Oral LD_{50} Toxicity Modeling of Per- and Polyfluorinated Chemicals on Rat and Mouse

Barun Bhhatarai, Paola Gramatica
Department of Structural and Functional Biology, University of Insubria, QSAR Research Unit in Environmental Chemistry and Ecotoxicology,
Via Dunant 3, 21100 Varese, paola.gramatica@uninsubria.it

INTRODUCTION

Quantitative structure-activity relationship (QSAR) and chemometric methods were applied to Perfluorinated Chemicals (PFCs) – fluorinated carbon chain (C4 to C16) containing linear or cyclic chemicals, which are considered as 'emerging pollutants'. They are found widely distributed in the environment, released due to their widespread use in different household and industrial products as cleansers, fire-fighting foams, micelles, oil and water repellants for leather, paper, and textiles etc. Continues exposure of these chemicals is found to be the source of bio-accumulation in body parts of human, wildlife and is ultimately becoming the cause of toxic reactions. However, there are more than 650 PFCs, linear and cyclic, that are found in ECHA (European CHemical Agency) preregistration list of compounds and these chemicals needs to be identified, if, they belong to Substances of Very High Concern (SVHC). Experimental data for majority of these compounds are unavailable or are proprietary and a need to use the existing available data to predict the activity of these compounds is necessary. Thus, a dataset of oral lethal dose 50% (LD_{50}) was compiled for short and long chain PFCs on two species of rodents –Rat (*Rattus*) and Mouse (*Mus*). The oral exposure analysis was chosen as it is an important indicator of food and accidental domestic poisonings, and/or occupational poisonings. QSAR was then applied to model the available data and predict the oral LD_{50} toxicity of other chemicals including those listed in ECHA for which toxicity data is not available. The set of descriptors which best describes the structure-toxicity relationship, the similarities, and the differences observed related to two species are discussed. Principal Component Analysis (PCA) was used to select most toxic compounds from those within the structural applicability domain (AD) of both the models. QSAR study on LC_{50} inhalation data of PFCs on rodents has been published earlier and combining with the result of current LD_{50} oral study, a comparative toxicity analysis of two different end-points on rodents and consensus prediction and prioritization of hazardous PFCs is performed. The prioritized chemicals will be further subjected to experimental test under the EU-FP7 funded CADASTER project.

MATERIALS AND METHODS

Data Set: 58 Mouse and 50 Rat LD50 oral data were used. Training and prediction sets were prepared *a priori* from available experimental datasets in terms of structure (SOM) and random by response approach and these sets were used to derive statistically robust and predictive (both internally and externally) models. 26% to 37% splitting were used. Structural applicability domain (AD) of the models were verified on 376 per- and polyfluorinated chemicals including those in REACH preregistration list

Molecular Descriptors: More than 600 molecular descriptors (0D–3D) were calculated by the software DRAGON [1] from the XYZ coordinates in Hyperchem using AM1 [2].

Multiple Linear Regression (MLR) and Genetic Algorithm-Variable Selection (GA-VSS) were performed by the software MOBY DIGS [3] using the Ordinary Least Square regression (OLS) method.

Validation: The robustness of the models and their internal predictive ability were evaluated by both Q^2 based on leave-one-out cross-validation and bootstrap. The proposed models were also checked for reliability and robustness by permutation testing [4]: new models were recalculated for randomly reordered response ($R^2_{Y\text{-scrambling}}$).

The external validation was performed by developing the model on the training set and then using those models to predict the test set [5].

RESULTS AND DISCUSSION

$$\text{Mus } log\ 1/LD_{50} = 4.543 - 2.450\ (\pm 0.312)\ \text{HATS2u} + 1.362\ (\pm 0.203)\ \text{B09[C-O]} - 0.142\ (\pm 0.032)\ \text{F01[C-O]} - 0.486\ (\pm 0.174)\ \text{B04[C-F]}$$

$n=58$, $r^2=75.93$, $q^2=71.89$, $s=0.41$, $F=41.8$

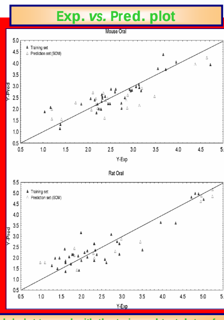
Equation 1

Response (Compounds)	Descriptors		Splitting criteria	Compounds	R^2	Q^2_{LOO}	Q^2_{BOOT}	$Q^2_{ext\ F1}$	$Q^2_{ext\ F3}$	$RMSE_{TR}$	$RMSE_{EXT}$	R^2_{YS}
	Input	Selected										
Mouse Oral LD_{50} (58)	690	HATS2u; B09[C-O]; FO1[C-O]; B04[C-F]	SOM 37.9 %	Train: 36, Predict: 22	*82.8*	*75.7*	*74.2*	*65.6*	*55.6*	*0.32*	*0.51*	*11.5*
			Random by Activity 32.7%	Train: 39, Predict: 19	*81.4*	*75.4*	*74.1*	*63.0*	*64.5*	*0.35*	*0.48*	*12.2*
			Full model		*75.9*	*71.9*	*67.9*	-	-	*0.39*	*$0.42_{(CV)}$*	*7.0*

