

Computational estimation of logarithm of octanol/air partition coefficients and subcooled vapour pressures for each of 75 chloronaphthalene congeners

Tomasz PUZYN¹, Jerzy Falandysz¹, Paweł Rostkowski¹, Sławomir Piliszek¹,
Agata Wilczyńska¹

¹University of Gdańsk, Gdańsk

Introduction

Polychlorinated naphthalenes (PCNs, CNs) are known persistent organic pollutants, contaminating natural ecosystems in effect of technical human activity¹⁻⁶. Toxic effects induced by individual congeners of PCNs are reported elsewhere^{7,8}. Great risk of these chemical compounds is additionally connected with their excellent ability to be transported *via* atmosphere from a source to the remote regions on the Glob. Chloronaphthalene congeners had been found in Arctic regions at significant level in spite of the fact, that they had never been synthesized there, and also thermal processes like municipal waste incineration or domestic heating (other possible sources of PCNs in the environment) were not so intensive there^{1,3-6,8}.

In 1996 F. Wania and D. Mackay have formulated some empirical rules, which have been very useful in estimation and modeling of environmental transport processes of persistent organic pollutants like PCNs. Two very important physico-chemical parameters in the theory of global distillation and cold condensation are: logarithm of n-octanol/air partition coefficient ($\log K_{OA}$) and logarithm of subcooled vapour pressure ($\log P_L$)⁹.

Values of $\log K_{OA}$ and $\log P_L$ in standard procedures are determined by means of chromatographic methods¹⁰⁻¹². In order to reduce costs and number of experiments, we have proposed simple computational method of estimation $\log K_{OA}$ and $\log P_L$.

Materials and Methods

The method combined quantum-chemical and chemometric tools. In the first step, the matrix of quantum-chemical, thermodynamic, and topological (X-matrix) descriptors was computed. The role of the X-matrix was to describe significant variability (similarities and differences) between individual congeners.

In this study we used about 190 descriptors, so there is unable to present the full list of the descriptors. 16 quantum-chemical descriptors, such as: average molecular polarizability, dipole moment, the most positive and negative partial charge on atoms, ionization potential, electron affinity, molecular volume were computed on the level of density functional theory (DFT) using B3LYP hybrid functional in Popple-type triple zeta basis 6-311++G** (polarization and diffuse functions were added). In each case, polarizability tensors were diagonalized, and average

polarizability was calculated as a mean of eigenvalues from the polarizability tensor. Ionization potential was determined as a difference between total energy of molecular cation (+1) and the neutral molecule. Analogically, electron affinity in this study was a difference between energy of molecular anion (-1) and the neutral molecule¹³.

Thermodynamic descriptors (4): standard heat of formation in gas phase, standard Gibbs free energy, head capacity in constant volume, entropy were calculated based on frequency analysis on the level of B3LYP/6-311++G** using the algorithm proposed by Otherski^{14,15}. All quantum-mechanical and thermo-chemical computation were done using GAUSSIAN 03 package¹⁶.

Topological descriptors (170), such as: connectivity indices, topological indices, indices of neighborhood symmetry were calculated by means of DRAGON software. The full list of those descriptors is available on the DRAGON website^{17,18}.

After preparing of X-matrix, the matrix was autoscaled, and total number of descriptors was reduced using principal component analysis (PCA). In PCA, new orthogonal descriptors (called "principal components") were formed as a linear combination of original, strongly correlated descriptors. Principal components were extracted in order of decreasing variability. In details, the method is described in many chemometric handbooks and publications¹⁹⁻²¹. In our study, we additionally used normalized Varimax rotation¹⁹, so there are used varivectors (Vs) instead of principal components (PCs).

In the final step, the linear modeling method – multiple regression of principal components (PCR) was used to define quantitative relationships between structural information (X-matrix) and physico-chemical properties ($\log K_{OA}$ and $\log P_L$). These relationships might be described by the following equation (1).

$$Y = b_0 + b_1 V_1 + b_2 V_2 + \dots + b_n V_n$$

where: Y – is estimated property, b_0 – intercept, b_i – i-th parameter of the model, x_i – i-th varivector.

Within the model, independent variables (Vs) were taken from principal component analysis, while the values of dependent variable were collected from published experimental data^{10-12,22}. Models were defined using backward stepwise technique²³. The regression coefficients were found by means of the least squares method^{20,23}. After identification, efficiency of the model was evaluated by cross-validation (leave-by-one)¹⁹, and finally values of $\log K_{OA}$ and $\log P_L$ was estimated for each of 75 chloronaphthalene congeners.

