Available online at www.tox.or.kr

Exploring Structure-Activity Relationships for the *In vitro* Cytotoxicity of Alkylphenols (APs) toward HeLa Cell

Myung-Gil Kim[1], Hye-Seoung Shin[2] & Jae-Hyoun Kim[3]

[1]Gyeonggi-do Public Health and Environment Research Institute 324-1 Pajang-dong, Jangan-gu, Suwon-si, Gyeonggi, Korea
[2]Hankyong National University Analysis Center, 67 Seokjungdong Anseongsi, Gyeonggi-do, Korea
[3]Department of Health Science, School of Natural Science, Dongduk Women's University, Wolgog-dong 23-1, Seongbuk-gu, Seoul 136-714, Korea
Correspondence and requests for materials should be addressed to J. H. Kim (kjhyon@dongduk.ac.kr)

Accepted 15 November 2008

Abstract

***In vitro* cytotoxicity of 23 alkylphenols (APs) on human cervical cancer cell lines (HeLa) was determined using the lactate dehydrogenase (LDH) cytotoxicity assay. Two different sets of descriptors were used to construct the calibration model based on Genetic Algorithm-Multiple Linear Regression (GA-MLR) based on the experimental data. A statistically robust Structure-Activity Relationships (QSAR) model was achieved (R^2=95.05%, Q^2_{LOO}=91.23%, F=72.02 and SE=0.046) using three Dragon descriptors based on Me (0D-Constitutional descriptor), BELp8 (2D-Burden eigenvalue descriptor) and HATS8p (3D-GETAWAY descriptor). However, external validation could not fully prove its validity of the selected QSAR in characterization of the cytotoxicity of APs towards HeLa cells. Nevertheless, the cytotoxicity profiles showed a finding that 4-*n*-octylphenol (4-NOP), 4-*tert*-octylphenol (4-TOP), 4-*n*-nonylphenol (4-NNP) had a more potent cytotoxic effect than other APs tested, inferring that increased length and molecular bulkiness of the substituent had important influence on the LDH cytotoxicity.**

Keywords: HeLa cell, LDH, Cytotoxicity, Alkylphenols, QSAR

In vitro short-term cytotoxicity assays are potentially useful tools that have been used to screen compounds not only to assess the reproducibility and the test conditions but also as aids in establishing priorities for regulation[1].

Reactive oxygen species (ROS) damage various cell components such as unsaturated lipids, proteins and nucleic acids. Direct cell damage occurs, since the ROS oxidize the lipid moiety in the plasma cell membrane, leading to the initiation of a lipid peroxidation cascade, followed by complex biochemical processes[2].

It is well known that hydrophobic compounds can disturb membrane energetization by non-specific membrane perturbation. APs are known to cause most severe changes in cell membrane structures and lipid droplets by such an oxidative stress[3]. Relationship between lipophilicity of C6-10 hydrocarbon compounds and their ROS-inducing potency in animal cells was established[4]. Both nonylphenol and octylphenol showed a significant cytotoxic effect on splenocytes *in vitro*[5]. It was found that nonylphenol cause apoptosis (both *via* the intrinsic and extrinsic pathway) and that ROS generation and Ca^{2+} play a fundamental role in the process[6]. 4-NNP caused suppressive effects on cell growth and cellular respiration in different cells and that these effects were associated with 4-NNP-induced reactive oxygen species (ROS) generation[7-9]. In addition, nonylphenol and octylphenol induced cell death by inhibiting testis endoplasmic reticulum Ca^{2+} pumps[10]. 4-*tert*-octylphenol (4-TOP) or *p*-Nonylphenol (4-NNP)-induced apoptosis increased in a concentration-dependent manner in cultured rat Sertoli cell[11], and in human embryonic stem cells[12].

The QSAR analyses have demonstrated that many physico-chemical properties were dominantly important for the prediction of cytotoxicity. LogP was of critical importance in describing the cytotoxicity for phenols with both electron-withdrawing and releasing groups[13-15]. Hydrophobicity and acidity as well as steric effects and charge distribution within the molecule govern the partitioning of for the intrinsic uncoupling activity of phenolic compounds[16]. LogP, the constant of Hammett ($\Sigma\sigma$), pKa, the first order valence molecular connectivity index (1Xv) and the perimeter of the efficacious section (^{P}Dg), representing lipophilic,

electronic and steric effects, were chosen to predict cytotoxicity of chlorophenols[17].

A QSAR analysis of cytotoxicity by a set of di- and trisubstituted phenols with mostly alkyl substituents in the *ortho* position was found with the combination of the Taft steric parameter for the larger of the two *ortho* substituents and radical parameter[18]. The QSAR models obtained for structurally diverse electron-releasing phenols were characterized by σ^+ (the Brown variation of the Hammett σ) and logP, or BDE and logP for each model, implying that highly electron-releasing substituents (negative coefficient of σ^+) enhance stabilization of the phenoxy radical and increase cytotoxicity[19]. As described above, authors provided basic information and a clear picture of mechanistic basis of substituted phenols.

