Environmental ChemOinformatics (ECO)
Marie Curie Initial Training Network

FINAL REPORT

Short-term project:

Ecotoxicological Property Evaluation by Read-across Methodologies

Actual Submission Date	August 15, 2012
Project Dates	Start Date: September 1, 2011 End Date: July 31, 2012 Project Duration: 11 months

HOSTING INSTITUTION:
UNIVERSITY OF MILANO BICOCCA

SUPERVISOR:
PROF. ROBERTO TODESCHINI

SUBMITTED BY:
EVANTHIA GIAGLOGLOU

1. Introduction

Substances with ecotoxicological properties can cause damage with possible irreversible consequences to ecosystems. For this reason, the European REACH regulation requires that all substances produced or imported in more than one ton per year (REACH, annex VII) need information regarding aquatic ecotoxicological data. The regulation recommends ecotoxicity tests, as the short term acute toxicity test to *Daphnia* magna. In particular the acute immobilization test. EC_{50} 48 hours, which consists in the determination of the effective concentration of a chemical that immobilises the 50% of the daphnids in 48 hours, is required. The principles of this test are of the same as test No 202 of OECD guideline document.

Daphnids are as the preferred test animal for ecotoxicological studies for two reasons: 1) they are aquatic organisms which are very important for the stability of the ecosystem because they are intermediate consumers and 2) they are small and can be reared in a small space giving birth to young daphnids, genetically identical to the mother, (parthenogenesis) within their first week.
In the last years ecotoxicological tests have been commonly conducted with Daphtoxkit, which contains Ephippia (eggs of daphnids). Ephippia can be hatched on demand. This facilitates the test because cost and time can be saved by culturing and maintaining the test organisms. REACH regulation and OECD guidance do not specify which test should be undertaken for ecotoxicological evaluation. Nevertheless, there is evidence of different sensitivity between tests on cultured organisms and kit organisms *(Persoone et al. 2000).*

REACH regulation (Annex XI) foresees the use of alternative *in silico* methods, such as Quantitative Structure Activity Relationships (QSAR), in order to avoid experiments on animals. In the latest years, QSAR studies have been demonstrated to be reliable on predicting toxicity and they are recommended as an alternative to animal testing (OECD 1994, US EPA 1994, ECETOX 2003).
During the last decades a number of studies have been performed on the relationships between toxicity and chemical structure. The choice of molecular descriptors to be used is one of the most crucial parts of QSAR modelling. The selected descriptors must encode the structural features responsible for the toxicological activity of the molecule. Key feature for the aquatic toxicity is idrophobicity, which can be quantified by the measure of octanol/water partition (logP) of a substance. The first developed QSARs for predicting acute aquatic toxicity were based on logP. Könemann introduced the general narcosis *(Könemann, 1981)*, which is a toxicity mechanism based on logP. There is evidence for two different mechanisms of narcosis: *polar* and *general*, which are distinguished by different QSAR models based on logP. Today many software and studies still use just logP for development of toxicological models; an example is ECOSAR, which is among the most used QSAR programs *(Reuschenbach et al.,2008).* Reuschenbach demonstrated that ECOSAR gives pour correlation while using a diverse dataset, but gives good correlation only for some chemical classes.

Several QSAR models are based on the narcotic level and determine the toxicity on *Daphnia Magna* defining the chemical mode of action. These approaches are based on the fact that the mode of action is associated to specific chemical structures *(von der Ohe et. al., 2004).* The study conducted by von der Ohe was a two step classification approach: initially a definition of the chemicals as narcotics or excessively narcotics was carried out and then a classification scheme based on the mode of action was applied. Results demonstrated that there is good predictivity on categorising the toxicants in narcotic effect level or excess toxicity level toxicants, with best model predictivity of 100% for categorising industrial toxicants in excess toxicity level category, even if this could be detected only for specific chemical classes.

Read-across is a QSAR strategy that estimates unknown toxicity of a query substance by means of known toxicity values of some of structurally similar analogues. This approach involves identification of structural features or molecular properties in order to define similarity towards a number of substances (analogues). The read-across methodology can be then summarised as the investigation of similarity among the chemical substances. Therefore, one has to define the kind of similarity index and the mathematic algorithm for measuring this similarity. The methodology can be approached either qualitatively or quantitatively. In quantitative read-across, the known value of a property for one or more source chemicals is used to estimate by averaging or regression models the unknown value of the same property for the target chemical. In qualitative read-across, the purpose is a qualitative characterization of the query molecule, i.e. its classification in a toxicity class or category *(JRC, Chemical Categories and read across, EUr 21898)*. In the present study, both qualitative read-across for classifying query chemicals into toxicity classes and quantitative read-across in order to estimate the unknown toxic endpoint based on experimental values of LC_{50} 48 hours to *Daphnia* magna were carried out.

2.0 Materials and methods

2.1 Collection and screening of experimental data

The experimental toxicity test consists in exposing young daphnids for 48 hours to varying concentrations of the substance under testing for the estimation of the LC50. LC50 is the lethal concentration to 50% of the population in 48 hours. Although the regulation demands EC50 48 hours to *Daphnia magna* some studies use EC50 and LC50 as identical endpoints (*von der Ohe et al., 2005*). For the present project it was preferred LC50. Mortality is a more defined endpoint than immobilization because some immobilized organisms may recover after the exposure of the toxic substance and others may not.

The experimental values of aquatic acute toxicity were collected by screening available databases, namely, US-EPA, ECOTOX, OCHEM, and scientific literature. In the first part of the project, a preliminary screening on these data was carried out. REACH and OECD guidelines recommend experimental values from laboratories with GLP (Good Laboratory Practice). For this reason data were evaluated for their test conditions. Only data presented with their source article and only tests in accordance with the OECD principles were selected (Table I, Appendix). Once the data had been selected and evaluated, they were classified under the GHS (Global Harmonised System). GHS is a system for standardizing and harmonizing the classification of chemicals. The European Union has introduced the CLP (Classification, Labelling and Packaging) regulation 1272/2008, which entered into force in June 2009, in order to be compliant with GHS criteria for classification and packaging. The CLP is not aligned completely with the GHS but the purpose is to be harmonised step by step. The aquatic acute toxicity classes defined by GHS are: acute category I (48h LC50 $\leq$ 1mgL^{-1}), acute category II (1 mgL^{-1} < 48h LC50 $\leq$10 mgL^{-1}) and acute category III (10 mgL^{-1} < 48h LC50 $\leq$100 mgL^{-1}). In order to classify the chemicals into toxicity classes, in the present study we adopted the GHS criteria for defining three toxicity classes and an additional class was considered for the classification of the chemicals that are not toxic towards *Daphnia magna*.

2.2 Data pretreatment

A total of 511 substances with their toxicity values were selected and classified into toxicity categories according to their experimental values of 48h LC50. Some of these substances had more than one experimental value of lethal concentration. It is a common problem in toxicity studies that replicate values, i.e. toxicity values assessed on the same molecule with the same test conditions, can vary a lot. In order to harmonise the data to be modelled in the present project, the median of replicates was considered as the value to be used for building the QSAR models. Data were transformed into molarity units since molarity is preferred in toxicological studies than other units of concentration, the activity of a molecule being related to the number of moles and not to its weight.

Experimental toxicological values of the data set in analysis were characterised by a significant variance since a large set of diverse chemicals was considered. For this reason a logarithmic transformation was applied to the experimental values before modelling. To implement the read-across methodology the dataset was randomly divided into a training set, with 358 compounds, and a test set, with 153 compounds (Tables 1 and 2, respectively)

Table 1. Classification into defined toxicity classes for the training set

Class	Toxicity range	No. of chemicals
Category **I** (Very toxic)	$LC_{50} \leq 1 mgL^{-1}$	104
Category **II** (Toxic)	$1 < LC_{50} \leq 10\ mgL^{-1}$	77
Category **III** (Harmful)	$10 < LC_{50} \leq 100\ mgL^{-1}$	91
Category **IV** (Not harmful)	$LC_{50} > 100\ mgL^{-1}$	86

Table 2. Classification into defined toxicity classes for the test set

Class	Toxicity range	No. of chemicals
Category **I** (Very toxic)	$LC_{50} \leq 1 mgL^{-1}$	35
Category **II** (Toxic)	$1 < LC_{50} \leq 10\ mgL^{-1}$	41
Category **III** (Harmful)	$10 < LC_{50} \leq 100\ mgL^{-1}$	44
Category **IV** (Not harmful)	$LC_{50} > 100\ mgL^{-1}$	33

2.3 DRAGON molecular descriptors and their selection

"The molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment" (Todeschini and Consonni, 2009). Two dimensional molecular descriptors were calculated by means of DRAGON software (Talete srl, version 6.0-2012, http://www.talete.mi.it/). Ten blocks of 2D descriptors were initially calculated: Constitutional indices, Ring descriptors, Topological indices, Connectivity indices, Information indices, Burden eigenvalues, CATS 2D, Atom-centred fragments, Molecular properties and Drug-like indices. The total number of the descriptors obtained was 537. Descriptors with missing and constant values were discarded. Since reliable QSAR models are usually based on few descriptors and EU and REACH regulators require simple models to be used to asses toxicity, a variable selection strategy was applied in order to find the best optimal set of descriptors for modelling aquatic toxicity. Genetic Algorithms (GAs) are an evolutionary computation technique based on the principles of genetics and natural selection. In the last decades, they have been applied for variable selection purposes. Genetic Algorithms were used in order to remove irrelevant, noisy, redundant variables or variables not needed for modelling *(Leardi et. González, 1998)*. Calculations of GAs were performed in MATLAB 6.5 (Mathworks) by means of routines built by Milano Chemometrics and QSAR Research Group.

2.4 Binary descriptors

Structural keys and fingerprints are binary descriptors that indicate the presence or the absence of a particular structure in the molecule. The structural keys are represented as boolean arrays in which the presence or absence of a particular predefined chemical structure is indicated by 1 or 0. Unlike the structural keys with their pre-defined patterns, the patterns for a molecule's fingerprint are generated from the molecule itself. A molecule can generate a huge number of patterns based on the atoms (a pattern for each atom), type of the bonds (a pattern for each bond) in a way that every pattern of the molecule is generated. The final number of the patterns produced in this way can be too big so the software uses only the patterns presented in all the molecules of the data set. (Daylight, Chemical Information Systems Inc, www.daylight.com). The output of fingerprints calculation is a matrix containing bits of 0 and 1 values, but there is no assigned meaning to each bit. Two blocks of fingerprints (fingeprints and extended fingerprints), and three blocks of structural keys (PUBCHEM, MACCS and Substructural) were used. These descriptors were calculated by means of PaDel software (National University of Singapore http://padel.nus.edu.sg/software/padeldescriptor/).

