBCF was set to $r^2 \leq 0.2$, yielding 51 and 4 descriptors, respectively (Var = 0 was not found). Variable selection resulted in a set of 522 descriptors including 467 DRAGON + 51 VolSurf + 4 semiempirical quantum chemical parameters. Table 2 shows the descriptors displaying $r^2 \geq 0.65$ correlation with BCF_{BBGR}. In the next step, cross-correlation between the 522 descriptors was calculated, and from the pairs that showed $r^2 > 0.9$ correlation, one of the descriptors has been discarded. Finally, a set of 496 descriptors has been selected and used in the present modeling study. The BCF data for BB and GR showed high correlation and small standard deviation ($r^2 = 0.982$, $s = 0.0037$), indicating that the experimental error of the BCF values must have been very small.

In our study the average of the BB and GR data (BCF_{BBGR}) and the logarithm of their reciprocal values ($pBCF_{BBGR}$) were used as

Table 3. Relative Contributions in Equations 2−4 and r^2, q^2_{LOO}, and F Values of the Descriptors Included

	contribution (%)				correlation with $pBCF_{BBGR}$ calculated for 34 compounds		
	set 1		set 2				
descriptor	eq 2	eq 3	eq 4	eq 5	r^2	q^2	F
VolSurf							
VOH2	10.5				0.264	0.165	11.5
W4O	26.0	26.3			0.423	0.361	23.4
D3DRY	21.9				0.623	0.549	52.9
D6DRY				39.7	0.704	0.646	76.3
BV31OH2		24.2	20.7	22.6	0.319	0.205	11.5
HB5O::			20.6				
DRAGON GETAWAY							
H4p	41.6	32.3	33.5		0.725	0.678	84.6
H5e				22.5	0.373	0.284	19.0
quantum chemical							
LUMO energy					0.695	0.646	72.9
hardness			25.2		0.453	0.368	27.5
Z-component of dipole moment		17.2		15.2	0.229	0.121	9.5

dependent variables. Two sets of 27 compounds each (sets 1 and 2) were generated by randomly omitting 7−7 compounds from the complete set of 34 compounds and used for model building. The two 7-compound sets were used for external validation of the regression equations developed.

RESULTS AND DISCUSSION

The calculations yielded statistically significant linear regression models for the $pBCF_{BBGR}$ values of the PCDD, PCDF, and PCB derivatives shown in Table 1. Stepwise regression analyses based on two 27 × 496 data matrices yielded four regression models, eqs 2 and 3 for set 1 and eqs 4 and 5 for set 2. The stepwise regression analysis has been repeated several times, each time leaving out the next descriptor with the highest contribution to the equation that could not be interpreted. In the next step PLS analysis of the selected descriptor set of the final regression models has been performed, yielding leave-one-out (LOO) cross-validated correlation coefficients (q^2_{LOO}). The following representative models (eqs 2−5) have been selected on the basis of the statistical parameters of the equations and the interpretability of the descriptors.

$$pBCF_{BBGR} = 7.356H4p - 0.002VOH2 + 0.016D3DRY + 0.072W4O - 0.320 \quad (2)$$

$$n = 27, R^2 = 0.921, q^2 = 0.886, SE = 0.186, F = 89.4$$

$$\text{external validation set: } n = 7, R^2 = 0.831, q^2 = 0.563, SE = 0.273, F = 24.5$$

$$pBCF_{BBGR} = 5.519H4p + 0.092BV31OH2 + 0.070W4O + 0.257Z\text{-component} - 0.175 \quad (3)$$

$$n = 27, R^2 = 0.940, q^2 = 0.922, SE = 0.155, F = 392.1$$

$$\text{external validation set: } n = 7, R^2 = 0.739, q^2 = 0.477, SE = 0.338, F = 14.2$$

$$pBCF_{BBGR} = 5.347H4p + 0.077BV31OH2 - 0.017HB5O:: - 0.588(LUMO-HOMO) + 4.869 \quad (4)$$

$$n = 27, R^2 = 0.921, q^2 = 0.898, SE = 0.161, F = 140.4$$

$$\text{external validation set: } n = 7, R^2 = 0.826, q^2 = 0.530, SE = 0.388, F = 23.7$$

$$pBCF_{BBGR} = 0.043D6DRY + 1.253H5e + 0.084BV31OH2 + 0.197Z\text{-component} - 1.214 \quad (5)$$

$$n = 27, R^2 = 0.918, q^2 = 0.880, SE = 0.164, F = 134.3$$

$$\text{external validation set: } n = 7, R^2 = 0.921, q^2 = 0.795, SE = 0.261, F = 58.6$$

The values of the determination coefficient (R^2) for eqs 2−5 are above 0.9, so these models explain >90% of the variance of the pBCF. All equations show excellent internal predictive capabilities displaying $q^2_{LOO} \approx 0.9$ values as calculated by PLS analysis.

