



soluzioni informatiche

Contract: Preparation of a report on applicability of physicochemical data, QSARs and read-across in Threshold of Toxicological Concern assessment

S-IN, Soluzioni Informatiche

Manuela Pavan, Arianna Bassan, Elena Fioravanzo, and Matteo Stocchero

9 June 2011, Brussels

Workshop on the Threshold of Toxicological Concern: Scientific challenges and approaches
BOG 1: Non-cancer - Database and chemical domain

1. TTC datasets analysis

- Investigation and characterisation of the chemical space of the TTC datasets.
- Comparison of the chemical space of the TTC datasets with a large reference dataset.

2. Investigation of possible improvements by means of *in silico* methods.

- Applicability of QSAR approaches
- Critical evaluation of the Cramer classification scheme
- Refinement of the Cramer classification scheme by statistically based methods

3. Summary

1. TTC datasets analysis

- Investigation and characterisation of the chemical space of the TTC datasets.
- Comparison of the chemical space of the TTC datasets with a large reference dataset.

2. Investigation of possible improvements by means of *in silico* methods.

- Applicability of QSAR approaches
- Critical evaluation of the Cramer classification scheme
- Refinement of the Cramer classification scheme by statistically based methods

3. Summary

- **Datasets**

1. **Carcinogenicity dataset**: it was originally based on 343 carcinogens from animal studies compiled in the Carcinogen Potency Database (CPDB) (*Gold et al., Environmental Health Perspectives, 58, 9-319, 1984*). Subsequently, the dataset has been expanded to **651 carcinogens** taking into account the continuously updated CPDB.
2. **The non-cancer toxicological endpoints dataset** of Munro et al. (*Munro et al., Food Chem Toxicol, 34, 829-867, 1996*) consists of **613 organic chemical** substances tested for a variety of non-cancer endpoints in rodents and rabbits in oral toxicity tests. It includes the chemical structures and the distribution of No Observed Effect Levels (NOEL) for chronic, subchronic, and reproductive toxicity after oral administration.

The two datasets were retrieved and importantly verified for the chemical specification !

Three different classes of MOLECULAR DESCRIPTORS were employed:

- **74 Structural molecular descriptors**
 - QikProp descriptors (constitutional descriptors and functional groups)
 - ChemAxon descriptors (geometrical molecular descriptors)
 - Adriana descriptors (shape and global descriptors)
- **21 Descriptors of physicochemical properties**
 - QikProp physchem and ADME descriptors of properties
 - ChemAxon protonation and partitioning
- **Fingerprints**
 - CANVAS dendritic and Molprint2D fingerprints
 - MOSES fingerprints (structural fragment fingerprints)

CONCLUSIONS OVER THE DESCRIPTORS.

