THE USE OF MULTIVARIATE EXPERIMENTAL DESIGN IN THE SELECTION OF POLYCHLORINATED NAPHTHALENES FOR SCREENING OF BIOCHEMICAL AND TOXICOLOGICAL EFFECTS

Mats Tysklind , Kristina Elg[1], Patrik L. Andersson, Bert van Bavel and Peter Haglund
Institute of Environmental Chemistry, Umeå University, S-901 87 Umeå, Sweden
[1] Present address: MoDo Research and Development, S-891 80 Örnsköldsvik, Sweden

1. INTRODUCTION

Polychlorinated naphthalenes (PCNs) are one group of halogenated persistent aromatic compounds consisting of 75 congeners with different degree of chlorination and substitution patterns. PCNs were extensively used in the manufacture of electric insulation in the 1930s to 1950s, but the production had significantly fallen in the late 1970s due to the replacement of PCNs by a variety of substitutents [1]. Today, PCNs are found in basically all compartments of the abiotic and biotic systems, including soil, sediment, human blood and human mothers' milk [2,3]. However, in biota a selective and structure-dependent metabolism and retention of specific PCN congeners are found [3].

In this paper we present the use of multivariate experimental design in combination with a multivariate physico-chemical characterisation of the 49 tetra- through octachlorinated PCNs. A well-balanced set of 20 congeners are presented and suggested to be used in future screening of biochemical and toxicological effects of the PCNs.

2. MATERIALS AND METHODS

Physico-chemical descriptors. Semi-empirical parameters were generated using the Austin Model 1 (AM1) semi-empirical quantum mechanical method [4] included in the HyperChem^{TM} program package [5]. See Table 1. The parameters and calculations are presented elsewhere [6]. In addition, physico-chemical descriptors were captured from the literature and other parameters generated depicting different substitution patterns. The numbering of the PCNs is described in Table 2 [10].

Data analysis. Multivariate projection methods, such as principal component analysis (PCA)[11], provide a tool by which the physico-chemical data may be analysed and interpreted. PCA approximate the systematic variation within the included variables by a few orthogonal vectors, the principal components (PCs). PCA gives an overview of the dominant patterns and trends in the data, easily interpreted when visualized in pictures. In this study, the primary scope of the PCA model is to express the chemical properties in the original data set into a few new properties, so-called principal properties.

Experimental design. Statistical experimental design is a tool for selection of appropriate sets of compounds. Statistical designs, such as factorial design (FD) and fractional factorial design (FFD) produce well-balanced so-called training sets through distribution of the compounds

Table 1. Physico-chemical parameters used in the multivariate characterisation of the tetra- through octachlorinated PCNs.

Semi-empirical parameters:	*Literature data:*
Total energy (ET)	Melting point (Mp)[7]
Binding energy (EB)	GC retention times - DB5 (RRT1 & 2)[8,9]
Isolated atomic energy (EIA)	
Electronic energy (EE)	*Other parameters:*
Core-core interaction energy (ECCI)	Molecular weight (Mw)
Heat of formation (Hf)	Bit pattern[a] (C1-C8)
Dipole moments - point charge (PCx, PCy, PCz, PCtot)	Vicinal hydrogen atoms[b] (Vic)
Dipole moments - hybridization (Hybx, Hyby, Hybz, Hybtot)	
Ionization potential - HOMO (Ip)	
Electron affinity - LUMO (Ea)	
Absolute hardness (h)	
Absolute electronegativity (c)	

[a)] Variables depicting the substitution pattern, "1"or "0" depending on chlorine atom substitution or not . [b)] Variables describing the number of vicinal hydrogen atoms, viz. the total number (vic), the number on the 1st ring (vic1), and the number the 2nd ring (vic2).

(here: PCNs) over the entire chemical domain, because it generates sets of compounds by varying all design factors (here: principal properties) simultaneously. Both FD and FFD have previously been used for selection of training sets of polychlorinated dibenzo-*p*-dioxins (PCDDs) and dibenzofurans (PCDFs)[11], and polychlorinated biphenyls (PCBs)[12].

