Bull Environ Contam Toxicol (2011) 87:134–137
DOI 10.1007/s00128-011-0316-x

Toxicity Prediction of Dioxins and Dioxin-like Compounds Based on the Molecular Fragments Variable Connectivity Index

Qiang Chen · Jingmin Sun · Jing Liu

Received: 6 March 2011 / Accepted: 13 May 2011 / Published online: 28 May 2011

Abstract The toxicity of 95 doxins and dioxin-like compounds was investigated by quantitative structure–activity relationship (QSAR) with the molecular fragments variable connectivity index (mfVCI). For each of the four homologues, the models have good fitting ($R^2 > 0.80$) and predictive ($Q^2_{EXT} > 0.80$) ability. The models developed from more than one homologues are also satisfactory with $R^2 > 0.80$ and $Q^2_{EXT} > 0.77$. The molecular fragments weights have the ability to diagnose the contribution of the molecular fragments to the toxicity of the compounds. The mfVCI may play an important role in the development of molecular descriptors in further QSAR research.

Keywords Quantitative structure–activity relationship (QSAR) · Molecular fragments variable connectivity index (mfVCI)

Dioxins and dioxin-like compounds, often coming from combustion or the industrial uses or synthesis of other chemicals, such as polychlorinated dibenzo-*p*-dioxins and dibenzo-*p*-furans (PCDD/Fs) and polychlorinated biphenyls (PCBs), are typical persistent organic pollutants (Quaß et al. 2004). They are widespread toxic pollutants in soil, sediment, air and water (Jones and de Voogt 1999) through atmospheric transport, and in other ways (Lohmann et al. 2007). Most of dioxins are toxic to human via exposure pathways, such as dermal contact, inhalation and consumption of contaminated food (Hays and Aylward 2003). The toxicity of dioxins largely depends on the number and position of the chlorine atoms (Haws et al. 2006). The compounds must have at least three or four halogen atoms in the lateral positions and have at least one hydrogen atom (Zhao et al. 2008). Generally speaking, the toxicity of dioxins decreases with the addition of nonlateral chlorines or removal of lateral chlorines from their structures (Safe 1990; Zhao et al. 2008). Therefore, quantitative toxicity prediction of dioxins and dioxin-like compounds would be helpful in assessing the threat that these compounds pose to human beings and the ecosystems.

Quantitative structure–activity relationships (QSARs) have been widely used to predict the toxicity and biological activity of dioxins (Ashek et al. 2006; Papa et al. 2008). Constructing suitable structural descriptors relevant to the variation of different activity is the critical process in QSAR research. Topological indices of all compounds can be easily obtained from topological structural information without experiments. Significant progress in the development of topological indices was made during the QSAR studies (Randić 1975, 1991, 2001; Pompe 2005), but there are still some limitations and drawbacks (Randić 2001; Pompe and Randić 2007). The molecular fragments variable connectivity index (mfVCI) was proposed for the toxicity (*Vibrio fischeri* pT_{50}) prediction of substituted-benzenes (Chen et al. 2009). The mfVCI combines the advantage of topological index with molecular fragment. The molecular fragments were used to distinguish the positive and negative contributions to the toxicity. In this study, the mfVCI was applied to the toxicity prediction of dibenzo-*p*-dioxins, dibenzofurans, biphenyls and naphthalene. The aim of this study is providing basic method for the toxicity evaluation and toxicological mechanism research.

Q. Chen (✉) · J. Sun · J. Liu
College of Atmospheric Sciences, Lanzhou University,
222 Tianshui South Road, Lanzhou 730000,
People's Republic of China
e-mail: chenqqh@163.com

Materials and Methods

The toxicity data (pEC_{50}) of the 95 dioxins and dioxin-like compounds were collected from Ashek et al. (2006). The data are the molar concentration of chemical necessary to displace 50% of radio labeled TCDD from the Ah receptor.

