QSAR PREDICTION OF THE ENDOCRINE ACTIVITY OF PERFLUORINATED COMPOUNDS

Simona Kovarich, Paola Gramatica, Ester Papa

QSAR Research Unit in Environmental Chemistry and Ecotoxicology – DBSF - University of Insubria (Varese, Italy)
e-mail: simona.kovarich@uninsubria.it; 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)

INTRODUCTION Perfluorinated compounds (PFCs) are a class of emerging pollutants still widely used in different materials as non-adhesives, waterproof fabrics, fire-fighting foams, etc. Their toxic effects include potential for endocrine disrupting (ED) activity among others. Unfortunately, the available amount of experimental data for these pollutants is limited. Therefore the use of predictive strategies such as QSAR/QSPR is recommended under the REACH regulation, to fill the data gaps and also to allow the screening and prioritization of chemicals for experiments, with a consequent reduction of costs and of the number of tested animals. In this study the T4-TTR competing potency of 24 PFCs has been modelled by two different QSAR approaches: multiple linear regression, by Ordinary Least Squares (OLS), and classification, by K-NN method. Models were developed taking into account the OECD principles for QSAR validation for regulatory purposes [1].

EXPERIMENTAL DATA SET

24 Perfluorinated compounds with different carbon chain length (4-14 C), fluorination degree and functional groups (carboxylates, sulfonates, sulfonamides, alcohols, etc.) [2].

PFBA Perfluorobutyric acid	**PFUnA** Perfluoroundecanoic acid
PFHxA Perfluorohexanoic acid	**PFDoA** Perfluorododecanoic acid
PFHpA Perfluoroheptanoic acid	**PFTdA** Perfluorotetradecanoic acid
7H-PFHpA 7H-Perfluoroheptanoic acid	**FTUA** 2H-Perfluoro-2-octenoic acd
PFOA Perfluorooctanoic acid	
PFNA Perfluorononanoic acid	
PFDcA Perfluorodecanoic acid	

PFAA Perfluorinated alkyl acids

PFAS Perfluorinated alkyl sulfonates

PFBS Nonafluorobutane sulfonate
PFHxS Perfluorohexane sulfonate
PFOS Perfluorooctane sulfonate
L-PFDSI Perfluorooctane sulfiniate
L-PFDS Perfluorodecane sulfonate

PFAS were converted into the respective sulfonic acids and used as an additional VALIDATION SET.

Perfluorinated alkyl sulfonamides

FOSA Perfluorooctane sulfonamide
N-MeFOSA N-methyl perfluorooctane sulfonamide
N-EtFOSA N-ethyl perfluorooctane sulfonamide
N,N-Me2FOSA N,N-dimethyl perfluorooctane sulfonamide

Perfluorinated telomer alcohol

FTOH (6:2) 2-Perfluorohexyl ethanol
FTOH (8:2) 2-Perfluorooctyl ethanol

N-MeFOSE 2-(N-methylperfluoro-1-octane sulfonamido) ethanol
N-EtFOSE 2-(N-ethylperfluoro-1-octane sulfonamido) ethanol

Perfluorinated alkyl sulfonamido-ethanols

MOLECULAR DESCRIPTORS

The Semi-empirical method AM1 in HYPERCHEM program (ver. 7.03 for Windows, 2002) was used to draw and optimize (minimum energy conformation) the structures of the studied Perfluorinated compounds.

444 molecular descriptors, which encode the mono-, bi- and tri-dimensional structural information, were calculated from the optimized structures by using the software DRAGON (ver. 5.5 for Windows, 2007).

METHODS – REGRESSION MODELS

ENDPOINT: IC_{50} T4-TTR COMPETING POTENCY [2]. To obtain increasing trends of toxicity, the experimental values were transformed into the logarithm of the inverse nM concentrations ($Log1/IC_{50}$).

ALGORITHM: Multiple linear regression was performed by *Ordinary Least Squares* regression (OLS) method. *All Subset Selection* method was applied to select the best variables [3].

APPLICABILITY DOMAIN: verified by leverage approach.

TOOLS of VALIDATION: goodness-of-fit and internal stability were verified by Q^2_{LOO}, Q^2_{BOOT}, R^2/Q^2_{YS} and RMSE; external predictivity was measured by calculating Q^2_{EXT} on the additional validation set (3 PFAS) [4-5].

METHODS – CLASSIFICATION MODELS

CLASSES: C1=INACTIVE (no T4-TTR_{comp} potency detected); C2=ACTIVE (low to high T4-TTR_{comp} potency). Classification criteria according to Hamers *et al.*, 2006 [6].

