Combination of QSAR models and chemical clustering - application to the PFAS toxicology assessment

Management of Environmental & Health Risks

June 13 & 14, 2023 – Paris

Dr Philippe Oudin, CEO SEMACO
Dr Sinan Karaboga, Scientific and Commercial Director, HARMONIC PHARMA

HARMONIC PHARMA

SAFETY BY DESIGN®
for a world with
safer compounds

SEMACO
ENVIRONNEMENT

HARMONIC PHARMA
SAFETY BY DESIGN

SEMACO ENVIRONNEMENT

MINESnancy ARTEM

Training room with twenty workstations running
SAFETY BY DESIGN® software

Ecole des Mines de Nancy, at the heart of the Artem Campus
and a stone's throw from Nancy's new thermal baths

Outline

- Toxicological information source
  - State of the Environmental Issues
- NAM (New Approach Methodology):
  - Application to a dataset of 74 PFAS recommended by the EPA
  - Combination of approaches : Clustering, QSAR Predictions, TRV Mapping, Read Across with experimental data (TRV)
- Demonstration of Harmonic Pharma's Safety By Design® software
- Take-Home message
- Perspectives

Toxicological information source

State of the Environmental Issues

PFAS issue

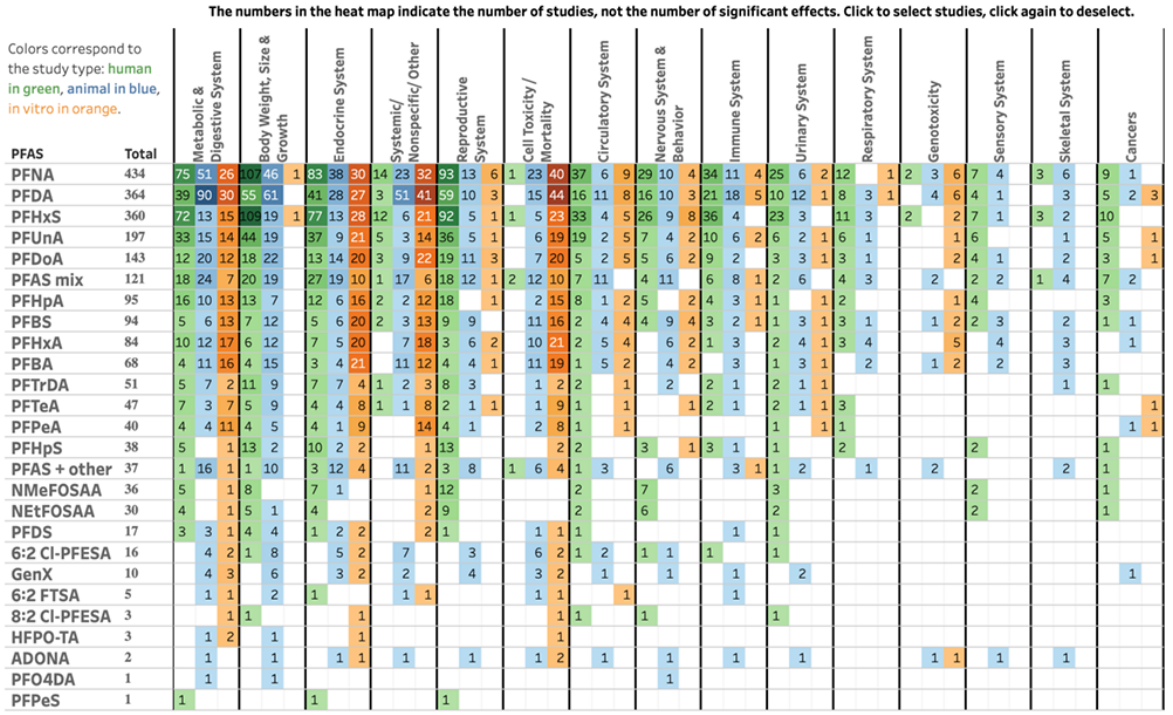
- BEFORE
  - 10 000 Molecules
  - Lack of TRV
  - Lack of studies

- TODAY
  - Many sites to investigate
  - Multiple Uncertainty
    - Sampling, Analysis, TRV

Environment International Volume 167, September 2022, 107408

- A systematic evidence map of health studies on 29 per- and polyfluoroalkyl substances

Contrairement à l'idée fabriquée par les industriels d'un manque de données sur les effets sanitaires des PFAS actuellement utilisés, le projet PFAS-Tox a identifié 742 études sur 29 PFAS, justifiant encore plus le besoin de gérer de tous les PFAS, environ 5000 substances, comme une seule classe de produits chimiques.



742 studies on 29 PFAs out of 10,000

Real issues

- Need to conduct health risk assessments based on validated toxicological methodology
- Impossible to evaluate each molecule with classical toxicological approach
- In line with European REACH regulation
  - Validation of “In silico” for toxicological evaluation by QSAR Models as an alternative to experimental testing

- QSAR : Quantitative Structure Activity Relationship
  - Validated By
    - OECD
    - European Chemical Agency (ECHA)

NAM
"New Approach Methodology"

Combining clustering approaches, predictive toxicology (QSAR)

and experimental data (reference TRV value)

HARMONIC PHARMA

SAFETY BY DESIGN®
for a world with
safer compounds

SEMACO
ENVIRONNEMENT

Case study

- **CONTEXT** :

Characterize the toxicological risk associated with a dataset of 74 PFAS selected by the EPA and representative of the ~10,000 PFAS

- **OBJECTIVES** :
  - Evaluation of each PFAs toxicity
  - Classification of 74 PFAS into clusters
  - Prioritize PFASs clusters to be tested regarding existing or non existing TRV
  - Gap-Filling of TRV by a Read-Across approach

