Three-dimensional Holographic Vector of Atomic Interaction Field (3D-HoVAIF) for the QSPR/QSAR of Polychlorinated Naphthalenes

LI Zheng-Hua  CHEN Gang  CHEN Zhi-Tao  XIA Zhi-Ning  CHENG Fan-Sheng  CHEN Hua
(College of Chemistry and Chemical Engineering, Chongqing University, Chongqing 400030, China)

ABSTRACT  Three-dimensional holographic vector of atomic interaction field (3D-HoVAIF) is used to describe the chemical structures of polychlorinated naphthalenes (PCNs). After variable screening by stepwise multiple regression (SMR) technique, the liner relationships between gas-chromatographic relative retention time (RRT), 298 K supercooled liquid pressures (logP_L), n-octanol/air partition coefficient (logKOA), n-octanol/water partition coefficient (logKOW), aqueous solubilities (logS_W), relative in vitro potency values (–logEROD) of PCNs and 3D-HoVAIF descriptors have been established by partial least-square (PLS) regression. The result shows that the 3D-HoVAIF descriptors can be well used to express the quantitative structure-property (activity) relationships of PCNs. Predictive capability of the models has also been demonstrated by leave-one-out cross-validation. Moreover, the predicted values have been presented for those PCNs which are lack of experimentally physico-chemical properties and biological activity by the optimum models.

Keywords: polychlorinated naphthalenes, three-dimensional holographic vector of atomic interaction field, QSAR

1 INTRODUCTION

Polychlorinated naphthalenes (PCNs) are an important type of environmentally persistent pollutants and have always attracted great attention. PCNs have been frequently found in several matrices including sediments, soils, water, air, biota and even food and dietary exposure. The toxicologic studies have shown that PCNs have similar toxic properties to polychlorinated dibenzo-p-dioxins (PCDDs), polychlorinated dibenzofurans (PCDFs) and polychlorinated biphenyls (PCBs), such as the induction of aryl hydrocarbon hydroxylase (AHH) and 7-ethoxyresorufin-O-deethylase (EROD) activity which are important for the hepatic drug-metabolising activity. In addition, the effect of PCNs on GABA-metabolizing enzymes and cytochrome P-450 has been investigated.

Physico-chemical properties of an organic chemical compound are often the key role in assessing its distribution and transport in the global environment, such as vapour pressures (P_L), water solubility (S_W), n-octanol/air partition coefficients (KOA) and n-octa-
nol/water partition coefficients ($K_{OW}$). Moreover, various biological activities play an important role in evaluating the integrated risk for adverse human health effect and environment risk assessment. There is the presence of 76 theoretically possible isomers for PCNs depending on the number and substitution pattern of chlorine atoms. Not only owing to the time consuming and high expense, but also to the lack of PCNs standards, it is hard to determine experimentally the physico-chemical properties and various biological activities for all PCNs. Therefore, alternative approaches are needed. Many previous studies[13-16] indicated that it is a feasible and effective approach to predict the physico-chemical proprieties and biological activity of many organic compounds by quantitative structure-property (activity) relationship (QSPR/QSAR) models. Actually, the QSPR/QSAR studies on PCNs have been reported in recent literatures[17-20], but most of these studies just focused on one or few properties.

In the QSPR and QSAR studies, people always have a profound interest in the choice of appropriate structural parameters. Presently, the common molecular structure descriptors can be divided into 2D and 3D types. However, 2D descriptors are impossible to take reappearance of molecular actual spatial structures and regardless of the molecular interactions. Under such circumstances, 3D descriptors have been progressively developed in structural representation. For example, CoMFA and the similar method, CoMSIA, have been applied to study the persistent organic pollutants (POPs) screening for atmosphere persistence[14] and the toxicity of polybrominated diphenylethers (PBDEs)[21]. However, these methods suffer from the intrinsic unfavorable drawback of strong reliance on molecular conformation, complicated calculation, and so on. Three-dimensional Holographic Vector of Atomic Interaction Field (3D-HoVAIF) is a new 3D descriptor based on the 2D molecular structure characterization method proposed by Li and his co-workers[22-24], including merits of both traditional 2D and 3D descriptors (2D are easy and rapid for calculations and 3D is more amenable to physico-chemical interpretations). Based on the two space invariants (atomic relative distance and atomic properties) and three kinds of non-bonded interaction modes, the 3D-HoVAIF method derives multidimensional vectors to represent molecular structural characteristics and is independent of experiments. In this paper, we report six quantitative structure-property (activity) relationships of PCNs established by using the 3D-HoVAIF descriptors. It can help us gain insight into how the physico-chemical properties and biological activity are affected by the 3D-HoVAIF descriptors, and provide a new method for predicting physicochemical properties and biological activities of PCNs.