Table 2. The numbering of the tetra- through octachlorinated PCNs [10].

No.	PCN	No.	PCN	No	PCN
27	1,2,3,4-TcCN	44	1,3,6,7-TcCN	61	1,2,4,6,8-PcCN
28	1,2,3,5-TcCN	45	1,3,6,8-TcCN	62	1,2,4,7,8-PeCN
29	1,2,3,6-TcCN	46	1,4,5,8-TcCN	63	1,2,3,4,5,6-HxCN
30	1,2,3,7-TcCN	47	1,4,6,7-TcCN	64	1,2,3,4,5,7-HxCN
31	1,2,3,8-TcCN	48	2,3,6,7-TcCN	65	1,2,3,4,5,8-HxCN
32	1,2,4,5-TcCN	49	1,2,3,4,5-PcCN	66	1,2,3,4,6,7-HxCN
33	1,2,4,6-TcCN	50	1,2,3,4,6-PcCN	67	1,2,3,4,6,8-HxCN
34	1,2,4,7-TcCN	51	1,2,3,5,6-PcCN	68	1,2,3,5,6,8-HxCN
35	1,2,4,8-TcCN	52	1,2,3,5,7-PcCN	69	1,2,3,5,7,8-HxCN
36	1,2,5,6-TcCN	53	1,2,3,5,8-PcCN	70	1,2,3,6,7,8-HxCN
37	1,2,5,7-TcCN	54	1,2,3,6,7-PcCN	71	1,2,4,5,6,8-HxCN
38	1,2,5,8-TcCN	55	1,2,3,6,8-PcCN	72	1,2,4,5,7,8-HxCN
39	1,2,6,7-TcCN	56	1,2,3,7,8-PcCN	73	1,2,3,4,5,6,7-HpCN
40	1,2,6,8-TcCN	57	1,2,4,5,6-PcCN	74	1,2,3,4,5,6,8-HpCN
41	1,2,7,8-TcCN	58	1,2,4,5,7-PcCN	75	1,2,3,4,5,6,7,8-OcCN
42	1,3,5,7-TcCN	59	1,2,4,5,8-PcCN		
43	1,3,5,8-TcCN	60	1,2,4,6,7-PcCN		

3. RESULTS AND DISCUSSION

The PCA of the 49 PCNs examined resulted in a 8-dimensional model, significant according to cross-validation, explaining 93 % of the variation. The first four PCs described 38%, 22%, 11%, and 7% of the variation, respectively. As seen in the score plot in Figure 1A, where the first PC is plotted vs. the second, the 49 PCNs are distributed in five groups, with the tetrachlorinated PCNs to the left and successively followed by the penta-, hexa-, hepta-, and octachlorinated congeners. Thus, the first PC reflects the degree of chlorination among the compounds, whereas the second and higher PCs resolve differing substitution patterns. See Figure 1B showing PC3 vs. PC4.