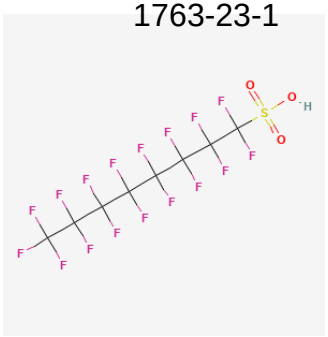
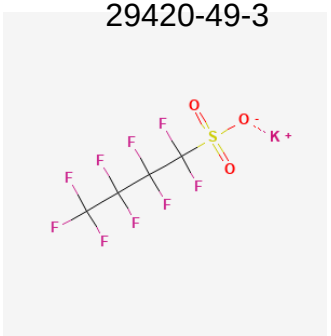
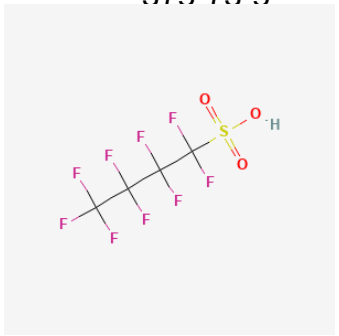
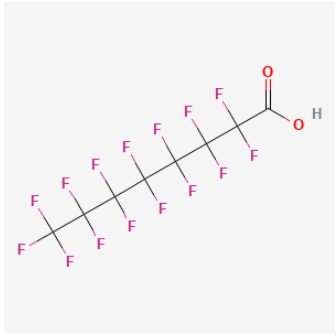
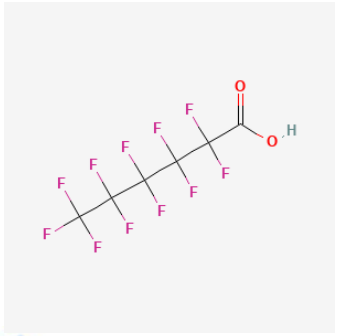
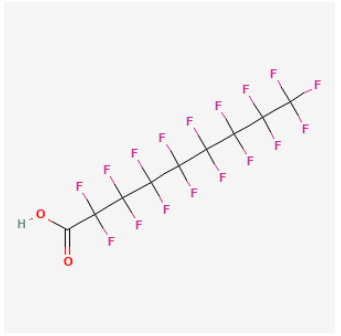
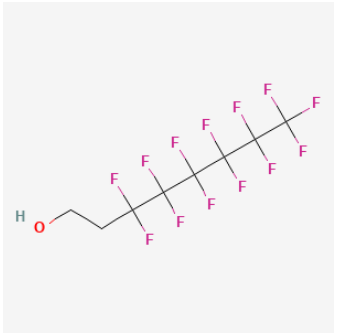
- **SOLUTION** :

A NAM "New Approach Methodology" combining clustering approaches, predictive toxicology (QSAR) and experimental data (reference TRV value)

Dataset of 74 representative PFAS

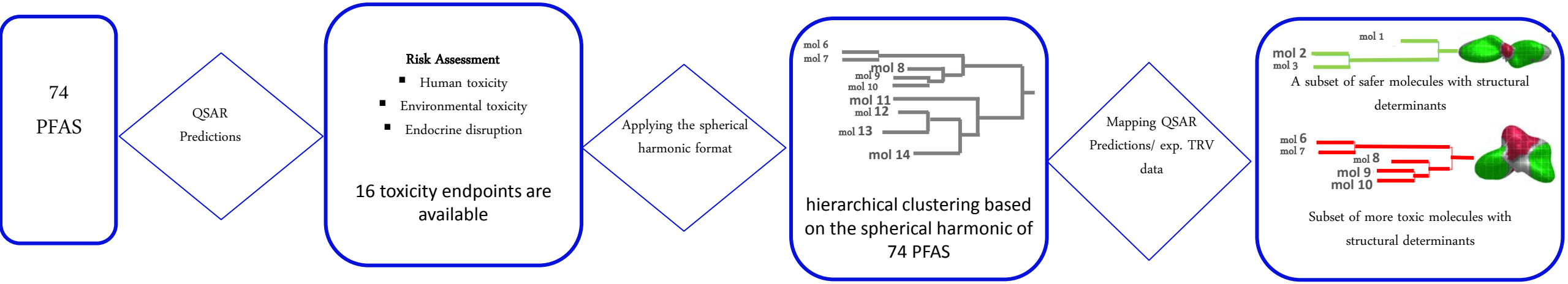
- The EPA has made available a subset of ~74 to support analytical method development and toxicity testing (https://www.epa.gov/chemical-research/pfas-chemical-lists-and-tiered-testing-methods-descriptions).

- This dataset of 74 PFAS was used in our study to propose a NAM combining in silico approaches and experimental data (TRVs).
- Only 7 out of 74 PFAS have experimental TRV available.

Perfluorooctanesulfonic 1763-23-1	Perfluorobutanesulfonate 29420-49-3	Perfluorobutanesulfonic 375-73-5	Perfluorooctanoic acid 335-67-1	Perfluorohexanoic acid 307-24-4	Perfluorononanoic acid 375-95-1	2-(Perfluorohexyl)ethan 647-42-7
						
0,00002 mg/kg/day (DSSTox)	0,02 mg/kg/day (DSSTox)	0,02 mg/kg/day (DSSTox)	0,00002mg/kg/day (DSSTox)	0,32mg/kg/day (Anses)	0,0000003 mg/kg/day (DSSTox)	0,000115 mg/kg/day (DSSTox)

Case Study: NAM for predicting PFAS toxicity

- Clustering, QSAR Predictions, experimental data and read-across as four pillars of a new methodological approach (NAM) to classify and predict PFAS toxicity



