# Investigation of the Possible Use Of *in Silico* Methods to Improve the Threshold of Toxicological Concern Approach

*Elena Fioravanzo[1], Sue Barlow[2], Arianna Bassan[1], Alan Boobis[3], Manuela Pavan[1], Andrew Worth[4], Chihae Yang[5].*

*1- S-IN Soluzioni Informatiche, Vicenza, Italy (www.s-in.it) elena.fioravanzo@s-in.it;*
*2- Consultant, Brighton, UK;*
*3- Imperial College of London, London UK;*
*4- European Commission - Joint Research Centre, Ispra, Italy;*
*5- Altamira, LLC, 1455 Candlewood Dr., Columbus, Ohio 43235.*

## INTRODUCTION

The Threshold of Toxicological Concern (TTC) approach is an assessment tool based on the principle of establishing a human exposure threshold value for chemicals, below which the probability of causing adverse effects to human health and/or the environment is very low. According to this approach, such a level of exposure can be identified for many chemicals based on their chemical structure and known toxicity. Starting with the generic approach ('exposure threshold') used by the US Food & Drug Administration (FDA) in the 80s, the TTC concept has evolved over the years to take into account extensive analysis of available data on mainly the oral toxicity of substances and their intake/exposures. This led to a structure-based decision tree which has been used mainly in the food area. The approach has been used to evaluate flavouring substances[1], food contact materials[2], genotoxic impurities in pharmaceuticals[3] and has been proposed for use in the risk assessment of chemicals more generally[4]. In 2010 the European Food Safety Authority (EFSA) funded the present research project aimed at investigating how the applicability of TTC schemes can be improved by incorporating physicochemical data as well as toxicity data generated by non testing methods. In the current study we explore *in silico* methods that can give further insight into the Cramer classification scheme[5] which is the best known approach for structuring chemicals in order to make a TTC estimation. The work presented is in progress and the results are preliminary and do not represent the opinion of the EFSA.

## MATERIALS AND METHODS

### Datasets

**CPDB dataset** (Carcinogenic Potency Database)[6,7]. Carcinogenicity dataset currently under revision by Cheeseman. 651 chemicals with corresponding TD50 values.

**Munro dataset**[8]. Non-cancer toxicological endpoints dataset. 613 organic chemicals tested for a variety of non-cancer endpoints in rodents and rabbits in oral toxicity tests.

**Dataset compilation and quality check**. Electronic copies of the datasets were compiled. The datasets were verified for their chemical specification: chemical name; CAS RN (Identification of structures missing the CAS registry numbers); CAS RN and name correspondence; chemical and structural information. CPDB dataset: 609 structures; Munro dataset: 596 structures confirmed.

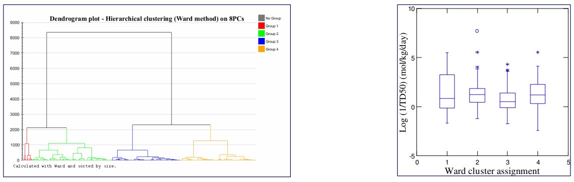
**Generation of QSAR ready datasets**. Identification and removal of missing structures and duplicates[9]; treatment of ionized species applying neutralization and desalting[10]; generation of 3D structures[10]; generation of descriptors; TD50 and NOEL original values expressed in mg/kg/day were transformed into mol/kg/day and a logarithmic transformation was applied.

### Characterization of the chemical space

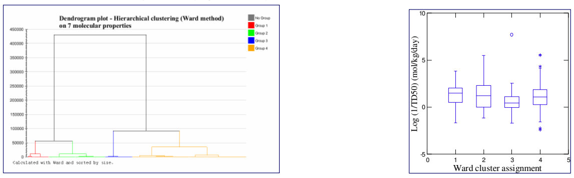
**Descriptors**
Structural Molecular descriptors: QikProp[11], ChemAxon[12] and Adriana[13] descriptors
Fingerprints: Canvas[14] Dendritic fingerprints and Molprint2D fingerprints.
Physicochemical properties: QikProp[9] and ChemAxon[10].

**Statistical Tools**
X-matrix analysis: Principal component analysis and O2PLS[15].
Classification techniques: Cluster Analysis[15], Naïve Bayesian còassifier[16], Decision tree[16].

## RESULTS - CPDB DATASET



Cluster analysis on structural molecular descriptors and box-and-whisker diagram. Structures are clustered according to the structural information rather than on their carcinogenicity potency.