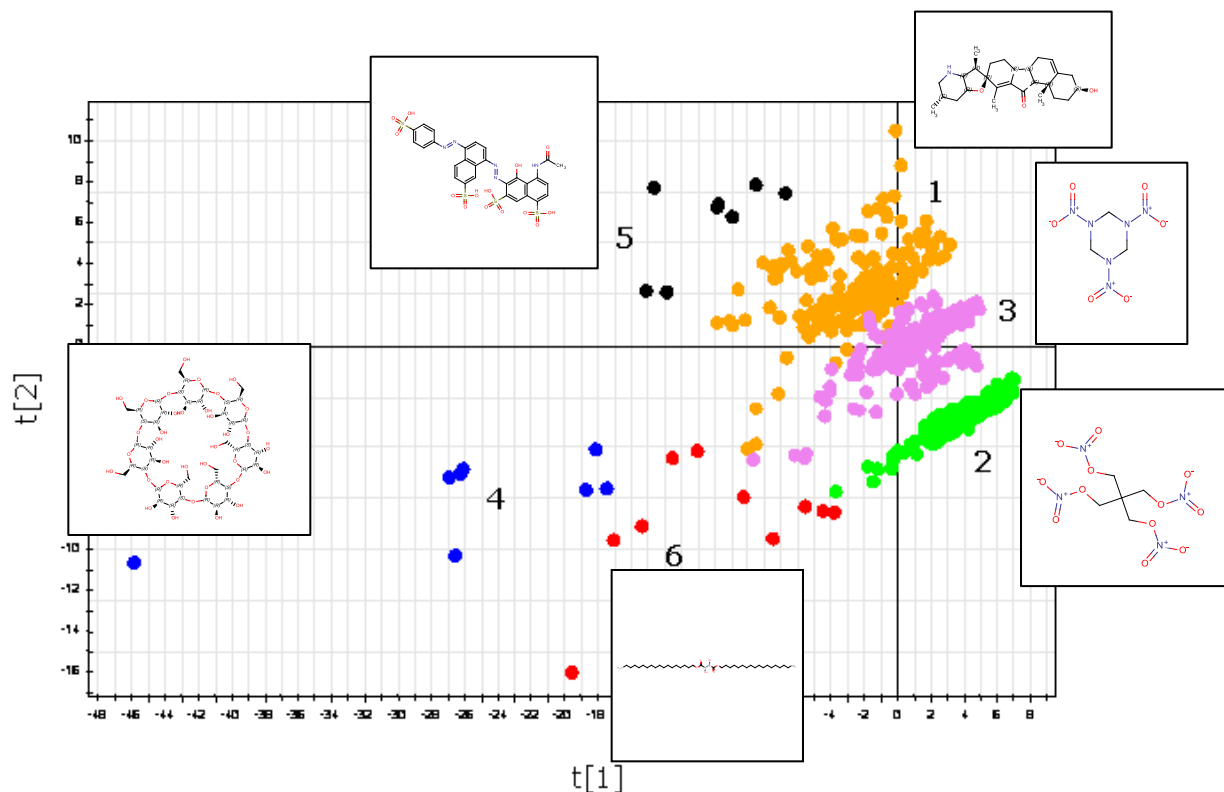
- The representation obtained by the **molecular descriptors is able to better discriminate the structures.**
- The representation obtained with the **molecular descriptors contains the one obtained with the physico-chemical properties** and that the two representations are not mutually independent.
- **Mainly the molecular descriptors were employed for further analysis as they include the physico-chemical properties information.**
- **The fingerprint turned out to be less informative than the molecular descriptors.** In fact they identified rather confused, not well defined groups of similar structures, named clusters.

Aims

- Assess structural similarities and dissimilarities in the datasets
- Assess which approach – statistical methods & molecular descriptor – may improve the TTC approach
- Provide a comprehensive view of selected data mining approaches
- Explore the possibilities of future development.

Munro dataset results

- Six clusters were identified and characterised in terms of individual descriptors
 - **560 relative homogeneous structures** → 3 clusters
 - **27 diverse structures** → 3 clusters



$R^2X[1] = 0.38$ $R^2X[2] = 0.14$

- Cluster 1
 - High #ringatoms
- Cluster 2
 - High branching
- Cluster 3
 - intermediate values of #ringatoms and branching
- Cluster 4
 - High no. aliphatic rings
- Cluster 5
 - High no. aromatic rings
- Cluster 6
 - High no. rotatable bonds

The chemical characterization of the TTC datasets was also performed by comparing them against the wider universe of chemicals as represented by the **DSSTox dataset** (about 10,000 chemicals) as a reference.