Read-accross

Demonstration of Harmonic Pharma's SAFETY BY DESIGN® software

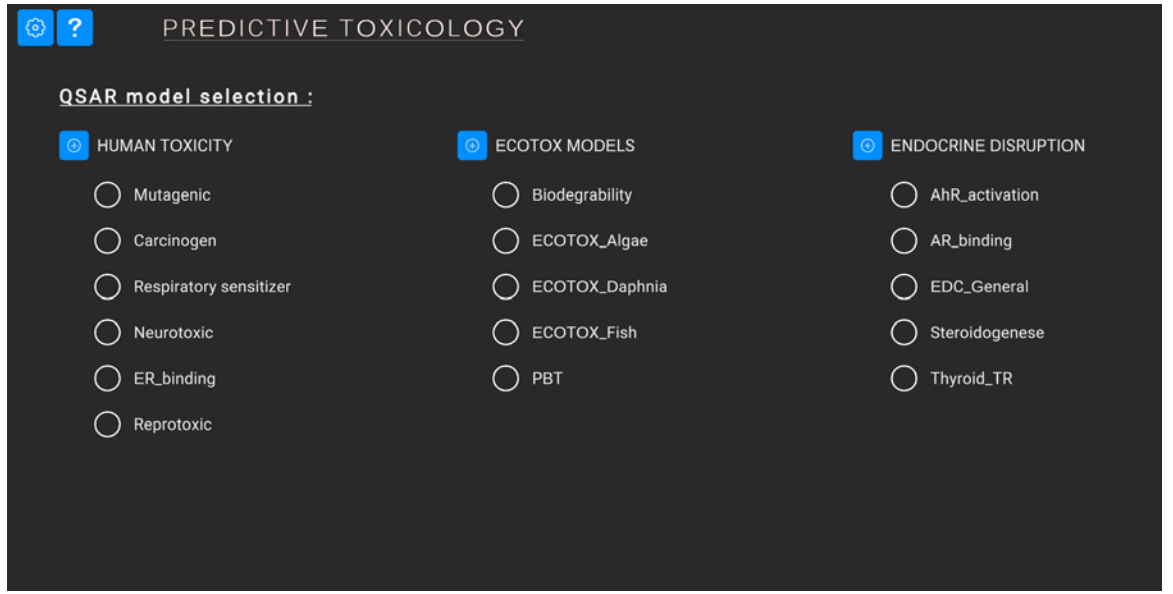
HARMONIC PHARMA

SAFETY BY DESIGN®
for a world with
safer compounds

SEMACO
ENVIRONNEMENT

QSAR Models

- 16 QSAR models are available through SAFTEY BY DESIGN® Software

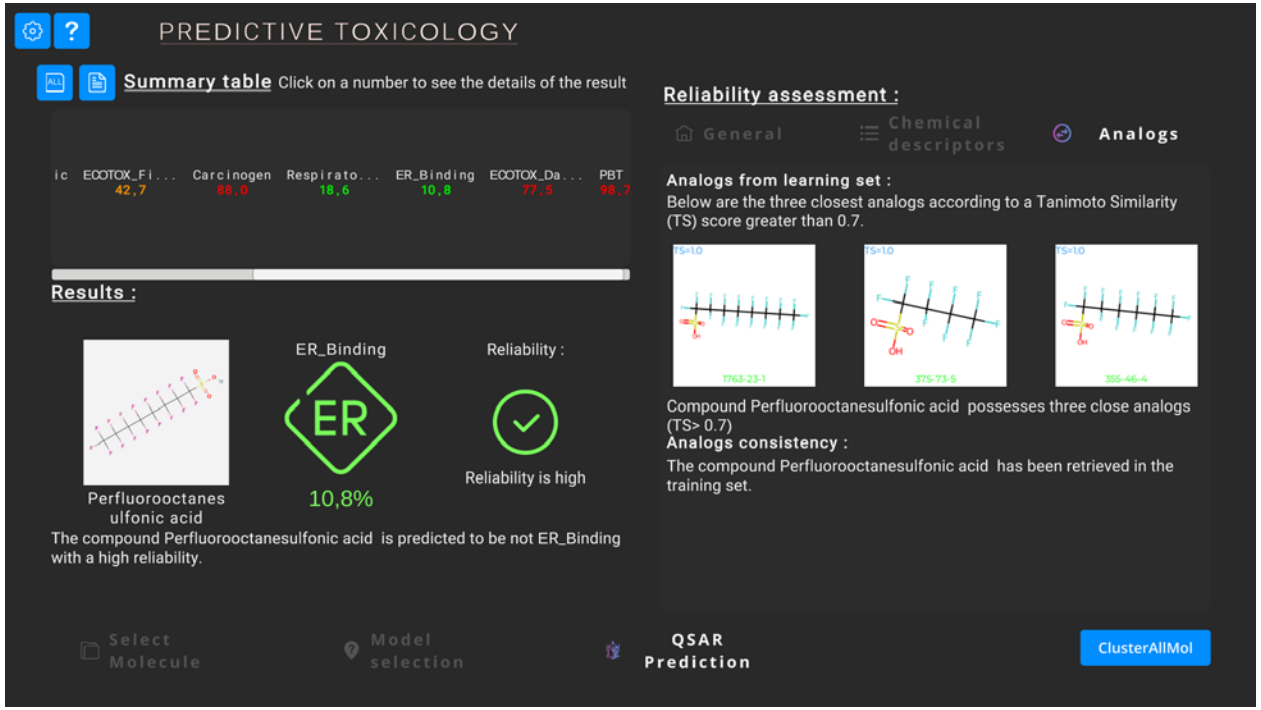


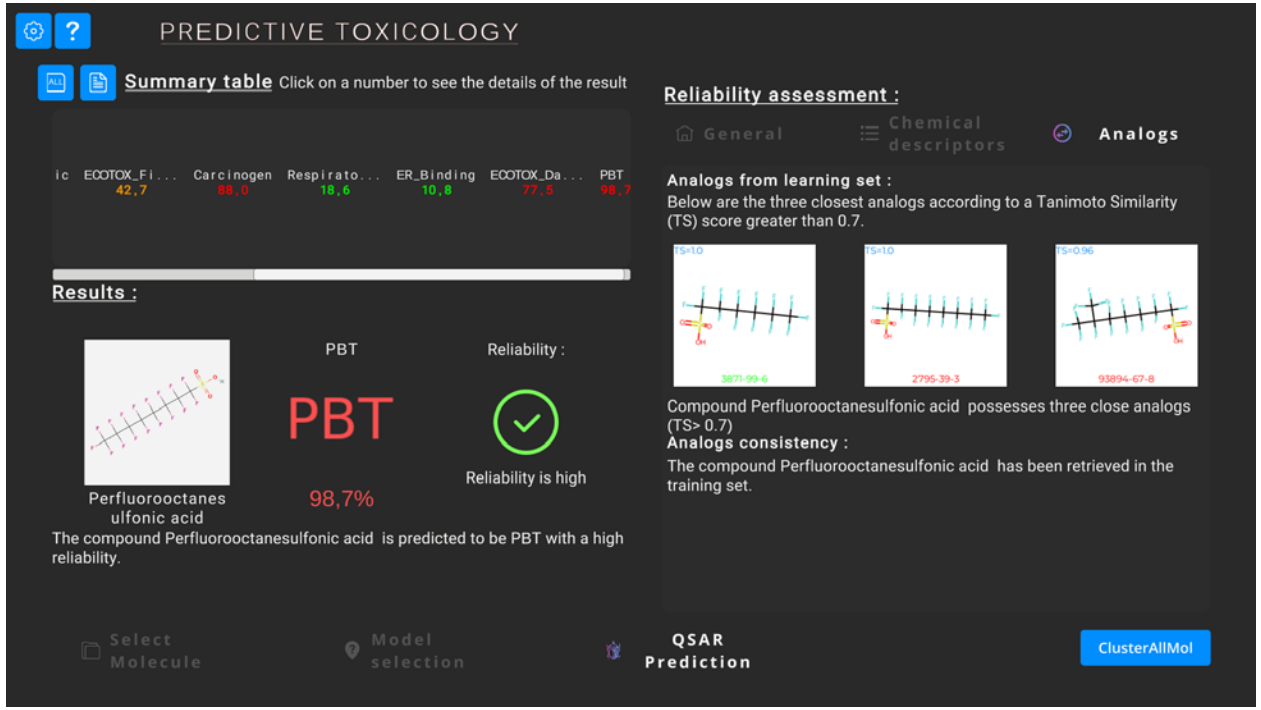
- Demo version is available at :