2 MATERIALS AND METHODS

2.1 Data set

All experimentally determined physico-chemical properties and biological activities of PCNs were taken from previous publications, including $RRT^{[17]}$, $\log P_L^{[25]}$, $\log K_{OA}^{[26]}$, $\log K_{OW}^{[27]}$, $\log S_W^{[28]}$ and $-\log EROD^{[29-30]}$ of PCNs. All the values in this study are listed in Table S1 of the Supplementary materials.

2.2 Three-dimensional holographic vector of atomic interaction field

2.2.1 Atomic types and interactions

As is well known, ordinary atoms of organic molecules including H, C, N, P, O, S, Cl, Br and I are partitioned into 5 types in the Periodic Table of elements. According to hybridization state of atoms, these atoms are furthermore subdivided into 10 tapes. Therefore, there are 55 interatomic interactions in a molecule (Table 1). In this paper, three kinds of potential energy, electrostatic, steric and hydrophobic, take part in the representation of different interactions, producing $3 \times 55 = 165$ interaction items for the organic compounds.

2.2.2 Electrostatic interaction

Electrostatic interaction, an important non-bonded interaction, could be expressed by classical Coulomb
2.2.3 Steric interaction

Steric interaction describing the interatomic spatial nondipole-dipole or dipole-induced interactions is expressed by Lennard-Jones (Eq. 2).

$$ E_{st}(S) = \sum_{i=m}^{n} \sum_{j=n}^{m} \left( \frac{Z_i Z_j}{r_{ij}} \right)^{2} - 2 \left( \frac{R_{ij}^{*}}{r_{ij}} \right)^{6} $$

(1 ≤ m ≤ 10, m ≤ n ≤ 10) (2)

where $\epsilon_{ij} = (\epsilon_{ii} \epsilon_{jj})^{1/2}$ is potential well of atomic pairs cited from literatures [31-32]. $D$ is empirical atomic interaction correction constant (0.01)$^{[39]}$. $R_{ij}^{*} = (C_h R_{ii}^{*} + C_h R_{jj}^{*})/2$ is van der Waals' radius for modified atom-pair, with corrected factor Ch of 1.00 in case of sp$^3$ hybridization, 0.95 sp$^2$ hybridization and 0.90 sp hybridization$^{[33]}$.

2.2.4 Hydrophobic interaction

Hydrophobic interaction force field is defined as interatomic hydrophobic interaction in hint method proposed by Kellogg$^{[34]}$ (Eq. 3).

$$ E_{hy}(H) = \sum_{i=m}^{n} \sum_{j=n}^{m} S_{ij} S_{aj} e^{T_{ij}} $$

(1 ≤ m ≤ 10, m ≤ n ≤ 10) (3)

where $H$ is the solvent accessible surface area for atoms$^{[35]}$, indicating formation on surface area when water-molecule probes the roiling sphere at the atomic surface; $a$ is the atomic hydrophobic constant cited from the reference$^{[36]}$; $T$ is the sign function, indicating entropy change resulting from different types of atomic interaction$^{[34]}$.
The variable selection by SMR and the statistic of the models between 3D-HoVAIF descriptors and the physico-chemical properties and biological activity established by PLS were collected in Table S3-S8 of the supplementary materials. Table 2 summarizes the optimum QSPR and QSAR models for PCNs, in which N represents the number of data points submitted to the regression, PC the number of principal constituents, R the correlation coefficient, Q the cross-validated correlation coefficient, and RMSEE the fitted root-mean-square error of estimation.

3.2.1 QSPR model of RRT

The optimum QSPR model of RRT includes four variables: hydrophobic interaction items $V_{113}$, $V_{120}$, $V_{137}$, and electrostatic interaction item $V_{3}$. This model adopts one significant principal component explaining 99.01% variance ($R^2$) of dependent variable, whose cross-validation variance of dependent variable ($Q^2$) is 98.95%, and the fitted root-mean-square error of estimation (RMSEE) is 23.8923. The result shows the model exhibits excellent prediction ability and stability, which is similar to other estimation models found by Oliveron et al.[17], Zhai et al.[18] or Xu et al.[19]. The plot predicted by the optimum QSPR models and observed RRT values is shown in Fig. 1(a).

