Multivariate chemical characterization of polychlorinated dibenzo-p-dioxins and dibenzofurans.

Tysklind, M.[A], Lundgren, K.[A], Rappe, C.[A], Eriksson, L.[B], Sjöström, M.[B]
[A] Institute of Environmental Chemistry,University of Umeå, S-90187 Umeå, Sweden.
[B] Research Group for Chemometrics, University of Umeå, S-90187 Umeå, Sweden.

ABSTRACT
By principal component analysis of a battery of physicochemical variables, an overview of similarities and differences between congener groups and substitution patterns of the PCDDs and PCDFs is obtained. This approach indicates some partially 2,3,7,8-substituted congeners to be chemically related to the high toxic 2,3,7,8-substituted.

INTRODUCTION

Quantitative structure-activity relationships (QSARs) for groups of related compounds, such as the polychlorinated dibenzo-p-dioxins (PCDDs) and dibenzofurans (PCDFs), can give important knowledge of related compounds for which less information exists. The primary step toward the construction of QSAR models for a series of compounds, such as the PCDDs and PCDFs, is the characterization of their chemical and structural properties, preferably by compilation of as many relevant physico-chemical variables as possible[1]. Depending on which method is used, different requirements regarding proportion of objects and variables have to be met.

One method especially suitable for chemical characterization is principal component analysis (PCA) [2]. This projection-based method can deal with a large number of chemical variables, which may in fact exceed the number of objects. PCA can be used to delineate general trends and irregularities in the physicochemical variables. This may serve not only as a basis to understand which chemical factors are correlated with a certain behavior but also to identify compounds that are likely to behave differently.

The aim of this work is to report on the multivariate characterization of the PCDDs and PCDFs. By interpretation of the principal properties obtained in the principal component analysis a selection of representative congeners can be made, covering the whole chemical domain of the two groups of compounds. These congeners are proposed to be tested in future biological and toxicological as well as physical evaluations of the behavior of the PCDDs and PCDFs.

METHODS

Physicochemical variables. The chemical characterization of the PCDDs and PCDFs is based on literature data and data obtained at our laboratory[3,4]. The analysis is restricted to variables that has been obtained for larger sets of congeners with four to eight chlorine atoms. The variation in chemical structures between the different congeners of PCDDs and PCDFs is described by two types of variables. The first type comprises global molecule property variables such as the number of chlorine atoms and the log K_{ow} - value. The second type of variables depict the chlorine substitution pattern with variables such as IR spectroscopy data (PCDDs) and UV absorption data (PCDFs). See Table 1. The substitution pattern is also described by eight variables obtained by assigning to the eight positions open for chlorine substitution an indicator variable "1" or "0" depending of chlorine substitution or not. In total, the analysis is based on 22 physico-chemical variables for the PCDDs and 18 variables for the PCDFs.

Table 1. Physico-chemical variables used in the multivariate characterization of the PCDDs and PCDFs.

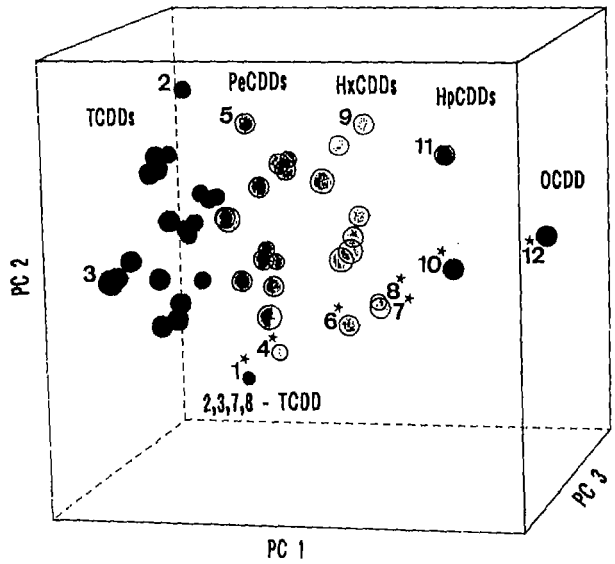
PCDDs and PCDFs:	Var. no.	**PCDDs:**	Var. no.
Ionization potential	1	Heat of vaporization	16
Heat of formation	2	Molar volume	17
Dipole moment	3	Solubility parameter	18
LUMO	4	FTIR-spectroscopy	19-22
Log Kow	5		
GC-retention time	6	**PCDFs:**	
Number of chlorine atoms	7	UV-absorption	16-18
Substitution pattern indicators	8-15		