https://safetybydesign.software.harmonicpharma.com

Toxicological Evaluation by QSAR models

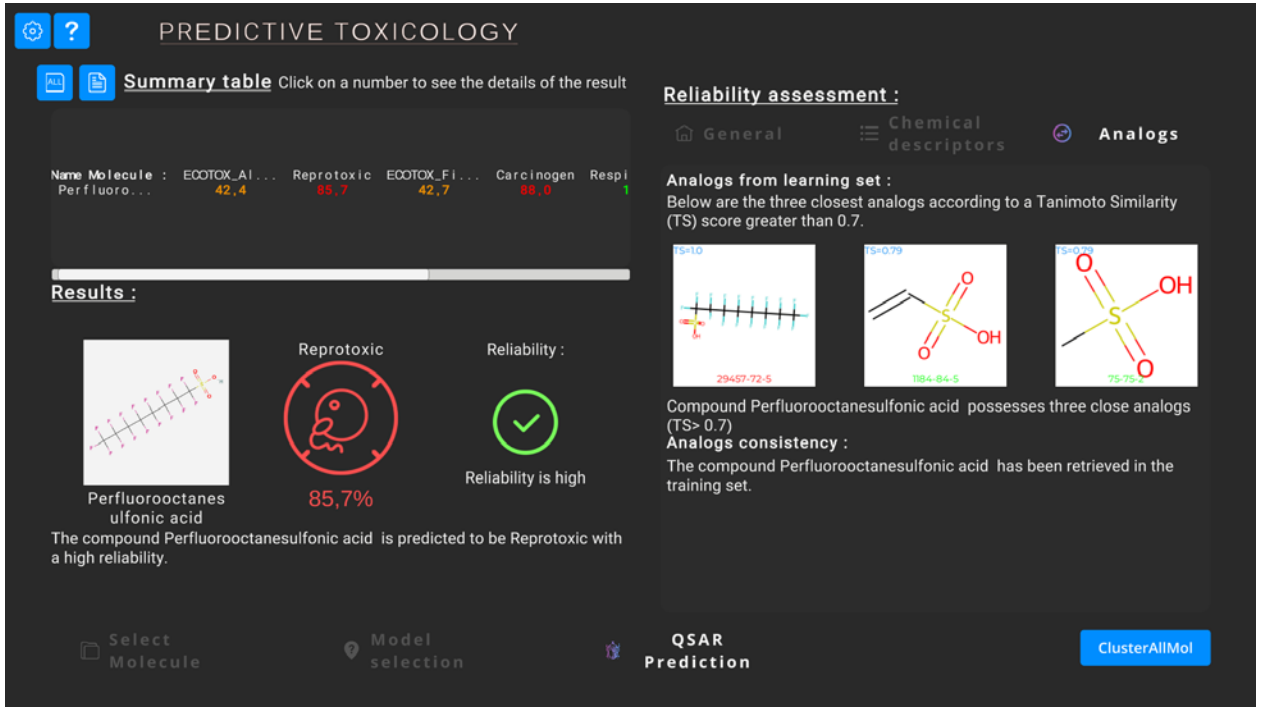
- Example of prediction for Perfluorooctanesulfonic acid (1763-23-1)

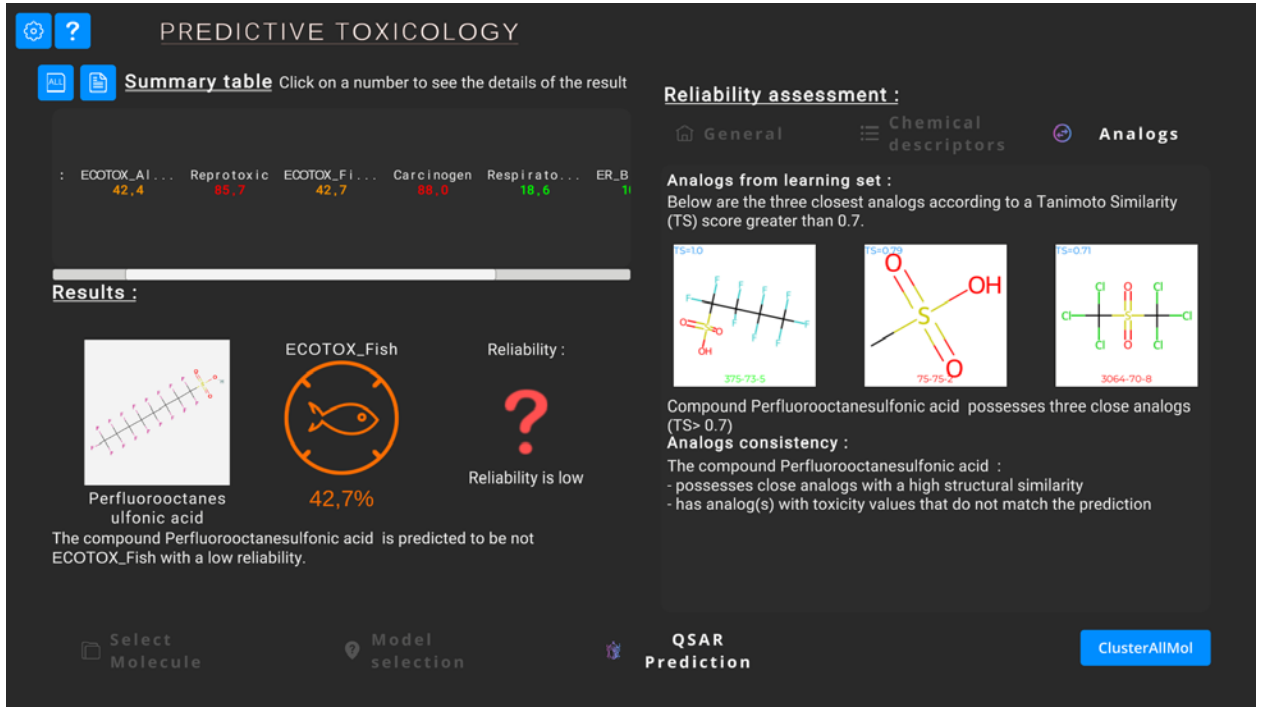




Toxicological Evaluation by QSAR models

- Example of prediction for Perfluorooctanesulfonic acid (1763-23-1)





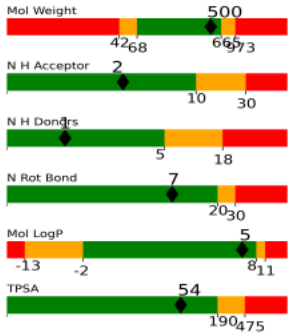
Automatic Reports Generation

The QSAR model

Endpoint : ER_Binding
Model creation date : 2021-07-01
Version : 1.0.0

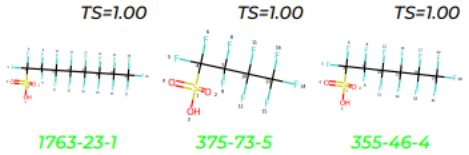
Applicability Domain

Compound Perfluorooctanesulfonic acid possesses chemical descriptors within the applicability domain, therefore the QSAR model is suited to make consistent predictions.