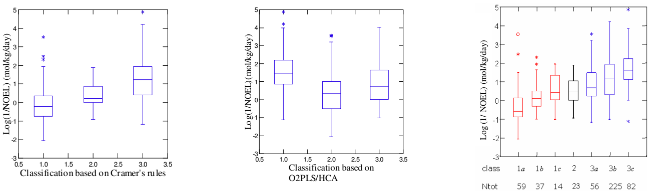


Cluster analysis on structural physicochemical properties and box-and-whisker diagram. Clusters 2 and 4 are very close in terms of Log(1/TD50) values (medians very similar) and are characterized by structures that are moderately carcinogenic. Cluster 1 is populated by the most carcinogenic structures, while cluster 3 by the less carcinogenic structures.

**CONCLUSIONS**. There is a clear relationship between molecular descriptors and potency ranges. To explore the feasibility of higher threshold levels for sub-groups of chemicals, it would be of interest to develop a classification model, able to predict whether an untested substance is likely to be a potent carcinogen.

## RESULTS - MUNRO DATASET

The autoscaled O2PLS model provided 5 parallel components and 2 orthogonal components for each block. The 5 components were employed to perform a Hierarchical Cluster Analysis. The eight variables identified as characterizing the three clusters were used to develop a simplified Naïve Bayesian classification model. The three identified clusters and the three Cramer classes were compared with respect to the Log(1/NOEL) (mol/kg/day) by the box plots below (a, b). The Kruskal-Wallis test demonstrated in both cases that with a confidence level greater than 99% the three medians are not equal. Moreover the two classifications, compared by the confusion matrix, were very different. The three Cramer classes were then analyzed separately to highlight subgroups of structures. The main variables characterizing each class were used to develop a decision tree which provided a simple but effective set of rules that allows correct prediction the classification. The NOEL dissimilarities within the three subclasses 1a, 1b and 1c of class I and 3a, 3b and 3c of class III are shown in Box plot (c).



(a) Cramer (b) Bayesian (c) Decision tree

**CONCLUSIONS**. There appears to be a clear relationship between structure and NOEL, which suggests the feasibility of applying a complete QSAR approach. The subclasses identified describe the NOEL differences with greater discrimination than the Cramer classification scheme. However, this needs to be confirmed in further work.

## REFERENCES

**1**. Schnabel, Regulatory and Scientific Affairs, Givaudan Flavours. *The TTC Concept: Past Present and Future Developments* Eurotox CEC 5,13 September 2009.
**2**. Hirose et al., *Toxicology Letters*, Volume 172, Supplement 1, 7 October **2007**, Page S195.
**3**. Müller et al., *Regulatory Toxicology and Pharmacology*, Volume 44, Issue 3, April **2006**, Pages 198-211.
**4**. Kroes et al., *Toxicological Sciences* **2005** 86(2):226-230.
**5**. Cramer et al, *Food Cosmet Toxicol*, **1978**, 16, 255-276.
**6**. Gold et al., *Environmental Health Perspectives*, **1984**, 58, 9-319.
**7**. Kroes et al., *Food and Chemical Toxicology*, **2004**, 42, 65-83.
**8**. Munro et al., *Food Chem Toxicol*, **1996**, 34, 829-867.
**9**. KNIME v2.0.3.0021120, Konstanz Information Miner Copyright, 2003 – 2009.
**10**. LigPrep version 2.3, Copyright 2009 Schrödinger, LLC.
**11**. Qikprop version 3.2, Copyright 2009 Schrödinger, LLC
**12**. ChemAxon, Jchem for Excel version 5.3.1.1885, Copyright 2008-2009 ChemAxon Ltd.
**13**. Adriana, Algorithms for the Encoding of Molecular Structures, version 2.2.2, Molecular Networks GmbH Computerchemie, December 2009.
**14**. Canvas version 1.2.211, Schrödinger Software Modules, Copyright Schrödinger, LLC.
**15**. Simca, version 12.0.1.0, April 21 2009, Copyright Umetrics.
**16.** WEKA version 3.4.11, University of Waikato, New Zealand.

*This poster is the output from an ongoing scientific project that EFSA has funded to support its work in accordance with EFSA public procurement rules. It was produced by the contractor (S-IN) of an EFSA tender following a strict tendering procedure. The present poster is published complying with the transparency principle to which EFSA is subject and cannot be considered as an output adopted by EFSA. EFSA reserves its rights, view and position as regards the issues addressed and the conclusions reached, in the present document, without prejudice to the rights of the authors*