The membrane toxicity of linear alcohol ethoxylates increased strongly with increasing length of the alkyl chain[20]. The cytotoxicity of APs increases with the number of carbon atoms and with the partition coefficient[21]. LogP is clearly linked with membrane perturbation[22,23] as well as enzymatic interactions[24] in combination with bioaccumulation. Cytotoxicity is closely related to logP for electron releasing phenols such as *p*-octyl- and *p*-nonylphenols[25]. Another study indicated that logP and σ^+ were important descriptors in modeling the cytotoxicity (QSAR) of the simple alkylphenols[26]. However, the cytotoxicity of the sterically encumbered phenols (*para* alkyl substituted phenols) was found to be dependent primarily on electronic and radical effects, not on steric factor[27]. Cytoxicity in human hepatocytes strongly correlate with hydrophobicity (logP), ease of oxidation (E_{HOMO}) and dipole moment (the asymmetric charge distribution according to arrangement of halogen substituents)[28]. Nevertheless, there is insufficient information on the cytotoxicity of APs in HeLa cell.

The LDH leakage assay is based on the measurement of lactate dehydrogenase activity in the extracellular medium[29]. The release of LDH is the toxicological endpoint associated with irreversible cell death due to cell membrane damage[30,31]. Determination of LDH release from cultured cells shows values that were significantly different, depending on cell type (tumor or normal)[32].

QSARs for the toxic and metabolic effects of aliphatic alcohols on the perfused rat liver were developed[33], where cytotoxic activity of the LDH was assessed for development of QSARs using descriptors, hydrophobicity (logP) and electrophilicity (LUMO). It was shown that the cytotoxicity of halogenated aliphatic hydrocarbons critically depended on their hydrophobic and steric properties[34]. There is not much information available in the literature concerning the structure-activity relationship of APs on the HeLa cell-mediated cytotoxicity.

Although octyl- or nonylphenols have been known to be linked to strong cytotoxicity, there is still insufficient knowledge about the whole spectrum of the cytotoxic potential of APs to HeLa cells. The objective of this study is to create a QSAR model able to to predict the cytotoxic activities of APs that mediate cell-damaging activity, and to understand the mechanistic concepts of the activity in complex HeLa cells.

QSAR Model Derived Using Dragon Descriptors

In order to investigate the possible existence of outliers from the 26 compounds which constitute the entire data set. The *LOO* procedure indicated that 3 chemicals, no. 4 (4-Ethylphenol), 15 (4-*tert*-Butylphenol) and 17 (3,5-Di-*tert*-butylphenol), were shown to be outliers fell beyond ± 2SD and were removed from the procedure of regression analysis to evaluate the cytotoxicity of 23 compounds (Table 1). The best model for 23 APs was chosen on the basis of experimental data with 3 descriptors, Me, BELp8, and HATS8p. It results then in the following Eq. 1 with increasing R^2 and Q^2_{LOO}:

$$\begin{aligned}\mathrm{Log(CT)} = &-13.34(\pm 4.13)+19.87(\pm 4.20)\mathrm{Me}\\ &-0.78(\pm 0.12)\mathrm{BELp8}\\ &+8.06(\pm 0.73)\mathrm{HATS8p}\end{aligned} \quad \text{(Eq. 1)}$$

$n=23$,	$R^2=89.07\%$,	$Q^2_{LOO}=84.24\%$,
$R^2_{adj}=87.35\%$,	$Q^2_{boot}=83.91\%$,	$SDEC=0.064$,
$SDEP=0.077$,	$K_X=82.09\%$,	$K_{XY}=61.21\%$,
$SE=0.071$,	$PRESS=0.137$,	$F=51.61$

All 23 compounds lies within $\pm 2\sigma$ standard deviation. Multiple regression generated a statistically significant predictive model with the highest predictivity, that could predict ($Q^2_{LOO}=84.24\%$) and explain ($R^2_{adj}=87.35\%$) the variance to a significant extent. The QSAR also gave F value of 51.61 and SE value of 0.071. Standardized correlation coefficients of Me, BELp8, and HATS8p are 0.637, −1.375, and 1.929 respectively for this model (Table 2). The QSAR yielded high Q^2_{boot} value 83.91% with a very low standard error of 0.071 to simulate the reliability of this model. However, the descriptors used for internal validation of predictive models show high correlation among themselves indicating high probability of chance correlation as K_{XY}-K_X values of all compounds showed negative.

The results of the cytotoxicity predictions are shown in Table 1 and in Figure 1 as a scatter plot of the predicted *versus* experimental values.

List of molecular descriptors selected by Genetic

Table 1. Values of experimental, calculated and predicted log(CT) using the cross validation method with Leave-One-Out procedure.