2.5 Read-across

The kNN (k Nearest Neighbour) method was used for the development of read-across models, both for regression and classification purposes.

The k-Nearest Neighbour (kNN) method *(Cover et al., 1967)* is conceptually quite simple: an object is classified according to the classes of the k closest objects, i.e. it is classified according to the majority of its k-nearest neighbours in the data space. In case of ties, the closer neighbours can acquire a greater weight. From a computational point of view, it is necessary to calculate and analyse a distance matrix between all of the pairs of objects. The distance of each object from all the other objects is computed, and the objects are then sorted according to this distance. Hence, kNN develops models based on local information, since only the nearest samples are used to assign an untested sample to a predefined class.

kNN can be also used to estimate quantitative responses: the response value was calculated as the average of the experimental responses of the k neighbours. This method was used for assigning LC_{50} value to query compounds based both on binary and global descriptors. kNN estimates were derived as the weighted averages on the basis of the distances of the nearest neighbours. In this way, the nearest neighbour had a greater weight in defining the predicted response, the second nearest neighbour a smaller weight and so on.

While dealing with binary data, the similarity between two objects is evaluated from a pairwise comparison among the bits of the molecules. In Table 3, the similarity indices used in this study are listed. Then, similarity was transformed into a distance measure to implement kNN. While dealing with real data as in the case of global molecular descriptors, kNN was implemented on the selected distance measures shown in table 4.

Two validation methods were used for estimating the predicting ability of all the classification and regression models: cross-validation with Venetian blocks and external validation. The Venetian blocks method divides the training set in groups (in the present project the training set was divided in 10 groups), each group is then removed and used as the test set just once.

Table 3. Similarity indices for binary data

Similarity Index	Abbreviation	Mathematical type
Consonni-Todeschini	CT4	$S = \frac{log2(1+a)}{log2(1+a+b+c)}$
Gleason-Dice	GLe	$S = \frac{2a}{2a+b+c}$
Jaccard-Tanimoto	JT	$S = \frac{a}{a+b+c}$
Sokal-Michener	SM	$S = \frac{(a+d)}{p}$
Austin-Colwell	AC	$S = \frac{2}{\pi} sin\sqrt{\frac{a+d}{p}}$

a: number of common 1 bits between two molecules
b: number of 1 bits in molecule A and 0 bits in molecule B
c: number of 0 bits in molecule A and 1n bits in molecule B
d: number of common 0 bits between two molecules
p: total number of bits

Table 4. Distance measures for real data

Distance measure	Mathematical Type
Euclidean	$d_{rs}^2 = (\boldsymbol{x}_r - \boldsymbol{x}_s)(\boldsymbol{x}_r - \boldsymbol{x}_s)'$
Cityblock	$d_{rs} = \sum_{j=1}^{p} \lvert \boldsymbol{x_{rj}} - \boldsymbol{x_{sj}} \rvert$
Mahalanobis	$d_{rs}^2 = (\boldsymbol{x}_r - \boldsymbol{x}_s)\boldsymbol{V}^{-1}(\boldsymbol{x}_r - \boldsymbol{x}_s)'$
Minkowski	$d_{rs} = \left\{ \sum_{j=1}^{p} \lvert \boldsymbol{x_{rj}} - \boldsymbol{x_{sj}} \rvert^{q} \right\}^{\frac{1}{q}}$

d_{rs}: distance between samples r and s, defined by p-dimensional vectors $\mathbf{x}_r$ and $\mathbf{x}_s$
V: the sample covariance matrix
q: a user defined parameter which can take integer positive values

2.6 Global QSAR models

As stated before, kNN uses only local information in order to predict new molecules. A global QSAR model, i.e. a regression model derived from the whole training set, was built for the sake of comparison. Global models might be less sensitive than local models to minor features which are relevant to a small group of molecules. It is interesting to see how a global model performs with this dataset which contains diverse chemical structures.

The global model was calculated by PLS (Partial Least Squares) regression method and the variable selection was undertaken by Genetic Algorithms *(Leardi et González, 1998).* One of the methods for improving the performance of global QSARs is the elimination of potential outliers. Outliers can be defined as molecules with a significantly different chemical structure, when compared to the whole training set of molecules. Many times outliers indicate a systematic error or an error of calculation and in this case they must be eliminated. Usually, more chemical structures a data set encompasses more likely to contain extreme samples. For the present study a PCA model based on

the 537 calculated molecular descriptors (Figure 1) was carried out for the screening and possible detecting of the potential outliers. The six compounds eliminated are highlighted by ellipse: cyclosporine, tylosin, erythromycin, hexabutyldistannoxane, digoxin and digitoxin. Then, descriptor selection based on Genetic Algorithms coupled with PLS regression was performed on the reduced set of molecules.

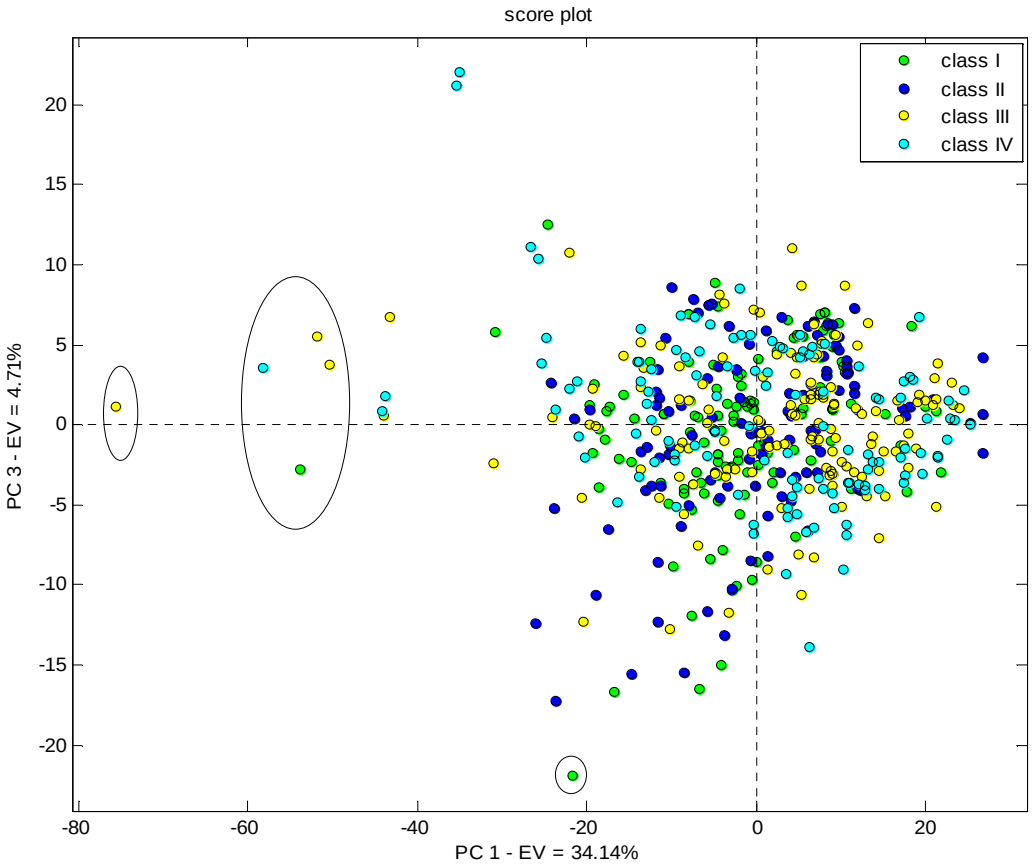

Figure 1. PCA based on 537 variables.

3.0 Results

3.1. QSAR model based on Octanol/Water partition coefficient

A lot of studies on QSAR use the octanol/water partition coefficient as unique variable for aquatic toxicity models. This approach is based on the concept that if other toxic mechanisms are absent then the compounds are at least as toxic as determined by their bioconcentration. This concept opened the way for predicting the baseline toxicity of aquatic pollutants through the QSAR equations. In the present study a linear regression QSAR model based on logP and the log(LC_{50}) to *Daphnia magna* was performed.

The regression model can be judged for its goodness of fit with the correlation coefficient R^2. The R^2 measures in fact how well the mathematical model approximates the real toxicity values. It can have values from 0 to 1, where R^2 =1 indicates that the regression model fits perfectly the experimental values of toxicity. For evaluating the capability of a model to predict unknown data (external validation), i.e. data which have not been used for building the model, the Q^2 coefficient is used. High value of Q^2 indicates good predictivity. The distribution of the toxicity values into training and test sets and their correlation with logP, is shown in Figure 2.

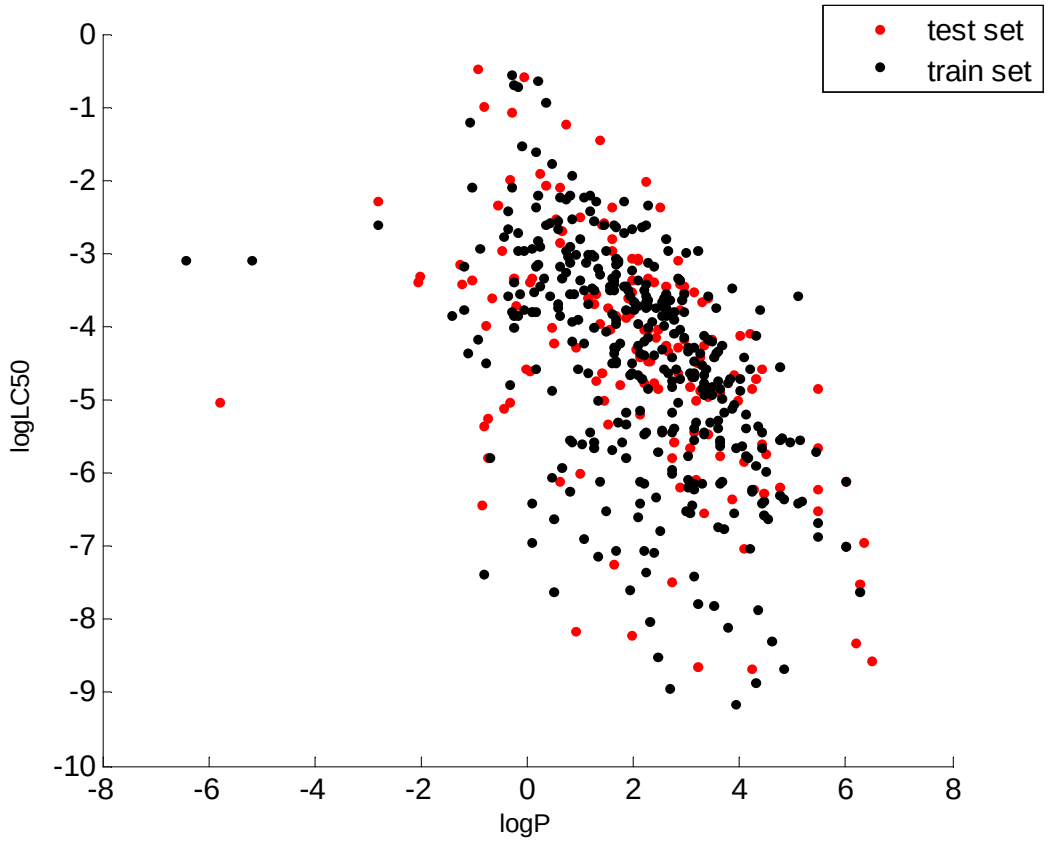

Figure 2: Distribution of the training and test set based on octanol/water partition and molarity of median LC_{50} 48h to *Daphnia magna*

The regression equation and the regression statistics are shown in the Table 5.