Database Analogs

Below are the three closest analogs according to a Tanimoto Similarity (TS) score greater than 0.7.



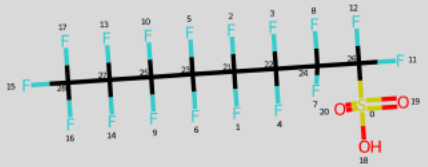
Three are reported to be safe (CAS in green) which support and strengthen the confidence of the current prediction.

Conclusion

The compound Perfluorooctanesulfonic acid is predicted to be not ER_Binding with a high reliability. Indeed, the reliability of the prediction considers that Perfluorooctanesulfonic acid :

- is within the applicability domain
- possesses close analogs with a high structural similarity
- has analog(s) with a coherent toxicity with the prediction
- has a statistical RF-reliability score of 0.78

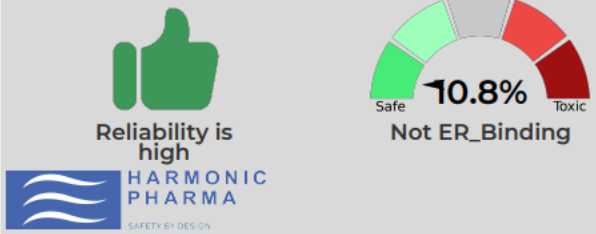
The molecule



Perfluorooctanesulfonic acid

O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C...

Summary result :



Reliability is high

Not ER_Binding

HARMONIC PHARMA
SAFETY BY DESIGN

Harmonic Pharma QSAR report

Model Name :	HarmonicPharma ER_Binding
Version :	1.0.0
Date model :	2021-07-01
QMRF:	
Endpoint description :	ER_Binding
Additional information :	The QSAR model has been developed using a proprietary dataset of 2934 compounds, 451 reported to be ER_Binding and 2483 as safe.

Substance information :

Substance name	Perfluorooctanesulfonic acid
Substance Smile	O=S(=O)(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C...

Contact information

Company :	HARMONIC PHARMA Campus Artem 92 rue du Sergent Blandan 54000 Nancy Tel. +33(0) 6 69 02 94 94 www.harmonicpharma.com
Contact person :	Sinan A. Karaboga, CSO/CBD karaboga@harmonicpharma.com

Prediction

Model outcome :	Random forest QSAR model is predicting a probability of toxicity of 10.8% for compound Perfluorooctanesulfonic acid

Assessment of the prediction :

Model algorithm and results interpretation:	Random Forest, the result is the probability of toxicity
Applicability domain definition :	The applicability domain of a QSAR model is the physico-chemical descriptors space on which the training set of the model has been developed, and for which it is applicable to make predictions for new compounds. Six descriptors are used : Molecular Weight, Number Hydrogen bond donor, Number hydrogen bond acceptor, Number rotatable bond, Molecular LogP and TPSA
Applicability domain result :	Compound Perfluorooctanesulfonic acid possesses chemical descriptors within the applicability domain, therefore the QSAR model is suited to make consistent predictions.
Structural Analogs definition :	Below are the three closest analogs according to a Tanimoto Similarity (TS) score greater than 0.7.
Structural Analogs :	TS=1.00 TS=1.00 TS=1.00 1763-23-1 375-73-5 355-46-4
Is the compound part of training set ?	No
Are the descriptors within boundaries ?	Compound Perfluorooctanesulfonic acid is is within the applicability domain
Does Molecule have neighbours ?	Compound Perfluorooctanesulfonic acid possesses three close analogs (TS> 0.7)
Have neighbours a coherent toxicity ?	Three are reported to be safe (CAS in green) which support and strengthen the confidence of the current prediction.
Is RF consensus strong ?	has a statistical RF-reliability score of 0.78

Conclusion

The compound Perfluorooctanesulfonic acid is predicted to be not ER_Binding with a high reliability. Indeed, the reliability of the prediction considers that Perfluorooctanesulfonic acid :

- is within the applicability domain
- possesses close analogs with a high structural similarity
- has analog(s) with a coherent toxicity with the prediction
- has a statistical RF-reliability score of 0.78