ID	Compound	Symbol	CAS RN	Y-Exp.	Y-Calc.	Y-Pred.	Hat	Std. Err. Pred.
1	2-Methylphenol	2-MP	95-48-7	6.08	6.14	6.15	0.107	0.99
2	4-Methylphenol	4-MP	106-44-5	6.18	6.13	6.12	0.171	−0.92
3	2-Ethylphenol	2-EP	90-00-6	6.2	6.13	6.12	0.171	−1.29
4	4-Ethylphenol	4-EP	123-07-9					
5	2-Propylphenol	2-PP	644-35-9	6.38	6.33	6.32	0.209	−1.01
6	4-Propylphenol	4-PP	645-56-7	6.23	6.33	6.36	0.209	2
7	2-Isopropylphenol	2-IPP	88-69-7	6.31	6.41	6.43	0.173	1.84
8	3-Isopropylphenol	3-IPP	618-45-1	6.29	6.35	6.36	0.085	0.97
9	4-Isopropylphenol	4-IPP	99-89-8	6.38	6.34	6.34	0.142	−0.65
10	2,6-Di-isopropylphenol	2,6-DIPP	2078-54-8	6.07	6.01	5.99	0.198	−1.23
11	2-*sec*-Butylphenol	2-SBP	89-72-5	6.37	6.33	6.32	0.209	−0.81
12	4-*sec*-Butylphenol	4-SBP	99-71-8	6.47	6.51	6.52	0.195	0.72
13	2-*tert*-Butylphenol	2-TBP	88-18-6	6.18	6.19	6.2	0.243	0.28
14	3-*tert*-Butylphenol	3-TBP	585-34-2	6.29	6.28	6.28	0.168	−0.23
15	4-*tert*-Butylphenol	4-TBP	98-54-4					
16	2,6-Di-*tert*-butylphenol	2,6-DTBP	128-39-2	6.06	6.12	6.13	0.166	1.14
17	3,5-Di-*tert*-butylphenol	3,5-DTBP	1138-52-9					
18	2-Cyclopentylphenol	2-CPP	1518-84-9	6.41	6.32	6.31	0.134	−1.59
19	4-Cyclopentylphenol	4-CPP	1518-83-8	6.22	6.23	6.24	0.207	0.26
20	4-*n*-Pentylphenol	4-NPP	14938-35-3	6.08	6.19	6.2	0.066	1.7
21	4-*tert*-Pentylphenol	4-TPP	80-46-6	6.05	5.95	5.92	0.244	−2.06
22	4-*n*-Hexylphenol	4-NHP	2446-69-7	5.96	6.05	6.07	0.171	1.78
23	4-*n*-Heptylphenol	4-NHTP	1987-50-4	6.51	6.45	6.44	0.136	−1.04
24	4-*n*-Octylphenol	4-NOP	1806-26-4	6.59	6.58	6.58	0.168	−0.22
25	4-*tert*-Octylphenol	4-TOP	140-66-9	6.66	6.59	6.58	0.182	−1.28
26	4-*n*-Nonylphenol	4-NNP	104-40-5	6.65	6.65	6.65	0.244	0.01

3*HAT value is 0.522.
Y-Exp data represent the mean value of triplicate measurements.
CT indicates cytotoxicity/(mg L^{-1}).
Data represent the means and standard error in triplicate.

Table 2. List of molecular descriptors selected by Genetic Algorithm, descriptor type and standardized regression coefficient.

Descriptor	Descriptor type	Std. Reg. Coeff.
Me	0-D (Constitutional)	0.637
BELp8	2-D (Burden eigenvalue)	−1.375
HATS8p	3-D (GETAWAY)	1.929

Algorithm, descriptor type and standardized regression coefficients are shown in Table 2.

The 2D plot of the correlation between experimental log(CT) and predicted log(CT) using the cross validation method with *LOO* procedure was presented (Figure 1). The linear regression model was found to be statistically significant although a compound 15 (4-*tert*-Butylphenol) was slightly over the two standard deviation line on the Williams plot (data not shown). The chemical domain of applicability was verified by the leverage approach in method section, thus all 23 chemicals belonged to the applicability domain and the predicted data were reliable.

After the model was internally validated, the cyto-

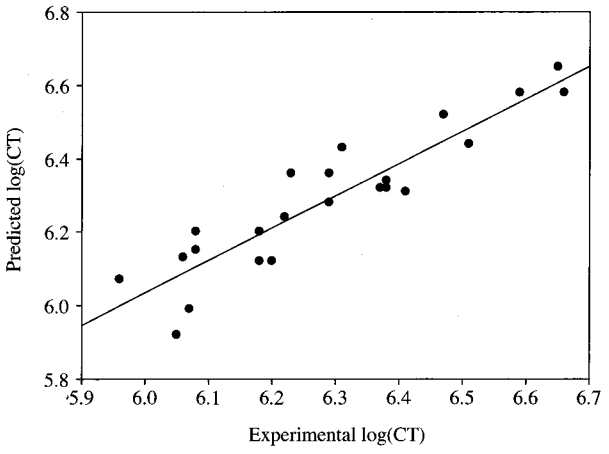

Figure 1. The correlation plot between experimental log(CT) and predicted log(CT) using the cross validation method with Leave-One-Out procedure.

toxicity was externally predicted for the same set of molecules. Even though the external validation is the way to establish a reliable QSAR model for predic-

Table 3. The statistical results of the MLR models of the cytotoxicity on 23 AP.