Data analysis. Multivariate projection methods, as PCA [2], provide a means by which compiled physico-chemical data may be analyzed and interpreted. PCA combines the included variables to a few underlying descriptive dimensions, summarizing the systematic information present in the data matrix. The primary scope of PCA is to get an overview of the dominant patterns or major trends in the data. Interpretation of such patterns is possible, since the general features are easy to survey when visualized in pictures. Thus, this procedure reveals information concerning relationships between objects and variables. In this study the PCDDs and PCDFs constitute the objects.

The first calculated PC reflects the linear combination associated with the main variation in the data, the second PC explains the next largest variance, and so on. The resulting object plot, the score plot, is an optimal projection showing the relation between the different objects, here congeners. Objects close to each other in the score plot have similar physico-chemical characteristics. The corresponding variable plot, the loading plot, shows how the variables are related to each other and how they influence the different PCs. The number of meaningful principal components is determined with a significance test called cross- validation[2]. The cross-validation technique provides a means of obtaining optimal predictive power without overfitting of the model. The calculations were performed on an IBM PS2 computer by the SIMCA 4.3R program package (Umetri AB, S-90124 Umeå, Sweden).

RESULTS AND DISCUSSION

Dioxins. The PC analysis of the 49 dioxins resulted in a four-dimensional model, significant according to cross-validation, explaining 61% of the variation in the data. The separate principal components (t_1 - t_4) described 33%, 13%, 8% and 7% respectively. Figure 1 shows the plot of PC1, PC2 and PC3 versus each other.The first PC separates the 49 dioxin congeners in five groups from the tetrachlorodioxins, followed successively by the penta-, hexa-, hepta- and octachloro congeners. Thus it appears that the first principal component reflects the degree of chlorination among the compounds, while the second and third components separates different substitution patterns. It is noteworthy that the striking tendency of all the 2,3,7,8-substituted congeners shows up beside all the other congeners in each of the different congener groups. The first dimension is described by size related variables such as retention time and $\log K_{ow}$. The second and third dimension are mainly dominated by substitution pattern related variables such as IR_2, chlorine position three and nine (PC2) and dipole moment and IR_4 (PC3).

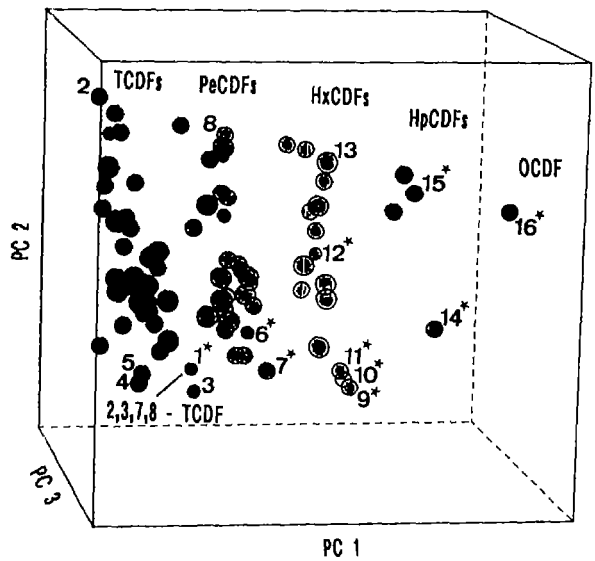


1. 2,3,7,8-TCDD
2. 1,4,6,9-TCDD
3. 1,2,3,9-TCDD
4. 1,2,3,7,8-PeCDD
5. 1,2,4,6,9-PeCDD
6. 1,2,3,7,8,9-HxCDD
7. 1,2,3,6,7,8-HxCDD
8. 1,2,3,4,7,8-HxCDD
9. 1,2,4,6,7,9-HxCDD
10. 1,2,3,4,6,7,8-HpCDD
11. 1,2,3,4,6,7,9-HpCDD
12. OCDD

