WORKSHOP ON QSAR MODELS FOR REACH

Mario Negri Institute, Milan, Italy - March 10-11, 2009

Qasim Chaudhry

Central Science Laboratory, UK

The CAESAR Model for Skin Sensitization

http://www.caesar-project.eu/

Skin Sensitization

- Skin sensitization is the term used to refer to a human risk called Allergic Contact Dermatitis (ACD) that can be caused by skin contact with a wide range of chemicals.
- Experimental tests (*in vivo*):
  - OECD 429 The Local Lymph Node Assay (LLNA)
  - OECD 406 The Magnusson Kligman Guinea Pig Maximisation Test (GPMT)
  - OECD 406 Buehler test
- Cost in the range of 30,000 euros/ compound

Skin Sensitization: Dataset

- Extracted from Gerberick et al. (2005)
- Tests carried out according to official guidelines (LLNA assay)
- 211 compounds with EC3 values and activity classes (binary and 5 classes)
- Data toxicity and structures quality check remaining 209 compounds

  - **Checking Names, structures, CAS etc** by online databases: ChemFinder (http://chemfinder.cambridgesoft.com), ChemIDPlus (http://chem.sis.nlm.nih.gov/chemidplus/);

  - **Searching duplicate chemicals and isomers**;

  - **Removing ions and neutralizing molecules;**

  - **Cross-checking** by at least 2 different partners.

CAESAR Modeling for Skin Sensitization

- Descriptors
  - 2D desc.: DRAGON, CODESSA, ACD, PALLAS, MDL
- Individual classification models
  - AFP (Adaptive Fuzzy Partition)
  - MLP (MultiLayer Perceptron)
  - GMDH (Self-organising networks of active Neurons based on the Group Method of data Handling)
- Combined classification models (GMDH)
- Mechanisms of action (read across approach)

EC3 (%)	LLNA Class	Binary class	Total compounds
NC	NC	Non sensitizers	42
≥10	Weak		66
1-10	Moderate	Sensitizers	68
0.1-1	Strong		21
< 0.1	Extreme		12
			209

Classification Ranges

EC3 (%)	Official LLNA Class			LLNA Binary class	
NC	Class1	NC	42	Class1 Non sensitizers	108
≥ 10	Class2	Weak	66		
≥ 1	Class3	Moderate	68	Class2 Sensitizers	101
≥ 0.1	Class4	Strong	21		
< 0.1	Class5	Extreme	12		

ECETOC

Weak sensitizers in the non sensitizers class

EC3 (%)	Official LLNA Class			Binary class	
NC	Class1	NC	42	Class1 Non sensitizers	42
≥ 10	Class2	Weak	66	Class2 Sensitizers	167
≥ 1	Class3	Moderate	68		
≥ 0.1	Class4	Strong	21		
< 0.1	Class5	Extreme	12		

CAESAR ranges

Classification Models – ECETOC Ranges

ECETOC ranges: **Non Sensitizers (NC/weak)**
Sensitizers (moderate/strong/extreme)

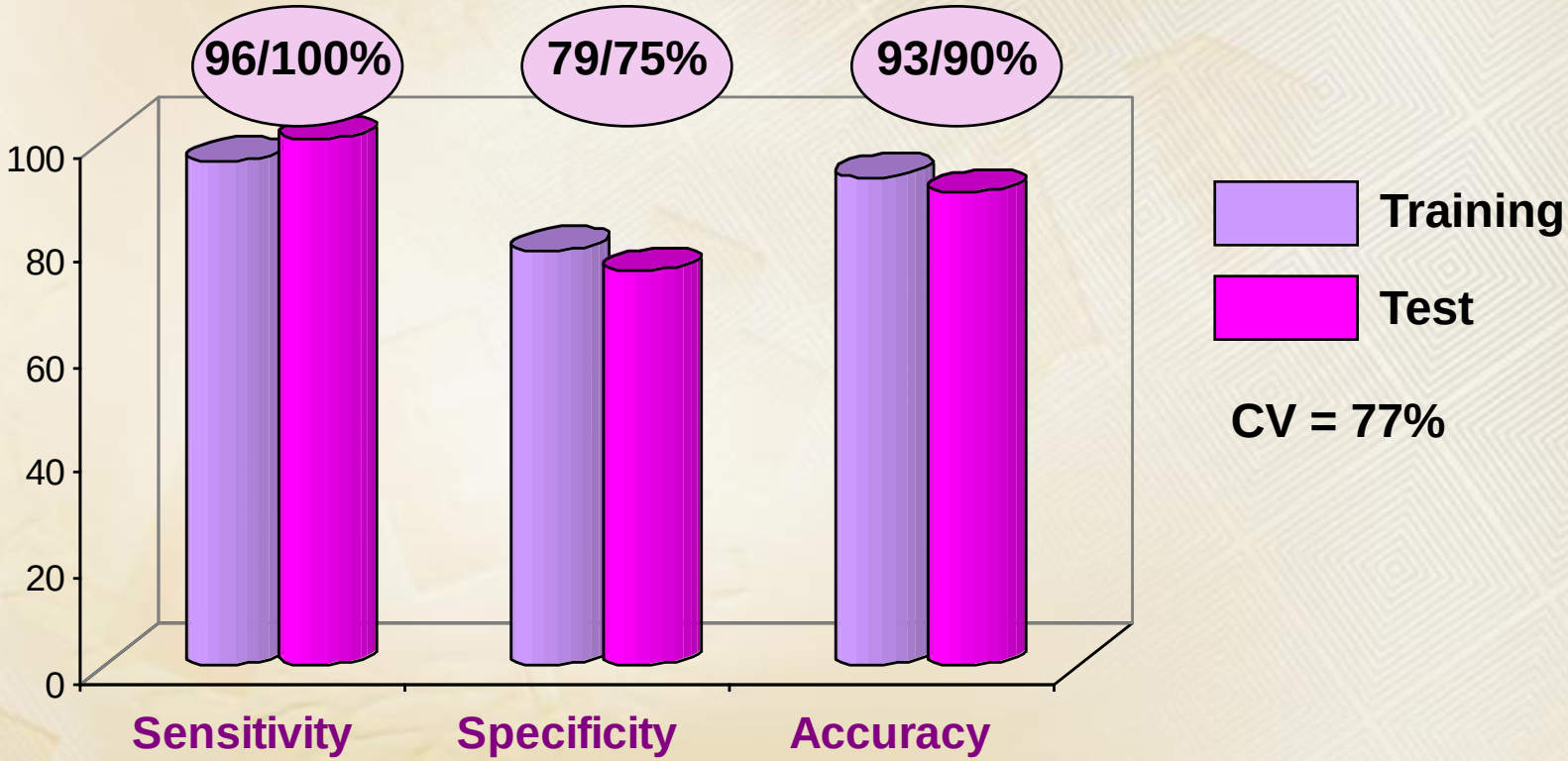
		Training set			Test set		
Modelling method	**Nb of des.**	**Acc.**	**Sen.**	**Spe.**	**Acc.**	**Sen.**	**Spe.**
MLP-NN	7	84	80	87	71	76	67
Combined model (7 models)	**74**	**91**	**100**	**83**	**83**	**100**	**67**