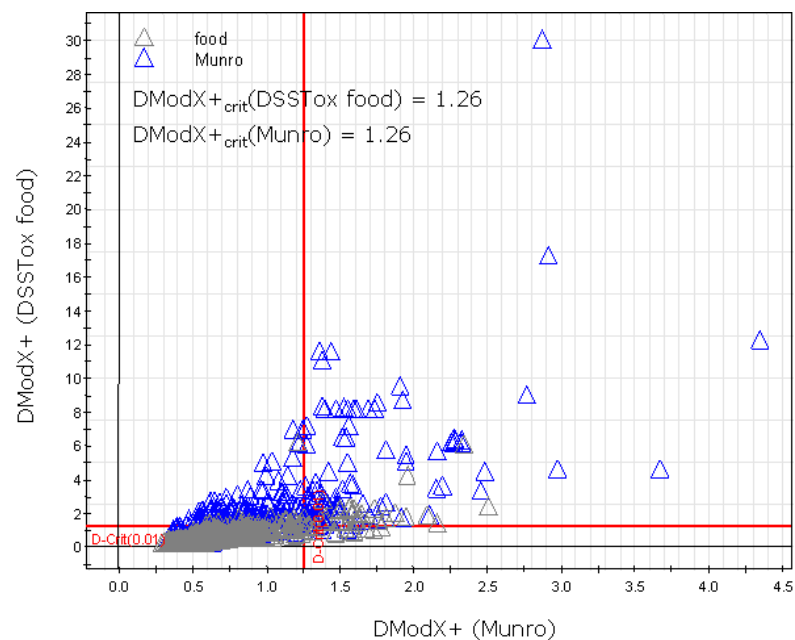
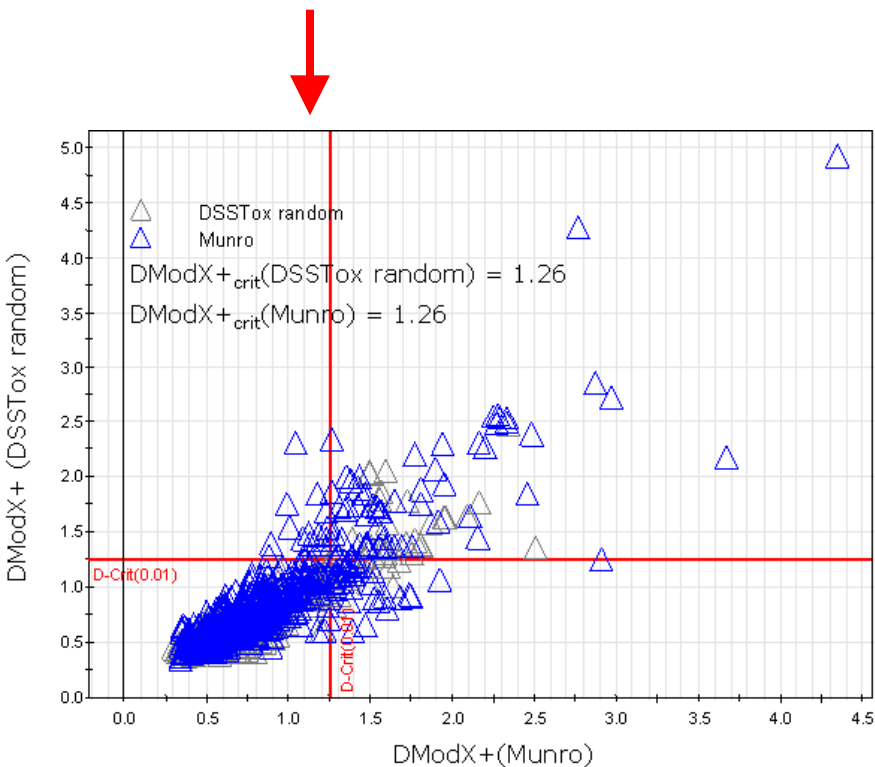
Two reference datasets:

1. **Food DSSTox subcategory** → 495 food related ingredients
2. **Random DSSTox subcategory** → 502 structures (taken as representation of the entire dataset)

The comparison was performed according to:

- **Structural and physico-chemical property descriptors**
- **MOSES Fingerprints**

Munro versus random DSSTox dataset: same space !



Munro versus food DSSTox dataset: approximately 30% of the Munro dataset is characterized by higher molecular weight with respect to food DSSTox dataset compounds.

1. TTC datasets analysis

- Investigation and characterisation of the chemical space of the TTC datasets.
- Comparison of the chemical space of the TTC datasets with a large reference dataset.

2. Investigation of possible improvements by means of *in silico* methods.

- Applicability of QSAR approaches
- Critical evaluation of the Cramer classification scheme
- Refinement of the Cramer classification scheme by statistically based methods

3. Summary

Outcome on the applicability of QSARs

It was **NOT possible** to develop a global QSAR model to predict either NOEL values employing the Munro dataset or TD50 values employing the CPDB dataset.

Possible reasons:

- intrinsic high **heterogeneity** of the datasets.
- **little correlation** between the examined biological activities and the calculated descriptors.
- chronic toxicities (NOEL and TD50) are the result of **complex biological mechanisms** of action probably composed by several steps with no apparent unifying concept.

Outcome on the applicability of QSARs

An alternative approach should be employed:

- 1) grouping of chemicals according to their **mechanism of action**;
- 2) development of **local** QSAR models.

?

No information about the mechanism of actions of the chemicals included in the two studied datasets were available, **this approach could not be applied.**

Recommendation: local QSAR models could be developed for the **structural clusters** identified in the above described analysis.

Conclusion:

- Based on the analysis performed employing a wide variety of statistical methods and molecular descriptors it turned out that the Cramer scheme **well fits the regulatory needs being highly conservative.**
- The percentage of experimentally high hazard structures classified as low hazard is less than 5%.

Aim

Assess the possible refinement of the Cramer scheme in order to improve its applicability to the TTC approach.

Several statistical techniques and different molecular descriptors were explored.