Results and Discussion

Principal component analysis of the structural matrix: Principal component analysis (PCA) of the matrix of 190 descriptors with normalized Varimax rotation resulted in five factors describing together 89 % (69 % + 8 % + 4 % + 3 % + 2 %) of the total structural variability in chloronaphthalene congener set.

Full list of the factor loadings are not shown. The first varivector (**V1**) described molecular size. The second varivector (**V2**) was connected with atomic composition and symmetry of zero and first order neighborhood in molecules of individual CNs. The third (**V3**) was constructed from indices describing neighborhood symmetry of higher orders. Molecular properties depended on

sigma electrons were cumulated in the fourth varivector (**V4**). The fifth (**V5**) factor described ionization potential of the molecule.

Analysis of the score plots (not shown) suggested applicability of the used domain into whole set of chloronaphthalene congeners.

Modeling and estimation of $\log K_{OA}$: Varivectors extracted in the pervious step were next used as the independent variables in modeling of $\log K_{OA}$. Backward stepwise regression resulted in the linear model described by the following equation:

$$\log K_{OA} = 8.26 (0.03) + 1.11 (0.03) V1 + 0.15 (0.02) V2 + 0.15 (0.02) V4$$

$$n = 27; R^2 = 0.999; Q^2_{CV} = 0.988; p < 0.0000; s = 0.15; F = 736$$

As it could be observed, there were very good values of determination and cross-validation coefficients, which suggest good stability of the model. The plot of observed vs. predicted values of $\log K_{OW}$ (figure 1) also confirmed linear character of the model and very good fit.

From this model unknown values of logarithm of n-octanol / air partition coefficients for each of 75 possible chloronaphthalene congeners were estimated. Results of the estimation are presented in the table 1.

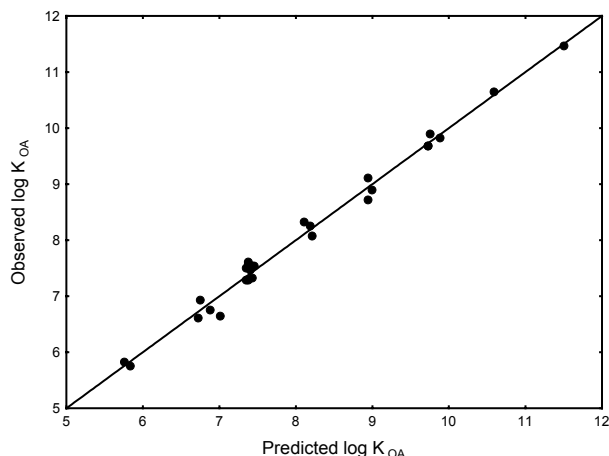


Figure 1. Plot of observed vs. predicted values of $\log K_{OA}$ for the test set.

Modeling and estimation of $\log P_L$: Analogous algorithm was used in order to estimate values of logarithm of subcooled vapour pressures for individual CNs.

The same varivectors were chosen by means of backward stepwise regression to construct a model:

$$\log P_L = -1.72 (0.03) - 0.93 (0.03) V1 - 0.09 (0.02) V2 - 0.16 (0.02) V4$$

$$n = 17; R^2 = 0.994; Q^2_{CV} = 0.990; p < 0.0000; s = 0.12; F = 739$$

The plot of observed vs. predicted values of $\log P_L$ is presented in the figure 2. The model also are characterized by high values of R^2 and Q^2_{CV} , which determined use of the model in prediction of $\log P_L$ for non-tested chloronaphthalenes.

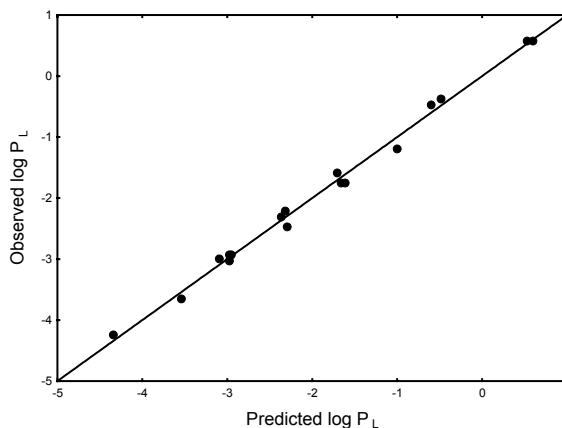


Figure 2. Plot of observed vs. predicted values of $\log P_L$ for the test set.