Figure 1. The "chemical space" of the PCDDs as described by the plot of PC1, PC2 and PC3 versus each other. 2,3,7,8-substitution marked with an asterisk.

Dibenzofurans. The analysis of the 87 dibenzofurans is analogous to the above analysis of the dioxins. The PC analysis, based on the18 physico-chemical variables listed in Table1, resulted in four significant principal components which explained 54% of the total variance. The first component (PC1) described 31%, PC2 9%, PC3 8% and PC4 6% of the variance. Figure 2 shows the score plot, PC1, PC2 and PC3 versus each other. As seen, the first dimension resolves the different congener groups with the TCDFs to the far left up to the OCDF to the far right. The second dimension discriminates between different substitution patterns. In the upper part of the plot, 1,2,9-substituted congeners can be found. The 2,3,7,8-

substituted congeners show up in the lower part of the plot. Close to the highly toxic 2,3,7,8-TCDF also other substitution patterns appears, such as 2,3,6,8-TCDF, 1,3,6,7-TCDF and 2,3,4,8-TCDF .The interpretation of the corresponding loadings shows that variables like ionization potential, log K_{ow}, retention time, number of chlorine atoms and LUMO dominate the first dimension. The second dimension is regulated by the quantitative substitution pattern variables position seven and nine together with UV_2 and UV_3 . The third dimension is influenced by the variables heat of formation, UV_1 and chlorine position six.



1. 2,3,7,8-TCDF
2. 1,2,8,9-TCDF
3. 2,3,6,8-TCDF
4. 1,3,6,7-TCDF
5. 2,3,4,8-TCDF
6. 1,2,3,7,8-PeCDF
7. 2,3,4,7,8-PeCDF
8. 1,2,4,8,9-PeCDF
9. 2,3,4,6,7,8-HxCDF
10. 1,2,3,6,7,8-HxCDF
11. 1,2,3,4,7,8-HxCDF
12. 1,2,3,7,8,9-HxCDF
13. 1,2,4,6,8,9-HxCDF
14. 1,2,3,4,6,7,8-HpCDF
15. 1,2,3,4,7,8,9-HpCDF
16. OCDF

Figure 2. The "chemical space" of the PCDFs as described by the plot of PC1, PC2 and PC3 versus each other. 2,3,7,8-substituted congeners are marked with an asterisk.

ACKNOWLEDGEMENT

This work has been funded by The Environmental Science Center (CMF), Umeå, Sweden, which is gratefully acknowledged.

REFERENCES

1. Eriksson L. *A strategy for ranking of environmentally occurring chemicals.* Thesis, University of Umeå, Sweden, 1991.

2. Wold S, Albano C, Dunn W J, Edlund U, Esbensen K, Geladi P, Hellberg S, Johansson E, Lindberg W, Sjöström M. *Data analysis in chemistry.* In: Kowalski B R Ed. Chemometrics: Mathematics and statistics in chemistry. Riedel, Dordrecht, 1984; 4-96.

3. Tysklind M, Lundgren K, Rappe C, Eriksson L, Jonsson J, Sjöström M, Ahlborg U G. Multivariate characterization and modeling of polychlorinated dibenzo-p-dioxins and dibenzofurans. *Environ. Sci. Technol.* 1992. In press.

4. Tysklind M, Lundgren K, Rappe C, Eriksson L, Jonsson J, Sjöström M. Multivariate quantitative structure-activity relationships for polychlorinated dibenzo-p-dioxins and dibenzofurans. Submitted to *Environ. Tox. Chem.* 1992.