Mouse Oral descriptors:

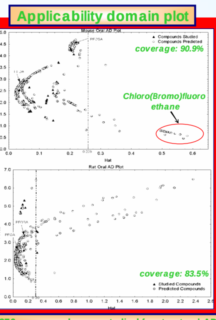
HATS2u (-0.589), 3D GETAWAY
→ leverage-weighted autocorrelation of lag 2/unweighted.
B09[C-O] (0.478), 2D binary finger-print
→ the presence of [C-O] at n=9
F01[C-O] (-0.303), 2D frequency finger-print
→ the frequency of [C-O] at n=1
B04[C-F] (-0.198), 2D binary finger-print
→ the presence of atom pair [C-F] at n=4

Exp. *vs.* Pred. plot



Full model plot tagged with the train and test data of SOM split to compare distribution of compounds

Applicability domain plot



376 compounds were studied for structural AD study
308 compounds within the AD of both the models

$$\text{Rat } log\ 1/LD_{50} = -2.277 + 0.041\ (\pm 0.003)\ \text{D/Dr09} + 2.943\ (\pm 0.580)\ \text{MATS1e} + 8.838\ (\pm 1.712)\ \text{E1u} + 1.166\ (\pm 0.211)\ \text{H8m}$$

$n=50$, $r^2=88.28$, $q^2=85.50$, $s=0.44$, $F=84.78$

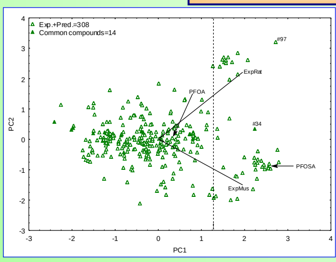
Equation 2

Response (Compounds)	Descriptors		Splitting criteria	Compounds	R^2	Q^2_{LOO}	Q^2_{BOOT}	$Q^2_{ext\ F1}$	$Q^2_{ext\ F3}$	$RMSE_{TR}$	$RMSE_{EXT}$	R^2_{YS}
	Input	Selected										
Rat Oral LD_{50} (50)	635	D/Dr09; MATS1e; E1u; H8m	SOM 28.0 %	Train: 36, Predict: 14	*85.5*	*80.3*	*70.9*	*91.1*	*81.4*	*0.41*	*0.46*	*12.2*
			Random by Activity 26.0%	Train: 37, Predict: 13	*90.7*	*87.5*	*85.6*	*80.7*	*75.1*	*0.36*	*0.59*	*11.3*
			Full model		*88.3*	*85.5*	*82.2*	-	-	*0.42*	*$0.47_{(CV)}$*	*8.0*

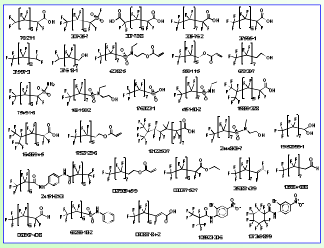
Rat Oral descriptors:

D/Dr09 (0.765), 2D topological
→ distance/detour ring index of order 9
MATS1e (0.331), 2D Moran autocorrelation
→ electronegativity
E1u (0.316), 3D WHIM
→ 1st component accessibility directional index/unweighted
H8m (0.294), 3D GETWAY
→ H autocorrelation of lag 8 (weighted by atomic mass)

Toxicity Trend – LD50 Oral



61.4% data variance explained by PC1, cumulative (PC1+PC2)= 100%



30 long chain PFCs beyond PC1 cutoff (1.25) are prioritized

3 domains of PCA plot:

top right → more toxic to Rat = fluorinated benzimidazoles,
bottom right → more toxic to Mouse = long chain PFCSs including PFOSA
center → neutral to Rodents = PFCs including PFOA

Overview – LC50 Inhalation [6]

Chem. Res. Toxicol. **2010**, *23*, 528–539

Per- and Polyfluoro Toxicity (LC_{50} Inhalation) Study in Rat and Mouse Using QSAR Modeling