The molecular fragments are defined as the molecular structure units which have different characteristics due to different chemical bonding for the major influence of toxicity. The sorts of the fragments are illustrated in Table 1 for the case of 2,3,7,8-Tetrachlorodibenzo-*p*- dioxin.

The fragments were given different optimized weights (w_i) respectively. The valence of molecular fragment (δ_i^{mf}) is calculated as the sum of Randić valence (δ_i) and the fragment weight (w_i). Then, the molecular fragments connectivity index of order 1 ($^1\chi^{mf}$) is expressed as follows:

$$^1\chi^{mf} = \sum_{p=1}^{m} {}^1\chi_p^{mf} = \sum_{p=1}^{m} \prod_{i=1}^{2} \left(\delta_i^{mf}\right)_p^{-0.5} \quad (1)$$

where $^1\chi_p^{mf}$ is the contribution of the adjacent vertices connected by one edge to the connectivity index. In this study, the linear relationship was assumed between mfVCI and the toxicity, which is $pEC_{50} = a \times {}^1\chi^{mf} + b$. The $^1\chi^{mf}$ was optimized by using *solver* in Microsoft Office Excel. The target function is the root mean calculated residual sum of squares for training set.

The fitting ability of the QSAR model is mainly characterized by determination coefficient (R^2), standard error (SE) and Fischer variance ratio (F). Leave-one-out and leave-many-out methods with internal cross-validated correlation coefficients (Q^2_{LOO} and Q^2_{LMO}) are used to evaluate the model's predictive ability and robustness. The external predictive power of the model is validated by external cross-validation correlation coefficient (Q^2_{EXT}).

Results and Discussion

The model for the 25 dibenzo-*p*-dioxins is as follows:

$$pEC_{50} = 0.152\,{}^1\chi^{mf} - 24.72$$
$$R^2 = 0.948,\ SEE = 0.328,\ F = 422.6 \quad (2)$$

It can be seen that the model has good fitting ability. According to the algorithm of the first order of mfVCI Eq. (1), the contribution of the adjacent two vertices connected by one edge to the index of $^1\chi^{mf}$ is expressed as follows:

$$^1\chi_p^{mf} = \frac{1}{\sqrt{\delta_i^{mf}\delta_j^{mf}}} = \frac{1}{\sqrt{(\delta_i + w_i)(\delta_j + w_j)}} \quad (3)$$

where i and j are the two adjacent molecular fragments connected by edge p. For each compound of the dibenzo-*p*-dioxins, the part C_2–C_3 or C_7–C_8 is the most important contributor to the pEC_{50}, which indicates that the sites 2,3,7,8 are the key positions that affect the toxicity of dibenzo-*p*-dioxins. For the central symmetrical structure of the compounds, only the values of part C_2–C_3 ($^1\chi^{mf}_{C_2-C_3}$) were analyzed. The values of part C_2–C_3 are different when the hydrogen atom at 1 and/or 4 sites is substituted by chlorine or bromine atom (Table 2).

The toxicity of compounds pair which are different only in structures at site 1 or 4 is compared. For example, the $^1\chi^{mf}_{C_2-C_3}$ of compound 1 (44.32) is greater than that of compounds 2 and 3 (35.54 and 30.21), as the hydrogen atoms at site 1 of compound 2 and site 6 (the equivalent of site 4) of compound 3 are substituted by chlorine atoms (Table 2). The toxicity of compound 1 ($pEC_{50(1)} = 8.00$) is stronger than that of compounds 2 and 3 ($pEC_{50(2)} = 7.10$, $pEC_{50(3)} = 6.80$). The relations between the toxicity and substitution position of 41 comparable pairs comply with the above mentioned conclusion except for compounds 3 and 9, 8 and 11, 15 and 23, and compounds 19 and 20. Hence, any substitution in lateral position of the ring enhances the toxicity relative to medial substitutions.