The most interesting results were obtained with two different approaches:

1. Identification of **structural subclasses** within the Cramer classes (I and III)
2. Development of a **ranking classification model**

Development of a **ranking classification model**.

Approach:

Different ranking methods were used to sequentially order the Munro dataset structures by using the molecular descriptors mostly correlated to the toxicological end point (Log(1/NOEL)).

Descriptor-based rank

Chemical x15
Chemical x23
Chemical x12
Chemical x41
...
...



Experimental Log(1/NOEL) rank

High Log(1/NOEL)
...
...
...
...
Low Log(1/NOEL)

- Cramer classification**

Experimental Hazard classes	Cramer hazard classes			Total
	Class 1 (low hazard)	Class 2 (medium hazard)	Class 3 (high hazard)	
Class 1 (low hazard)	80	11	77	168
Class 2 (medium hazard)	37	16	177	227
Class 3 (high hazard)	10	3	179	192
Total	127	27	433	587

- Refinement: Cramer & Ranking model**

Experimental Hazard classes	Refinement hazard classes			Total
	Class 1 (low hazard)	Class 2 (medium hazard)	Class 3 (high hazard)	
Class 1 (low hazard)	47	19	102	168
Class 2 (medium hazard)	17	8	202	227
Class 3 (high hazard)	4	1	187	192
Total	68	28	491	587

Conclusions:

- As the Cramer scheme could be improved combining it with a statistically based classification model, it **can be concluded that the employed statistically based methods and molecular descriptors encode useful information not already included in the Cramer rules.**

1. TTC datasets analysis

- Investigation and characterisation of the chemical space of the TTC datasets.
- Comparison of the chemical space of the TTC datasets with a large reference dataset.

2. Investigation of possible improvements by means of *in silico* methods.

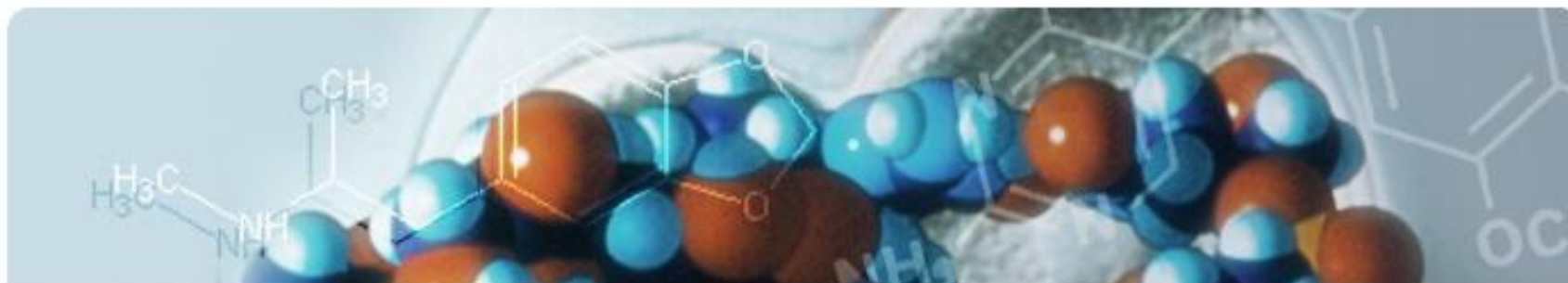
- Applicability of QSAR approaches
- Critical evaluation of the Cramer classification scheme
- Refinement of the Cramer classification scheme by :
 - Identification of structural subclasses
 - Development of a ranking classification model

3. Summary

1. From the analysis performed employing a wide variety of statistical methods and molecular descriptors it turned out that the Cramer scheme well fits the regulatory needs being highly conservative. **Additional analysis could be performed on a wider group of datasets** to evaluate its goodness in the TTC framework.
2. In addition the opportunities of developing QSARs for **individual clusters within the Munro** dataset is recommended. It is also suggested to focus on a **MOA-based approach** to grouping and QSAR development.

- **European Food Safety Authority** that has sponsored the work as part of the “Opinion on the applicability of the TTC in the different areas of food and feed risk assessment” to be published by the end of 2011.
- **Alan Boobis** ,**Sue Barlow** and **Andrew Worth** as part of the steering committee of the project.
- **Dr Chihae Yang**, who supported S-IN during the execution of the project in light of her experience in the field and of the similar work she carried out for FDA.





Services

- REACH
- Consulting activities

Training

- Courses
- Seminars

Products

- Software
- Applications

Scientific activity

- Publications
- Poster and presentations



www.s-in.it

manuela.pavan@s-in.it