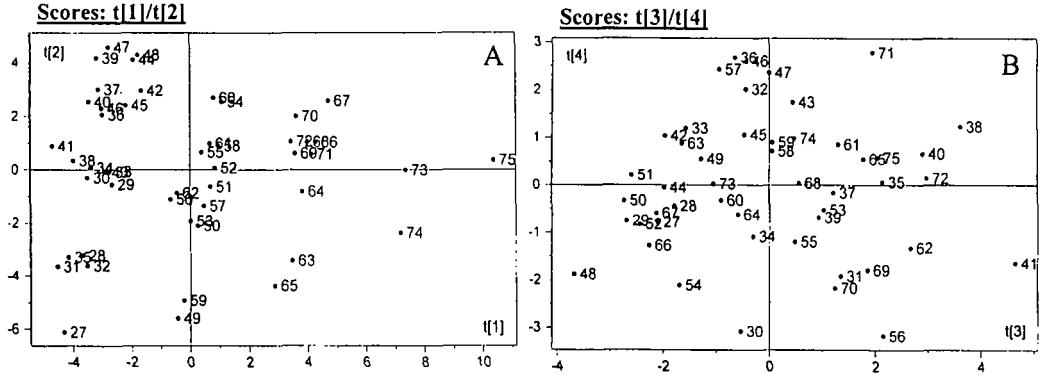

Figure 1. Score plots of PC1 (t1) vs. PC2 (t2) (Fig. 1A) and PC3 (t3) vs. PC4 (t4) (Fig.1B) for the 49 tetra- through octachlorinated PCNs. For the numbering of the congeners, see Table 2.

The interpretation of the corresponding loading plots in Figure 2 shows that the variables EB, ECCI, Ip, c, Mw, RRT1, RRT2, ET, EIA, EE and h dominate the first dimension, all describing molecular size/bulk. The second, third, and fourth dimensions are regulated by chlorine substitution related variables, viz. PC2: C7, Hyby, Hybz, vic2, PCy, PCz and Hybtot, PC3: C8, Hf, Hybx, C3, h, PCx, and Ea, and PC4: C4, C5, Pcx, PCtot, C2, C3 and C7.

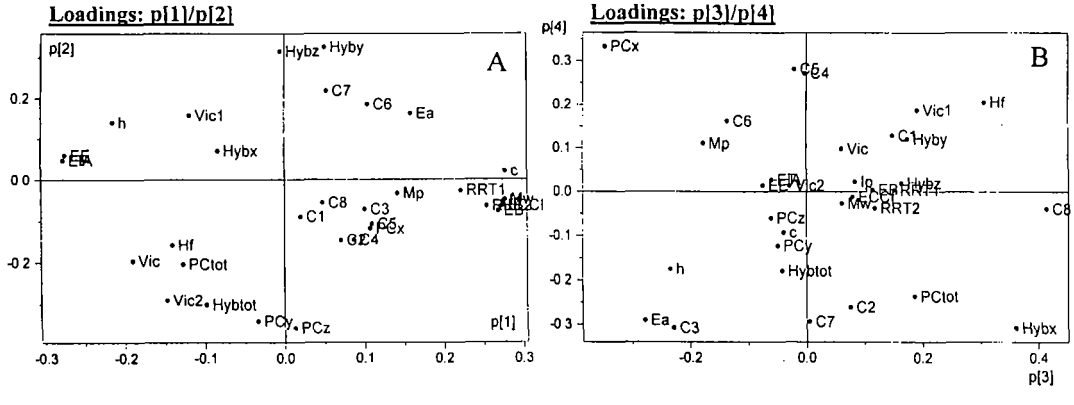

Figure 2. Loading plots of loading vector 1 vs. 2 (Fig. 2A) and loading vector 3 vs. 4 (Fig. 2B). For the identification of the variables, see Table 1.

In order to minimize the number of congeners for biological testing but include the major trends in the physico-chemical properties, the first four PCs were used as design variables in a 2^4-factorial design. The four PCs correspond to a representation of the 49 tetra- through octachlorinated PCNs in a sub-space with four dimensions. These four PCs account for as much as 78% of the variance in the data, i.e. the major part of the chemical information extracted from the original data matrix. The 2^4-FD generates 16 design levels and the congeners were grouped acccording to their sign and level of PC1 to PC4. As seen in Table 3, in most cases several congeners fit to one specific design level except level +/-/+/- with only one candidate and design level +/+/-/+/ with no candidate. The congeners on each design level were ranked, the optimal choise to the left, followed by the next best choice etc. (marked with I to VI in Table 3).

Table 3. Principal properties of the PCNs in combination with a 2^4-FD design.