The model for 35 dibenzofurans was obtained and showed as follows:

$$pEC_{50} = 16708.5\,{}^1\chi^{mf} - 87802.5$$
$$R^2 = 0.829,\ SEE = 0.571,\ F = 159.7 \quad (4)$$

For all of the 35 compounds, the maximum value of $^1\chi^{mf}_{C_i-C_j}$ will appear when the hydrogen atom at sites 2,3,7,8 is substituted by chlorine atom. For example, the values of parts C_1–C_2, C_3–C_4, C_6–C_7, C_8–C_9 are the greatest ones of 0.442 when the chlorine atoms are substituted for hydrogen

Table 1 Molecular fragments for 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin

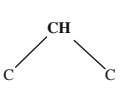
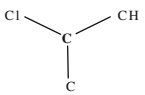
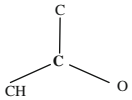
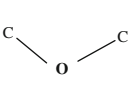
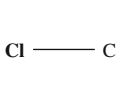
Name	Diagrams				
Molecular Fragments					
Expressions	2(C)***CH***	(C)(CH)***C***(Cl)	(C)(CH)***C***(O)	2(C)***O***	(C)***Cl***
Weights	w_1	w_2	w_3	w_4	w_5

Table 2 The values of part C_2–C_3 ($^1\chi^{mf}_{C_2-C_3}$) for dibenzo-*p*-dioxins

No.	1	2	3	4	5	6	7
C_2, C_3 substituted	C_2, C_3^a	C_2, C_3^b	C_2, C_3^c	C_2, C_3^a	C_2, C_3^a	C_2/C_3^b	C_2/C_3^a
C_1, C_4 substituted	–	–	–	C_1/C_4^a	C_1,C_4^a	–	–
values	44.32	41.92	39.64	35.54	28.5	36.48	36.24

[a] Only chlorine atom substitutes for hydrogen atom

[b] Chlorine and bromine atoms substitute for hydrogen atoms

[c] Only bromine atom substitutes for hydrogen atom

atoms at positions 3,2,8,7. This also indicates that the substituted chlorine atoms at sites 2,3,7,8 play a primary role in the toxicity of dibenzofurans.

The QSAR model for all of the 30 biphenyls based on mfVCI is as follows:

$$\begin{aligned} &\mathrm{pEC}_{50} = 0.006\,{}^1\chi^{mf} - 0.697 \\ &R^2 = 0.963,\ SEE = 0.141,\ F = 721.6 \end{aligned} \tag{5}$$

In order to distinguish the different roles of the molecular fragments, compounds which are different only in the fragment connected with the lateral position (labeled as $C_{4'}$) were selected for analysis (Table 3). Table 3 shows that the calculated pEC_{50} values are in good agreement with the observed values except for compound 75. For all of the 16 compounds, the Randić valence of $C_{4'}$ is 3. The orders of experimental pEC_{50} values and the valence of $C_{4'}$ $(w_i + \delta_i)$ show the same tendency. Hence, the closer the value of the molecular fragment weight gets to the corresponding $-\delta_i$, the greater the molecular fragment contributes to the toxicity.

Here, naphthalene only includes five compounds. The model is as follows:

$$\begin{aligned} &\mathrm{pEC}_{50} = 1.08\,{}^1\chi^{mf} - 0.31 \\ &R^2 = 0.979,\ SEE = 0.134,\ F = 138.0 \end{aligned} \tag{6}$$

So, the model also has good fitting ability.

In order to validate the models' predictive ability, all kinds of homologues (except for naphthalene) were splitted into training set and test set. A good splitting requires that the whole range of both descriptors and activity of compounds in test set are included into the training set (Golbraikh and Tropsha 2002). So, any of compounds in Table 3 is not included in the test set for biphenyls. The results for each of the three homologues are shown in Table 4.

There is a good correlation between the observed and calculated pEC_{50} for all of the compounds ($R^2 > 0.80$). Validation results indicate that the models are robust and have good prediction ($Q^2_{EXT}, Q^2_{LOO}, Q^2_{LMO} > 0.77$) (Papa et al. 2005).