HARMONIC PHARMA

SAFETY BY DESIGN®
for a world with
safer compounds

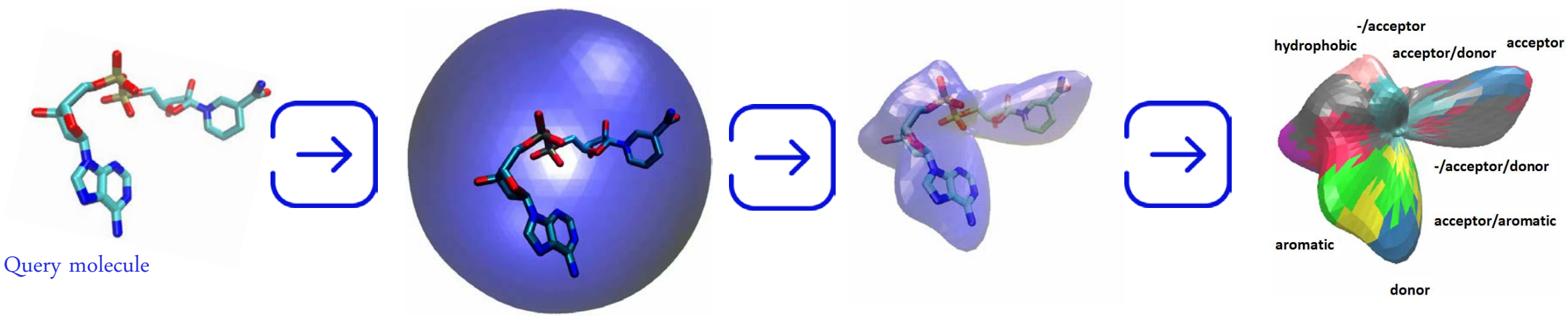
SEMACO ENVIRONNEMENT

First Step: Applying the spherical harmonic format

Placing a molecule inside a virtual sphere

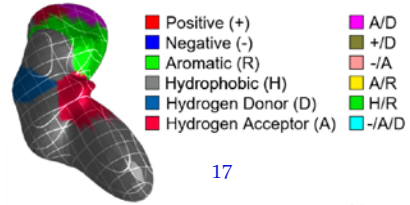
Deflating the sphere

Mapping properties onto the resulting surface

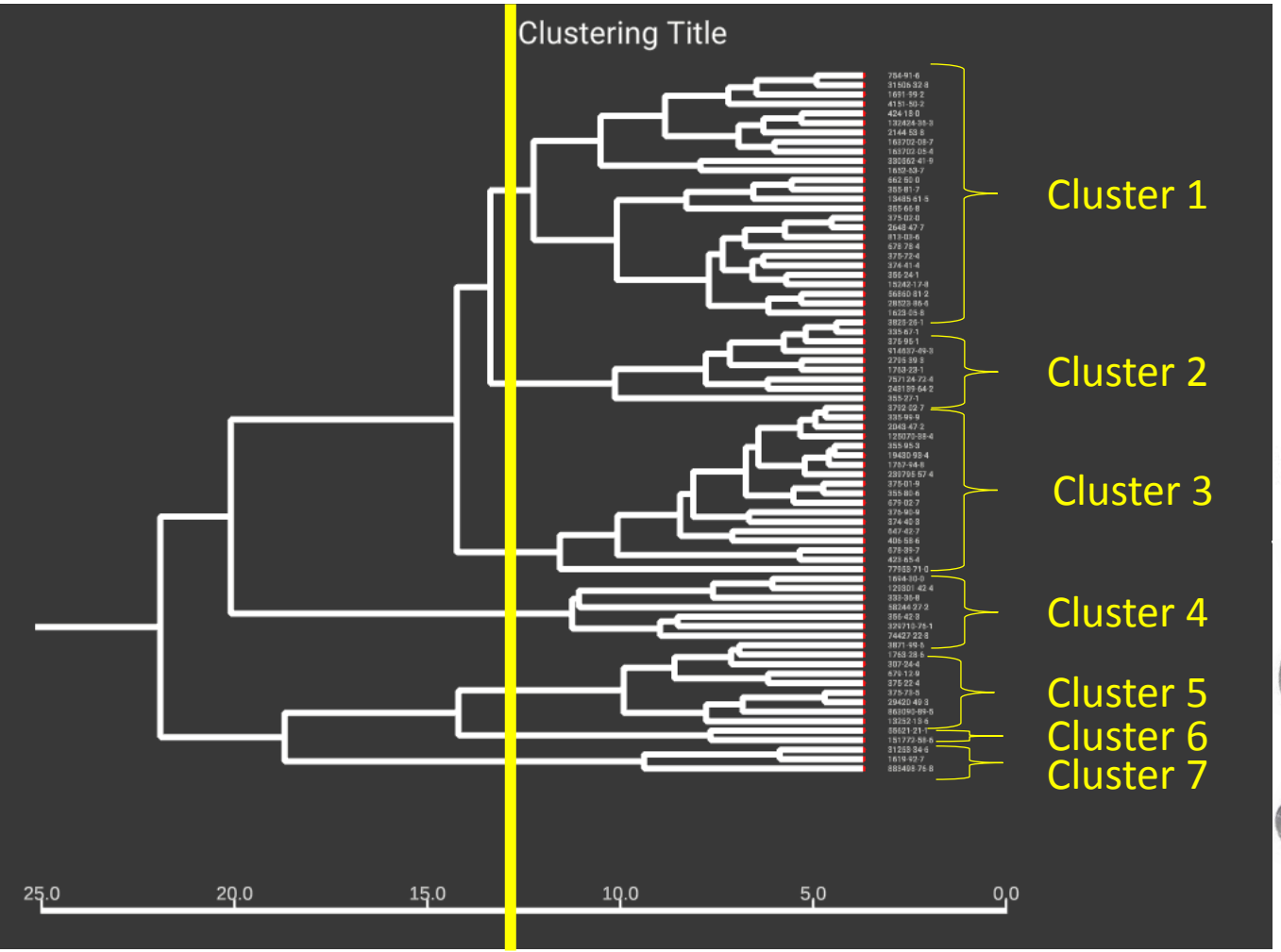


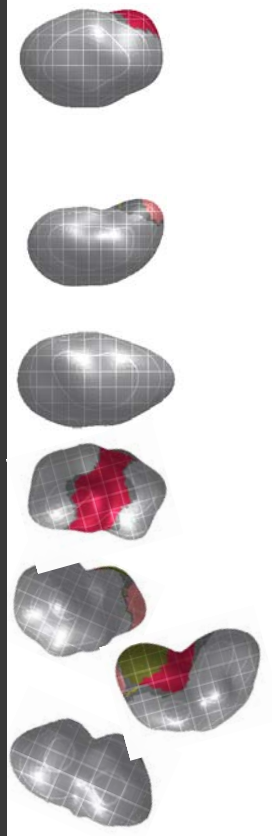
Query molecule

Unique 3D molecular descriptors



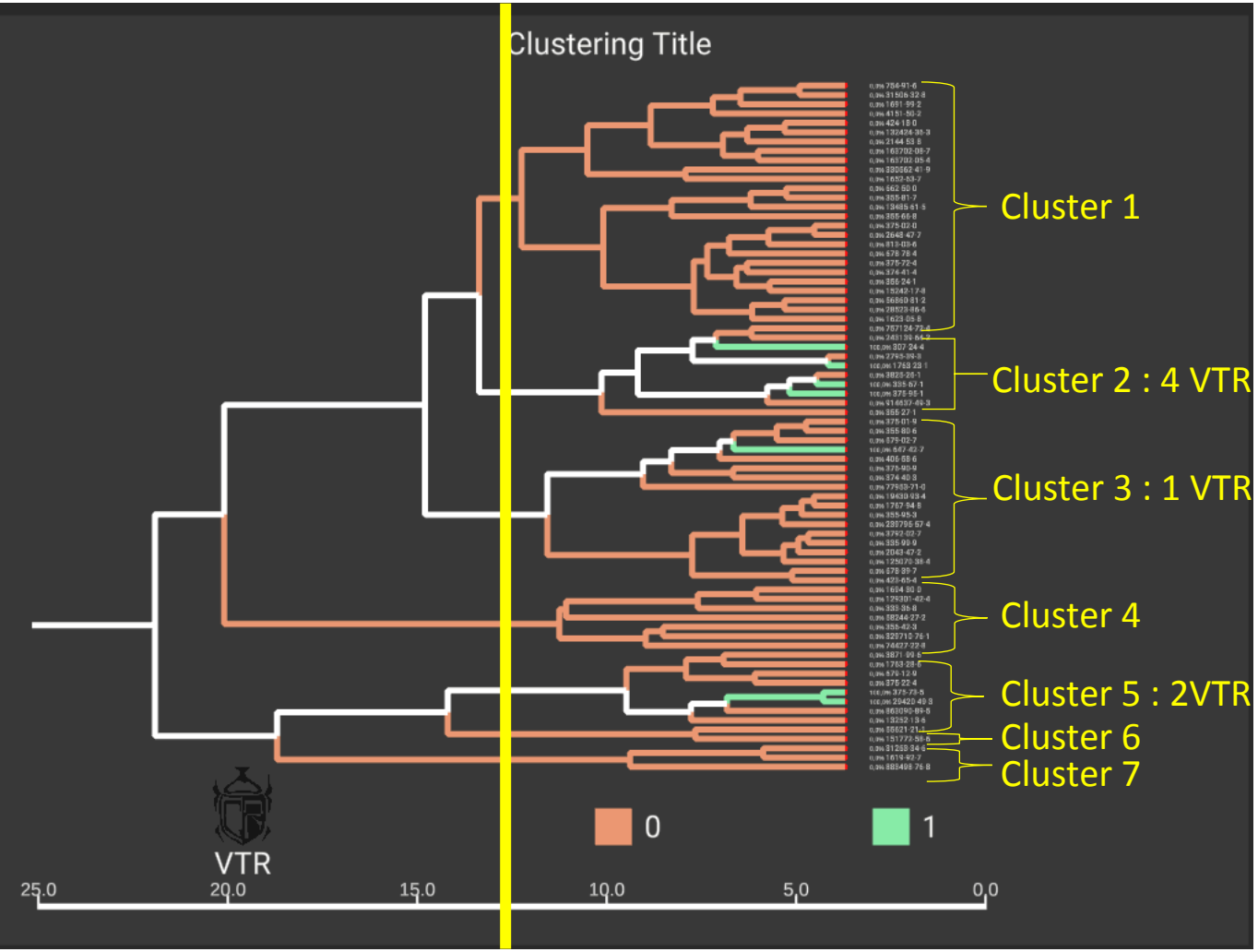
First Step: Clustering based on spherical harmonics



