Chemometrical approaches for the characterization of the environmental behavior of fragrances

Ester Papa, Mara Luini, Paola Gramatica
QSAR Research Unit in Environmental Chemistry and Ecotoxicology – DBSF- University of Insubria (Varese, Italy)
Web: http://www.dipbsf.uninsubria.it/qsar/ ; www.qsar.it
e-mail: ester.papa@uninsubria.it; paola.gramatica@uninsubria.it

Department of Structural and Functional Biology
University of Insubria
via J.H. Dunant, 3
21100 Varese (Italy)

Abstract

Fragrances represent a new class of emerging pollutants. These compounds, which are used in many consumer products, including cleaning and washing agents, and personal care products, are often released directly to the environment, posing a risk for the exposed organisms. Concerns from the environmental presence of these compounds is a result of their potential persistence and/or toxicity in environmental media and of their possible effects on humans (asthma, allergies, headaches…), as well as of their ability to accumulate in adipose tissue such as breast milk. Moreover, since the fragrances are often volatile compounds, they are connected to both indoor and outdoor air pollution.
Unfortunately little is known about the environmental occurrence and fate of these substances which potential effects on humans and aquatic ecosystems are not yet clearly understood. The use of predictive approaches based on chemometrical techniques, such as Quantitative Structure Activity (Property) Relationships (QSA(P)R), can help in filling this data gap and characterizing the environmental and ecotoxicological profile of these substances.
In this study different Ordinary Least Squares (OLS) regression-based QSA(P)R models were developed for toxicological and physico chemical endpoints. Theoretical molecular descriptors were calculated by DRAGON software, and the best modeling variables were selected also by applying Genetic Algorithms (GA).
The developed models could be particularly useful for characterization, screening and prioritization of widely used fragrances compounds, and also *a priori*, in the design of new products as safer alternative to the existing dangerous.

Materials and Methods

Data Set: The experimental data related to toxicological and physico-chemical data were taken partially from literature [1,2,3,4,5,6] and partially from available online databases [7,8,9].
Toxicological endpoints modeled are LD50 Oral Mouse [1,2,7], inhibition of NADH oxidase (EC50 NADH-Ox) and the effect on mitochondrial membrane potential (EC50 ΔΨm)[3].
All the responses have been transformed to logarithmic units and, if necessary, multiplied by -1 to obtain positive values. Physico-chemical properties modeled are LogKow, Water Solubility (WS) and Vapor Pressure (Vp); like for toxicological endpoints, the responses have been transformed to logarithmic units.

Molecular Descriptors:
452 molecular descriptors (OD,1D,2D,3D) were calculated for Physico-chemical properties and LD50 Oral Mouse and 459 for the others toxicological endpoints by the software DRAGON [10].
4 quantum-chemical descriptors [Highest Occupied Molecular Orbital (HOMO), Lowest Unoccupied Molecular Orbital (LUMO), HOMO-LUMO gap (DHL) and the ionisation potential (P ion)] were calculated. The input files for descriptor calculation were obtained by the Semi empirical (AM1) Hamiltonian for the geometry optimisation method available in the HYPERCHEM package [11].

MULTIPLE LINEAR REGRESSION MODELS and Variable Selection were performed by *Ordinary Least Squares* regression (OLS) and Genetic Algorithms (GA) method [12].

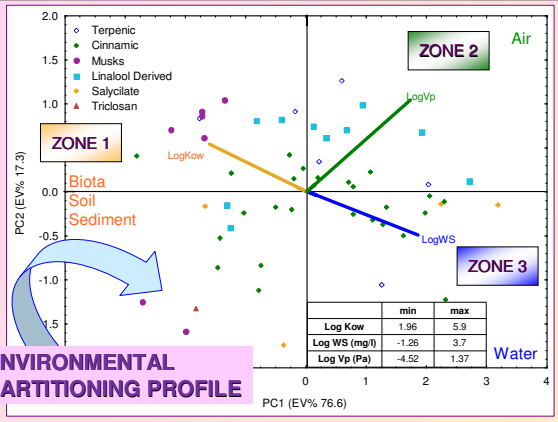
EXTERNAL VALIDATION
Prediction set selection was based on the molecular structure (by Kohonen Maps - Artificial Neural Networks (K-ANN) or using the Random by response approach.

