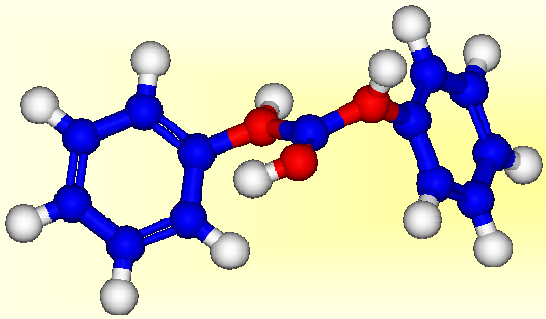
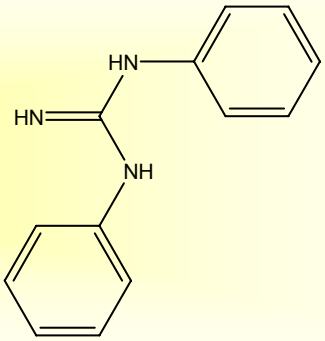




PRELIMINARY ANALYSIS OF AN AQUATIC TOXICITY DATASET AND ASSESSMENT OF QSAR MODELS FOR NARCOSIS

Manuela Pavan, Andrew Worth and Tatiana Netzeva

2005 **EUR 21749 EN**

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

ihcp
Institute for Health and Consumer Protection

EUROPEAN COMMISSION
DIRECTORATE GENERAL
JOINT RESEARCH CENTRE

Institute for Health and Consumer Protection
Toxicology and Chemical Substances Unit
European Chemicals Bureau
I-21020 Ispra (VA) Italy

PRELIMINARY ANALYSIS OF AN AQUATIC TOXICITY DATASET AND ASSESSMENT OF QSAR MODELS FOR NARCOSIS

Manuela Pavan, Andrew Worth and Tatiana Netzeva

2005 **EUR 21749 EN**

LEGAL NOTICE

A great deal of additional information on the
European Union is available on the Internet.
It can be accessed through the Europa server
(http://europa.eu.int)

EUR 21749 EN

Printed in Italy

CONTENTS

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
E-state	Electrotopological index
F	Fisher statistics.
FIT	Kubinyi function
GETAWAY	GEometry, Topology, and Atom-Weights AssemblY
LC50	Concentration of a compound that causes 50% lethality of the animals in a test batch
LOO	Leave-one-out cross-validation
OLS	Ordinary Least Squares
PCA	Principal component analysis
QSAR	Quantitative Structure-Activity Relationships
Q^2_{Boot}	average predictive power calculated by boot-strapping validation
Q^2_{ext}	explained variance in prediction calculated by external validation
R2	Coefficient of determination
R^2_{cv}	Cross-validated R^2
R^2_{adj}	Adjusted R2
RMS	Residual Mean Square
s	Standard error of estimate
SDEC	Standard Deviation Error in Calculation,
SDEP	Standard Deviation Error of Prediction
SDEPext	External Standard Deviation Error of Prediction
WHIM	Weighted Holistic Invariant Molecular descriptors

1. INTRODUCTION

The purpose of the analyses presented in this report was to contribute to an evaluation of the possibility of using QSAR predictions for regulatory purposes. To this end QSAR predictions were compared with SIDS test data. Furthermore, the models were also assessed according to the extent to which they meet OECD principles for QSAR validation (OECD ENV/JM/Mono(2004)24). It is emphasized that the comparisons are not intended to be scientific validations, because the SIDS test chemicals were not selected to ensure that they are sufficiently diverse and representative for the entire applicability domain of the individual models. Nevertheless, many of the analyses presented here form the basis for scientific validation.

1.1. Danish dataset

The "Danish dataset" (OECD ENV/JM/TG(2004)26) contains 177 SIDS test data and (Q)SAR predictions for various SIDS endpoints for these substances. The predictions in the Danish database are based on models available at the DK-EPA. The SIDS data include three selected end points:

1. Biodegradability
2. Acute toxicity to aquatic organism:
   - fish
   - algae
   - Daphnia
3. Mutagenicity

The aquatic toxicities (LC50 fish, EC50 for Daphnia and algae) are not very well defined, due to variations in test species, test method, time of exposure. Therefore, data processing was preceded by a preliminary analysis to check data consistency and to arrange data for further processing. In order to compare QSAR predictions with the SIDS test data, all the measured effect concentrations expressed as ">" were disregarded. The reason for excluding measured > values was to keep the comparison as simple as possible, even though it is recognized that a comparison of toxicity with the water solubility is important information for decision making.