Eqn.	Equations	n	R^2	Q^2_{LOO}	Q^2_{boot}	SE	F	$PRESS$	Q^2_{EXT}
	A set of descriptors for internal validation								
Eq. 1	Log(CT)=−16.36(±5.36)+19.060(±5.45)Me −0.715(±0.16)BELp8+7.34(±0.98)HATS8p	23	89.07%	84.24%	83.84%	0.071	51.61	0.137	
	A set of descriptors for 30% external validation								
Eq. 2	Log(CT)=−7.140(±6.67)+13.590(±6.78)Me −0.887(±0.21)BELp8+7.83(±1.16)HATS8p	23	79.48%	68.79%	82.27%	0.100	34.06	0.083	84.43%
	A second set of descriptors for internal validation								
Eq. 3	Log(CT)=4.603(±0.38)+5.803(±1.85)$\sigma_{(m)}$ −3.282(±1.00)Log P+0.14(±0.04)MV	23	53.07%	37.33%	25.68%	0.126	7.16	0.88	
	A second set of descriptors for 30% external validation								
Eq. 4	Log(CT)=5.81(±0.17)−0.80(±0.15)X0 −0.09(±0.02)TE−0.00019(±0.00004)MOMI-R1	23	80.66%	70.54%	60.90%	0.100	18.07	0.316	0.00%

PRESS : the predicted residual sum of squares, Q^2_{LOO} : leave-one-out cross-validated R^2, *SE* : standard error of estimation, *F* : sequential Fischer test, Q^2_{EXT} : externally cross-validated R^2, Me : the mean atomic Sanderson electronegativity (scaled on Carbon atom), BELp8 : lowest eigenvalue n. 8 of Burden matrix/weighted by atomic polarizabilities, HATS8p : the leverage-weighted autocorrelation of lag 8/weighted by the atomic polarizabilities, MV : molecular volume, ^{0}X : connectivity (Kier, order 0), $\sigma_{(m)}$: Hammet substituent parameter (meta-), TE : total energy, MIR3-R1 : moment of inertia-rotational constants component 3.

tive purposes. The obtained model exhibited good external predictivity (Table 3). However, 50% of the whole 94 models generated a low external predictivity ranging from 0.00% to 48.44%.

QSAR Model Derived using a Second Set of Descriptors

A QSAR model for cytotoxicity was constructed for a same training set of 23 APs using a second set of descriptors. Eq. 2 shows a three-descriptor model including $\sigma_{(m)}$, logP and MV selected with the highest R^2 and Q^2_{LOO} :

$$\begin{aligned}\text{Log(CT)}=&4.60(\pm0.38)+5.80(\pm1.85)\sigma_{(m)}\\&-3.28(\pm1.00)\text{logP}\\&+0.14(\pm0.04)\text{MV}\end{aligned}\qquad\text{(Eq. 2)}$$

n=23, R^2=53.07%, Q^2_{LOO}=37.33%,
R^2_{adj}=45.66%, Q^2_{boot}=25.21%, $SDEC$=0.169,
$SDEP$=0.196, K_X=56.47%, K_{XY}=48.81%,
SE=0.186, $PRESS$=0.088, F=7.16

The standardized regression coefficients of $\sigma_{(m)}$, logP and MV were 1.137, −11.709 and 12.524, respectively, so that logP and MV are regarded as relatively more important than $\sigma_{(m)}$ parameter.

As the QSAR model shown by Eq. 2 was statistically weak, we did not choose this Eq. 2 as a final predictive model. But this model provides valuable information regarding the types of descriptors and mechanistic interpretation of a QSAR by Eq. 1 constructed by using Dragon descriptors as well.

The descriptors used in this QSAR Eq. 2 were divided into three categories: (i) the molecular hydrophobicity descriptor, logP as octanol/water partitioning coefficient, (ii) electronic parameter, $\sigma_{(m)}$, as a Hammet substituent parameter, (iii) physicochemical parameter, MV as a molecular volume. These physico-chemical parameters are used to characterize QSAR models in the cytotoxicity study, as described above. A parameter for radical reactivity by activated oxygen species or phenoxy radical of APs is considered as the gap energy between E_{HOMO} and E_{LUMO}. The shape of the molecule is defined by the moment of inertia which is proportional to length-to-breadth ratio (L/B). The principal moments of inertia, a spatial descriptor that provides information about the orientation and conformational rigidity of a molecule and are related to Rotational constants (MHz)[35]. Thereby, *n*-nonylphenol possesses a larger moment of inertia of the phenol moiety due to its longer chain length. The Taft parameter[36,37] was used to correlate the field/inductive effects of alkyl substituents. The Hammett constant[38] is an electronic substituent descriptor reflecting the electron-donating or -accepting properties of a substituent at the *m*- and *p*- position.

The resulting regression equations including statistics are summarized in Table 3.

Discussion

In the present study, it was confirmed that octylphols or nonylphenol with 8-9 carbon side chain elicited a high cytotoxicity and selectivity of all APs

against HeLa cells, suggesting that the cytotoxicity is very dependent on the alkyl chain length, and degree of branching of the alkyl group and its position of substitution on the phenol molecule. Additionally, 4-*n*-nonylphenol induced a highest cytotoxicity of tested compounds. However, no clear differences in the degree of activity between 2- (*ortho*-) and 4- (*para*-) substituted APs were found, although many 4- substituted APs have shown higher activities compared to 2- substituted APs (Table 1).

A combination of Me, BELp8 and HATS8p descriptors were used for encoding the structural information of the studied compounds: (i) Me is the mean atomic Sanderson electronegativity(scaled on Carbon atom) characterized the molecular polarity (ii) BELp8: lowest eigenvalue n. 8 of Burden matrix/weighted by atomic polarizabilities (2D-Burden eigenvalue descriptor), encoding molecular branching, position and its length. BELp8 is the one of the BCUT (Burden-CAS-University of Texas eigen values), topological descriptors which are the eigen values of a modified connectivity matrix known as the Burden matrix. (iii) HATS8p is the leverage-weighted autocorrelation of lag 8/weighted by the atomic polarizabilities (3D-GETAWAY descriptor).