With further analysis of this model, $V_{3}$ means the electrostatic interaction between $sp^2$ hybrid C and H atoms, $V_{113}$ represents the hydrophobic interaction between $sp^2$ hybrid C and H atoms, $V_{120}$ is the hydrophobic interaction of H and Cl atoms, and $V_{137}$ shows the hydrophobic interaction of $sp^2$ hybrid C and Cl atoms. The values of variables important in projection (VIP) of these four variables are $V_{113}$ 1.11898, $V_{137}$ 1.1172, $V_{2}$ 1.11226 and $V_{120}$ 0.512417. It makes clear that hydrophobic interaction has the most effect on the gas-chromatographic relative retention time of PCNs, and electrostatic interaction has the secondary impact.

3.2.2 QSPR model of log$PL$

The optimum QSPR models of log$PL$ are deemed to be created by SMR-PLS method with 3 independent variables, i.e., electrostatic interaction item $V_{20}$ and steric interaction items $V_{56}$, $V_{58}$. This model just adopts two principal components explaining 99.72% square error of Y variable and 99.66% in cross-validation, which suggests that it uses 3D-HoVAIF descriptors superior in both internal estimation ability and external predictabilities compared with the models established by Puzyn et al.[20] ($R^2 =$
The plot predicted by the optimum QSPR models and observed log$P_L$ values are shown in Fig. 1(b).

The best SMR-PLS model equation (Eq. 2) was selected for further analysis. In this model, $V_{58}$ means steric interaction between the $sp^2$ hybrid C and H atoms, $V_{56}$ represents the steric interaction between H and H atoms, $V_{20}$ stands for the electrostatic interaction of $sp^2$ hybrid C and $sp^2$ hybrid C atoms, and so on. The values of VIP of these three variables are $V_{58}$ 1.02292, $V_{20}$ 1.00199 and $V_{56}$ 0.974501, respectively, so the steric interaction has the most effect on the vapour pressures of PCNs, and electrostatic interaction has the secondary impact.

### 3.2.3 QSPR model of log$K_{OA}$

The optimum QSPR model of log$K_{OA}$ by SMR-PLS includes 4 independent variables: $V_3$, $V_{10}$, $V_{58}$, $V_{165}$, and 2 principal components explaining 99.35% variance of the Y variables in contrast with 99.10% by the cross-validation. It exhibits a good prediction ability and stability, similar to the model found by Puzyn et al.$^{[20]}$ ($R^2 = 0.999, Q^2 = 0.988$). Fig. 1(c) shows the plot predicted by the optimum QSPR models and observed log$K_{OA}$ values.

We select the best SMR-PLS model equation (Eq. 3) for further analysis. In this model, $V_3$ means the electrostatic interaction between $sp^2$ hybrid C and H atoms, $V_{10}$ represents the electrostatic interaction between H and Cl atoms, $V_{58}$ is the steric interaction between $sp^2$ hybrid C and H atoms, $V_{165}$ stands for the hydrophobic interaction of Cl and Cl atoms, etc. The values of VIP of these four variables are $V_{165}$ 1.03554, $V_{58}$ 1.00156, $V_3$ 0.995346 and $V_{10}$ 0.966346, as we can find that hydrophobic interaction has the most effect on vapour pressures of PCNs, and steric and electrostatic interactions have the secondary impact.

### 3.2.4 QSPR model of log$K_{OW}$

The optimum QSPR model of log$K_{OW}$ includes six variables, i.e., electrostatic interaction items $V_1$ and $V_3$ together with steric interaction items $V_{82}$ and $V_{110}$. This model adopts three significant principal components to explain 93.11% variance of Y variables in contrast with 91.05% by cross-validation. The result shows this model exhibits quite satisfactory prediction ability and stability. In fact, this model is similar to the models constructed by Puzyn et al.$^{[38]}$. The plot predicted by the optimum QSPR model and observed log$S$ values are presented in Fig. 1(e).