1.2 Outline of the method

The work was based on the following main steps:

1. Preliminary analysis of SIDS acute fish toxicity data.
2. Generation of molecular structure files for the SIDS chemicals (Smiles, mol files), for further calculation of both two-dimensional molecular descriptors and three-dimensional descriptors. An excel file containing chemical names, CAS numbers and SMILES for 177 chemicals was kindly provided by Eva Wedebye (DK).
3. Development of a list of literature-based models to make predictions of SIDS endpoints. The focus was on models for fish toxicity.
4. Selection of transparent and reproducible models: recovery of the training set used to develop the models and checking of the test method used to generate it; identification of the molecular descriptors used and assessment of the transparency of the algorithm.

5. Estimation of predictive ability by internal validation techniques (cross-validation, bootstrap, response randomization).
6. Evaluation of QSAR applicability domains by making predictions of SIDS test data: checking the domain of applicability with respect to descriptor ranges and any structural rules defining the group of substances for which the models are valid.
7. Application of the models to the SIDS chemicals
8. Evaluation of predictive performance in terms of explained variance (Q^2_{ext}) and the prediction reliability (order of magnitude between estimated and experimental data). Predictive performance was assessed for the full set of SIDS substances, and for subsets based on different hypotheses about the applicability domain.
9. Comparative analysis of the model quality.

2. DATA SCREENING TO ASSESS VALIDITY/QUALITY OF INPUT

The SIDS fish toxicity data include short term aquatic toxicity on *Pimephales promelas* (fathead minnow) expressed as the chemical concentration at which 50% lethality is observed in a test batch of fish within a 96 h exposure period (LC50, in mg/l). 96 h LC50 (mg/l).

Prior to the main analysis, SIDS data were analyzed to evaluate the effects they could have upon the results. Screening of the input data helped assess the appropriateness of the using the SIDS data set, by identifying data peculiarities and adjusting data in advance of the further multivariate analysis.
The following sequence for screening has been performed:

1. SIDS toxicity data selection
2. Univariate descriptive statistics for accuracy of input
   a) check skewness and kurtosis
   b) variable transformation (if desirable)
   c) check results of transformations

2.1 SIDS toxicity data selection

The experimental toxicity values were available for 32 SIDS chemicals; interval values were provided for 4 chemicals and open intervals (>) for 6 chemicals. All the measured effect concentrations expressed as ">" were disregarded, since these values were difficult to compare with QSAR predictions.

In order to provide a deeper and more realistic further evaluation/validation of the selected models the AQUIRE (**AQU**atic toxicity **I**nformation **RE**trieval) database developed by the U.S. EPA Mid-Continent Ecology Division, Duluth, MN (MED-Duluth) (http://www.epa.gov/ecotox/) was investigated to fill in the experimental missing values of the SIDS data.

The AQUIRE database provided experimental toxicity values of 25 SIDS missing values. Since the database gave more than one value for each chemical the average value was used to fill in the data gaps. Thus the final integrated SIDS dataset was made of 57 experimental toxicity data out the 177 SIDS chemicals. The 177 SIDS chemicals investigated in this study, their toxicity in

terms of LogLC50(mol/l), their logKow values and their mechanism of action are listed in Table I.

2.2 Univariate descriptive statistics for accuracy of input

Simple descriptive statistics analysis was performed for testing the shape of the experimental toxicity distribution, looking at the frequency of values from different toxicity ranges to see how well the distribution could be approximated by the normal distribution. Screening for normality was performed by examining skewness and kurtosis.
Skewness is a measure of the asymmetry of the data around the sample mean. A positive skewness reveals that the data are spread out more to the right. The skewness of the normal distribution (or any perfectly symmetric distribution) is zero. Kurtosis is a measure of how "peaked" outlier-prone a distribution is. The kurtosis of the normal distribution is 3. A kurtosis greater than 3 characterizes distributions that are more outlier-prone than the normal one.
The skewness (which measures the deviation of the distribution from symmetry) and the kurtosis (which measures "peakedness" of the distribution) reveal the non-normality of the data.