Table 5. Linear regression model based on logP

log(LC_{50}) = -3.442 -0.493logP					
R^2	RMSEC	Q^2 cv	RMSEC cv	Q^2 ex	RMSEP ex
0.278	1.320	0.274	1.324	0.208	1.382

R^2: 1-$\frac{RSS}{TSS}$, where RSS is the residual sum of squares, TSS is total sum of squares

Q^2: 1-$\frac{PRESS}{TSS}$, where PRESS is the residual sum of squares in prediction

RMSEC: $\sqrt{RSS/n}$

RMSEP: $\sqrt{PRESS/n}$

3.2 Qualitative read-across

The results for classification (Table 6) show the best models obtained after applying the similarity indexes of the table 4. The number of the nearest neighbours k was calculated by means of cross validation for each similarity index and each block of binary descriptors.

Table 6. Classification models based on binary descriptors

Model	Descriptors	Bit	k	Similarity Index	NER	Sn Class I	Sn Class II	Sn Class III	Sn Class IV
Fitting	Fp1_2	2048	4	AC	0.623	0.500	0.355	0.435	0.442
CV	Fp1_2	2048	4	AC	0.623	0.490	0.276	0.391	0.407
Prediction	Fp1_2	2048	4	AC	0.636	0.629	0.268	0.477	0.455
Fitting	Fp1_2	2048	6	CT4	0.634	0.654	0.237	0.424	0.477
CV	Fp1_2	2048	6	CT4	0.64	0.683	0.197	0.478	0.465
Pred	Fp1_2	2048	6	CT4	0.702	0.743	0.268	0.636	0.576
Fitting	Fp1_2	1024	4	GLe	0.634	0.625	0.263	0.402	0.500
CV	Fp1_2	1024	4	GLe	0.623	0.635	0.184	0.391	0.512
Prediction	Fp1_2	1024	4	GLe	0.679	0.743	0.268	0.591	0.485
Fitting	MACCS	166	1	CT4	0.678	0.664	0.329	0.587	0.477
CV	MACCS	166	1	CT4	0.67	0.692	0.276	0.533	0.500
Prediction	MACCS	166	1	CT4	0.65	0.800	0.220	0.500	0.394
Fitting	PUBCHEM	881	3	SM	0.625	0.664	0.211	0.391	0.465
CV	PUBCHEM	881	3	SM	0.613	0.664	0.211	0.348	0.442
Prediction	PUBCHEM	881	3	SM	0.643	0.714	0.293	0.409	0.455
Fitting	Substructural	307	6	JT	0.623	0.606	0.184	0.457	0.477
CV	Substructural	307	6	JT	0.61	0.625	0.105	0.413	0.500
Prediction	Substructural	307	6	JT	0.626	0.686	0.073	0.523	0.485

Fp1_2: Fingerprints and extended fingerprints

NER: Non error rate, NER=$\frac{\sum_g C_{gg}}{n}$, c_{gg}: number of correctly assigned objects.

Sn: Sensitivity, Sn= $\frac{C_{gg}}{n_g}$, n_g: total number of the objects of the class *g*.

3.3 Quantitative read-across

kNN quantitative models based on binary descriptors are shown in Table 7. The best model was calculated with fingerprints and extended fingerprints, for k=4 and applying the similarity index of Gleason-Dice.

Table 7. kNN quantitative models based on binary descriptors

Descriptors	Bit	k	Distance	R^2	RMSEC	Q^2cv	RMSECcv	Q^2	RMSEP
PUBCHEM	881	4	JT	0.502	1.096	0.485	1.115	0.389	1.214
PUBCHEM	881	3	SM	0.443	1.160	0.449	1.153	0.331	1.270
Fp	1024	4	Gle	0.501	1.097	0.479	1.121	0.576	1.011
Fp	1024	4	JT	0.500	1.099	0.476	1.125	0.577	1.010
Fp1_2	2048	6	CT4	0.512	1.085	0.504	1.094	0.548	1.045
Fp1_2	2048	4	Gle	0.525	1.070	0.507	1.091	0.581	1.006
MACCS	166	4	JT	0.457	1.145	0.446	1.156	0.440	1.163
Substructural	307	4	AC	0.350	1.253	0.397	1.207	0.292	1.307
MACCS	166	4	CT4	0.448	1.154	0.429	1.174	0.420	1.183

The results for kNN based on DRAGON descriptors are shown in Table 8. The best model was obtained by using the City Block distance and k=8.

Table 8. kNN quantitative models based on DRAGON descriptors

Variable Selection	Descriptors DRAGON	K	Distance	R^2	RMSEC	Q^2cv	RMSECcv	Q^2	RMSEP
GA	25	6	Euclidean	0.562	1.028	0.584	1.002	0.566	1.023
GA	25	6	Minkowski	0.562	1.028	0.584	1.002	0.566	1.023
GA	25	5	Mahalanobis	0.504	1.095	0.536	1.059	0.409	1.195
GA	25	8	City Block	0.586	1.000	0.590	0.995	0.623	0.953

3.4 Global QSAR model

The results of GA-PLS regression are shown in Figure 3 and Table 9. The 17 selected variables are listed in Table 10.

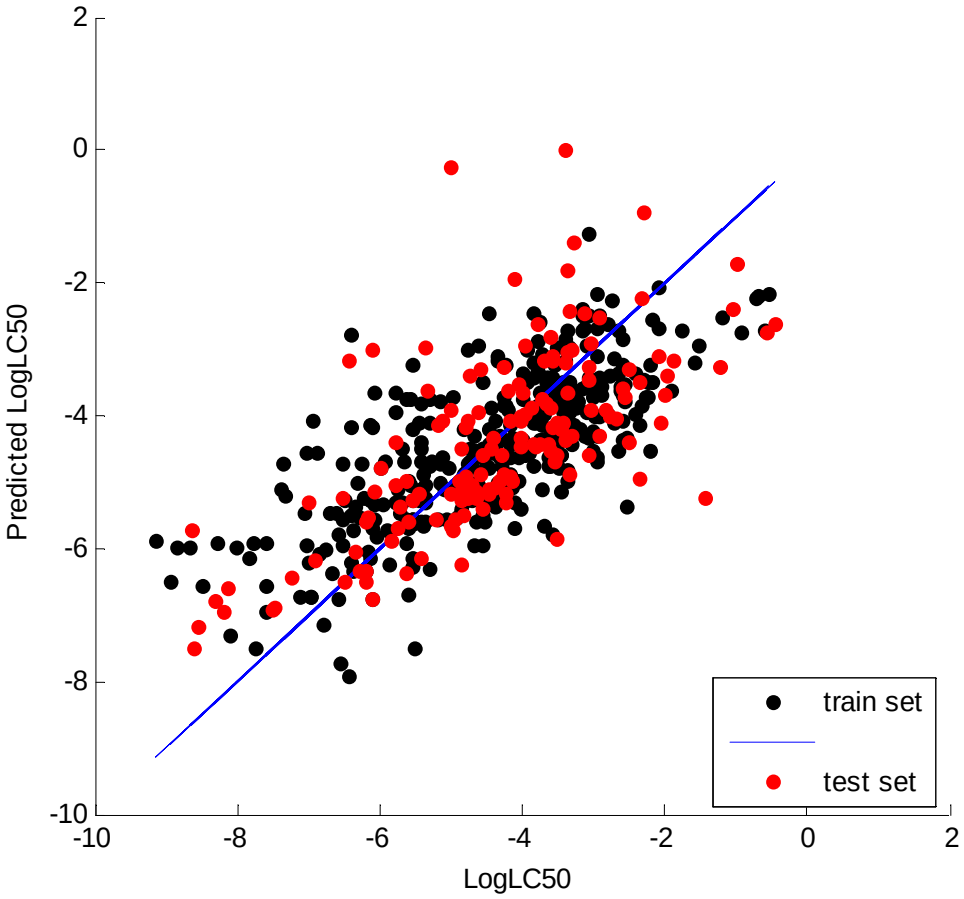

Figure 3. **PLS_GA** model built with 17 **DRAGON** descriptors

Table 9. PLS-GA model .

No of Descriptors	Variable Selection	R^2	RMSEC	Q^2 cv	RMSEC cv	Q^2 ex	RMSEP ex
20	GA	0.530	1.065	0.489	1.110	0.412	1.191

Table 10. Variables selected by PLS-GA

Descriptor	Regression Coefficient	Description
CATS2D_04_NL	0.1071	CATS2D are pharmacophore based descriptors, non lipophilic at lag 04
nP	-0.2317	Constitutional Indices, Number of phosphorous atoms
P-117	0.1611	Atom Centred Fragments, X3-P=X phosphate
C-041	-0.2002	Atom Centred Fragments, X-C(=X)-X
IC0	0.0724	Information Indices, Information Content index (neighborhood symmetry of 0-order)
ICR	0.0534	Topological Indices, radial centric information index
nS	-0.0987	Constitutional Indices, number of Sulfur atoms
C-044	-0.0940	Atom-Centred Fragments, X--CX..X
CATS2D_04_DA	-0.0609	CATS2D, CATS2D Donor-Acceptor at lag 04
C-006	0.2409	Atom Centred Fragments, CH2RX
C-021	-0.0842	Atom Centred Fragments, #CH
nHM	-0.0613	Constitutional Indices, number of heavy atoms
MLOGP2	-0.1294	Molecular Properties, logP2 Moriguchi octanol-water partition coefficient
DLS_02	-0.0740	Drug-like Indices
CATS2D_02_PL	0.0549	CATS2D Positive-Lipophilic at lag 02
nCsp	0.1204	Constitutional Indices, number of sp hybridized Carbon atoms
Mi	0.0339	Constitutional Indices, mean first ionization potential (scaled on Carbon atom)

The results of the obtained model without outliers are shown in Table 11. The seven selected descriptors are listed in Table 12.