Table 3 pEC_{50}, fragments and weights of $C_{4'}$ for compounds 75–90

ID	Chemical name	pEC_{50}		Fragments	Weights (w_i)	$w_i + \delta_i$
		Exp.	Cal.			
85	4′-Trifluoromethyl-2,3,4,5-tetrachlorobiphenyl	6.43	6.43	2(CH)***C***(CF_3)	−2.9992	0.0008
84	4′-Isopropyl-2,3,4,5-tetrachlorobiphenyl	5.89	5.89	2(CH)***C***($CHCH_3CH_3$)	−2.9987	0.0013
83	4′-Iodo-2,3,4,5-tetrachlorobiphenyl	5.82	5.82	2(CH)***C***(I)	−2.9986	0.0014
82	4′-Bromo-2,3,4,5-tetrachlorobiphenyl	5.60	5.60	2(CH)***C***(Br)	−2.9984	0.0016
81	4′-Ethyl-2,3,4,5-tetrachlorobiphenyl	5.46	5.46	2(CH)***C***(CH_2CH_3)	−2.9979	0.0021
80	4′-Cyano-2,3,4,5-tetrachlorobiphenyl	5.27	5.27	2(CH)***C***(CN)	−2.9971	0.0029
88	4′-Phenyl-2,3,4,5-tetrachlorobiphenyl	5.18	5.18	2(CH)***C***(C_6H_5)	−2.9964	0.0036
89	4′-*t*-Butyl-2,3,4,5-tetrachlorobiphenyl	5.17	5.17	2(CH)***C***($CCH_3CH_3CH_3$)	−2.9963	0.0037
79	4′-Acetyl-2,3,4,5-tetrachlorobiphenyl	5.17	5.17	2(CH)***C***($CH_2CH_2CH_2CH_3$)	−2.9962	0.0038
90	4′-*n*-Butyl-2,3,4,5-tetrachlorobiphenyl	5.13	5.13	2(CH)***C***(C=OCH_3)	−2.9962	0.0038
87	4′-N-Acetylamino-2,3,4,5-tetrachlorobiphenyl	5.09	5.09	2(CH)***C***(NHC=OCH_3)	−2.9958	0.0042
86	4′-Nitro-2,3,4,5-tetrachlorobiphenyl	4.85	4.85	2(CH)***C***(N=OO)	−2.9938	0.0062
78	4′-Methoxy-2,3,4,5-tetrachlorobiphenyl	4.80	4.80	2(CH)***C***(OCH_3)	−2.9924	0.0076
77	4′-Fluoro-2,3,4,5-tetrachlorobiphenyl	4.60	4.60	2(CH)***C***(F)	−2.9853	0.0147
76	4′-Methyl-2,3,4,5-tetrachlorobiphenyl	4.51	4.51	2(CH)***C***(CH_3)	−2.9782	0.0218
75	4′-Hydroxy-2,3,4,5-tetrachlorobiphenyl	4.05	4.11	2(CH)***C***(OH)	70.2512	73.2512

Table 4 Results of the models after splitting the data into two sets

Compounds	n	R^2	SEE	F_{tr}	Q^2_{LOO}	Q^2_{LMO}	Q^2_{EXT}	SEP	F_{test}
Dioxins	25	0.953	0.329	344.8	0.938	0.938	0.894	0.434	30.17
Furans	35	0.840	0.596	125.3	0.816	0.818	0.806	0.474	50.1
Biphenyls	30	0.892	0.252	174.4	0.869	0.876	0.967	0.113	85.27
Dioxins and furans	60	0.848	0.603	234.1	0.834	0.834	0.778	0.629	61.57
Dioxins furans and biphenyls	90	0.855	0.557	389.6	0.847	0.848	0.826	0.515	119.6
All	95	0.804	0.623	274.7	0.792	0.792	0.912	0.498	106.5

For further studies of mfVCI, QSAR models were established by more than one homologues. The results indicate that the more homologues the training set includes, the lower fitting ability the model gets. The models have better fitting ability after removing the outliers from the data set (Table 4). It should be mentioned that the outlier removal and training set selection have an important effect on the QSAR model. And optimum size of the training set should be set based on a particular data set, types of descriptors and statistical analysis being used (Roy et al. 2008).