Equation 2 contains three VolSurf (VOH2, W4O, and D3DRY) and one DRAGON GETAWAY (H4p) descriptor; eq 3 contains two VolSurf (W4O and BV31OH2), one DRAGON GETAWAY (H4p), and one quantum chemical descriptor (Z-component); eq 4 contains two VolSurf (BV31OH2 and HB5O::), one DRAGON GETAWAY (H4p), and one quantum chemical descriptor (LUMO−HOMO); eq 5 contains two VolSurf (D6DRY and BV31OH2), one DRAGON GETAWAY (H5e), and one quantum chemical descriptor (Z-component). The relative contributions, as well as the determination coefficients (r^2), q^2, and F values with $pBCF_{BBGR}$ of these descriptors for the complete set of 34 compounds are shown in Table 3.

Definitions of the selected descriptors are the following: VOH2, molecular volume given as water solvent excluded volume (in Å^3), that is, the volume contained within the water-accessible surface of the molecule computed at 0.20 kcal/mol; W4O, calculated at −2.0 kcal/mol energy level with carbonyl

Table 4. Validation Set Compounds and the Differences between Their Experimental and Predicted Values (Residuals) Calculated by the Regression Equations (Equations 2−5)

		residuals	
	validation set compound	eq 2	eq 3
set 1	1368TCDD	−0.404	−0.381
	1234678HCDD	0.424	0.499
	12378PCDF	−0.227	−0.267
	1234678HCDF	−0.088	−0.130
	344′5TCB	0.092	0.239
	233′44′5HCB	−0.072	−0.047
	22′344′55′HCB	−0.140	−0.223
		residuals	
	validation set compound	eq 4	eq 5
set 2	1379TCDD	−0.013	0.056
	123789HCDD	−0.084	−0.006
	12378PCDF	−0.294	−0.260
	OCDF	0.419	0.571
	33′44′TCB	0.425	0.365
	23′44′5PCB	−0.476	−0.201
	233′44′55′HCB	0.408	0.063

oxygen probe atom bonded to a single other atom, which accepts two hydrogen bonds in the direction of its lone pairs; this descriptor represents the molecular envelope accessible by the probe atom; D3DRY and D6DRY, hydrophobic descriptors calculated at −0.6 and −1.2 kcal/mol energy levels, respectively; BV31OH2, best volume descriptor representing one of the best hydrophilic volumes generated by a water probe calculated at −1 kcal/mol energy level; HB5O::, a hydrogen bond descriptor calculated with carboxyl oxygen probe; H4p, GETAWAY descriptor H autocorrelation of lag 4/weighted by atomic polarizabilities; H5e, GETAWAY descriptor H autocorrelation of lag 5/weighted by atomic Sanderson electronegativities; (LUMO−HOMO) (chemical hardness), quantum chemical descriptor, the difference of LUMO and HOMO energies, where LUMO and HOMO refer to the energy of the lowest unoccupied and highest occupied molecular orbitals, respectively;[18] Z-component, Z-component of the dipole moment; VolSurf Log*P*, directly calculated according to internal equations by the program.

The predictive power of the regression equations (eqs 2−5) has been tested using two sets of randomly selected compounds, each containing seven substances, for external validation. Table 4 shows the validation set compounds and the differences between their experimental and predicted values (residuals) calculated by the regression equations (eqs 2−5). The magnitude of the residuals for the external validation set compounds remained close to the range of the standard deviation (SE) calculated by the regression equations for the compounds in sets 1 and 2. The magnitude of the residuals indicated that the $pBCF_{BBGR}$ values of POPs not used in model building can be predicted with reasonable accuracy. Figure 2 shows the plot of the calculated versus experimental $pBCF_{BBGR}$ values for training set (open circles) and external validation set (solid circles) compounds for set 1 calculated using eq 2.