ALGORITHM: K-NN method was applied to model the two classes of T4-TTR_{comp} [7]. The selection of the best subset of variables was realised by the *All Subset Selection* method. SPLITTING: data were split into training and prediction set by Random selection (50%).

APPLICABILITY DOMAIN: verified by descriptor's range.

TOOLS of VALIDATION: Internal stability was verified by Sn, Sp, NER_{CV}. For the external validation, NER_{EXT} was calculated for the prediction set and for the additional validation set (5 PFAS) [4].

⇨ PARAMETERS [8]: Sn = TP/(TP+FN) Sp = TN/(TN+FP) NER = (TP+TN)/Tot

RESULTS

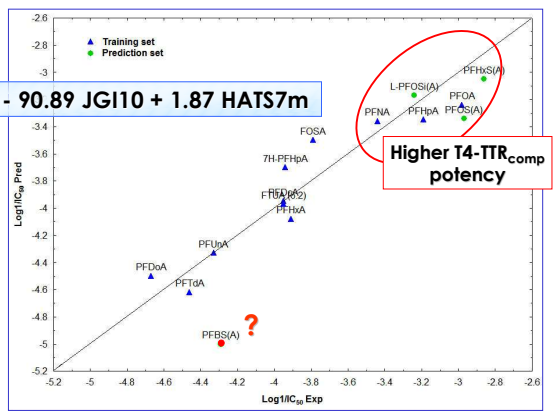
1

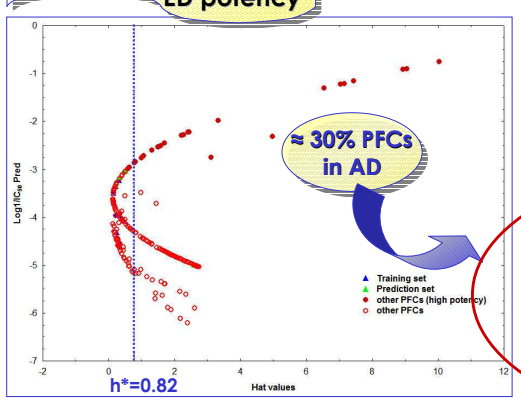
Variables	N_{TR}	N_V	R^2	Q^2_{LOO}	Q^2_{BOOT}	R^2_{YS}	Q^2_{YS}	$RMSE_{TR}$	$RMSE_{CV}$	Q^2_{EXT}	$AD\%_{24}$
JGI10 HATS7m	11	3*	0.88	0.77	0.72	0.19	0.15	0.17	0.24	0.75	87.5

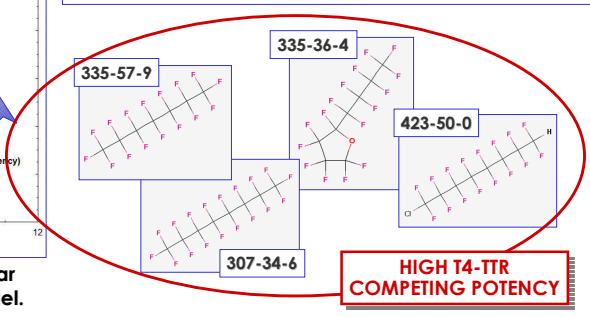
Molecular Size & Functional Groups

$Log1/IC_{50}T4\text{-}TTR_{comp} = -5.04 - 90.89\ JGI10 + 1.87\ HATS7m$

Screening of a set of 358 PFCs according to their ED potency







- PFBS was identified as an outlier by a preliminar PLS model [2] and it falls out of the AD of our model. Therefore it was excluded from the validation data-set.

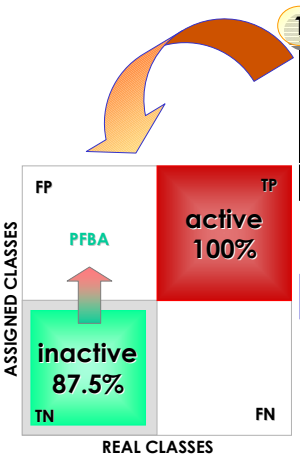
Comparison with QSAR model by Weiss *et al.* 2009 [1]

Model	Method	N_{TR}	N_V	No. Descriptors	No. PLS Components	R^2	Q^2_{LOO} / R^2_{CV}	Q^2_{EXT}
This study (1)	MLR	11	3*	2	-	0.88	0.77	0.75
This study (2)	MLR	14*	-	2	-	0.89	0.84	-
Weiss 2009	PLS	14*	-	56	2	0.61	0.41	-