Table 1. Observed and predicted values of $\log K_{OW}$ and $\log P_L$ for PCNs

No.	Congener	Varivectors			Log K_{OA}		Log P_L	
		V1	V2	V4	Obs.	Pred.	Obs.	Pred.
1	1-CN	-1.61	0.90	-5.62	5.81	5.76	0.57	0.60
2	2-CN	-1.53	0.79	-5.63	5.76	5.84	0.58	0.53
3	1,2-diCN	-1.40	0.31	0.92	6.76	6.89	-0.48	-0.60
4	1,3-diCN	-1.33	0.24	0.89		6.96		-0.65
5	1,4-diCN	-1.47	0.15	0.67	6.92	6.75	-0.38	-0.47
6	1,5-diCN	-1.50	0.21	0.69	6.62	6.73		-0.45
7	1,6-diCN	-1.32	0.26	0.79		6.95		-0.64
8	1,7-diCN	-1.33	0.24	0.94		6.96		-0.66
9	1,8-diCN	-1.40	0.31	0.72		6.86		-0.56
10	2,3-diCN	-1.41	0.16	0.73		6.83		-0.54
11	2,6-diCN	-1.37	0.05	0.71		6.86		-0.57
12	2,7-diCN	-1.22	0.05	0.72	6.66	7.02		-0.71
13	1,2,3-triCN	-0.86	0.29	0.28	7.62	7.39	-1.19	-0.99
14	1,2,4-triCN	-0.87	0.18	0.32	7.27	7.36		-0.97
15	1,2,5-triCN	-0.84	0.18	0.31	7.45	7.41		-1.01
16	1,2,6-triCN	-0.92	0.34	0.40	7.50	7.35		-0.96
17	1,2,7-triCN	-0.80	0.07	0.45	7.54	7.45		-1.05
18	1,2,8-triCN	-0.78	0.14	0.34		7.47		-1.06
19	1,3,5-triCN	-0.82	0.15	0.34	7.32	7.42		-1.02
20	1,3,6-triCN	-0.83	0.05	0.37		7.41		-1.02

PHYSICO-CHEMICAL PROPERTIES, DISTRIBUTION AND MODELLING

21	1,3,7-triCN	-0.86	0.24	0.44		7.41		-1.01
22	1,3,8-triCN	-0.78	0.11	0.39		7.46		-1.06
23	1,4,5-triCN	-0.82	0.08	0.23	7.56	7.39		-1.00
24	1,4,6-triCN	-0.86	0.12	0.32	7.27	7.37		-0.99
25	1,6,7-triCN	-0.81	0.18	0.34	7.54	7.44		-1.04
26	2,3,6-triCN	-0.76	0.11	0.27		7.47		-1.06
27	1,2,3,4-tetraCN	-0.06	-0.40	-0.12	8.33	8.11	-1.76	-1.61
28	1,2,3,5-tetraCN	-0.06	-0.16	0.08	8.26	8.18	-1.75	-1.67
29	1,2,3,6-tetraCN	-0.03	-0.11	0.04		8.21		-1.68
30	1,2,3,7-tetraCN	0.03	-0.27	0.11		8.27		-1.75
31	1,2,3,8-tetraCN	0.01	-0.10	0.09		8.27		-1.73
32	1,2,4,5-tetraCN	-0.03	-0.23	0.14		8.21		-1.69
33	1,2,4,6-tetraCN	-0.12	-0.06	0.14		8.14		-1.62
34	1,2,4,7-tetraCN	-0.04	-0.08	0.18	8.08	8.23	-1.61	-1.70
35	1,2,4,8-tetraCN	-0.06	-0.25	0.09		8.17		-1.66
36	1,2,5,6-tetraCN	-0.18	-0.25	-0.01		8.02		-1.53
37	1,2,5,7-tetraCN	-0.07	-0.03	0.15		8.19		-1.67
38	1,2,5,8-tetraCN	-0.07	-0.22	0.07		8.16		-1.65
39	1,2,6,7-tetraCN	-0.05	-0.19	0.17		8.20		-1.68
40	1,2,6,8-tetraCN	-0.06	0.02	0.14		8.22		-1.69
41	1,2,7,8-tetraCN	-0.06	-0.26	0.04		8.16		-1.65
42	1,3,5,7-tetraCN	-0.24	-0.24	0.03		7.96		-1.48
43	1,3,5,8-tetraCN	-0.06	-0.22	0.10		8.17		-1.66
44	1,3,6,7-tetraCN	-0.09	-0.08	0.06		8.15		-1.63
45	1,3,6,8-tetraCN	-0.14	-0.25	0.12		8.09		-1.59
46	1,4,5,8-tetraCN	-0.17	-0.61	-0.09		7.97		-1.50
47	1,4,6,7-tetraCN	-0.24	-0.32	0.04		7.95		-1.47
48	2,3,6,7-tetraCN	-0.18	-0.48	-0.09		7.98		-1.50
49	1,2,3,4,5-pentaCN	0.67	-0.35	-0.07		8.94		-2.30
50	1,2,3,4,6-pentaCN	0.74	-0.47	-0.08	8.90	9.00	-2.32	-2.35
51	1,2,3,5,6-pentaCN	0.76	-0.49	-0.10		9.02		-2.37
52	1,2,3,5,7-pentaCN	0.69	-0.43	-0.04	8.72	8.95	-2.23	-2.31
53	1,2,3,5,8-pentaCN	0.69	-0.48	-0.09	9.12	8.94	-2.48	-2.30
54	1,2,3,6,7-pentaCN	0.65	-0.37	-0.14		8.91		-2.27
55	1,2,3,6,8-pentaCN	0.71	-0.43	-0.04		8.98		-2.34
56	1,2,3,7,8-pentaCN	0.77	-0.46	-0.06		9.04		-2.39
57	1,2,4,5,6-pentaCN	0.66	-0.42	0.00		8.93		-2.30
58	1,2,4,5,7-pentaCN	0.74	-0.67	-0.03		8.97		-2.34
59	1,2,4,5,8-pentaCN	0.62	-0.49	0.00		8.88		-2.26
60	1,2,4,6,7-pentaCN	0.64	-0.40	-0.10		8.90		-2.27
61	1,2,4,6,8-pentaCN	0.70	-0.52	-0.06		8.95		-2.31
62	1,2,4,7,8-pentaCN	0.71	-0.61	0.04		8.97		-2.34
63	1,2,3,4,5,6-heksaCN	1.60	-0.78	-0.27		9.88		-3.09
64	1,2,3,4,5,7-heksaCN	1.61	-0.69	-0.38		9.89		-3.10
65	1,2,3,4,5,8-heksaCN	1.48	-0.88	-0.22		9.73		-2.98
66	1,2,3,4,6,7-heksaCN	1.51	-0.93	-0.50	9.69	9.73	-2.92	-2.96