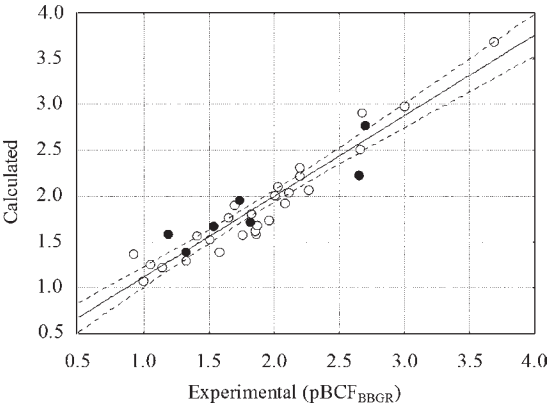

Figure 2. Calculated versus experimental $pBCF_{BBGR}$ values for training set (○) and external validation set (●) compounds for set 1 calculated using eq 2. Dashed lines denote 95% confidence interval.

Table 5. Correlations (r^2) between BCF_{BBGR} of POPs and Some of Their Molecular Descriptors Related to Hydrophobicity[a]

	r^2			
POP	N_{Cl}	MLog*P*	ALog*P*	VolSurf Log*P*
dioxins	(+) 0.833	(+) 0.832	(+) 0.833	(+) 0.733
dibenzofurans	(−) 0.872	(−) 0.866	(−) 0.872	(−) 0.846
dioxins and benzofurans	(−) 0.793	(−) 0.511	(−) 0.827	(−) 0.065
biphenyls	(+) 0.036	(−) 0.068	(+) 0.036	(+) 0.088
all	(+) 0.391	(−) 0.026	(+) 0.363	(+) 0.078

[a] (+) and (−) symbols denote the sign of the slope of the regression equations.

So far, only a few QSAR studies on the BCF values of POPs in plants have been published,[22] and this is the first study to use VolSurf descriptors for modeling the ADME behavior of POPs in a plant.

In classical QSAR analysis studies of ADME phenomena the relationship between lipophilicity (usually log*P*) and the experimental dependent variable (e.g., BCF) is routinely examined. In this study the log*P* values of the compounds were estimated by three different methods and correlated with the BCF_{BBGR} values of the individual PCDD, PCDF, and PCB compound sets and the united set of the PCDDs and PCDFs, as well as the complete set of all 34 compounds. The calculated correlations with log*P* values are shown in Table 5.

The BCF_{BBGR} values of PCDDs and PCDFs separately and together show certain correlations with their log*P* values and the number of Cl atoms in the molecule (N_{Cl}) (Table 5). In contrast, neither the PCBs nor the united set of all 34 compounds shows statistically significant correlation with log*P*.

Pavan et al.[6] gave a comprehensive review of QSAR models reported for bioconcentration. Typically, fish are the target organisms of BCF assessments due to the importance of fish as a human food source and the availability of standardized testing protocols. BCF models on fish can be based on experimentally derived and theoretical molecular descriptors. In general, most of the QSAR models reported for the prediction of BCF in regulatory context are based on the correlation of logBCF with $\log K_{OW}$ (log*P*). In the European Union Technical Guidance Document (TGD) on Risk Assessment,[23] the QSAR models

suggested for estimating the bioconcentration factor for aquatic organisms are based on the relationship between logBCF and logK_{OW}. For terrestrial plants no such guidance is available.

The most widely used expert system for BCF prediction, BCFWIN, developed by Syracuse Research Corp., is freely downloadable from the U.S. Environmental Protection Agency (EPA) Web site.[24] It incorporates the QSARs developed by Meylan et al.[25]

The five non-VolSurf descriptors selected by us for modeling of the BCF_{BBGR} values of POPs for the uptake and bioaccumulation to zucchini were H4p, H5e LUMO energy, (LUMO−HOMO), and Z-component. Some of these descriptors have already been used in QSAR studies for modeling of BCF values in fish.[26,27]

LUMO is the lowest energy level in the molecule that contains no electrons. When a molecule acts as a Lewis acid (an electron-pair acceptor) in bond formation, incoming electron pairs are received in its LUMO. Molecules with low LUMO energies are more able to accept electrons than those with high LUMO energies; thus, the LUMO descriptor should measure the electrophilicity of a molecule. It is important in governing molecular reactivity and properties.[28]

H5e is one of the GETAWAY descriptors that have been proposed with the aim of matching 3D molecular geometry, atom relatedness, and chemical information. The GETAWAY descriptors are extensively used for QSPR and BCF modeling[29] of diverse compounds.