No.	Design level[a]				PCNs[b,c]					
	t_1	t_2	t_3	t_4	I	II	III	IV	V	VI
1	-	-	-	-	**27**	28	29	30		
2	+	-	-	-	**64**	50				
3	-	+	-	-	**48**	34	44			
4	+	+	-	-	**66**	67	54	60	52	
5	-	-	+	-	**31**	56	62			
6	+	-	+	-	**53**					
7	-	+	+	-	**39**	37	41			
8	+	+	+	-	**70**	69	55			
9	-	-	-	+	**32**	33	49			
10	+	-	-	+	**63**	57	51	73		
11	-	+	-	+	**36**	45	42	46		
12	+	+	-	+	-					
13	-	-	+	+	**35**	43	59			
14	+	-	+	+	**65**	74				
15	-	+	+	+	**40**	38	47			
16	+	+	+	+	**72**	68	71	58	61	75

[a]/ The 2^4-factorial design generating a training set of 16 tetra- through octachlorinated PCNs. [b]/ the congeners to the left the optimal choice, followed by the next best choice etc. [c]/ for the numbering of the individual congeners, see Table 2[10].

In Table 4, the optimal choice of 16 congeners is shown. Design level +/+/-/+ with no candidate, as shown in Table 3, was substituted with PCN#67 from design level +/+/-/-. PCN#67, which fulfills the selection criteria in PC1 - PC3, has a low score value in PC4. The training set has been supplemented with three center points chosen to represent the chemical properties of congeners located in the interior part of the design. In addition, octachloronaphthlene was included due to "extreme" properties. In some cases, e.g. reduce the number of research animals, it can be appropriate to reduce the number of congeners for testing even more. Using a 2^{4-1} FFD, 10 congeners can be selected as a half-fraction of the 2^4-FD, see Table 4. Thus, in total 20 and 10 tetra- through octachlorinated PCNs, respectively, can be selected, for which future screening of biochemical and toxicological effects may be carried out.

Table 4. The optimal selection of PCNs according to the design are the following 20 congeners. Three center points and octaCN are included.

No.	t_1	t_2	t_3	t_4	PCNs
1	-	-	-	-	1,2,3,4-TeCN (27)[a]
2	+	-	-	-	1,2,3,4,5,7-HxCN (64)
3	-	+	-	-	2,3,6,7-TeCN (48)
4	+	+	-	-	1,2,3,4,6,7-HxCN (66)[a]
5	-	-	+	-	1,2,3,8-TeCN (31)
6	+	-	+	-	1,2,3,5,8-PeCN (53)[a]
7	-	+	+	-	1,2,6,7-TeCN (39)[a]
8	+	+	+	-	1,2,3,6,7,8-HxCN (70)
9	-	-	-	+	1,2,4,5-TeCN (32)
10	+	-	-	+	1,2,3,4,5,6-HxCN (63)[a]
11	-	+	-	+	1,2,5,6-TeCN (36)[a]
12	+	+	-	(+)	1,2,3,4,6,8-HxCN (67)
13	-	-	+	+	1,2,4,8-TeCN (35)[a]
14	+	-	+	+	1,2,3,4,5,6,8-HpCN (74)
15	-	+	+	+	1,2,6,8-TeCN (40)
16	+	+	+	+	1,2,4,5,7,8-HxCN (72)[a]
	Center points + octaCN				
17	**0**	**0**	**0**	**0**	1,3,5,8-TeCN (43)[a]
18	**0**	**0**	**0**	**0**	1,2,4,5,7-PeCN (58)
19	**0**	**0**	**0**	**0**	1,2,4,6,8-PeCN (61)[a]
20	**x**	**x**	**x**	**x**	1,2,3,4,5,6,7,8-OcCN (75)

[a] Congeners selected according to a 2^{4-1} fractional factorial design

4. CONCLUSIONS

The selection of the PCNs is facilitated using statistical experimental design, viz. a 2^4-factorial design and a 2^{4-1} fractional factorial design, which are informationally optimal protocols for designing experimental testing. This procedure also minimizes the number of congeners that should be tested. However, testing only a few representative congeners does not sacrifice achieving relevant information of the relationship between chemical structure and biological activity of the group of PCNs. In future, we suggest that biochemical and toxicological effects should be investigated for the selected PCNs, preferably in several different test systems in order to capture different biological effects of the PCNs.