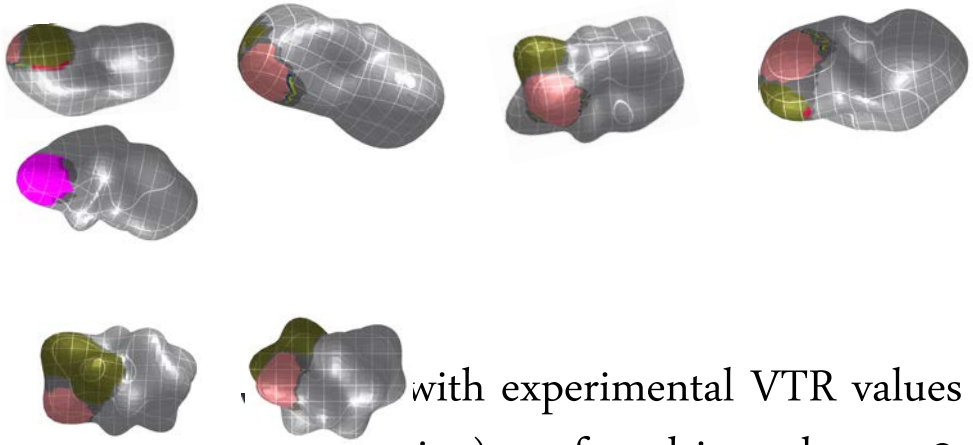


- 74 PFAS are classified into 7 clusters based on structural similarity derived from Spherical harmonics descriptors
- Each cluster is represented by a "metamolecule" which is the sum of each underlying spherical harmonics

Step 2: Mapping experimental TRV data

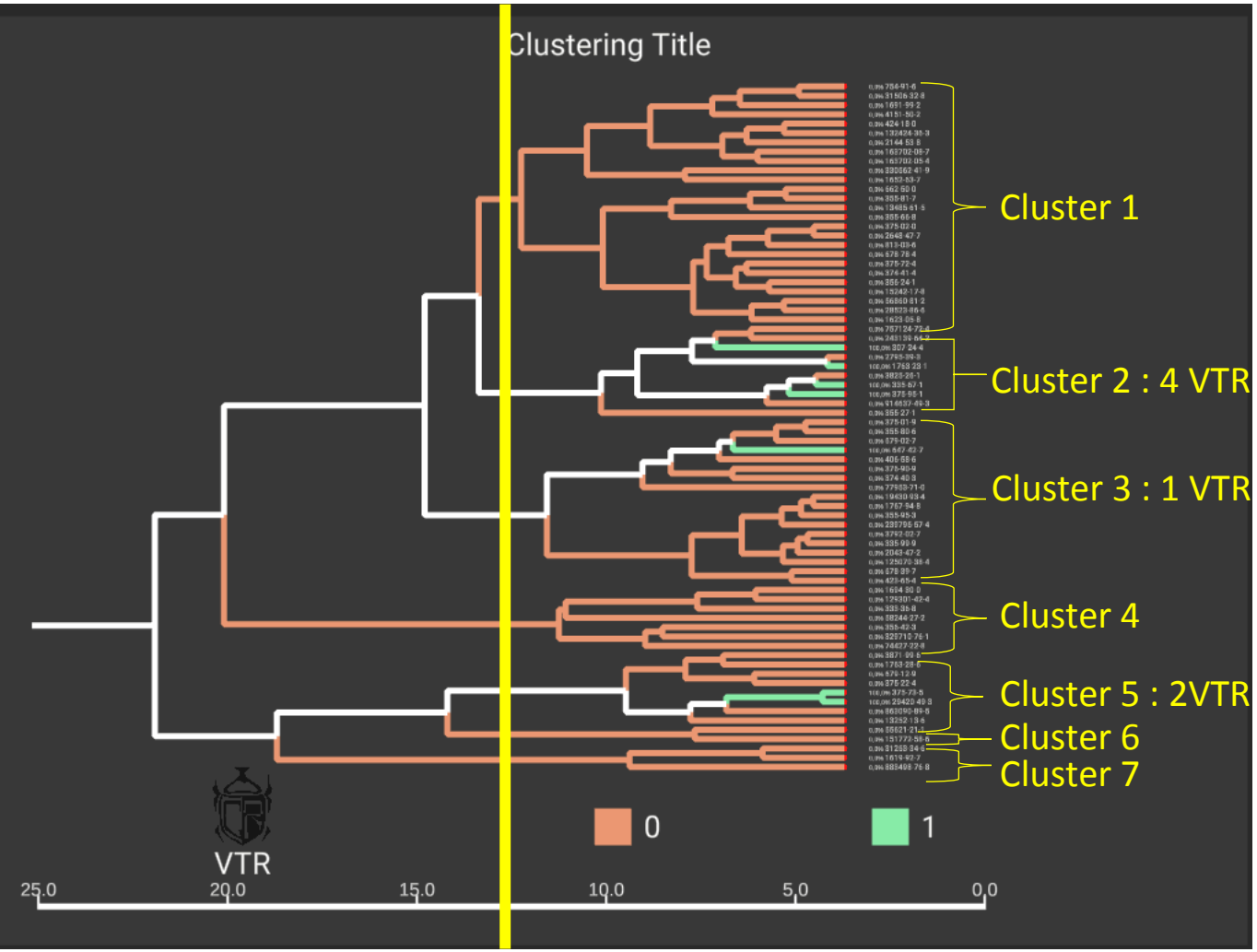


- PFAS found in Clusters 1, 4, 6 and 7 show lack of experimental TRV and should therefore be tested as a priority