The interpretation of VolSurf descriptors is relatively straightforward. The local values of the VolSurf descriptors on the 3D structure of a molecule can be visualized by color code contours. These contours highlight certain interaction capabilities of the molecules in 3D that have significant importance in explaining the variance in the dependent variable modeled, that is, $pBCF_{BBGR}$. Polarizability, hydrophobic effects, solvent-excluded volume, hydrophobic volume, and hydrogen bonding seem to be all important molecular properties governing molecular interactions of the POPs in the plant, and as such influencing uptake and accumulation. In our case the W4O descriptor accounts for polarizability; D3DRY and D6DRY represent hydrophobic energy calculated with the DRY probes; VOH2 quantifies molecular volume given as water solvent excluded volume (Å^3), that is, the volume contained within the water accessible surface calculated at a certain energy level (kcal/mol); BV31OH2 is the "best volume" descriptor representing the local hydrophobic volumes calculated at a certain energy level (kcal/mol); and HB5O:: is a descriptor for hydrogen bonding, which is one of the most important intermolecular interaction types in living matter. The regression equations developed (eqs 2−5) containing these descriptors can be used as filters for identifying those compounds in compound libraries of any size that have the same molecular interaction properties in 3D as the training set POPs used for developing the model equations.

In conclusion, the regression models (eqs 2−5) developed for the uptake and accumulation of representative PCDDs, PCDFs, and PCBs (Figure 1) in zucchini are statistically highly significant. Predictive capabilities of the models have been tested using two external validation sets showing that the $pBCF_{BBGR}$ values of POPs not used in model building can be predicted with reasonable accuracy. The regression models revealed that the POPs taken up preferentially from soil by the two zucchini subspecies (BB and GR) are characterized by low H4p, high VOH2, low D3DRY, low W4O, low BV31OH2, low Z-component, high HB5O::, high LUMO−HOMO, low D6DRY, and low H5e descriptor values (see eqs 2−5). VolSurf descriptors, originally developed for modeling pharmacokinetic properties of drugs in mammals, were found to be useful also in modeling the ADME properties of POPs of types I−III in plants. The use of these descriptors for filtering compound libraries may help in identifying chemical substances preferentially taken up by the zucchini plants. We suggest that in a wider context the use of these types of chemical descriptors might serve an alerting purpose by identifying those environmental contaminants that might be preferentially taken up from the soil by certain plant species. We suggest that the present study contributes to the development of viable soil remediation strategies.

AUTHOR INFORMATION

Corresponding Author

*Phone: +36 1 4877553; fax: +36 1 4877555; e-mail: barna.bordas@gmail.com.

Funding Sources

This study was financially supported by the National Office for Research and Technology (NKTH) and the Hungarian Scientific Research Fund (OTKA) No. K 67737, as well as the Economic Competitiveness Operative Program, Project GVOP-3.2.1-2004-04-0068/03.

REFERENCES

(1) Jones, K. C.; de Voogt, P. Persistent organic pollutants (POPs): state of the science. *Environ. Pollut.* **1999**, *100*, 209–221.

(2) Boethling, R.; Fenner, K.; Howard, P.; Klecka, G.; Madsen, T.; Snape, J. R.; Whelan, M. J. Environmental persistence of organic pollutants: guidance for development and review of POP risk profiles. *Integr. Environ. Assess. Manag.* **2009**, *5*, 539–556.

(3) McLachlan, M. S. Bioaccumulation of hydrophobic chemicals in agricultural food chains. *Environ. Sci. Technol.* **1995**, *30*, 252–259.

(4) Galiulin, R. V.; Bashkin, V. N.; Galiulina, R. A. Review: Behavior of persistent organic pollutants in the air−plant−soil system. *Water, Air Soil Pollut.* **2002**, *137*, 179–191.

(5) Collins, C.; White, J. C.; Rock, S. Plant uptake of organic chemicals: current developments and recommendation for future research. *Environ. Toxicol. Chem.* **2007**, *26*, 2465–2466.

(6) Pavan, M.; Worth, A. P.; Netzeva, T. I. *Review of QSAR Models for Bioconcentration*; EUR 22327 EN; European Commission, Directorate − General, Joint Research Centre, 2006.

(7) Uegaki, R.; Seike, N.; Otani, T. Polychlorinated dibenzo-*p*-dioxins, dibenzofurans, and dioxin-like polychlorinated biphenyls in rice plants: Possible contaminated pathways. *Chemosphere* **2006**, *45*, 1537–1543.