These descriptors encode information about molecular shape, size, and atom distribution, and that application of the atomic polarizabilities[39] as weighting coefficients, takes into account, to some degree, charge distribution inside a molecule.

The most relevant descriptor according to the stdardized regression coefficients is HATS8p. A negative K_{XY}-K_X value, however, represents high degree of collinearity of descriptors in this model. According to the K (QUIK) rule, only those models with a global correlation of the $[X+y]$ block (K_{XY}) greater than the global correlation of the X block (K_{XX}) variable (X being the molecular descriptors and y the response variable) were considered acceptable. In this study, most of 94 models calculated were shown to have negative values.

The problem of variable selection remains as in GA-MLR technique even though PLSR (Partial Least Squares Regression) or PCR (Principle Component Regression) modeling solves the collinearity problem of the descriptors[40]. Nonetheless, use of both PCR and PLSR techniques using same data showed similar predictive power and high chance correlation (data not shown) when compared to GA-MLR result so that future studies would be carried out to solve the collinearity problems for further studies.

The QSAR approach achieved the prediction accuracy of 84.43% for the 30% external validation set (70% training set) by random sampling. But, more than 50% of the 95 predictive models generated by GA-MLR technique showed low R^2 and Q^2_{LOO}. The finding seems to be due to the small data size and inaccuracy of trained data set though their reliability are reasonable. The results of the statistical analysis are shown in Table 3.

It is postulated that cytotoxicity mechanism of cell damage by APs follows three steps. Physico-chemical, steric (shape area, volume), structural (branching) factors are likely to be involved in the penetration process into membrane. Molecules may then permeate into the cell by lipophilic diffusion both across and along the membranes so that higher values of log P (elongation of substituent) can allow the molecule to cross membranes more rapidly[41-43].

The three-descriptor model provided a clear mechanistic basis for the cytotoxicity. Thus, the coefficient of Me in the QSAR is positive, hence polar AP molecules with logKo/w higher than 4.0 will exhibit greater cytotoxicity, that is, cell damage[44]. The negative regression coefficient of BELp8 leads to a decrease in the value of cytotoxicity, implying that the decrease in bulkiness and size inhibits permeation into the membrane. The positive coefficient of HATS8 suggests that they account for favorable steric effects that polarizabilities are increasing nearly linearly with the elongation of the alkyl chain. The lipophilicity of a chemical is very relevant to cytotoxicity through nonpolar-nonpolar interactions.

Additional use of a second set of descriptors confirmed that the predicted model, which has poor statistical quality (R^2=53.07%, Q^2_{LOO}=37.33%, SE=0.255), demonstrates that electronic ($\sigma_{(m)}$) and chemical (MV, logP) parameters are major factors controlling the binding of the phenol derivatives to HeLa cells.

4-*n*-Octylphenol, 4-*tert*-octylphenol and 4-*n*-nonylphenol were determined to be the most potent cytotoxic compounds within this group of chemicals (Table 1). Therefore, a long straight chain AP, *n*-nonylphenol exhibited the highest cytotoxic activity among the tested compounds. Straight chain APs appear to exert higher activities compared to branched APs. Other small alkylphenols were found to be less toxic to HeLa cell distinctively. An additional finding indicated that di-substituted APs such as 2,6-di-isopropylphenol and 2,6-di-*tert*-butylphenol were slightly less active than we anticipated, reflecting that steric hindrance caused by the double *ortho* substitution in 2,6-dialkylphenols might limit its penetration into the membrane.

The QSAR model was validated for predictivity by internal validation and was capable of achieving high predictive power within their domain of applicability, providing a predictive model of the HeLa cell cytotoxicity by 23 APs. Three important parameters for

the final predictive model were Me, HATS8p and BELp8 encoding significant structural information such as shape, branching and elongation etc. The finding was also validated by the QSAR equation using a different set of descriptors ($\sigma_{(m)}$, logP and MV) in terms of mechanistic interpretation. Further, the model serves as a tool to elucidate the mechanism of cell damage assessed by LDH release on APs for HeLa cells (epitheloid cervix carcinoma).

In summary, the molecular polarity and topology are two key characteristics that appear to explain the observed cell viability of 23 APs. The choice of these descriptor classes, and particularly these three molecular descriptors, indicates a plausible connection between the proposed QSAR model and molecular stereospecificity of the cytotoxic potential, where the 2D and 3D atomic polarizability (elongation of the alkyl chain) information about an AP is the most critical component of the AP-HeLa cell interaction.

Materials & Methods

Test Chemicals

Alkylphenols (purity > 97%) and other reagents of analytical grade were purchased from Sigma Chemical Company (St. Louis, MO).

Preparation of Cell and Test Solutions

The stock solution of each of 20 APs was made using DMSO/water mixture (50 : 50; v/v). HeLa cells were exposed to 10^{-5} mol of each AP. The pH of AP solutions were in a range from 6.21 to 6.52 in a 50/50 (v/v) ethanol/H_2O mixture. The human cervical cancer cells (HeLa # CCL-2) were obtained from the American Type Culture Collection (Manassas, VA, USA). The cells were cultured in RPMI 1640 (Gibco, Grand Island, NY, USA) supplemented with 5% FBS (fetal bovine serum) (Gibco), 15 mmol HEPES (N-2-hydroxyethylpiperazine-N′-2-ethanesulfonic acid), and maintained at 37°C in 5% CO_2 in atmosphere.

