Coupling carcinogenicity with mutagenicity models to improve its predictive power

Rodolfo Gonella Diaza[a], Chiara Porcelli[a], Alessandra Roncaglioni[a], Nadège Piclin[b], Marco Pintore[b], Jacques Chrètien[b], Emilio Benfenati[a]

[a] Istituto di Ricerche Famacologiche Mario Negri, Milan, Italy
[b] BioChemics Consulting SAS, Olivet, France

Aim & Scope

- Often mutagenicity and carcinogenicity endpoints are addressed together although this implies confounding factors in dealing with epigenetic carcinogens.
- The aim of this investigation was to elaborate an architecture for model combination to better address non-genotoxic carcinogens.

METHODS: TWO LAYER CLASSIFIER

Tox data: **DSSTox CPDBAS**[1] (Ames Mutagenicity data + Carcinoganicity TD50 on rat), merged with **Bursi dataset**[2] (Ames Mutaganicity); N_{tot} = 806
The dataset was split into training and test set on a chemical similarity basis.

Molecular descriptors: **DRAGON** (*v5.4*) 2D descr. (n = 790) **+ Bursi Fragments** (n = 27)

Modelling methods: Classification trees under **MATLAB** (v7.0), fuzzy classification approach with in house software (**Adaptive Fuzzy Partition**)

First step: classification model for mutagenicity. All ***mutagens*** are also classified ***as carcinogen.***

Second step: classification models based on carcinogenicity. In this step only non-mutagen compounds have been considered (nm/nc + ***epigenetics***). See Figure 1 for details.

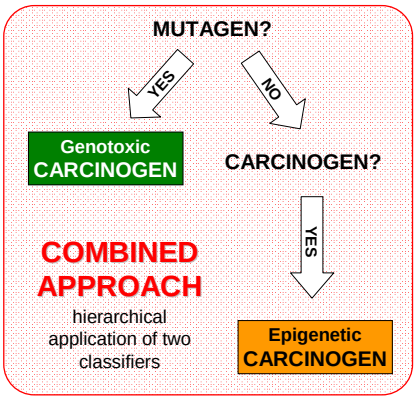

Fig. 1: *Representation of the two-classifier combined approach*

First Step: MUTAGENICITY CLASSIFIER

Bursi Fragments: performances for the assignment

M1	Acc.	Sens.	Spec.
Training set (437)	0.76	0.81	0.70
Test set (106)	0.71	0.73	0.68

Performances on mutagenicity

M2	Acc.	Sens.	Spec.
Training set (645)	0.64	0.64	0.63
Test set (161)	0.65	0.65	0.65

Performances on carcinogenicity

Classification Tree Results

M3	Acc.	Sens.	Spec.	Model info
Training set (437)	0.85	0.82	0.88	18 descr. 19 terminal nodes
Test set (106)	0.74	0.63	0.89	

Performances on mutagenicity

M4	Acc.	Sens.	Spec.	Model info
Training set (645)	0.64	0.58	0.70	18 descr. 19 terminal nodes
Test set (161)	0.58	0.50	0.68	

Performances on carcinogenicity

Second Step: CARCINOGENICITY CLASSIFIER

Classification Tree Results

M5	Acc.	Sens.	Spec.	Model info
Training set (208)	0.83	0.81	0.84	33 descr. 39 terminal nodes
Test set (44)	0.59	0.47	0.67	

AFP Results

M6	Acc.	Sens.	Spec.	Model info
Training set (208)	0.75	0.83	0.70	7 DRAGON descr. AFP parameters: 25 rules; p=1.05; q=0.75; occurrences=3; cuttings=8;
Test set (44)	0.75	0.59	0.85	

COMBINED MODEL

Performances of the model coupling M3 with M5
including externally predicted compounds

	Acc.	Sens.	Spec.
Training set (645)	0.57	0.77	0.36
Test set (161)	0.63	0.79	0.44

Conclusions & Future Perspectives

- Modelling firstly mutagenicity and then carcinogenicity allows to focus the attention to non genotoxic carcinogens but overall there is a lack in the overall accuracy.
- The most critical step is the second one (Carcinogenicity) and the hierarchical combination of models obtained so far caused a propagation of the errors.
- Further analysis will be conducted to enlarge and increase reliability of data available for non genotoxic carcinogens.
- A complementary study will be performed to explore the quantitative relationship accounting for carcinogenic potency of epigenetic carcinogens.

[1] DSSTox CPDBAS dataset: http://www.epa.gov/ncct/dsstox/sdf_cpdbas.html
[2] Kazius J, McGuire R, Bursi R., 2005, Derivation and validation of toxicophores for mutagenicity prediction., *J Med Chem* 48(1), 312-20

**Authors acknowledge financial support from EC for the CAESAR project.
Contract nr. SSPI-022674.**