Table 11. PLS-GA results on the reduced set of molecules

No of Descriptors	Variable Selection	R^2	RMSEC	Q^2 cv	RMSEC cv	Q^2 ex	RMSEP ex
9	GA	0.517	1.070	0.481	1.109	0.245	1.337

Table 12. Variables selected by PLS-GA on the reduced set of molecules

Descriptor	Regression Coefficient	Description
nS	-0.2059	Constitutional Indices, number of Sulfur atoms
nP	-0.2153	Number of phosphorous atoms
CIC3	0.1838	Information Indices, Complementary Information Content index (neighborhood symmetry of 3-order)
MLOGP2	-0.5344	Molecular Properties, logP2 Moriguchi octanol-water partition coefficient
nHM	-0.1411	Constitutional Indices, number of heavy atoms
nTB	-0.0962	Constitutional indices, number of triple bonds
O-060	-0.2091	Atom Centred Fragments, Al-O-Ar / Ar-O-Ar / R..O..R / R-O-C=X

3.5 Discussion

In the present project classification of chemicals into predefined GHS toxicity classes with the kNN method was evaluated. The best model was obtained with MACCS structural keys and the similarity index Consonni-Todeschini CT4 with k=1, which gave a percentage of correct predictions equal to 65% (Table 6). Models had even more predictive results for class I (very toxic) and class IV (not harmful). This was probably due to the fact that these classes are at the edge of the scale and miss-predictions are only possible to one side *(Reuschenbach et al.,* 2008*)*. Performance of this classification model is very satisfactory if compared with the other published ECOSAR models. The QSAR programme ECOSAR is a freely available software developed by US EPA. It applies SARs that are based on linear regression between logarithmic octanol-water partition and aquatic toxicity for more than 50 different chemical classes. The ECOSAR first assigns a chemical into a chemical class based on its SMILES notation and after finds its corresponding SAR. The study of Reuschenbach proves a percentage of correct predictions for classification into the four GHS aquatic toxic classes equal to 51.5% with ECOSAR. The good results obtained in this project by using kNN classification can be probably explained by the methodology itself. kNN does not assign chemicals into classes a priori but on the basis of their structural similarity. Some chemicals could be assigned to more than one chemical class, for example an organic aromatic acid could be assigned as acid or aromatic structure, and probably this could be a problematic case with the ECOSAR approach. On the other hand, kNN assigns the molecule to its most similar compounds without complication based on the chemical domain. In addition, predictions achieved by means of kNN are not based only on the octanol-water partition coefficient. The qualitative read-across was reviled as a very flexible technique since this methodology relies on the choices of the user for selecting appropriate variables and similarity measures. Although in the present project only binary descriptors waere used for classification purposes, there is evidence for good classification models based on global molecular descriptors since they gave slightly better results than the binary descriptors in quantitative kNN .

The quantitative read-across based on binary descriptors did not give good predictive results. The best model was obtained with fingerprints and extended fingerprints by using the *Gleason-Dice* similarity index (Table 7). It seems understandable that classification can give better predictive results than regression. The range of toxicity for assigning a chemical into a toxicity class is wider than predicting a unique toxicity endpoint.

The kNN models with DRAGON molecular descriptors gave better results than the kNN models based on binary variables but still the models were not characterised by good predictive performance. The best model obtained was the one based on City Block distance with 25 molecular descriptors (Table 8). Almost all the models based on DRAGON molecular descriptors are better, in terms of fitting and prediction ability, than those based on the binary descriptors. One possible explanation is that DRAGON descriptors encompass structures and molecular proprieties so they can encode more chemical information than fingerprints and structural keys. Nevertheless, the predictive capability of the kNN models based on DRAGON descriptors was not acceptable.

Global regression models were also developed for purposes of comparison. They were calculated by PLS-GA regression applied to DARGON molecular descriptors. The results are shown in Table 10. The predictive ability of these global models is not satisfactory but it is comparable with performance of the local models. A global QSAR was built also without potential outliers (Table 11). The existence of outliers is an open question in the field of statistics, their detection could be problematic and subjective since it is not so easy to define a compound as outlier. The chosen method for detecting outliers to the present project was based on optical observation of the data set projected on the first and the third Principal Component. Deletion of outliers did not significantly improved results. Nevertheless, the two obtained global models gave comparable results to other QSAR models based on heterogeneous data. Faucon's study was based on 96 compounds and the best obtained model had $R^2 = 0.65$ and $Q^2 = 0.50$ and the worst model had $R^2 = 0.42$ and $Q^2 = 0.38$

(Faucon et al., 2001). High variety of chemical structures does not seem to be an ideal choice for predicting aquatic toxicity by global QSARs.
Generally the kNN models had a better predictive ability than the global models.

4.0 Conclusions

In the present research project, we have developed qualitative read-across by means of kNN method applied on binary descriptors and quantitative read-across by applying kNN method both on binary descriptors and DRAGON molecular descriptors, after GA variable selection. Several similarity/ distance measures were evaluated in combination with different sets of molecular descriptors.
In addition, global QSAR models by GA-PLS regression and a QSAR model based on logP were calculated. Most of the models did not give significantly better results than the already published QSAR models on aquatic toxicity. Possible reasons of the poor predictivity can be:

- The quality of the experimental data. The data collection is a very important step for a QSAR model since good quality of data decides on the robustness and quality of the final model. Combining data from different sources is always a risk for the quality of data.
- The perception of chemical similarity strongly depends on the selected structural features and it cannot be the only criterion for the biological activity of a chemical substance. Toxicity is not only a factor of a chemical structure or a molecular property. It is a reaction between the toxic substance and the organism, it can depend on too many factors and the results are not always easy interpretable. The mistake is that we look at the test organism as a black box (*Berenbaum, 1985*).
- The variable selection method can influence the results of the model and its predictive capacity. Choosing the right variables is the core of QSAR methods. Although there are many chemometric methods for variable selection, the final decision on the type and number of the important descriptors remains always a problem.
- Large data sets with big heterogeneity can have better results with kNN methodology than with global regression models. In addition the methodology kNN can treat the whole data set without elimination of potential outliers because its basic approach is the similarity among a small number of molecules and not a global regression in which every compound influence the final model.
- The kNN regression gives better results with molecular descriptors than with binary descriptors probably because molecular descriptors encompass properties and structures.
- The kNN classification based on binary descriptors gives better results than the ECOSAR classification, probably because only the logP variable is not able to distinguish the acute aquatic toxicants.

The development of kNN models is a very active field of QSAR methodologies. The methodology is fast and can be applicable to large and heterogeneous data. Further steps such as other variable selection methods and classification based on molecular descriptors can be done for improving the fit and the predicitivity of the kNN models.

5.0 References

1. Berenbaum M.C.; The expected effect of a combination of agents: the general solution, *Journal of Theoretical Biology* 1985, (114) 413-431
2. Bernot R.J.; Brueseke A.M.; Evans-White M.A.; Lamberti G.A.;Acute and chronic toxicity of imidazolium-based ionic liquids on *Daphnia magna*, *Environmental Toxicology and Chemistry* 2005, (24) 87-92

3. Brausch J.M.; Smith P.N.; Development of resistance to cyfluthrin and naphthalene among *Daphnia magna*, *Ecotoxicology* 2009, (19) 600-609
4. Costanzo S.D.; Watkinson A.J.; Murby E.J.; Kolpin D.W.; Sandstrom M.W.; Is there a risk associated with the insect repellent DEET (N,N-diethyl-m-toluamide) commonly found in aquatic environments? *Science of the Total Environment* 2007, (384) 214-220
5. Cover T.M.; Hart P.E.; Nearest neighbor pattern classification, *IEEE Transactions on Information Theory* 1967, (13) 21-7
6. Dojmi di Delupis G.; Macrì A.; Civitareale C.; Migliore L.; Environmental risk assessment of six human pharmaceuticals: are the current environmental risk assessment procedures sufficient for the protection of the aquatic environment? *Environmental Toxicology and Chemistry* 2004 (23) 1344-1354
7. Faucon J.C.; Bureau R.; Faisant J.; Briens F.; Rault S.; Prediction of the Daphnia acute toxicity from heterogeneous data, *Chemosphere* 2001, (44) 407-422
8. Ferrari B; Garric J.; Mons R.; Vollat B.; Fraysse B.; Paxéus N.; Lo Giudice R.; Pollio A.; Garric J.; Competition increases toxicant sensitivity and delays the recovery of two interacting populations, *Aquatic Toxicology* 2012 (106-107) 25-31
9. Foit K.; Kaske O.; Liess M.; Antibiotics for zootechnical use: effects of aute high and low dose contamination on *Daphnia magna* Strauss, *Aquatic Toxicology* 1992, (22) 53-60
10. Guilhermino L.; Ribeiro R.; Gonqalves F.; Soares A.M.V.M.; METIER (Modular Ecotoxicity Tests Incorporating Ecological Relevance) for difficult substances-III. Effects of medium renewal and use of a carrier on the bioavailability of parathion, *Environmental Pollution* 1996 (92) 97-99
11. Guilhermino L.; Diamantino T.; Silva M.C.; Soares A.M.V.M.; Acute toxicity test with *Daphnia magna*: an alternative to mammals in the pre-screening of chemical toxicity? *Ecotoxicology and Environmental Safety* 2000 (46) 357-362
12. Hirano M.; Ishibashi H.; Matsamura N.; Nagao Y.; Watanabe N.; Watanabe A.; Onikura N.; Kishi K.; Arizono K.; Acute toxicity responses of two crustaceans, Americamysis bahia and *Daphnia Magna* to endocrine disrupters, *Journal of health science* 2004 (50) 87-100
13. Horn O.; Nalli S.; Cooper D.; Nicell J.; Plasticizer metabolites in the environment
Water Research 2004 (38) 3693-3698
14. Jemec A.; Tisler T.; Drobne D.; Sepcic K.; Fournier D.; Trebs P.; Comparative toxicity of imidacloprid, of its commercial liquid formulation and of diazinon to a non-target arthropod, the microcrustacean *Daphnia magna*, *Chemosphere* 2007 (68) 1408-1418
15. Könemann, H.; Quantitative structure–activity relationships in fish toxicity studies. Part 1: Relationship for 50 industrial pollutants, *Toxicology* 1981 (19) 209–221
16. Kyriakopoulou K.; Anastasiadou P.; Machera K.; Comparative toxicities of fungicide and herbicide formulations on freshwater and marine species, *Bulletin of Environmental Contamination Toxicology* 2009 (82) 290-295
17. Leardi R.; Lupianez A.; Genetic algorithms applied to feature selection in PLS regression: how and when to use them, *Chemometrics and Inteligent Laboratory Systems* 1998 (41) 195-207
18. Liu W.; Chen S.; Toxicity of Chiral Pesticide Rac-Metalaxyl and R-Metalaxyl to *Daphnia magna*, *Bulletin of Environmental Contamination Toxicology* 2008 (81) 531-534
19. Martins J.C.; Saker M.L.; Oliva Teles L.F.; Vasconcelos V.M.; Oxygen consumption by *Daphnia magna* strauss as a marker of chemical stress in the aquatic environment, *Environmental Toxicology and Chemistry*, 2007 (26) 1987-1991
20. Nørgaard K.B.; Cedergreen N.; Pesticide cocktails can interact synergistically on aquatic crustaceans, *Environmental Science and Pollution Research* 2010 (17) 957-967
21. Ochoa-Acuña H.G.; Bialkowski W.; Yale G.; Hahn L.; Toxicity of soybean rust fungicides to freshwater algae and *Daphnia magna*, *Ecotoxicology* 2009 (18) 440-446
22. Okamura H.; Aoyama I.; Liu D.; Maguire R.J.; Pacepavicius G.J.; Lau L.Y.; Fate and ecotoxicity of the new antifouling compound Irgarol 1051 in the aquatic environment, *Water Research* 2000 (34) 3523-3530
23. Patlewicz G.; Chemical Categories and Read Across, *European Commission Joint Research Center* 2005 EUR 21898, Ispra Italy
24. Persoone G.; Janssen C.; de Coen W.; New microbiotests for routine toxicity screening and biomonitoring *Kluwer Academic/Plenum publishers* New York 2000
25. Randall W.F.; Dennis W.H.; Warner M.C.; Acute Toxicity of Dechlorinated DDT, Chlordane and Lindane to *Bluegill* (*Lepomis macrochirus*) and *Daphnia magna*, *Bulletin of Environmental Contamination Toxicology* 1979 (21) 849-854