Determination of Cytoxicity by LDH Activity of HeLa Cell

The experiment of cytotoxicity was performed according to CytoTox 96R Non-Radioactive Assay Kit (Promega, Madison, WI), which quantitatively measures LDH, a stable cytosolic enzyme that is released upon cell lysis. Cell culture medium (500 μL) was taken from different samples and centrifuged at 1,000 rpm for 4 min. Then, 90 μL of each supernatant (10^{-5} cell mL^{-1}) in triplicate were transferred to a 96-well enzymatic assay plate. To evaluate the target cell maximum LDH release, cells were incubated with 0.8% Triton X-100 (final concentration) in an atmosphere of 5% CO_2 at 37°C for 4 hours to yield complete lysis of cells. Then, 10 μL of the reconstituted substrate mix (alkylphenol) was added to each well. After incubation for 4 hours at room temperature in the dark, the reaction was stopped by addition of 50 μL stop solution (at pH 2.3) to each well. After removal of bubbles, the absorbance was measured at 490 nm in ELISA reader (Bio-Rad Laboratories, Inc.) within one hour after addition of stop solution. The cytotoxic responses were expressed as percent of LDH released by the testing compounds, compared with ones released by control cells as follow:

$$\text{Cytotoxicity (CT)} = [(S - N)/(P - N)] \times 100$$

S = Absorbance (A_{490}) of sample
N = Absorbance (A_{490}) of negative control
P = Absorbance (A_{490}) of positive control

Data Set

The training set included 20 APs (Table 1). Log(CT) values of the chemicals (4-ethylphenol, 4-*tert*-butylphenol, 2,6-di-*tert*-butylphenol, 3,5-di-*tert*-butylphenol, 4-*n*-pentylphenol and 4-*n*-hexylphenol) were predicted from the final QSAR model derived with 20 training data set.

Optimization and Descriptor Calculation

By using *Hyperchem* software 7.0 (Hypercube, Inc., Gainesville, FL, USA), chemical structures were drawn and named by CAS-number. Molecular mechanic force field (MM+) was selected for the geometry optimization using Polak-Ribiere algorithm with a maximum cycle (10,000) and a convergence limit of the 0.005 kcal/mol. After optimizing chemical structures from *Hyperchem* 7 software, *Dragon* 5.0 package (Milano Chemometrics and QSAR Research Group, University of Milano-Bicocca, Milan, Italy) is employed for the calculation of the Dragon molecular descriptors. The block description including a diverse set of descriptors are as follows: topological descriptors, connectivity indices, information indices, Burden eigenvalues, topological charge indices, eigenvalue-based indices, Randic molecular profiles, geometrical descriptors, WHIM descriptors, GETAWAY descriptors, atom-centered fragments and charge descriptors that are classified into the groups, i.e. 0D, 2D and 3D. The details of descriptor calculations and their meanings are described in the Handbook of Molecular Descriptors[45].

A second set of additional descriptors, Hartree total energy (kcal/mol), bending energy (kcal/mol), stretch bend energy (kcal/mol), van der Waal's energy (kcal/mol), torsional energy (kcal/mol), dipole interaction

energy (kcal/mol), moment of inertia (principal moment of inertia $-X$, Y and Z components (PMIx, PMIy and PMIz), moment of inertia rotational constants component 1, 2 and 3 (MIR1, MIR2 and MIR3)), Taft parameters sigma* (σ^*), the Hammett σ constants sigma (m) ($\sigma(m)$), sigma (p) ($\sigma(p)$) and summation of σ constants ($\sum\sigma$) were calculated by means of an MM2 minimizer and minimization with *MOPAC* 6.0 of the *Molecular Modeling Pro* (ChemSW, Inc.) program package. The highest occupied molecular orbital, E_{HOMO} (eV) and the lowest occupied molecular orbital, E_{LUMO} (eV) energies were calculated using PM5/H_2O geometry of *MOPAC* using version 6.1.10 of *CAChe's Windows Project Leader* interface (Fujitsu Ltd, OR, USA).

Chemometric Analysis

For collinearity, the collinearity among descriptors must be checked, and correlation between the X block and the Y response verified (the QUIK rule of Todeschini based on K: K_{XY} must be significantly higher than K_{XX}). From among models of similar performance, those with higher ΔK ($K_{XY}-K_{XX}$) were selected.

For appilicability of domain, the developed model was checked for the chemical domain to verify the prediction reliability. In the Williams plot, or OLS outlier and leverage plot, standardized cross-validated (or jackknifed) residuals were plotted versus leverages (hat diagonals). Outliers are considered to be those compounds with crossvalidated standardized residuals greater than 2.5 standard deviation units, while influential chemicals are those compounds with a leverage value (h) greater than the critical point $h^*=3p'/n$, where p' is the number of model variables plus one, and n the number of objects used to calculate the model.