Valid N	Mean	Minimum	Maximum	Std. Dev	Skewness	Kurtosis
57	0.0132	0.000001	0.230803	0.0413	3.9768	16.3988

LC50 (mol/l) statistics.

Moreover a visual examination of the data using a histogram (i.e., a graph that shows the frequency distribution of a variable) was inspected.

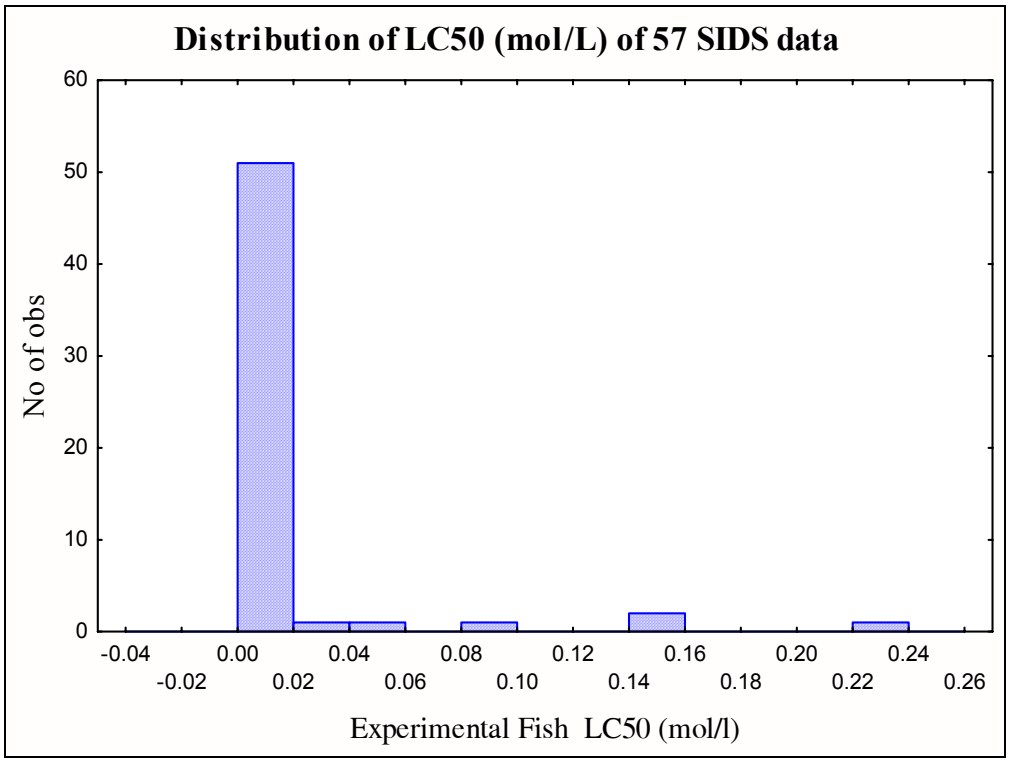


Figure 1 - Histogram of the non transformed LC50 (mol/l).

These data appear to be seriously non-normal. They are heavily asymmetric, right skewed, with a large mode near zero and some data to the right.
A simply logarithmic (Log) transformation was applied to remedy for outliers and failure of normality and obtain an approximately normal distribution.
Data were transformed in Log LC50 (mol/l).

Valid N	Mean	Minimum	Maximum	Std. Dev	Skewness	Kurtosis
57	-3.30661	-6.23592	-0.63676	1.2599	-0.08133	0.03962

Log LC50 (mol/l) statistics.