26. Reuschenbach P.; Silvani M.; Dammann M.; Warnecke D.; Knacker T.; ECOSAR model performance with a large test set of industrial chemicals, *Chemosphere* 2008 (71) 1986-1995
27. Sanderson H.; Thomsen M.; Comparative analysis of pharmaceuticals versus industrial chemicals acute aquatic toxicity classification according to the United Nations classification system for chemicals. Assessment of the (Q)SAR predictability of pharmaceuticals acute aquatic toxicity and their predominant acute toxic mode-of-action, T*oxicology Letters* 2009 (187) 84-93
28. Staples C.A.; Davis J.W.; An examination of the physical properties, fate, ecotoxicity and potential environmental risks for a series of propylene glycol ethers, *Chemosphere* 2002 (49) 61-73
29. Todeschini R.; Consonni V.; Molecular Descriptors for Chemoinformatics, *WILEY-VCH*, Weinheim Germany 2009
30. Von der Ohe P.C.; K.R.; Ebert R.; Altenburger R.; Liess M.; Schüürmann G.; Structural alertss-a new classification model to discriminate excess toxicity from narcotic effect levels of organic compounds in the acute daphnid assay, *Chemical Research in Toxicology* 2005 (18) 536-555
31. Wan M.; Kuo J.; Buday C.; Schroeder G.; Van Aggelen G.; Pasternak J.; Toxicity of a-, b-, (a+b)-Endosulfan and their formulated and degradation products to *Daphnia magna*, Hyalella azteca, Oncorhynchus mukiss, Oncorhynchus kisutch, and biological implication in streams, *Environmental Toxicology and Chemistry* 2005 (24) 1146-1154
32. Wan M.T.; Buday C.; Schroeder G.; Kuo J.; Pasternak J.; Toxicity to *Daphnia magna*, *Hyalella* azteca, *Oncorhynchus* kisutch, *Oncorhynchus* mykiss, *Oncorhynchus* tshawytscha, and Rana catesbeiana of Atrazine, Metolachlor, Simazine, and Their Formulated Products, *Bulletin of Environmental Contamination Toxicology* 2006 (76) 52-58
33. Williams E.S.; Berninger J.P.; Brooks B.W.; Application of chemical toxicity distributions to ecotoxicology data requirements under REACh, *Environmental Toxicology and Chemistry* 2011 (30) 1943-1954
34. Zou E.; Fingerman M.; Effects of estrogenic xenobiotics on molting of the *Water Flea Daphnia magna*, *Ecotoxicology and Environmental Safety* 1997 (38) 281-285

6.0 Dissemination

The results of this research project were presented in the following events:

- Oral communication: *"Ecotoxicological properties evaluation by read across methodologies, (first considerations)"* ECO Mid Term Review Meeting, Leiden Netherlands 26-27th September 2011
- Poster: *"Read-across methodology in aquatic ecotoxicology and ready biodegradation"* 2nd Winter School of the Marie Curie ITN "Environmental Chemoinformatics", INIA Spain, 27th -2nd March 2012
- Oral communication: *"Ecotoxicological property evaluation by read across methodologies, first kNN models"* 2nd Summer School 2012 of the Marie Curie ITN "Environmental Chemometrics", University of Milano Bicocca, Verona 11-15 June 2012

Finally, The results of the present project will be presented as oral communication at the "*Third International Symposium on Green Chemistry for Environment, Health and Development"* at Skiathos island, Greece, 3-5 October 2012

7.0 Training & Scientific Meetings

2nd Summer School of the Marie Curie Training Network "Environmental Chemoinformatics", Leiden University, The Netherlands, 19-30th September 2011
Topics covered:

- General introduction to REACH

- Integrated testing strategies
- Exposure modeling
- Multimedia Fate Modeling
- QSAR, in vivo-in vitro
- Fate-effect assessment nanoparticles

Internal training action "Chemoinformatics tools for eco-toxicology", University of Milano-Bicocca, Italy, 20, 21, 25, 27th October 2011
Lecturer: Alberto Manganaro
Topics covered:

- The KNIME platform: introduction and integration with software for Molecular calculation.
- Development of extension for the KNIME platform.
- The CDK library and its use inside chemoinformatic applications.
- The VEGA platform: an open-source tool for prediction of eco-toxicological endpoints.

ADME Toxicokinetics Workshop organized by Joint Research Centre: Potential for further integration of toxicokinetic modelling into the prediction of in vivo dose-response curves without animal experiments, Ispra, Italy, 13th October 2011.

HPC for Proteomics workshop organised by CINECA, Casalecchio di Reno, Italy, 12th December 2011.

RNBIO Advanced Training Course on Bioinformatics for Proteomics, Casalecchio di Reno, Italy, 13-14th December 2011.

School of Analytical Chemistry and Chemometrics, Dep. Of Chemistry, Pharmaceutical and Nutrition Technologies, University of Genova, Italy, 23- 26th January 2012.
Topics covered:

- Introduction in chemometrics
- Data structure, pre-treatment
- Cluster analysis with hierarchical methods and Kmeans
- Classification analysis
- Linear regression
- Multivariate regression

2nd Winter School of the Marie Curie Initial Training Network "Chemoinformatics", INIA Spain, 27th -2nd March 2012
Topics covered:

- General introduction to CADASTER projects
- In vitro and in silico toxicology: Biological activity and computational rationalization
- Electron microscopy of nanoparticles
- ICP-MS installation
- Use of DLS. Determination of nanoparticle size frequency distribution and Z-potential
- Training in patents and IP rights

Internal Training Action, "Variable Selection by the LASSO method", University of Milano Bicocca 5-9th March 2012.
Topics covered:

- Introduction to the variable selection methods in regression
- The LASSO method and related approaches in theory
- The LASSO method in practice
- Practical LASSO methods and discussion

Scientific meeting "Protecting the Mediterranean sea against pollution" Prince Albert II of Monaco Foundation, University of Milano Bicocca 6th March 2012.

School of Chemometric Methods for the Process Monitoring, University of Modena and Reggio Emilia, Modena Italy 17-20th April
Topics covered:

- PCA and PLS basis
- Control process charts methodology
- Illustration of multivariate process charts
- Continuous and batch process monitoring

6th SETAC World Congress, Berlin, Germany 20-24th May 2012

"Application of In Vitro methods for toxicological evaluation" SITOX (Italian Society of Toxicology), Milan, Italy 5th June 2012

3rd Summer School 2012 of the Marie Curie ITN "Environmental Chemometrics", Verona, Italy, 11-15th June 2012
Topics covered:

- QSAR: from molecular structure to models
- How to build a QSAR model
- QSAR toolbox
- Lectures of the Members of the *International Academy of Mathematical Chemistry*

Appendix

Table I. Median logLC_{50} 48 h. to *Daphnia magna* expressed in Molarity of 511 substances divided in training and test set.

TRAINING					TEST				
CAS	Name	Class	logLC50	References	CAS	Name	Class	logLC50	References
50-06-6	Phenobarbital	IV	-2.20	29	434-07-1	Oxymetholone	II	-4.49	19
50-28-2	Estradiol 17b	II	-4.96	12	503-87-7	2-Thioxo-4-Imidazolinone	III	-3.77	30
50-48-6	Amitriptyline	I	-5.55	29	525-79-1	N-(2-Furanylmethyl)-9H-Purin-6-Amine	II	-5.00	6,6
50-78-2	Aspirin	IV	-3.11	29	532-55-8	Benzoyl Isothiocyanate	II	-4.93	30
51-52-5	Propylthiouracil	III	-4.19	29	534-13-4	N,N'-Dimethylthiourea	III	-3.85	30
52-24-4	Thiotepa	IV	-2.54	29	536-90-3	3-Methoxybenzeneamine	I	-5.64	30
52-68-6	Trichlorphon	I	-6.61	33, 30	541-73-1	1,3-Dichlorobenzene	II	-4.18	30
54-85-3	Isoniazid	III	-3.78	29	542-75-6	1,3-Dichloropropene	II	-4.25	30
55-38-9	Fenthion	I	-6.79	30	542-85-8	Isothiocyanatoethane	I	-5.31	30
55-63-0	Nitroglycerin	III	-3.74	29, 30	554-00-7	2,4-Dichloroaniline	I	-5.43	30
56-23-5	Tetrachloromethane	III	-3.64	30, 2	556-61-6	Isothiocyanatomethane	I	-5.42	30
56-38-2	Parathion	I	-8.02	3, 10, 30, 33, 11	578-54-1	2-Ethylbenzenamine	II	-4.18	30
56-55-3	Benz(a)Anthracene	I	-6.37	3	589-16-2	4-Ethylaniline	I	-6.13	30
56-75-7	Chloramphenicol	IV	-3.03	29	592-82-5	1-Isothiocyanatobutane	I	-5.43	30
57-62-5	Chlortetracycline	IV	-3.57	29	602-01-7	2,3-Dinitrotoluene	I	-5.44	30
57-63-6	Ethinylestradiol	II	-4.72	29	609-19-8	3,4,5-Trichlorophenol	I	-5.46	30