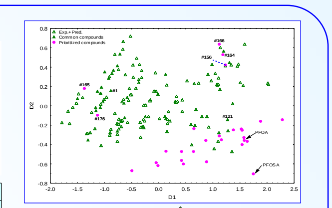
Barun Bhhatarai and Paola Gramatica*

QSAR Research Unit in Environmental Chemistry and Ecotoxicology, Department of Structural and Functional Biology (DBSF), University of Insubria, via JH Dunant 3, Varese 21100, Italy

Received July 24, 2009

Fully or partially fluorinated compounds, known as per- and polyfluorinated chemicals are widely distributed in the environment and released because of their use in different household and industrial products. Few of these long chain per- and polyfluorinated chemicals are classified as emerging pollutants,

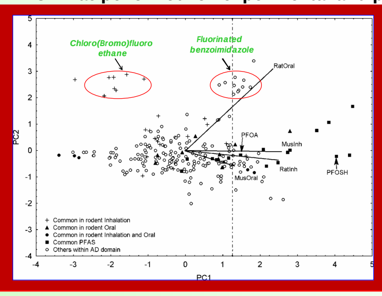
Endpoint	Descriptors	N_{obj}	R^2	Q^2_{LOO}	Q^2_{EXT}	$RMSE_{CV}$	AD% 250 PFCs
Mouse Inhalation	X3v, H-048, *M*log*P*, F01[C–C]	*56*	*79.8*	*76.31*	*71.62-85.11*	*0.74*	*75.6%*
Rat Inhalation	Jhetv, PCR, *MlogP*, B02[Cl–Cl]	*52*	*78.1*	*73.85*	*66.70-75.47*	*0.86*	*76.8%*

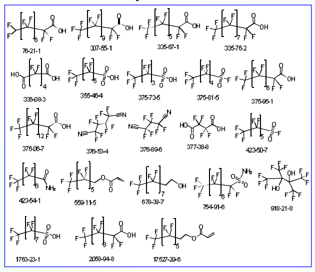


MDS plot of 7 molecular descriptors highlighting the structural diversity of prioritized compounds among 180 PFCs within the AD Rodents' LC50 models

- Two QSAR models each on Mouse and Rat LC50 inhalation data were published [6]
- Prioritization study on LC50 data was performed → 28 long chain PFCs predicted

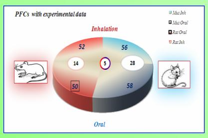
- All four models were applied individually to 376 PFCs and structural AD was verified
  → to understand the contribution of each toxicity end-points to PFCs
  - 204 compounds were found to be within the AD of all the four models
  - PCA was performed on experimental and predicted data and plotted



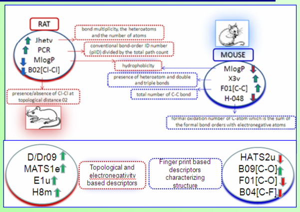


Long chain PFCs beyond PC1 cutoff (1.25) are prioritized

Combined toxicity trend: LC50 Inhalation and LD50 Oral



14 compounds with experimental value in inhalation dataset
28 compounds experimental value between Oral dataset
5 compounds experimental value in all the four dataset



CONCLUSIONS

- Toxicity of PFCs on rodents was studied by developing reliable, robust and predictive QSAR models
- Models for Rodents Oral LD50 shows combination of electronic (MATS1e, HATS2u) and fingerprint based descriptors (F01[C-O], Bo4[C-O])
- Models on Rodents Oral and Inhalation data shows importance of following main descriptors for overall toxicity
  - negative hydrophobicity (M*logP*),
  - positive electronegativity (Jhetv, X3v and MATS1e)
  - descriptors representing the position and the frequency of atom pairs like C-C, C-F and C-O that counts for the main functional groups of long chain PFCs
- Prioritized most toxic long chain PFCs will be suggested to CADASTER partners for the experimental design

REFERENCES
[1] Todeschini R., Consonni V. and Pavan M., 2001. DRAGON – Software for the calculation of molecular descriptors, rel. 1.12 for Windows.
[2] HYPERCHEM. Release 7.03 for Windows, 2002. Hypercube, Inc. Florida, USA;
[3] 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);
[4] Lindgren, F., Hansen, B., Karcher, W., Sjöström, M., Eriksson, L. Model validation by permutation tests: Applications to variable selection. J. Chemom. 1996, 10, 521-532.
[5] Gramatica P. 2007, QSAR Comb.Sci., 26,694-701.
[6] Bhhatarai, B., Gramatica, P. 2010, Chem. Res. Toxicol., 23, 528-539.

Financial support by European Union through the project CADASTER FP7-ENV-2007-1-212668