For the validation of the model, the combination GA-MLRA technique was utilized to select the appropriate descriptors and to generate different QSAR models as implemented in the *Mobydigs* program[46]. This algorithm allows to construct the models with the following statistical characteristics: the correlation coefficient squared R^2, low standard deviation S and the least number of descriptors involved. Therefore, the software package was employed for detailed statistical model analysis of the models obtained from the entire data set. The statistics include high Fisher coefficient F, Predictive Residual Error Sum of Squares (PRESS), Standard error of calibration (SEC), standard error of prediction (SEP), explained variance (R^2_{adj}) as additional statistical parameters. The predictive stability of the models was verified applying internal Leave-One-Out (*LOO*). The robustness of the model and its predictivity was guaranteed by both the stability of Q^2_{LOO}, and bootstrap (Q^2_{boot}). The model obtained on the first chemicals selected is used to predict the values for the excluded sample, then Q^2_{LOO} is calculated for each model. The bootstrapping was repeated 5,000 times for each validated model. The proposed model was also checked for reliability and robustness by permutation testing: new models were recalculated for randomly reordered response (Y scrambling). These chemometric validation techniques are well described in literature[47].

References

1. O'Brien, P. & Haskins, J. R. *In vitro* cytotoxicity assessment. *Methods Mol Biol* **356**:415-425 (2007).
2. Valko, M. *et al.* Free radicals and antioxidants in normal physiological functions and human disease. *Int J Biochem Cell Biol* **39**:44-84 (2007).
3. Haavisto, T. E. *et al.* Effects of 4-*tert*-octylphenol, 4-*tert*-butylphenol, and diethylstilbestrol on prenatal testosterone surge in the rat. *Reprod Toxicol* **17**:593-605 (2003).
4. Dreiem, A., Ring, A. & Fonnum, F. Organic solvent-induced cell death in rat cerebellar granule cells: Structure dependence of c10 hydrocarbons and relationship to reactive oxygen species formation. *Neurotoxicology* **26**:321-330 (2005).
5. Nair-Menon, J. U. *et al.* Toxic effects of octylphenol on cultured rat and murine splenocytes. *Toxicol Appl Pharmacol* **139**:437-444 (1996).
6. Pretorius, E. *et al.* Ultrastructural effects of low dosage endocrine disrupter chemicals on neural cells of the chicken embryo model. *Horm Metab Res* **38**:639-649 (2006).
7. Okai, Y. *et al.* Protective effect of antioxidants against para-nonylphenol-induced inhibition of cell growth in *Saccharomyces cerevisiae*. *FEMS Microbiol Lett* **185**: 65-71 (2000).
8. Okai, Y. *et al.* Enhancing effect of the endocrine disruptor para-nonylphenol on the generation of reactive oxygen species in human blood neutrophils. *Environ Health Perspect* **112**:553-660 (2004).
9. Li, H. M. *et al.* Protection against nonylphenol-induced cell. death by DJ-1 in cultured Japanese medaka (*Oryzias latipes*) cells. *Toxicology* **228**:229-238 (2006).
10. Hughes, P. J. *et al.* Estrogenic alkylphenols induce cell death by inhibiting testis endoplasmic reticulum Ca(2+) pumps. *Biochem Biophys Res Commun* **277**: 568-574 (2000).
11. Qian, J. *et al.* Octylphenol induces apoptosis in cultured rat Sertoli cells. *Toxicol Lett* **166**:178-186 (2006).
12. Kim, S. K. *et al.* Nonylphenol and octylphenol-induced apoptosis in human embryonic stem cells is related to Fas-Fas ligand pathway. *Toxicol Sci* **94**:310-321 (2006).
13. Leo, A. J. & Hansch, C. Role of hydrophobic effects in mechanistic QSAR. *Perspect Drug Discovery Des*