57-92-1	Streptomycin	IV	-3.08	29
58-08-2	Caffeine	IV	-3.03	29
58-14-0	Pyrimethamine	II	-4.63	29, 30
58-22-0	Testosterone	II	-5.17	29
58-89-9	Lindane	II	-5.39	34, 30, 25, 25
58-90-2	2,3,4,6-Tetrachlorophenol	I	-6.12	30
59-06-3	Ethopabate	IV	-3.07	30
59-87-0	Nitrofurazone	III	-3.84	29
62-56-6	Thiourea	III	-3.84	30
63-25-2	Carbaryl	I	-7.35	30, 33, 33
64-17-5	Ethanol	IV	-0.72	11, 30
67-20-9	Nitrofurantion	III	-3.51	29
67-64-1	Acetone	IV	-0.62	30
67-66-3	Trichloromethane	IV	-2.72	3, 30
67-72-1	Hexachloroethane	II	-4.83	30
68-12-2	N,N-Dimethylformamide	IV	-0.70	30
71-23-8	1-Propanol	IV	-0.93	30
71-43-2	Benzene	IV	-2.34	30, 2, 2
71-63-6	Digitoxin	III	-3.88	29
72-20-8	Endrin	I	-6.38	30
74-83-9	Methyl Bromide	II	-4.63	30
75-07-0	Acetaldehyde	IV	-0.55	30
75-08-1	Ethyl Mercaptan	I	-5.56	30
75-25-2	Bromoform	III	-3.74	30
75-35-4	1,1-Dichloroethene	III	-3.28	30
76-01-7	Pentachloroethane	IV	-2.97	30
77-47-4	Hexachlorocyclopentadiene	I	-6.72	30
78-83-1	2-Methyl-1-Propanol	III	-3.53	30
78-99-9	1,1-Dichloropropane	III	-3.57	30
79-00-5	1,1,2-Trichloroethane	III	-3.41	30
79-01-6	Trichloroethene	III	-3.35	30
79-06-1	Acrylamide	IV	-2.65	30
79-09-4	Propionic Acid	III	-3.17	30
83-41-0	1,2-Dimethyl-3-Nitrobenzene	II	-4.56	30
84-66-2	Diethyl Phtalate	III	-3.60	34, 30
85-01-8	Phenanthrene	I	-5.36	30
618-62-2	1,3-Dichloro-5-Nitrobenzene	II	-4.46	30
625-53-6	Ethylthiourea	III	-4.00	30
626-43-7	3,5-Dichloroaniline	II	-5.16	30
630-20-6	1,1,1,2-Tetrachloroethane	III	-3.84	30
632-22-4	1,1,3,3-Tetramethylurea	IV	-1.60	30
634-67-3	2,3,4-Trichloroaniline	I	-5.43	30
634-83-3	2,3,4,5-Tetrachloroaniline	I	-5.56	30
636-30-6	2,4,5-Trichloroaniline	II	-4.76	30
637-07-0	Clofibrate	III	-4.14	29
683-10-3	Dodecyl Dimethyl Betaine	III	-3.76	33
693-21-0	Diethylene Glycol Dinitrate	III	-3.34	30
732-11-6	Phosmet	I	-5.60	30
759-94-4	Dipropylcarbamothioic Acid,S-Ethyl Ester	II	-4.61	30
770-35-4	Propylene Glycol Mono-Phenyl Ether	IV	-2.61	210
786-19-6	Carbophenothion	I	-6.44	30
825-44-5	Benzo[B]Thiophene S,S-Dioxide	III	-4.07	30
859-18-7	Lincomycin	I	-5.78	29
877-43-0	2,6-Dimethylquinoline	III	-3.62	30
935-95-5	2,3,5,6-Tetrachlorophenol	I	-5.61	30
959-98-8	A-Endosulfan	II	-5.54	31
1014-70-6	Simetryn	III	-3.63	30
1016-05-3	Dibenzothiophene-5,5-Dioxide	II	-4.57	30
1024-57-3	Heptachlor Epoxide	I	-6.21	30
1031-07-8	Endosulfan Sulfate	II	-5.30	31
1141-88-4	Dithioaniline	I	-6.53	29
26914-33-0	2,2',4,4'-PCB	I	-6.10	30
1516-32-1	Butylthiourea	III	-3.85	30
1563-66-2	Carbofuran	I	-6.52	30
1570-64-5	4-Chloro- -Cresol	I	-5.69	30
1582-09-8	Trifluralin	I	-6.24	30
1622-61-3	Clonazepam	III	-4.02	29
1665-48-1	Metaxalone	III	-3.47	29
1806-26-4	4-Octylphenol	I	-6.36	34
1825-21-4	Pentachloroanisole	I	-7.02	30
1836-77-7	Chlornitrofen	I	-5.88	30
1897-45-6	Chlorothalonil	I	-6.21	30

85-68-7	Butyl Benzyl Phthalate	II	-5.19	30
86-74-8	Carbazole	II	-4.70	30
87-86-5	Pentaclorophenol PCP	I	-5.64	30
88-06-2	2,4,6-Trichlorophenol	III	-4.12	19
88-72-2	1-Methyl-2-Nitrobenzene	II	-4.14	30
88-73-3	1-Chloro-2-Nitrobenzene	III	-3.64	30
88-85-7	2-(1-Methylpropyl)-4,6-Dinitrophenol	I	-6.00	30
88-89-1	2,4,6-Trinitrophenol	III	-3.43	30
89-61-2	1,4-Dichloro-2-Nitrobenzene	III	-4.26	30
89-86-1	2,4-Dihydroxybenzoic (B-Resorcylic) Acid	IV	-3.12	30
90-02-8	Salicylaldehyde, 2-Hydroxybenzaldehyde	II	-4.45	30
90-04-0	O-Aminoanisole	III	-4.01	30
90-43-7	2-Phenylphenol	I	-5.38	30
91-20-3	Naphthalene	II	-4.17	30, 3
91-53-2	Ethoxyquin	II	-5.04	29
91-66-7	N N Diethylaniline	III	-3.32	30
91-94-1	3,3 -Dichlorobenzidine	II	-5.38	30
92-69-3	4-Phenylphenol	II	-4.67	30
94-75-7	2,4-Dichlorophenoxyacetic Acid	IV	-3.17	30
95-15-8	Benzo[ ]Thiophene	III	-3.36	30
95-47-6	O-Xylene	III	-3.78	30
95-50-1	1,2-Dichlorobenzene	II	-4.81	30
95-53-4	Ortho-Toluidine	I	-5.31	30
95-57-8	2-Chlorophenol	II	-4.34	30
95-76-1	Benzenamine, 3,4-Dichloro-	I	-5.95	30
95-82-9	2,5-Dichloroaniline	II	-4.74	30
96-18-4	1,2,3-Trichloropropane	III	-3.72	30
97-00-7	1-Chloro-2,4-Dinitrobenzene	I	-5.4	30
97-74-5	Bis(Dimethylthiocarbamyl)Sulfide	II	-4.86	30
977-7-8	Bis(Diethylthiocarbamoyl)Disulfide	I	-5.56	30
98-95-3	Nitrobenzene	III	-3.48	30
99-51-4	1,2-Dimethyl-4-Nitrobenzene	III	-3.98	30
99-87-6	Cymene	II	-4.32	30
1912-24-9	Atrazine	III	-3.53	30, 32
2032-65-7	Methiocarb	I	-7.07	33
2051-60-7	2-Chlorobiphenyl	I	-5.42	30
2051-61-8	3-Chlorobiphenyl	I	-5.64	30
2144083	Pyrogallolaldehyde	III	-3.67	29
2212-67-1	Molinate	II	-4.48	33
2257092	(2-Isothiocyanatoethyl)Benzene	I	-6.10	30
2303-17-5	Triallate	I	-6.19	33
2437-79-8	2,4,2 ,4 -Tetrachlorobiphenyl	I	-6.99	30
2489-77-2	Trimethylthiourea	IV	-2.19	30
2539-17-5	2-Methoxytetrachlorophenol	I	-6.08	30
2668-24-8	2-Methoxy-4,5,6-Trichlorophenol	I	-5.37	30
2782-91-4	Tetramethyl Thiourea	IV	-2.23	30
2809-21-4	Etidronic Acid	IV	-2.59	29
2921-88-2	Chlorpyrifos	I	-7.77	30
3056-17-5	Stavudin	IV	-2.36	29
3209-22-1	1,2-Dichloro-3-Nitrobenzene	II	-4.62	30
3332-27-2	Myristyl Dimethylamine Oxide	II	-4.58	33
3483123	Dithiothreitol	III	-3.76	30
3521-62-8	Erythromycin Estolate	I	-5.97	29
3547044	Dde	I	-6.86	30
3766-81-2	Methylcarbamate	I	-6.32	30
3930-20-9	Sotalol	IV	-2.96	29
4044-65-9	1,4-Diisothiocyanatobenzene	I	-6.40	30
4104-75-0	N-Methyl-N-Phenylthiourea	III	-3.36	30
5251-34-3	Cloprednol	III	-3.96	29
5417-35-6	Isopropyldioxepen	IV	-2.27	29
6317-18-6	Thiocyanic Acid, Methylene Ester	I	-6.25	30
6972-05-0	N,N-Dimethylthiourea	III	-3.39	30
7542-37-2	Aminosidine	IV	-3.09	11
9002-93-1	Alpha-Dodecyl-Omega-Hydroxypoly(Oxy-1,2-Ethanediyl)	II	-4.68	33
9016-45-9	Alpha-(Nonylphenyl)-Omega-Hydroxypoly (Oxy-1,2-Ethanediyl) (Nonylphenol Ethoxylate)	III	-4.52	33
9036-19-5	Alpha-[(1,1,3,3-Tetramethylbutyl)Phenyl]-Omega-	II	-4.70	33