5. ACKNOWLEDGEMENT

We gratefully acknowledge the Center for Environmental Research in Umeå (Grant 93257), the Swedish Environmental Protection Agency (Grant 30520), and the Faculty of Mathematics and Natural Sciences, Umeå University, for financial support.

6. REFERENCES

[1]) Brinkman, U.A.Th and H.G.M. Reymer (1976): Polychlorinated naphthalenes. *J. Chromatography* 127, 203-243.
[2]) Jakobsson, E. (1994): Synthesis and analysis of chlorinated naphthalenes. Biological and environmental implications. PhD. thesis, Stockholm University.
[3]) Falandyz J. and C. Rappe (1996): Spatial distribution in plankton and bioaccumulation features of polychlorinated naphthalenes in a pelagic food chain in the southern part of the baltic proper. Submitted for publication.
[4]) Stewart J.J.P (1990): MOPAC Manual, Version 6.0, Quantum Chemistry Exchange, Indiana, USA.
[5]) HyperChem^{TM} (1990): Release for Windows, Reference Manual, Sausalito, USA
[6]) Andersson P., P. Haglund, C. Rappe and M. Tysklind (1996): Ultraviolet absorption characteristics and calculated semi-empirical parameters as chemical descriptors in multivariate modelling of polychlorinated biphenyls. *J. Chemometrics* 10, 171-185.
[7]) Järnberg U., L. Asplund and E. Jakobsson (1994): Gas chromarographic retention behavior of polychlorinated naphthalenes on non-polar, polarizable, polar and ametic capillary columns. *J. Chromatography* 683, 385-396.
[8]) Nikiforov, V.A., S. Miltsov, V.S. Karavan, V.G. Tribulovich and R.H. Wightman (1994): Synthesis and characterization of polychlorinated naphthalenes, III. Gas chromatography. In: H. Fiedler, O. Hutzinger, R. Clement and S. Sakai (eds).Organhalogen Compounds, Vol. 19, 137-138, Kyoto, Japan.
[9]) Haglund P., E. Jakobsson, L. Asplund, M. Athanasiadou and Å. Bergman (1992): Determination of polychlorinated naphthalenes in polychlorianted biphenyl products via capillary gas chromatography-mass spectrometry after separation by gel permeation chromatography. *J. Chromatography* 634, 79-86.
[10]) Wiedmann T. and K. Ballschmiter (1993): Quantification of chlorinated naphthalenes with GC/MS using the molar response of electron impact ionization. *Fresenius J. Anal. Chem.* 346, 800-804.
[11]) Wold S., K. Esbensen and P. Geladi (1987): Principal component analysis. *Chemom. Intell. Lab. Syst.* 10, 191-193.
[12])Tysklind M., L. Lundgren, L. Eriksson, M. Sjöström and C. Rappe (1993): Identifying training sets of PCDDs and PCDFs for use in chemical and biological monitoring. *Chemosphere* 27, 47-54.
[13]) Tysklind M., K. Lundgren, C. Rappe, L. Eriksson, J. Jonsson and M. Sjöström (1993): Multivariate quantitative structure-activity relationships for polychlorinated dibenzo-*p*-dioxins and dibenzofurans. *Environ.Toxicol. Chem.* 12,659-672.
[14]) Tysklind M., P. Andersson, P. Haglund, B. van Bavel and C. Rappe (1995):Selection of polychlorinated biphenyls for use in quantitative structure-activity modelling. *SAR & QSAR Environ. Res.* 4, 11-19.
[15]) Tysklind M., D. Tillitt, L. Eriksson, K. Lundgren and C. Rappe (1994): A toxic equivalency factor scale for polychlorinated dibenzofurans. *Fundam. Appl. Toxicol.* 22, 277-285.