The robustness (Q^2_{LOO}>0.77) of QSAR models in this study is better than that from Ashek et al. (2006) (Q^2_{LOO}= 0.631). The influence of the number and position of the substitutional group on the toxicity of dioxins are consistent with the previous study (Ashek et al. 2006).

All of the above results show that the models developed from mfVCI are robust, and have good fitting and predictive performance. The mfVCI can be well used for explaining the property of compounds from different aspects, such as the fragments weights, the part values, the different substituted positions and so on. So, the mfVCI which is a very useful structural descriptor can be a good choice in QSAR research in the fields of environmental toxicology, biochemistry, etc.

Acknowledgments This reseach was supported by the Fundamental Research Funds for the Central Universities (lzujbky-2009-7), and State Key Laboratory of Environmental Chemistry and Ecotoxicology, and Research Center for EES of CAS (KF2008-03).

References

Ashek A, Lee C, Park H, Cho SJ (2006) 3D QSAR studies of dioxins and dioxin-like compounds using CoMFA and CoMSIA. Chemosphere 65:521–529

Chen Q, Kou YW, Wang Q, Chen H, Yuan JY (2009) A molecular fragments variable connectivity index for studying the toxicity (*Vibrio fischeri* pT_{50}) of substituted-benzenes. J Environ Sci Heal A 44:289–295

Golbraikh A, Tropsha A (2002) Beware of q2!. J Mol Graph Model 20:269–276

Haws LC, Su SH, Harris M, DeVito MJ, Walker NJ, Farland WH, Finley B, Birnbaum LS (2006) Development of a refined database of mammalian relative potency estimates for dioxin-like compounds. Toxicol Sci 89:4–30

Hays SM, Aylward LL (2003) Dioxin risks in perspective: past, present, and future. Regul Toxicol Pharm 37:202–217

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

Lohmann R, Breivik K, Dachs J, Muir D (2007) Global fate of POPs: Current and future research directions. Environ Pollut 150:150–165

Papa E, Villa F, Gramatica P (2005) Statistically validated QSARs, based on theoretical descriptors, for modeling aquatic toxicity of organic chemicals in pimephales promelas (Fathead Minnow). J Chem Inf Model 45:1256–1266

Papa E, Pilutti P, Gramatica P (2008) Prediction of PAH mutagenicity in human cells by QSAR classification. SAR QSAR Environ Res 19:115–127

Pompe M (2005) Variable connectivity index as a tool for solving the 'anti-connectivity' problem. Chem Phys Lett 404:296–299

Pompe M, Randić M (2007) Variable connectivity model for determination of pKa values for selected organic acids. Acta Chim Slov 54:605–610

Quaß U, Fermann M, Bröker G (2004) The European dioxin air emission inventory project-final results. Chemosphere 54:1319–1327

Randić M (1975) On characterization of molecular branching. J Am Chem Soc 97:6609–6615

Randić M (1991) Novel graph theoretical approach to heteroatoms in quantitative structure-activity relationships. Chemometr Intell Lab 10:213–227

Randić M (2001) The connectivity index 25 years after. J Mol Graph Model 20:19–35

Roy PP, Leonard JT, Roy K (2008) Exploring the impact of size of training sets for the development of predictive QSAR models. Chemometr Intell Lab 90:31–42

Safe S (1990) Polychlorinated biphenyls (PCBs), dibenzo-*p*-dioxins (PCDDs), dibenzofurans (PCDFs), and related compounds: environmental and mechanistic considerations which support the development of toxic equivalency factors (TEFs). Crit Rev Toxicol 21:51–88

Zhao YY, Tao FM, Zeng EY (2008) Theoretical study of the quantitative structure–activity relationships for the toxicity of dibenzo-*p*-dioxins. Chemosphere 73:86–91