In the optimum model, $V_1$ stands for the electrostatic interaction between H and H atoms, $V_3$ is the electrostatic interaction of $sp^2$ hybrid C and H atoms, $V_{82}$ is the steric interaction of $sp^2$ hybrid C and $sp^2$ hybrid C atoms, $V_{110}$ represents the steric interaction of Cl and Cl atoms, and $V_{165}$ shows the hydrophobic interaction of Cl and Cl atoms. The values of VIP of these six variables in sequence are $V_{165}$ 1.22202, $V_1$ 1.16178, $V_{85}$ 1.07558, $V_{58}$ 1.05911, $V_{165}$ 0.816118 and $V_{65}$ 0.460778. Conclusion could be drawn that the electrostatic interaction and steric interaction items are more significant and contribute more to the log$K_{OW}$ of PCNs, especially the electrostatic interaction items, and the hydrophobic interaction items are less influential for the log$K_{OW}$ of PCNs.

### 3.2.5 QSPR model of log$S$

The optimum model includes four variables, i.e., electrostatic interaction items $V_1$ and $V_3$ together with steric interaction items $V_{82}$ and $V_{110}$. This model adopts three significant principal components to explain 93.11% variance of Y variables in contrast with 91.05% by cross-validation. The result shows this model exhibits quite satisfactory prediction ability and stability. In fact, this model is similar to the models constructed by Puzyn et al.$^{[38]}$. The plot predicted by the optimum QSPR model and observed log$S$ values are presented in Fig. 1(e).
1.06147 and \( V_{110} \) 0.72191. The fact that the most contributive top four items including two electrostatic interaction items and two steric interaction items indicated an intimate relationship between the logarithm of aqueous solubility for PCNs, especially the steric interactions. Here the hydrophobic interaction items are not introduced.

3.2.6 QSAR model of \(-\log\text{EROD}\)

The optimum model just includes one variable, electrostatic interaction item \( V_{27} \), which can explain 81.94% variance of \( Y \) variables in contrast with 79.36% by cross-validation. Although low, the cross-validated \( Q^2 \) validation of this model is above 0.50, which indicates the model has an acceptable level of prediction ability and stability. This model is similar to that established by Falandysz et al.\(^{[39]} \). (\( R^2 = 0.823 \)). The plot predicted by the optimum QSAR model and observed \(-\log\text{EROD}\) values is presented in Fig. 1(f).

In this model, \( V_{27} \) means electrostatic interaction of \( sp^2 \) hybrid C and Cl atoms, so we can find the electrostatic interaction has most effect on the induction of ethoxyresorufin \( O\)-deethylase (EROD) for PCNs.

Hence, we take the optimum models to predict the \( RRT \), \( \log P_L \), \( \log K_{OA} \), \( \log K_OW \), \( \log S \) and \( -\log\text{EROD} \) values for all PCNs. The predicted values for all PCNs, including those whose experimentally determined physico-chemical properties and biological activity are unavailable, are listed in Table S1 of the supplementary materials.

### Table 1. Ten Atomic Types and Their 55 Types of Atomic Interactions in 3D-HoVAIF\(^a\)

<table>
<thead>
<tr>
<th>No.</th>
<th>Atoms types</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H</td>
<td>( V_{1=55n} )</td>
<td>( V_{2=55n} )</td>
<td>( V_{3=55n} )</td>
<td>( V_{4=55n} )</td>
<td>( V_{5=55n} )</td>
<td>( V_{6=55n} )</td>
<td>( V_{7=55n} )</td>
<td>( V_{8=55n} )</td>
<td>( V_{9=55n} )</td>
<td>( V_{10=55n} )</td>
</tr>
<tr>
<td>2</td>
<td>C(( sp^3 ))</td>
<td>( V_{11=55n} )</td>
<td>( V_{12=55n} )</td>
<td>( V_{13=55n} )</td>
<td>( V_{14=55n} )</td>
<td>( V_{15=55n} )</td>
<td>( V_{16=55n} )</td>
<td>( V_{17=55n} )</td>
<td>( V_{18=55n} )</td>
<td>( V_{19=55n} )</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>C(( sp^2 ))</td>
<td>( V_{20=55n} )</td>
<td>( V_{21=55n} )</td>
<td>( V_{22=55n} )</td>
<td>( V_{23=55n} )</td>
<td>( V_{24=55n} )</td>
<td>( V_{25=55n} )</td>
<td>( V_{26=55n} )</td>
<td>( V_{27=55n} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>C(( sp ))</td>
<td>( V_{28=55n} )</td>
<td>( V_{29=55n} )</td>
<td>( V_{30=55n} )</td>
<td>( V_{31=55n} )</td>
<td>( V_{32=55n} )</td>
<td>( V_{33=55n} )</td>
<td>( V_{34=55n} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>N/P(( sp^3 ))</td>
<td>( V_{35=55n} )</td>
<td>( V_{36=55n} )</td>
<td>( V_{37=55n} )</td>
<td>( V_{38=55n} )</td>
<td>( V_{39=55n} )</td>
<td>( V_{40=55n} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>N/P(( sp^2 ))</td>
<td>( V_{41=55n} )</td>
<td>( V_{42=55n} )</td>
<td>( V_{43=55n} )</td>
<td>( V_{44=55n} )</td>
<td>( V_{45=55n} )</td>
<td>( V_{46=55n} )</td>
<td>( V_{47=55n} )</td>
<td>( V_{48=55n} )</td>
<td>( V_{49=55n} )</td>
<td>( V_{50=55n} )</td>
</tr>
</tbody>
</table>