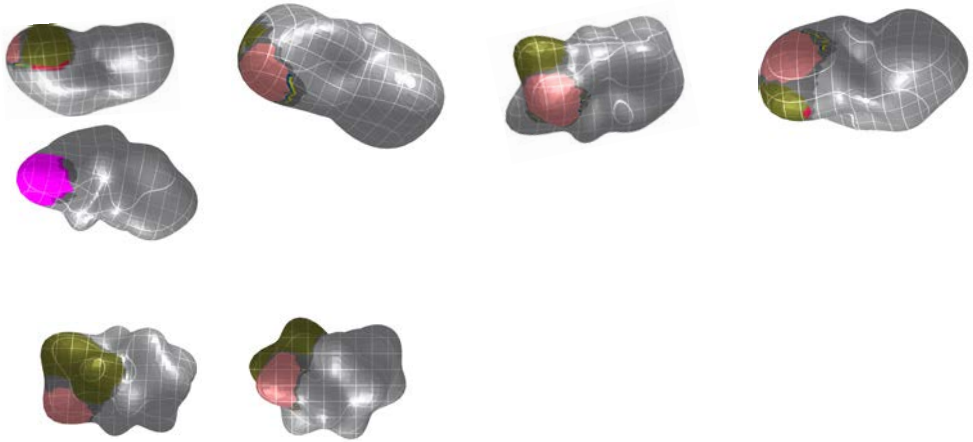


- with experimental VTR values (in green line) are found into clusters 2, 3 and 5

Step 3: Read-Across



- PFAS having VTR values could be used for read-across approach for other members of the respective cluster



Take-Home message

- We are able to provide a QSAR prediction for each PFAs on 16 toxicity endpoints

- We have classified 74 representative PFAS into 7 structural similarity clusters using spherical harmonics descriptors

- We pointed out lack of experimental TRV for some clusters and therefore recommend to test some members of those clusters in priority

- For clusters having existing experimental TRV data, we provide a read-across approach to attribute TRV values to PFAS devoid of experimental TRV

Perspectives

- The presented study could be extended to :
  - the whole ~10,000 PFAS dataset by providing support for gap-filling approaches by read-across using metamolecules identified from the 74 representative PFAS subset
  - to multiple toxicity endpoints such as Persisent Bioaccumulative and Toxic (PBT), Biodegradability, Eco-Toxicity and more according to your main interest

- All those functionalities are implemented in our software called **SAFETY BY DESIGN®** and is accessible upon request with several licensing options

https://safetybydesign.software.harmonicpharma.com

THANK YOU !

For any questions or requests, please contact us at

contact@harmonicpharma.com

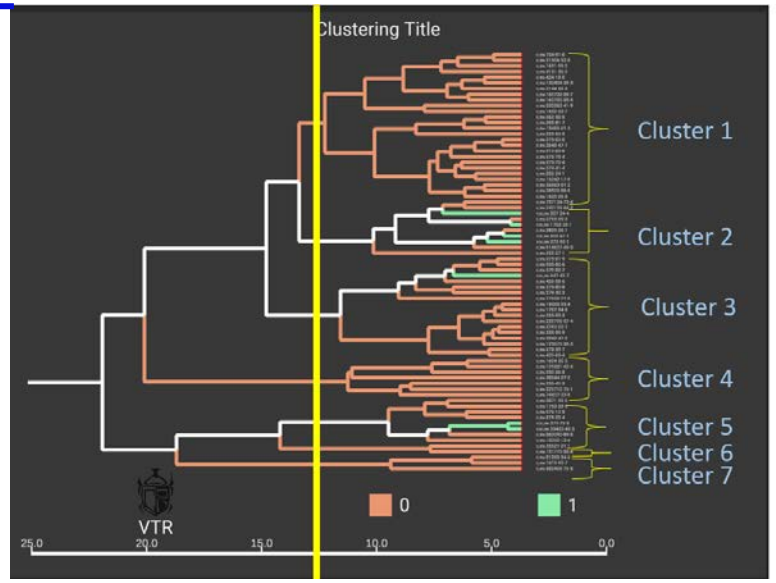
info@semaco.fr

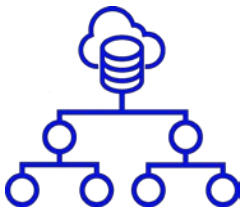
HARMONIC PHARMA
SAFETY BY DESIGN

SEMACO ENVIRONNEMENT

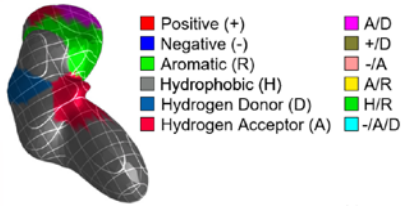
Full picture

Acute Toxicity
Carcinogenicity
Mutagenicity / Genotoxicity
Reprotoxicity
Neurotoxicity
Respiratory sensitization
Endocrine disruption – Full model
Endocrine disruption – Estrogen receptor
Endocrine disruption – Androgen receptor
Endocrine disruption – Thyroid Perturbation
Endocrine disruption – Steroidogenesis
Endocrine disruption – Aryl Hydrocarbon Receptor
Persistent, bioaccumulative and toxic (PBT) behavior
Eco-toxicity - Daphnea
Eco-toxicity - Algae
Eco-toxicity - Fish
Biodegradability
Skin sensitization – DPRA model
Skin sensitization – hCLAT model
Skin sensitization – KeratinnoSens® model
Skin sensitization – LLNA model
Skin irritation / Corrosion

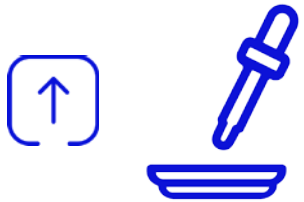




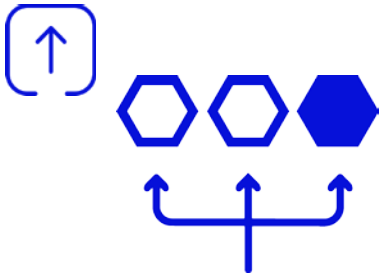
Predictive Toxicology



Unique 3D molecular descriptors



Empower experimental data



Understand the MOA with Activity profiling

SFT Société Française de Toxicologie
Best poster prize
Annual Congress 2022

HARMONIC PHARMA
SAFETY BY DESIGN®
for a world with safer compounds
SEMACO ENVIRONNEMENT