Satisfactory results with MLP method: accuracy (test)= 71%
Combined Model improves performances: sensitivity= 100%

Classification Models – CAESAR Ranges

CAESAR ranges: Non Sensitizers (NC)
Sensitizers (weak/moderate/strong/extreme)

AFP model by using 8 DRA descriptors



CV = 77%

Better results by using CAESAR ranges

The sensitizers compounds are the best predicted ones

AFP Model - Descriptors

The AFP model uses 8 DRAGON descriptors:
(nN; GNar; MDDD; X2v; EEig10r; GGI8; nCconj; O-058)

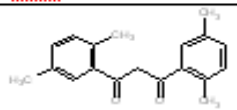
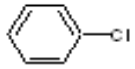
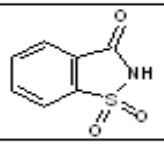
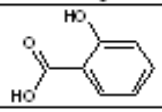
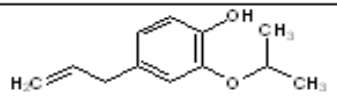
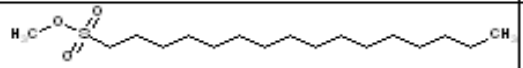
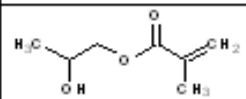
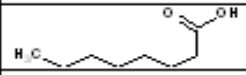
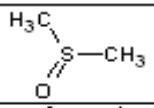
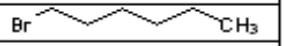
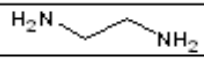
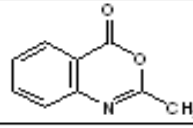
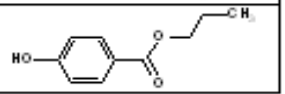
Name_descr	Definition
nN	Number of Nitrogen atoms
GNar	Narumi geometric topological index
MDDD	Mean distance degree deviation
X2v	valence connectivity index chi-2
EEig10r	Eigenvalue 10 from edge adj. matrix weighted by resonance integrals
GGI8	topological charge index of order 8
nCconj	number of non-aromatic conjugated C(sp2)
O-058	=O (atom-centred fragments)

AFP Model Performance Evaluation

Validation statistics derived from the AFP model

	Training	Test
Accuracy	**93**	**90**
Cross-validation	77	
Nb unpredicted compounds	0	2
Outliers False Positive	**7 outliers** 21 ; 54 ; 105 ; 188 ; 189 ;123 ; 145	**2 outliers** 112 ; 167
False Positive Rate	21	25
Outliers False Negative	**5 outliers** 84 ; 31 ; 90 ; 136 ; 185	NO
False Negative Rate	4	0
Postive Predictive Value (precision)	95	94
Negative Predictive Value	84	100
Sensitivity (class S)	96	100
Specificity (class NC)	79	75
F-measure	96	97

Outlier Compounds

ID	FP train	FP test	FN train
21			
54			
105			
188			
189			
123			
145			
112			
167			
84			
31			
90			
136			
185			

Conclusions

- New integrated models for skin sensitization have been developed.
- The models have been statistically evaluated using strict criteria.
- The final model will be implemented in the CAESAR applet.
- Focus on REACH:
  - Experimental data according to guidelines
  - Quality check (chemical structures)
  - Reproducibility
  - Transparency
  - False negatives minimized