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 cleaners, fire-fighting foams, micelles, oil and water repellants for leather, paper, and textiles etc. Continuous 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) pre-registration 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% (LD50) was compiled for short and long chain PFCs on two species of rodents - Rat (Rattus) and Mouse (Mus). This oral exposure study was chosen as it’s 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 LD50 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 between 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 oral LD50, inhalation data of PFCs on rodents had been published earlier and combining with the result of current oral LD50 study, a comparative toxicology 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 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 (SMI) 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 pre-registration list.

### Molecular Descriptors

More than 600 molecular descriptors (DDT-3D) were calculated by the software DRAGON [1] from the X12c coordinates in Hyperchem using AM1 [2].

### Multiple Linear Regression

#### Model validation by permutation tests: Application of Model Validation by Permutation Tests (OVM) algorithm, rel. 2.3 for Windows, Talete srl, Milan (Italy); and 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 Q2 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 (Rf reshuffling).

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

#### Prediction of activity relationship (QSAR) = 75.93 ± 0.486 (Equation 1)

#### Validation of activity relationship (QSAR) = 75.93 ± 0.486 (Equation 2)

#### Summary of results

- 308 compounds within the AD of both the models
- 376 compounds were studied for structural AD study
- 14 compounds with experimental value in Inhalation dataset
- 28 compounds experimental value between Oral dataset
- A data plot was obtained by using those models to predict the test set

###RESULTS AND DISCUSSION

#### Toxicity Trend - LD50 Oral

- Top right: more toxic to Rat = fluorinated benzenes and azoles, bottom right: more toxic to Mouse = long chain PFCs including PFOSA center
- Neutral to Rodents = PFCs including PROA

### Overview - LC50 Inhalation [6]

- MDS plot of 7 molecule descriptors highlighting the chemical diversity of prioritized components
- 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

### 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 (MATS 1e, HATS 2u) and fingerprint based descriptors (POIC 0, 0 ≤ 0.04, 0.04 ≤ C0 ≤ 0.08)
- Models on Rodents Oral and Inhalation data shows importance of following main descriptors for overall toxicity
  - Long chain of PFCs
  - Positive electronic density (Jbet, Xelv, MATS 1e)
  - 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**

[3] MOBY DIGS. Rel. 2.3 for Windows, Talete srl, Milan (Italy).

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

**Supporting information and Supplementary Information**

**Supporting Information**

**Supplementary Information**