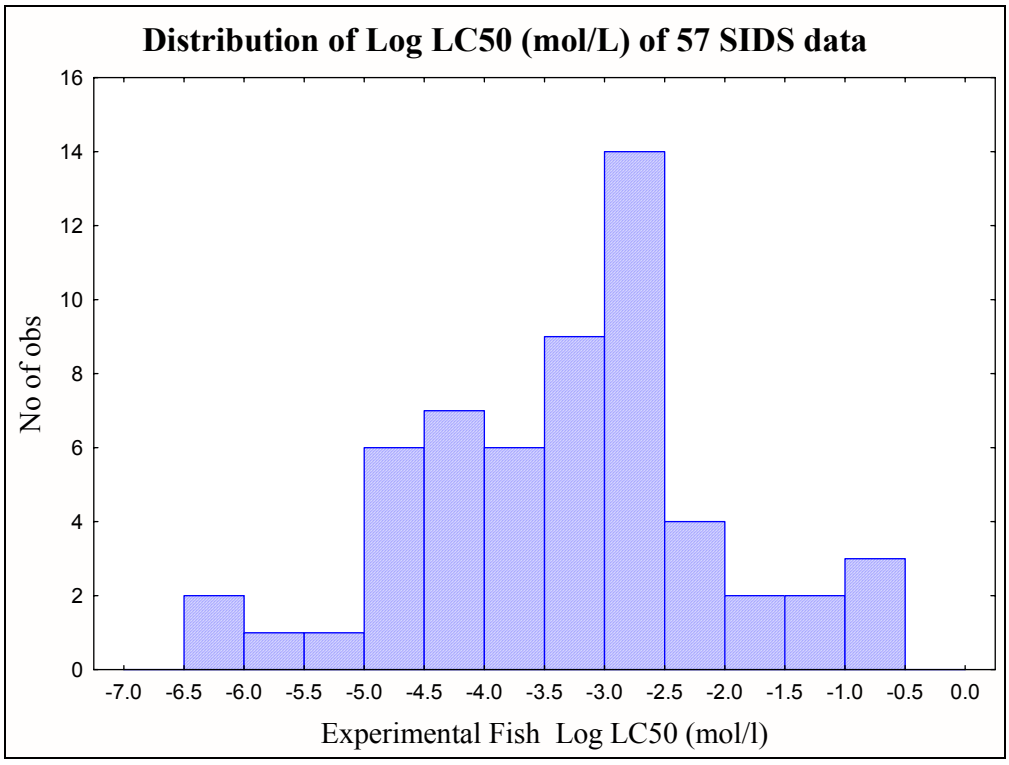


Figure 2 - Histogram of the transformed LogLC50 (mol/l).

2.3 Identification of SIDS mechanism of action

Quantitative structure-activity relationships (QSARs) rely on the paradigm that chemicals belonging to the same or similar chemical classes behave in a similar manner. In the field of aquatic toxicology, it is widely agreed that the QSARs are valid and suitable for prediction within the same applicability domain, i.e. for the same mechanism of toxic action (MOA). The aim of this study was to perform consensus classification according to MOA of the 177 SIDS chemicals.
For this purpose four classification schemes were compared. The first ("ECB") classification scheme was applied in-house and was used to classify chemicals into seventeen MOA. The second classification was done by an expert and included a similar number of mechanisms. The third classification scheme was provided by the ASTER (ASsessment Tools for the Evaluation of Risk) expert system classification. ASTER is an expert system developed by the U.S. EPA Mid-

Continent Ecology Division, Duluth, MN (MED-Duluth) to assist regulators in performing ecological risk assessments. ASTER is an integration of the AQUIRE (**AQU**atic toxicity **I**nformation **RE**trieval) toxic effects database and a structure activity based expert system. When empirical data are not available mechanistically-based predictive models are used to estimate ecotoxicology endpoints, chemical properties, biodegradation, and environmental partitioning (Russom, C.L., *et al.* 1997). ASTER was able to classify 176 SIDS chemicals out of 177 according to 16 diverse mechanisms of action. The fourth classification scheme is the well known proposed by Verhaar (Verhaar, H.J.M. *et al*. 1992; Verhaar, H.J.M. *et al*. 2000). This scheme provides a simple classification, based on only four modes of action; moreover, classifications were provided only for 89 SIDS chemicals.