TOOLS OF VALIDATION AND DIAGNOSTICS
Models were developed taking into account the recently proposed OECD principles for QSAR validation [13].
- Internal (by Q^2_{LOO} and Q^2_{boot}, Y-scrambling) and external validation (verified by Q^2ext) [14].
- Check of the quality of the best models by Residuals and Williams plot
- Applicability Domain verified by leverage approach.

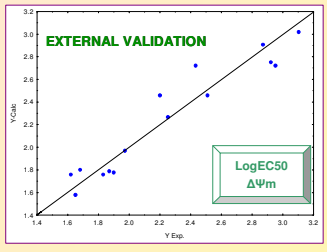
Results

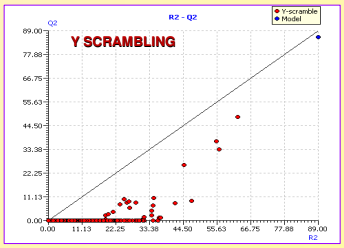
Physico chemical Properties	Model	Train Obj.	Test Obj.	Variables	R^2	Q^2_{LOO}	Q^2_{BOOT}	Q^2_{EXT}	RMSE Train	RMSE Test	R^2-YScr
Log Kow	Full Model	52	-	ATS1p H-047	81.98	80.1	76.34	-	-	-	4.2
	Random	42	10		79.9	76.43	76.53	84.47	0.56	0.36	-
	K-ANN	41	11		84.34	80.17	80.36	78.8	0.51	0.49	-
Log WS	Full Model	37	-	BEHm3 JGI3 nCconj	80.2	76.09	75.28	-	-	-	8.19
	Random	30	7		80.03	75.30	73.99	80.41	0.50	0.47	-
	K-ANN	29	8		80.69	75.44	73.25	75.00	0.54	0.34	-
Log Vp	Full Model	37	-	piPC01 nHDon	89.4	87.22	86.85	-	-	-	5.55
	Random	29	8		88.32	83.07	83.2	90.86	0.7	0.47	
	K-ANN	29	8		87.5	80.1	80.76	90.31	0.79	0.47	

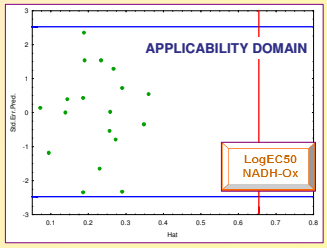
Toxicological Endpoints	Model	Train Obj.	Test Obj.	Variables	R^2	Q^2_{LOO}	Q^2_{BOOT}	Q^2_{EXT}	RMSE Train	RMSE Test	R^2-YScr
Log 1/LD50 Oral Mouse	Full Model	23	-	nR=Cs H-047	89	86.19	81	-	-	-	9.25
	Random	13	10		91.35	86.15	80.47	73.4	0.265	0.226	-
	K-ANN	15	8		88.78	83.41	72.2	90.23	0.256	0.186	-
LogEC50 ΔΨm	Full Model	15	-	ATS4v MATS2m	91.69	88.55	83.94	-	-	-	14.34
	Random	12	3		90.29	84.31	74.01	98.19	0.2	0.061	-
	K-ANN	13	2		90.88	85.8	78.78	97.26	0.19	0.07	-
LogEC50 NADH-Ox	Full Model	18	-	RBF HATS3m nCq	87.71	80.04	78	-	-	-	16.88
	Random	15	3		88.4	79.76	76.38	78.05	0.30	0.23	-
	K-ANN	14	4		88.97	78.04	66.47	76.78	0.32	0.22	-

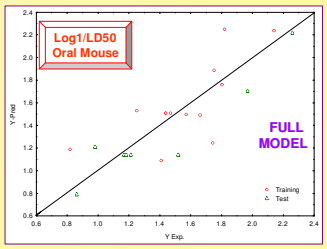


ENVIRONMENTAL PARTITIONING PROFILE









All the models were developed according to the OECD Principles for the acceptance of QSAR models in regulation