(8) Inui, H.; Wakai, T.; Gion, K.; Kim, Y.-S.; Eun, H. Differential uptake for dioxin-like compounds by zucchini subspecies. *Chemosphere* **2008**, *73*, 1602–1607.

(9) Todeschini, R.; Consonni, V.; Pavan, M. DRAGON. Software for the calculation of molecular descriptors, version 5, 2005; available at http://www.vcclab.org/lab/edragon.

(10) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors, Methods and Principles in Medicinal Chemistry*; Mannhold, R., Kubinyi, H., Timmerman, H., Eds.; Wiley-VCH: Weinheim, Germany, 2000; Vol. 11.

(11) Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V. A.; Radchenko, E. V.; Zefirov, N. S.; Makarenko, A. S.; Tanchuk, V. Y.; Prokopenko, V. V. Virtual computational chemistry laboratory − design and description. *J. Comput.-Aid. Mol. Des.* **2005**, *19*, 453–463.

(12) Consonni, V.; Todeschini, R. *Rational Approaches to Drug Design*; Hottje, H.-D. , Sippl, W. , Eds.; Prous Science: Barcelona, Spain, 2001; pp 235−240.

(13) Consonni, V.; Todeschini, R.; Pavan, M. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. Part 1. Theory of the novel 3D molecular descriptors. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 682–692.

(14) Consonni, V.; Todeschini, R.; Pavan, M.; Gramatica, P. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 693–705.

(15) Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. Simple method of calculating octanol water partition-coefficient. *Chem. Pharm. Bull.* **1992**, *40*, 127–130.

(16) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic physicochemical parameters for 3 dimensional structure directed quantitative structure−activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their applications for an automated superposition of certain naturally-occurring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172.

(17) Cruciani, G.; Pastor, M.; Guba, W. VolSurf, a new tool for the pharmacokinetic optimization of lead compounds. *Eur. J. Pharm. Sci.* **2000**, *11*, S29–S39.

(18) Parr, R. G.; Pearson, R. G. Absolute hardness: companion parameter to absolute electronegativity. *J. Am. Chem. Soc.* **1983**, *105*, 7512–7516.

(19) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure−Activity Analysis*; RSP-Wiley: Chichester, U.K., 1986.

(20) Schuur, J.; Gasteiger, J. *Software Development in Chemistry*; Gasteiger, J., Ed.; Fachgruppe Chemie-Information-Computer (CIC): Frankfurt am Main, Germany, 1996; Vol. 10, pp 67−80.

(21) Todeschini, R.; Lasagni, M. New molecular descriptors for 2D and 3D structures. *Theor. J. Chemom.* **1994**, *8*, 263–273.

(22) McKone, T. E.; Maddalena, R. L. Plant uptake of organic pollutants from soil: bioconcentration estimates based on models and experiments. *Environ. Toxicol. Chem.* **2007**, *26*, 2494–2504.

(23) TGD European Commission. Technical Guidance Document on Risk Assessment in support of Commission Directive 93/67/EEC on Risk Assessment for new notified substances, Commission Regulation (EC) No 1488/94 on Risk Assessment for existing substances, and Directive 98/8/EC of the European Parliament and of the Council concerning the placing of biocidal products on the market, 2003.

(24) Exposure Assessment Tools and Models, U.S. Environmental Protection Agency, http://www.epa.gov/oppt/exposure/pubs/episuite.htm.

(25) Meylan, W. M.; Howard, P. H.; Boethling, R. S.; Aronson, D.; Printup, H.; Gouchie, S. Improved method for estimating bioconcentration/bioaccumulation factor from octanol/water partition coefficient. *Environ. Toxicol. Chem.* **1999**, *18*, 664–672.

(26) Wei, D.; Zhang, A.; Wu, C.; Han, S.; Wang, L. Progressive study and robustness test of QSAR model based on quantum chemical parameters for predicting BCF of selected polychlorinated organic compounds (PCOCs). *Chemosphere* **2001**, *44*, 1421–1428.

(27) Gramatica, P.; Papa, E. QSAR modelling of bioconcentration factor by theoretical molecular descriptors. *QSAR Comb. Sci.* **2003**, *22*, 374–385.

(28) Honório, K. M.; Da Silva, A. B. F. An AM1 study on the electron-donating and electron-accepting character of biomolecules. *Int. J. Quantum Chem.* **2003**, *95*, 126–132.

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