17:1-25 (1999).
14. Hansch, C. *et al.* Comparative QSAR evidence for a free-radical mechanism of phenol-induced toxicity. *Chem Biol Interact* **127**:61-72 (2000).
15. Moridani, M. Y. *et al.* Quantitative structure-toxicity relationships for phenols in isolated rat hepatocytes. *Chem Biol Interact* **145**:213-223 (2003).
16. Escher, B. I. *et al.* Kinetic model to describe the intrinsic uncoupling activity of substituted phenols in energy transducing membranes. *Environ Sci Technol* **33**:560-570 (1999).
17. Liu, X. *et al.* Quantitative structure activity relationship (QSAR) for toxicity of chlorophenols on L929 cells *in vitro*. *Chemosphere* **64**:1619-1626 (2006).
18. Selassie, C. D. *et al.* QSAR for the cytotoxicity of 2-alkyl or 2,6-dialkyl, 4-X-phenols: the nature of the radical reaction. *J Chem Soc Perkin Trans* **2**:1112-1117 (2002).
19. Selassie, C. D. *et al.* Cellular apoptosis and cytotoxicity of phenolic compounds: a quantitative structure-activity relationship study. *J Med Chem* **48**:7234-7242 (2005).
20. Muller, M. T. *et al.* Membrane toxicity of alcohol ethoxylates. *Environ Toxicol Chem* **18**:2767-2774 (1999).
21. Boge, G. & Roche, H. Cytotoxicity of phenolic compounds on *Dicentrarchus labrax* erythrocytes. *Bull Environ Contam Toxicol* **57**:171-178 (1996).
22. Veith, G. D. & Broderius, S. J. Rules for distinguishing toxicants that cause type I and type II narcosis syndromes. *Environ Health Perspect* **87**:207-211 (1990).
23. Siraki, A. G. *et al.* Quantitative structure-toxicity relationships by accelerated cytotoxicity mechanism screening. *Curr Opin Drug Discov Devel* **7**:118-125 (2004).
24. Lewis, D. F. & Dickins, M. Quantitative structure-activity relationships (QSARs) within series of inhibitors for mammalian cytochromes P450 (CYPs). *J Enzyme Inhib* **16**:321-337 (2001).
25. Selassie, C. D. *et al.* On the toxicity of phenols to fast growing cells. A QSAR model for a radical-based toxicity. *J Chem Soc Perkin Trans* **2**:2729-2733 (1999).
26. Takahata, Y. *et al.* Core-Electron Binding Energy (CEBE) as descriptors in Quantitative Structure Activity Relationship (QSAR) analysis of cytotoxicities of a series of simple phenols. *QSAR Comb Sci* **26**:378 (2007).
27. Loader, R. J. *et al.* The cytotoxicity of *ortho* alkyl substituted 4-X-phenols: a QSAR based on theoretical bond lengths and electron densities. *Bioorg Med Chem Lett* **16**:1249-1254 (2006).
28. Chan, K. *et al.* Structure-activity relationships for halobenzene induced cytotoxicity in rat and human hepatoctyes. *Chem Biol Interact* **165**:165-174 (2007).
29. Decker, T. & Lohmann-Matthes, M. A quick and simple method for the quantitation of lactate dehydrogenase release in measurements of cellular cytotoxicity and tumor necrosis factor (TNF) activity. *J Immunol Meth* **15**:61-69 (1988).
30. Strubelt, O. *et al.* The toxic and metabolic effects of 23 aliphatic alcohols in the isolated perfused rat liver. *Toxicol Sci* **49**:133-142 (1999).
31. Fotakis, G. & Timbrell, J. A. *In vitro* cytotoxicity assays: comparison of LDH, neutral red, MTT, and protein assay in hepatoma cell lines following exposure to cadmium chloride. *Toxicol Lett* **160**:171-177 (2005).
32. Konjevic, G. *et al.* Correction of the original lactate dehydrogenase (LDH) release assay for the evaluation of NK cell cytotoxicity. *J Immunol Methods* **200**:199-201 (1997).
33. Cronin, M. T. D. *et al.* The importance of hydrophobicity and electrophilicity descriptors in mechanistically-based QSARs for toxicological endpoints. *SAR QSAR Environ Res* **13**:167-176 (2002).
34. Eriksson, L. *et al.* Modelling the cytotoxicity of halogenated aliphatic hydrocarbons. Quantitative structure-activity relationships for the IC50 to human HeLa cells. *Quant Struct-Act Relat* **12**:124-131 (1993).
35. Eisenberg, A. & Eu, B. C. Mechanical spectroscopy: An introductory review. *Ann Rev Mater Sci* **6**:335-359 (1976).
36. Hansch, C. & Leo, A. Exploring QSAR: Fundamentals and Applications in Chemistry and Biology. ACS Professional Reference Book. *American Chemical Society* **1**:557-1037 (1995).
37. Babij, C. & Poe, A. J. Deconstruction of Taft's σ^* parameter: QSAR meets QALE. *J Phys Org Chem* **17**:162-167 (2004).
38. Hansch, C. *et al.* A survey of Hammett substituent constants and resonance and field parameters. *Chem Rev* **91**:165-195 (1991).
39. Consonni, V. *et al.* Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors: 1. Theory of the novel 3D molecular descriptors. *Chem Inf Comput Sci* **42**:682-692 (2002).
40. Hemmateenjad, B. *et al.* A mechanistic QSAR study on the leishmanicidal activity of some 5-substituted-1,3,4-thiadiazole derivatives. *Chemical Biology and Drug Design* **69**:435-443 (2007).
41. Nakagawa, Y. & Tayama, S. Metabolism and cytotoxicity of bisphenol A and other bisphenols in isolated rat hepatocytes. *Arch Toxicol* **74**:99-105 (2000).
42. Lueken, A. *et al.* Synergistic DNA damage by oxidative stress (induced by H_2O_2) and nongenotoxic environmental chemicals in human fibroblasts. *Toxicol Lett* **147**:35-43 (2004).
43. Bashford, L. & Knox, P. Membrane-mediated cytotoxicity: From biophysics to medicine. *BioEssays* **5**: 134-135 (2005).
44. Harman, C. *et al.* Uptake rates of alkylphenols, PAHs and carbazoles in semipermeable membrane devices (SPMDs) and polar organic chemical integrative samplers (POCIS). *Chemosphere* **72**:1510-1516 (2008).
45. Todeschini, R. & Consonni, V. Handbook of Molecular Descriptors, Wiley-VCH Weinheim, Germany, p. 667 (2000).

46. Todeschini, R. *et al.* Software for Regression and Classification Models by Genetic Alghorithms. In Nature-inspired Methods in Chemometrics: Genetic Alghortims and Artificial Neural Networks (Leardi, R., ed.), Elsevier 374-385 (2003).
47. Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR Comb Sci* **26**:694-701 (2007).