99-99-0	4-Methylnitrobenzene	III	-4.01	30
100-00-5	4-Chloronitrobenzene	II	-4.31	30
100-41-4	Ethyl Benzene	III	-3.54	30
100-46-9	Benzenemethanamine	III	-3.33	29
100-61-8	N-Methylaniline	I	-5.79	30
100-66-3	Benzene, Methoxy-Anisol	III	-3.43	29
103-69-5	Ethylaniline	I	-5.46	30
103-90-2	Paracetamol	II	-4.22	29
103-72-0	Isothiocyanatobenzene	I	-6.13	30
104-76-7	2-Ethylhexanol	III	-3.12	13
104-90-5	2-Methyl-5-Ethylpyridine	III	-3.48	29
104-94-9	4-Methoxybenzenamine	I	-5.57	30
105-37-3	Ethyl Propionate	IV	-2.78	30
105-53-3	Malonic Acide Diethylester	IV	-2.90	29
106-47-8	P-Chloroaniline	I	-6.41	30
106-89-8	Epichlorohydrine	III	-3.77	29, 30
107-02-8	2-Propenal Acroleine	I	-6.41	29
107-03-9	1-Propanethiol	I	-6.10	30
107-06-2	1,2-Dichloroethane	IV	-2.29	30
107-07-3	2-Chloroethanol	IV	-2.61	30
107-11-9	Allylamine	III	-3.15	30
107-13-1	Acrylonitrile	II	-3.78	30
107-21-1	Ethylene Glycol	IV	-2.0875	30
107-92-6	N-Butyric Acid	III	-3.16	30
108-01-0	Ethanol, 2-(Dimethylamino)-	III	-2.9589	29
108-39-4	M-Cresol	III	-3.76	30
108-42-9	3-Chloroaniline	I	-6.11	30
108-44-1	M-Toluidine	I	-5.17	30
108-85-0	Bromocyclohexane	III	-3.89	30
108-88-3	Toluene	IV	-2.8	30
109-52-4	Pentanoic Acid	III	-3.36	30
109-89-7	Diethylamine	III	-3.12	30
110-02-1	Thiophene	IV	-2.42	30
110-16-7	Maleic Acid	IV	-2.76	29
10161-34-9	Alpha-Undecyl-Omega-Hydroxypoly(Oxy-1,2-Ethanediyl) Trenbolone Acetate	I	-5.54	29
12002-48-1	Trichlorobenzene	II	-4.40	30
13035-61-5	Betariboacetate	IV	-3.42	29
13194-48-4	Ethoprop	I	-6.58	33
13311-84-7	Flutamide	II	-5.30	29
13684-63-4	Phenmedipham	II	-4.65	33
15251-48-6	Oxytetracycline	III	-4.35	29
15263-53-3	Dithiocarbamate	I	-7.38	30
15687-27-1	Ibuprofen	II	-4.36	29
16752-77-5	Methomyl	I	-6.94	33, 33
18259-05-7	2,3,4,5,6-Pcb	I	-7.61	30
20485-39-6	Ethyl-4-Methyl-5-Oxazole Carboxylate	IV	-2.67	29
20830-75-5	Digoxin	III	-4.57	29
22204-53-1	Naproxen	III	-3.79	29
22224-92-6	Fenamiphos	I	-8.11	33
22664-55-7	Metipranolol	III	-3.92	29
22781-23-3	Bendiocarb	I	-6.88	33
23103-98-2	Pirimicarb	I	-7.04	33
23564-05-8	Thiophanate-Methyl	III	-4.57	30, 18
24579-73-5	Propamocarb	IV	-3.25	33
25154-52-3	Nonylphenol	I	-6.41	12, 30
25875-51-8	Robenidine	I	-6.61	30
28249-77-6	Thiobencarb	II	-4.67	33, 30
29122-68-7	Atenolol	III	-3.90	29
33813-20-6	5,6-Dihydro-3H-Imidazo[2,1-C]-1,2,4-Dithiazole-3-Thione	I	-5.92	30
34398-01-1	Alpha-Undecyl-Omega-Hydroxypoly(Oxy-1,2-Ethanediyl)	II	-4.76	33
35067-38-5	Diflubenzuron	I	-7.79	30
35693-99-3	2,2 ,5,5 -Tetrachloro-1,1' -Biphenyl	I	-6.99	30
37517-30-9	Acebutolol	III	-3.82	29
37680-65-2	2,2 ,5-Trichloro-1,1' -Biphenyl	I	-6.67	30
40487-42-1	Pendimethalin	III	-3.73	18
40596-69-8	Methoprene	II	-5.28	8
42200-33-9	Nadolol	IV	-3.28	29
51022-70-9	Albuterol Sulfate	IV	-2.99	29

110-83-8	Cyclohexene	II	-3.94	30
110-86-1	Pyridine	IV	-1.77	30
110-89-4	Piperidine	II	-3.93	29
111-42-2	Diethanolamine	IV	-2.93	29, 30
111-70-6	1-Heptanol	III	-3.22	30
111-90-0	2-(2-Ethoxyethoxy)Ethanol	IV	-1.53	30
111-91-1	Propoxur	IV	-2.94	30
114-07-8	Erythromycin	IV	-3.54	29
114-26-1	Propoxur	I	-5.67	33, 30, 33
115-20-8	2,2,2-Trichloroethanol	IV	-3.00	30
115-29-7	Endosulfan	I	-6.14	30
115-31-1	Isobornyl Thiocyanatoacetate	I	-6.50	30
115-86-6	Phosphoric Acid	II	-5.51	30
116-06-3	Aldicarb	I	-5.55	30, 33
118-96-7	2,4,6-Trinitrotoluene	II	-4.39	30
119-64-9	Naphthalene	II	-4.74	29
119-65-3	Isoquinoline	III	-3.71	30
120-78-5	Benzothiazole, 2,2-Dithiobis-	III	-3.62	29
120-83-2	2,4-Dichlorophenol	II	-4.80	30
120-93-4	Ethyleneurea	IV	-1.19	30
121-14-2	2,4-Dinitrotoluene	III	-3.72	30
121-29-9	Pyrethrine II	I	-7.40	30
121-69-7	N N-Dimethylaniline	II	-4.38	29
121-73-3	3-Nitrochlorobenzene	III	-3.84	30
121-75-5	Malathion	I	-7.61	33, 30, 33
121-87-9	2-Chloro-4-Nitroaniline	II	-4.49	30
122-14-5	Fenitrothion	II	-7.60	33, 30, 33
122-66-7	1,2-Diphenylhydrazine	II	-4.65	30
123-38-6	Propanal	III	-2.82	29
124-40-3	Dimethylamine	III	-2.96	30
129-06-6	Warfarin	IV	-2.96	29
130-80-3	Diethylstilbestrol	II	-5.54	29
131-11-3	Dimethyl Phthalate	III	-3.77	30
132-65-0	Dibenzothiophene	II	-5.06	30
135-19-3	2-Naphthol	II	-4.61	30
137-26-8	Thiram	I	-6.06	30, 19, 18
51218-45-2	Metolachlor	III	-4.34	32
51333-22-3	Budesonide	III	-4.36	29
51481-61-9	Cimetidine	IV	-2.53	29
51630-58-1	Fenvalerate	I	-8.67	30, 10
52645-53-1	Permethrin	I	-8.29	30
54910-89-3	Fluoxetine	I	-5.78	29
54965-21-8	Albendazole	I	-7.04	29
56392-17-7	Metoprolol	II	-4.48	29
59756-60-4	Fluridone	II	-4.86	30
59729-33-8	Citalopram	II	-4.92	29
59865-13-3	Cyclosporin	III	-4.78	29
60207-90-1	Propiconazole	II	-4.58	20, 21
61791-26-2	Tallow Alkyl Amines, Ethoxylated	II	-4.70	33
61869-08-7	Paroxetine	I	-5.75	29
63675-72-9	Nisolfipine	III	-4.07	29
64359-81-5	Sea Nine	I	-7.85	29
66455-14-9	C12–13 Alcohols, Ethoxylated	I	-5.56	33
67375-30-8	Alpha-Cypermethrin	I	-9.14	20
67564-91-4	Fenpropimorph	II	-5.10	20
68359-37-5	Cylfuthrin	I	-8.86	3
68439-46-3	Alcohols, C9–11, Ethoxylated	II	-4.83	33
68951-67-7	Alcohols, C14–15, Ethoxylated	I	-5.69	33
72956-09-3	Carvedilol	II	-5.13	29
73334-07-3	Lopromide	IV	-2.89	29
76470-66-1	Loracarbef	IV	-2.56	29
76824-35-6	Famotidine	IV	-2.93	29
77326-96-6	Flunisolide Hemihydrate	II	-4.65	29
82419-36-1	Ofloxacin	III	-3.67	6
83905-01-5	Azithromycin	IV	-3.80	29
85721-33-1	Ofloxacin	III	-4.28	29
87392-12-9	S-Metolachlor	II	-4.28	33
88768-40-5	Cilazapril	IV	-2.63	29
88917-22-0	Propanol, 1(Or 2)-(2-Methoxymethylethoxy)-, Acetate	IV	-2.24	28
91374-20-8	Ropinirole	IV	-2.95	29
98079-51-7	Lomefloxacin	IV	-3.43	29
98319-26-7	Finasteride	III	-4.25	29

138-59-0	Shikimic Acid	IV	-3.18	29	103628-46-2	Sumatriptan	IV	-3.01	29
142-28-9	1,3-Dichloropropane	IV	-2.61	30	103577-45-3	Lansoprazole	III	-4.23	29
142-96-1	Butyl Ether	III	-3.70	30	106266-06-2	Risperidone	II	-4.84	29
143-33-9	Sodium Cyanide	II	-4.16	29	107534-96-3	Tebuconazole	II	-5.61	20, 21
148-01-6	Dinitolmide	IV	-3.14	30	127779-20-8	Saquinavir	III	-4.27	29
150-19-6	3-Methoxyphenol	III	-3.48	30	138261-41-3	Imidacloprid	III	-3.48	14,14
152-11-4	Verapamil	II	-4.85	29	141517-21-7	Trifloxystrobin	I	-5.74	21
156-60-5	Trans-1,2-Dichloroethylene	IV	-2.64	30	154361-50-9	Capecitabine	IV	-2.61	29
206-44-0	Fluoranthene	I	-6.28	30	175013-18-0	Pyraclostrobin	I	-6.76	20
298-02-2	Phorate	I	-7.14	30	224452-66-8	Retapamulin	III	-4.11	29
311-45-5	Paraoxon	I	-8.93	11, 33	231277-92-2	Lapatinib	I	-6.53	29
333-41-5	Diazinon	I	-8.50	14, 14, 30, 33, 33	341-69-5	Orphenadrine HCL	III	-4.41	29
396-01-0	Triamterene	II	-4.40	29	15245-44-0	2,4,6-Trinitro-1,3-Benzenediol	IV	-1.93	30
79660-72-3	Fleroxacin	IV	-3.57	29	58-55-9	Theophylline	IV	-2.57	29
67-73-2	Fluocinolode Acetonide	IV	-3.66	29	470-90-6	Chlorfenvinfos	I	-6.56	30
60142-96-3	Gabaoentin	IV	-2.19	29	22994-85-0	Benzidazole	IV	-3.42	29
82410-32-0	Ganciclovir	IV	-2.41	29	65-85-0	Benzoic Acid	IV	-3.09	29
10238-21-8	Glinbenclamide	IV	-3.70	29	41859-67-0	Bezafibrate	IV	-3.56	29
115-19-5	Methylbutinol	IV	-2.23	29	1812-30-2	Bromazepan	IV	-3.50	29
98-92-0	Niacinamide	IV	-2.09	29	62571-86-2	Captopril	IV	-3.48	29
68-22-4	Norethrindrone	IV	-3.47	29	64544-07-6	Cefuroxime Axetil	IV	-2.71	29
2447-57-6	Sulfadoxine	IV	-3.49	29	81098-60-4	Cisapride	IV	-2.67	29
112410-23-8	Tebufenozide	IV	-4.55	34	50-50-0	Estradiol Benzoate	IV	-3.58	29
56211-40-6	Torasemide	IV	-3.54	29	56177-80-1	Ethoxylfluorouracil	IV	-3.20	29