67	1,2,3,5,6,7-heksaCN	1.51	-0.91	-0.43	9.69	9.73	-2.92	-2.97
68	1,2,3,5,6,8-heksaCN	1.60	-0.74	-0.38		9.87		-3.08
69	1,2,3,5,7,8-heksaCN	1.56	-0.70	-0.07	9.82	9.88	-3.00	-3.10
70	1,2,3,6,7,8-heksaCN	1.59	-0.77	-0.42		9.84		-3.06
71	1,2,4,5,6,8-heksaCN	1.47	-0.82	-0.15	9.88	9.74	-3.02	-2.99
72	1,2,4,5,7,8-heksaCN	1.55	-0.75	-0.10		9.86		-3.08
73	1,2,3,4,5,6,7-heptaCN	1.56	4.32	-0.18	10.63	10.61	-3.65	-3.53
74	1,2,3,4,5,6,8-heptaCN	1.55	4.33	-0.10		10.62		-3.54
75	1,2,3,4,5,6,7,8-oktaCN	2.13	4.80	1.20	11.48	11.52	-4.25	-4.32

Conclusions: QSPR prediction of log K_{OW} and log P_L for persistent organic pollutants like PCNs is an issue intensively study in recent time. Staikowa, Wania and Donaldson proposed mean polarizability as a single independent variable in linear regression model¹¹. Standard error of these models were 0.31 for log K_{OA} and 0.32 for log P_L . Adding ZZ Cartesian component of the quadrupole moment tensor to the equation, they have got a small improvement of the prediction accuracy ($s = 0.13$ for log K_{OA} and $s = 0.23$ for log P_L).

Presented PCR model for log K_{OA} predict with similar standard error ($s = 0.15$), while the model for log P_L predicts more precisely ($s = 0.12$) than models proposed by cited authors.

However, as it could be observed in the table 1, by means of QSPR procedure it is possible to simply describe differences in log K_{OW} and log P_L rather between homologue groups, than between individual congeners. In many cases standard error of the model is higher than real differences between individual congeners. More accurate models might require more precise measures of the experimental values for the training set and/or developing new and more adequate structural descriptors.

Acknowledgements

This study was supported by Ministry of Science and Informatics (grant no. KBN 1128/T09/2003/24). Computations were carried out using computers in TASK - Academic Computer Center in Gdańsk.