Principal Component Analysis (PCA) was performed on experimental and reliable predicted values of LogKow, LogWs and LogVp available for 53 compounds. This analysis was applied in order to have a multivariate view and a graphic representation of the potential overall environmental partitioning profile of the studied fragrances.
Chemicals are ranked along PC1 and PC2 (Tot E.V.%=93.9; PC1 E.V%=76.6%), according to their physico-chemical properties and (from right to left) according to increasing molecular weight (MW). Hydrophobic compounds, such as Musks, are located on the left side of the graph (zone 1 - accumulation potential in biota, soil and sediments); soluble and volatile fragrances are ranked on the right of PC1 and further along PC2. In particular volatile fragrances, such as linalool derivatives, are placed in zone 2 ; more soluble fragrances, such as low-MW salicilates and cinnamates are placed in zone 3.

Conclusions

1. **Limited availability of experimental data utilizable for QSAR (in particular for SIDS and toxicological endpoints).**
2. **New QSAR and QSPR models have been developed for the prediction of some physico-chemical (LogKow, LogWS, LogVp) and toxicological endpoints (acute oral mouse toxicity, and two endpoints related to mitochondrial toxicity [Inhibition of NADH-Oxidase and the effect on mitochondrial membrane potential (ΔΨm)).**
   Despite the limited amount of available data, all the models were carefully internally and externally validated. At our knowledge no other QSAR/QSPR models are available in literature for these endpoints.
3. **All the variables selected in the proposed models are 2D descriptors, independent of chemical conformation. Quantum chemical descriptors as well as LogP were not selected in the externally predictive models.**
4. **The combination of the predicted and experimental physico-chemical properties data by PCA allowed for the identification of a profile of the environmental behavior of fragrances. Three zones were identified which distinguish among volatile (linalool derivatives), soluble (salicilates and cinnamates) or adsorbed (musks) compounds.**

New data on the real persistence of these compounds could help in the refinement of the hazard profile of these fragrances.

REFERENCES
(1) Bickers, D.R., *et al.*,2003. *Regul.Toxicol.Pharm.*,37,218-273
(2) Belsito, D., *et al.*,2007. *Food Chem.Toxicol.*,45,S1-S23
(3) Griffiths,D.E., *et al.*, 2005. *ATLA-ALTERN.LAB.ANIM.*, 33, 471-486
(4) Simonich,S.L., *et al.*, 2000. *Environ.Sci.Technol.*, 34, 959-965
(5) Luckenbach,T., *et al.*, 2005. *Environ.Health Perspect.*, 113, 17-24
(6) Schramm,K.W. *et al.*, 1996. *Wat.Res.*, 30, 2247-2250
(7) PubChem: http://pubchem.ncbi.nlm.nih.gov/chemidplus/ProxyServlet?objectHandle=Search&acti_onHandle=clear&nextPage=chemidheavy.jsp
(8) ChemIDPlus: http://pubchem.ncbi.nlm.nih.gov/
(9) PhysProp: http://www.srcinc.com/what-we-do/databaseforms.aspx?id=386
(10) Todeschini R., Consonni V. and Pavan M., **2001**. *DRAGON – Software for the calculation of molecular descriptors*, rel. 1.12 for Windows. Free download available at http://www.disat.unimib/chm
(11) HYPERCHEM. Release 7.03 for Windows, 2002. Hypercube, Inc. Florida, USA;
(12) Todeschini, R., **2001**. *Moby Digs - Software for multilinear regression analysis and variable subset selection by Genetic Algorithm*, rel. 2.3 for Windows, Talete srl, Milan (Italy);
(13) Available online at: http://www.oecd.org/document/23/ (accessed March 2008)
(14) Gramatica P. 2007, *QSAR Comb.Sci.*, 26,694-701

Financial support, given by European Union through the project CADASTER (FP7-ENV-2007-1-212668), is gratefully acknowledged.