TEST SET

CAS	Name	Class	LogLC50	References	CAS	Name	Class	logLC50	References
50-29-3	4,4'-DDT	I	-8.32	3, 30, 26, 26	260-94-6	Acridine	II	-4.81	30
51-21-8	Fluorouracil	III	-3.72	29	298-00-0	Methyl Parathion	I	-7.24	30, 3
51-28-5	2,4-Dinitrophenol	II	-4.62	30	289-46-4	Carbamazepine	III	-4.23	29
54-11-5	Nicotine	II	-4.73	29	439-14-5	Diazepan	II	-4.82	29
56-35-9	Hexabutyldistannoxane	I	-6.93	19	534-52-1	Dinitro- -Cresol	II	-4.79	30
57-74-9	Chlordane	I	-6.51	26, 26	576-26-1	2,6-Dimethylphenol	III	-4.04	30
59-50-7	4-Chloro-3-Methylphenol	II	-4.85	30	598-16-3	Tribromoethene	III	-4.33	30
60-51-5	Dimethoate	II	-5.23	30, 33	598-52-7	Methylthiourea	II	-3.98	30
60-54-8	Tetracycline	I	-6.43	29	611-06-3	2,4-Dichloro-1-Nitrobenzene	II	-4.66	30
60-57-1	Dieldrin	I	-6.28	30	622-78-6	Benzylisothiocyanate	I	-6.54	30
62-53-3	Benzenamine Aniline	I	-5.33	30	657-24-9	Metformin	IV	-1.98	29

67-56-1	Methanol	IV	-0.99	11
69-72-7	Sakicylic Acid	IV	-3.07	29
75-05-8	Acetonitril	IV	-1.06	30, 2
75-15-0	Carbon Disulfide	II	-4.56	30
75-21-8	Ethylene Oxide	IV	-2.32	30
78-59-1	Isophorone	IV	-3.06	30
78-87-5	1,2-Dichloropropane	IV	-2.00	30
79-34-5	1,1,2,2-Tetrachloroethane	III	-3.45	30
79-43-6	Dichloroacetic Acid DCA	I	-6.11	11
80-05-7	Bisphenol A	III	-4.25	8
83-42-1	2-Chloro-6-Nitrotoluene	II	-4.61	30
84-74-2	Dibutyl Phthalate	II	-4.88	30
86-30-6	N-Nitrosodiphenylamine	II	-4.40	30
86-50-0	Azinphosmethyl	I	-8.15	30
89-59-8	4-Chloro-2-Nitrotoluene	II	-4.27	30
90-05-1	2-Methoxyphenol	III	-3.68	30
90-13-1	1-Chloronaphthalene	II	-5.01	30
91-64-5	Coumarin	III	-4.03	30
92-52-4	Biphenyl	II	-4.66	30
94-74-6	Acetic Acid, (4-Chloro-2-Methylphenoxy)-	IV	-3.05	111
95-48-7	O-Cresol	III	-3.87	30
95-51-2	2-Chloroaniline	I	-5.19	30
95-95-4	2,4,5-Trichlorophenol	II	-4.86	30
96-09-3	1,2-Epoxyethylbenzene	III	-4.02	30
96-45-7	Ethylene Thiourea	III	-3.59	30
98-82-8	Cumene	III	-3.64	30
99-08-1	1-Methyl-3-Nitrobenzene	III	-4.04	30
99-65-0	1,3-Dinitrobenzene	III	-3.59	30
100-02-7	P-Nitrophenol	III	-3.96	30
100-42-5	Styrene	III	-3.41	30
101-21-3	Chlorpropham	II	-4.76	33
101-55-3	4-Bromophenyl-Phenyl Ether	I	-5.84	30
101-84-8	Diphenyl Ether	I	-5.46	30
102-08-9	Diphenylthiourea	III	-3.53	30
103-85-5	Phenylthiourea	III	-3.54	30
680-31-9	Hexamethyl Phosphoramide	IV	-1.43	30
882-09-7	Clofibric Acid	III	-3.38	29, 8
935-92-2	Trimethylquinone	II	-5.00	29
944-22-9	Fonofos	I	-3.67	33, 30
999-97-3	Silanamine	IV	-2.94	29
1114-71-2	Pebulate	II	-4.47	33
1247-42-3	Prednisone	III	-3.84	29
1330-20-7	Xylene	II	-4.17	29
1401-69-0	Tylosin	IV	-3.13	29
1570-65-6	2,4-Dichloro-6-Methylphenol	I	-5.65	30
1918021	Picloram	III	-3.61	30
1982-47-4	Chloroxuron	II	-4.99	30
2008-58-4	2,6-Dichlorobenzamide	IV	-2.35	30
2051-62-9	4-Chloro-1,1 -Biphenyl	I	-5.60	30
2556-42-5	Tetrapropylthioperoxydicarbonic-Diamide	I	-6.19	30
2741062	1-Phenyl-3-Ethyl Thiourea	III	-3.35	30
2764-72-9	Diquat	II	-5.01	30
2921-88-2	Clorpyrifos	I	-8.63	33, 30, 33, 11
3282-30-2	Pivaloyl Chloride	IV	-2.58	29
3380-34-5	Triclosan	I	-6.35	29
7012-37-5	2,4,4' -Pcb	I	-6.21	30
7481-89-2	Zalcitabine	IV	-2.07	29
7664-41-7	Ammonia	II	-3.39	2, 2
8018017	Mancozeb	II	-5.02	30
10605-21-7	Carbendazim	I	-5.99	33, 30, 33
13071-79-9	Terbufos	I	-8.21	33
14080-23-0	GPS-Cyanopyrimidine	III	-3.34	29
15307-86-5	Diclofenac	III	-4.12	6, 29
15862-07-4	2,4,5-Trichlorobiphenyl PCB	I	-5.64	34
20324-32-7	2-Propanol, 1-(2-Methoxy-1-Methylethoxy)-	IV	-1.89	28
22071-15-4	Ketoprofen	III	-3.60	29
23135-22-0	Oxamyl	II	-5.12	33
25167-83-3	2,3,4,5-Tetrachlorophenol	I	-5.76	30
28159-98-0	Irgarol 1051	II	-4.48	22
30560-19-1	Acephate	III	-5.77	33, 33

105-55-5	1,3-Diethylthiourea	IV	-2.84	30
105-67-9	2,4-Dimethylphenol	II	-4.77	30
106-41-2	4-Bromophenol	II	-4.46	30
106-42-3	P-Xylene	III	-3.52	30
106-46-7	1,4-Dichlorobenzene	II	-4.17	30
106-48-9	4-Chlorophenol	II	-4.42	30
107-15-3	Ethylenediamine	III	-3.36	30
107-98-2	1-Methoxy-2-Propanol	IV	-0.59	27, 28
108-18-9	Bis(Isopropyl)Amine	IV	-2.35	30
108-38-3	M-Xylene	III	-3.43	30
108-45-2	M-Phenylenediamine	II	-4.26	29
108-65-6	Propylene Glycol Mono-Methyl Ether Acetate	IV	-2.51	28
108-90-7	Monochlorobenzene	III	-3.77	30
108-95-2	Phenol	III	-3.74	30, 2, 2
109-46-6	Dibutylthiourea	III	-3.52	30
111-30-8	Glutaraldehyde	III	-3.79	29
111-44-4	2,2 -Dichlorodiethyl Ether	IV	-2.78	30
112-27-6	Triethylene Glycol	IV	-0.46	30
122-34-9	Simazine	III	-3.33	30
123-54-6	2,4-Pentanedione	III	-3.32	30
126-07-8	Griseofulvin	IV	-2.51	29
126-73-8	Tributyl Phosphate	II	-4.86	30
127-18-4	Tetrachloroethene	III	-4.04	30
129-00-0	Pyrene	I	-6.17	3
134-62-3	Benzamide, N,N-Diethyl-3-Methyl-DEET	IV	-3.08	4
141-78-6	Ethyl Acetate	IV	-2.09	30
141-90-2	Thiouracil	II	-4.22	30
143-07-7	Dodecanoic Acid	III	-4.07	33
149-31-5	2-Methyl-1,3-Pentanediol	IV	-1.22	30
149-57-5	2-Ethylhexanoic Acid	IV	-3.08	13

33213-65-9	Beta-Endosulfan	II	-5.43	31
33820-53-0	Isopropalin	I	-7.01	30
37680-73-2	2,4,5,2 ,5 -Pcb	I	-7.51	30
38380-07-3	2,2 ,3,3 ,4,4 -Pcb	I	-8.56	30
53469-21-9	Aroclor 1242	I	-6.10	34
57057-83-7	3,4,5-Trichloroguaiacol	I	-5.56	30
57775-29-8	Carazolol	III	-4.30	29
59277-89-3	Acyclovir	III	-3.38	29
59467-70-8	Midazolam	I	-6.21	29
66357-35-5	Ranitidine	IV	-2.68	29
67306-00-7	Fenpropidin	I	-5.74	29, 20
67747-09-5	Prochloraz	II	-4.94	20
68131-39-5	Alcohols, C12–15, Ethoxylated	I	-5.79	33
68155-09-9	Cocoamidopropyl Dimethylamine Oxide	II	-4.83	33
73590-58-6	Omeprazole	III	-3.59	29
81103-11-9	Clarithromycin	III	-4.60	29
91465-08-6	Cyhalothrin	I	-8.65	30
96829-58-2	Orlistat	II	-4.85	29
70630-17-0	R-Metalaxyl	III	-3.82	18
104227-87-4	Famciclovir	IV	-2.59	29
106325-08-0	Epoxiconazole	II	-4.58	20
112281-77-3	Tetraconazole	II	-4.71	21
130926-19-9	Ibandronate	II	-5.36	29
131860-33-8	Azoxystrobin	I	-6.07	21
134308-13-7	Tolcapine Milled	III	-4.14	29
154-42-7	Thioguanine	III	-4.00	29
8048-52-0	Acriflavine	IV	-3.41	29
5329-14-6	Aminopropanol, Sulfamic Acid	IV	-2.29	29
2135-17-3	Flumethasone	IV	-3.61	29
59-92-7	Levodopa	IV	-3.29	29
110-91-8	Morpholine	IV	-2.94	29