References

- [1] Falandysz J., Kawano M., Ueda M., Matsuda M., Kannan K., Giesy J.P., Wakimoto T.J. Environ. Sci. Health 35, 281.
- [2] Falandysz J., Rappe C. (1996) Environ. Sci. Technol. 30, 3362.
- [3] Harner T., Lee R.G.M., Jones K.C. (2000) Environ. Sci. Technol. 34, 3137.
- [4] Helm P.A., Bidleman T.F. (2003) Environ. Sci. Technol. 37, 1075.
- [5] Horii Y., Falandysz J., Hanari N., Rostkowski P., Puzyn T., Okada M., Amano K., Naya T., Taniyasu S., Yamashita N., (2004) J. Environ. Sci. Health 39A, 587.
- [6] Yamashita N., Taniyasu S., Hanari N., Horii Y., Falandysz J. (2003) J. Environ. Sci. Health 38A, 1745.
- [7] Blankenship A., Kannan K., Villalobos S., Villeneuve D., Falandysz J., Imagawa T., Jakobsson E., Giesy J. (2000) Environ. Sci. Technol. 34, 3153.
- [8] Falandysz J. (1998) Environ. Pollut. 10, 77.
- [9] Wania F., Mackay D. (1996) Environ. Sci. and Technol. 30, 390A.

- [10] Lei Y.D., Wania F., Shiu W.Y. (1999) Chem. Eng. Data 44, 577.
- [11] Staikova M., Wania F., Donaldson, D.J. (2004) Atmos. Environ. 38, 213.
- [12] Su Y., Lei Y.D., Daly G. L., Wania F. (2002) J. Chem. Eng. Data 47, 449.
- [13] Piela L. (2003) Idee chemii kwantowej (*in Polish*), Wydawnictwo Naukowe PWN, Warszawa, Poland.
- [14] Otherski J.W. (2000). Thermochemistry in Gaussian. Gaussian Inc. <http://gaussian.com>
- [15] Curtis L.A, Redfern P.C. (1997) J. Chem. Phys. 106, 1063.
- [16] Frisch M.J., Trucks G.W., Schlegel H.B., Scuseria G.E., Robb M.A., Cheeseman J.R., Montgomery J.A., Vreven T., Kudin K.N., Burant J.C., Millam J.M., Iyengar S.S., Tomasi J., Barone V., Mennucci B., Cossi M., Scalmani G., Rega N., Petersson G.A., Nakatsuji H., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima Y., Honda Y., Kitao O., Nakai H., Klene M., Li X., Knox J.E., Hratchian H.P., Cross J.B., Adamo C., Jaramillo J., Gomperts R., Stratmann R.E., Yazyev O., Austin A.J., Cammi R., Pomelli C., Ochterski J.W., Ayala P.Y., Morokuma K., Voth G.A., Salvador P., Dannenberg J.J., Zakrzewski V.G., Dapprich S., Daniels A.D., Strain M.C., Farkas O., Malick D.K., Rabuck A.D., Raghavachari K., Foresman J.B., Ortiz J.V., Cui Q., Baboul A.G., Clifford S., Cioslowski J., Stefanov B.B., Liu G., Liashenko A., Piskorz P., Komaromi I., Martin R.L., Fox D.J., Keith T., Al-Laham M.A., Peng C.Y., Nanayakkara A., Challacombe M., Gill P.M.W., Johnson B., Chen W., Wong M.W., Gonzalez C., and Pople J.A. (2003) Gaussian 03, Gaussian Inc., Pittsburgh PA.
- [17] Todeschini R., Consonni V., Mauri A., Pavan M. (2003) DRAGON Software version 3. Milano Chemometrics. Milano, Italy.
- [18] Todeschini R., Consoni V. (2000) Handbook of molecular descriptors, Wiley, Weinheim.
- [19] Niemi G.J. (1990) Multivariate analysis and QSAR: applications of principal component analysis, In: Karcher W., Devillers J. (eds.) Principal applications of quantitative structure – activity relationships (QSAR) in environmental chemistry and toxicology, Kluwer Academic Publishers, Netherlands.
- [20] Mazerski J. (2000) Podstawy chemometrii (*in Polish*), Wydawnictwo PG, Gdansk, Poland.
- [21] Sharaf M.A, Illman D.H., Kowalski B.R. (1986) Chemometrics, John Wiley & Sons Inc., USA.
- [22] Harner T., Bidleman T.F. (1998) J. Chem. Eng. Data 43, 40.
- [23] Dobosz M. (2001) Wspomagana komputerowo statystyczna analiza wyników danych (*in Polish*), Akademicka Oficyna Wydawnicza EXIT, Warszawa, Poland.