*\( n = 0 \), the electrostatic interaction items; \( n = 1 \), the steric interaction items; \( n = 2 \), the hydrophobic interaction items

<table>
<thead>
<tr>
<th>No.</th>
<th>Equations</th>
<th>N</th>
<th>PC</th>
<th>( R^2 )</th>
<th>( Q^2 )</th>
<th>RMSEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( RRT = 8.829 + 0.314V_3 + 0.316V_{113} - 0.145V_{128} - 0.315V_{177} )</td>
<td>62</td>
<td>1</td>
<td>0.9901</td>
<td>0.9895</td>
<td>23.892</td>
</tr>
<tr>
<td>2</td>
<td>( \log P_L = -1.388 + 0.443V_{30} + 0.207V_{56} + 0.756V_{58} )</td>
<td>17</td>
<td>2</td>
<td>0.9972</td>
<td>0.9966</td>
<td>0.0788</td>
</tr>
<tr>
<td>3</td>
<td>( \log K_{OA} = 9.481 + 0.169V_1 + 0.055V_{13} + 0.215V_{36} - 0.579V_{165} )</td>
<td>24</td>
<td>2</td>
<td>0.9935</td>
<td>0.9910</td>
<td>0.0780</td>
</tr>
<tr>
<td>4</td>
<td>( \log K_OW = 6.271 - 3.354V_1 - 3.866V_{28} - 0.072V_{38} - 0.101V_{65} + 1.764V_{82} )</td>
<td>16</td>
<td>4</td>
<td>0.9791</td>
<td>0.9559</td>
<td>0.1362</td>
</tr>
<tr>
<td>5</td>
<td>( \log S = 1.380 - 0.516V_1 + 0.323V_3 - 2.097V_{42} - 0.458V_{110} )</td>
<td>15</td>
<td>3</td>
<td>0.9311</td>
<td>0.9105</td>
<td>0.3731</td>
</tr>
<tr>
<td>6</td>
<td>( -\log\text{EROD} = 3.183 + 0.905V_{27} )</td>
<td>17</td>
<td>1</td>
<td>0.8194</td>
<td>0.7936</td>
<td>0.7228</td>
</tr>
</tbody>
</table>

4 CONCLUSION

Quantitative structure-property (activity) relationships for \( RRT \), \( \log P_L \), \( \log K_{OA} \), \( \log K_OW \), \( \log S \) and \( -\log\text{EROD} \) of PCNs have been established with good correlation coefficients and cross-validated correlation coefficients. It has been shown that the models have good prediction capability and favourable stability and 3D-HoVAIF descriptors can be well used to characterize the molecular structure information and express the quantitative structure-property (activity) relationships of PCNs.
REFERENCES


(10) Engwall, M.; Brunström, B.; Jakobsson, E. Ethoxyresorufin o-deethylation (EROD) and aryl hydrocarbon hydroxyase (AHH)-inducing potency and lethality of chlorinated naphthalenes in chicken (Gallus domesticus) and eider duck (Somatera mollissima) embryos. *Arch. Toxicol. 1994, 68, 37–42.*


(23) Jing, J. H.; Liang, G. Z.; Mei, H.; Zhang, Q. X.; Li, Z. L.; Lv, F. L. QSAR studies on influenza neuraminidase inhibitors — acetylthiourea analogue.