RESULTS

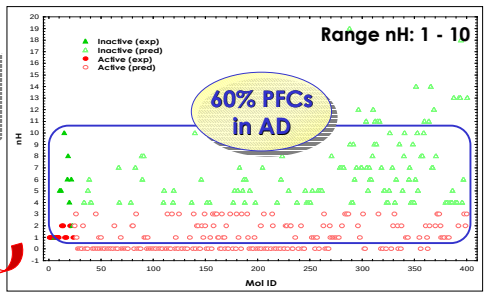
1

Model		Desc.	K	n	Sn	Sp	$NER_{CV}\%$	$NER_{EXT}\%$	$AD\%_{24}$
Full Model		nH	3	19	1	0.88	94.7	-	100
Split	Training set		3	10	1	0.75	90	-	95.8
	Prediction set		3	9	1	1	-	100	
Extra Validation set		nH	3	5				80	



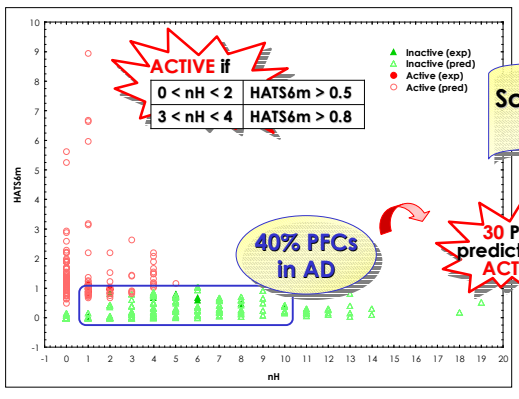
Screening of 358 PFCs

ACTIVE if 0 < nH < 4

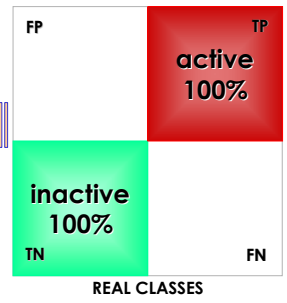


2

Model		Desc.	K	n	Sn	Sp	$NER_{CV}\%$	$NER_{EXT}\%$	$AD\%_{24}$
Full Model		nH HATS6m	1	19	1	1	100	-	100
Split	Training set		1	10	1	0.75	90	-	95.8
	Prediction set		1	9	1	1	-	100	
Extra Validation set		nH HATS6m	1	5				80	



Screening of 358 PFCs



- Linear carbon chains
- Average molecular size (6 < n°C < 11)
- Fully fluorinated chains
- Mainly with –COOH as a functional group

CONCLUSIONS

- ✓ The here proposed MLR model (1) is robust and predictive. However more experimental data would be necessary to develop QSARs with wider applicability.
- ✓ Interpretability of descriptors: JGI10 (2D) is mainly related to molecular size of PFCs (n° C), while HATS7m (3D) takes also into account the different functional groups.
- ✓ Both the here proposed MLR model (1) and the model developed using the same data-set as Weiss 2009 (2) show significantly higher performance than the existing model by Weiss et al. (2009) [2].

CONCLUSIONS

Two classification models are here proposed to predict T4-TTR competing potency of PFCs:

	M1 [nH]		M2 [nH HATS6m]	
Fitting & stability	NER_{CV} = 95% (1 false positive)	😐	NER_{CV} = 100%	😊
Sensitivity	Sn = 1 (no false negative)	😐	Sn = 1 (no false negative)	😐
External predictivity	NER_{EXT} = 100%	😐	NER_{EXT} = 100%	😐
Model dimension	Only one simple descriptor	😐	Two descriptors	😐
Simplicity & Interpretability of descriptors	nH: n° of hydrogen atoms (0D) → functional groups	😐	nH: n° of hydrogen atoms (0D) → functional groups	😐
			HATS6m: getaway desc., weighted by atomic masses (3D) → molecular size + functional groups	😐

The proposed regression and classification QSAR models are simple tools for the rapid screening of the T4-TTR competing potency of perfluorinated compounds and can be used for the prioritization of more hazardous chemicals.

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

REFERENCES

1. Available online at: http://www.oecd.org/document/23/ (accessed April 2009);
2. Weiss, JM *et al.*, 2009. *Tox.Sci.*, 109, 206-216;
3. Todeschini, R., 2001. *MOBY DIGS. Rel.* 2.3 for Windows, Talete srl, Milan (Italy);
4. Gramatica, P., 2007. *QSAR Comb. Sci.*, 26, 694-701;
5. Consonni, V. et al., 2009. *J. Chem. Inf. Model.*, 49, 1669-1678;
6. Hamers, T. et al., 2006. *Tox.Sci.*, 92, 157-173;
7. SCAN Software for Chemometric Analysis, 1995, ver. 1.1 for Windows, Minitab (USA);
8. Roncaglioni, A. et al., 2008. *SAR QSAR Environ. Res.*, 19, 697-733.