The four classification schemes and the corresponding number of mechanisms of action identified together with the number of chemicals classified are listed below:

Classification scheme	*N. MOA*	*N. SIDS classified*	*Reference*
ECB	17	161	European Commission - Joint Research Centre Institute for Health and Consumer Protection QSAR (European Chemicals Bureau)
Expert - Schultz	20	177	
EPA-ASTER	17	176	U.S. Environmental Protection Agency, Office of Research and Development, National Health and Environmental, Duluth, Minnesota
Verhaar	4	89	Verhaar, H.J.M. *et al*. 1992; Verhaar, H.J.M. *et al*. 2000

Comparing the first three classification schemes (ECB, Schultz, EPA-ASTER) a consensus classification (CONS1) was achieved comprising nine MOA. The consensus was based on a majority principle according to which each chemical has been classified belonging to the class most represented among the classifications compared. No classification was provided for those chemicals on which the three classification schemes were in disagreement. In this consensus the Verhaar classification was not considered being too simple with respect to the others and providing a classification for relatively few chemicals. However, a second and simpler consensus scheme (CONS2) based on 4 classes was determined and compared with the one of Verhaar.

ECB- MOA	Description	N.Chemicals
AChE	acetylcholinesterase (AChE) inhibition	3
AN	Amine narcosis	2
CNS	Central nervous system seizure action	5
EN	Ester narcosis	7
ISOCYA	Isocyanate based reactivity	1
MTA	Michael-type addition	25
NPN	Non polar narcosis	48
NPN_log D	Non polar narcosis based on log D	12
NUC	Nucleophile reaction	1
PE	Electrophile and proelectrophile reactivity	4
PE_RAD	Proelectrophile radical reaction	1
PN	Polar narcosis	39
PN_log D	Polar narcosis based on LogD	5
RAD	Radical reaction	2
SB	Schiff-base formation	1
SN2	SN2 reaction	4
WARE	Weak Acid Respiratory Uncoupler	1
UNK	Unknown mode of action	16

Schultz - MOA	Description	N.Chemicals
AMIN.ALCH	Aminoalcohol	1
CARB. ACID	Carboxylic acid	12
CNS	Central nervous system seizure agent	1
DICARB. ACID	Dicarboxylic acid	3
EPOX.	Epoxide	2
MTA	Michael-type addition	17
NON SPEC. ELECT	Non specific electrophile	5
NPN	Non polar narcosis	81
NTAS	Not toxic at saturation	1
PE	Electrophile and proelectrophile reactivity	4
PN	Polar narcosis	12
REAC.	Reactive	16
REAC. ACID	Reactive acid	4
REAC. HYD	Reactive hydrolysis	3
REAC. NON SPEC.	Non-specific reactivity	2
REAC. PHOSP.	Phosphoric reactive	1
SB	Schiff-base formation	4
SN2	SN2 reaction	5
SOFT ELECT	Soft Electrophile	2
STRONG ACID	Strong acid	1

EPA-ASTER- MOA	Description	N.Chemicals
ACRY	Acrylate toxicity	6
ACY	Acylation based reactivity	1
ALKY-ARYL	Alkylation / arylation based reactivity	13
CARB. Based	Carbonyl based reactivity	1
CARB. REAC.	Carbonyl reactivity (aldehyde eq. # 3)	2
CNS	Central nervous system seizure agent	1
DE	Diester toxicity	7
EN	Ester narcosis	11
ISOCYA	Isocyanate based reactivity	1
NPN	Non polar narcosis	108
OP-AChE	Organophosphate mediated AChE inhibition	1
PN	Polar narcosis	14
REAC.	Reactive	1
REAC.DIKE	Reactive diketone	3
REAC.DINITRO	Reactive dinitroaromatic group	3
SULPHY	Sulphydryl based reactivity	2
UNCOUPL	Uncoupler of oxidative phosphorylation	1
UNK	Unknown mode of action	1

Verhaar - MOA	Description	N.Chemicals
NPN	Non polar narcosis	30
PN	Polar narcosis	12
REAC.	Reactive	42
R/S	Reactive and specifically acting chemicals	5

CONS1- MOA	Description	N.Chemicals
AChE	Acetylcholinesterase (AChE) inhibition	1
CNS	Central nervous system seizure action	2
EN	Ester narcosis	5
MTA	Michael-type addition	16
NPN	Non polar narcosis	75
PE	Electrophile and proelectrophile reactivity	2
PN	Polar narcosis	12
SB	Schiff-base formation	1
SN2	SN2 reaction	1
UNK	Unknown mode of action	62

CONS2- MOA	Description	N.Chemicals
N	Narcosis	97
N*	Narcosis modeled by LogD	18
R	Reactive	44
S	Specifically acting	3
UNK	Unknown mode of action	15

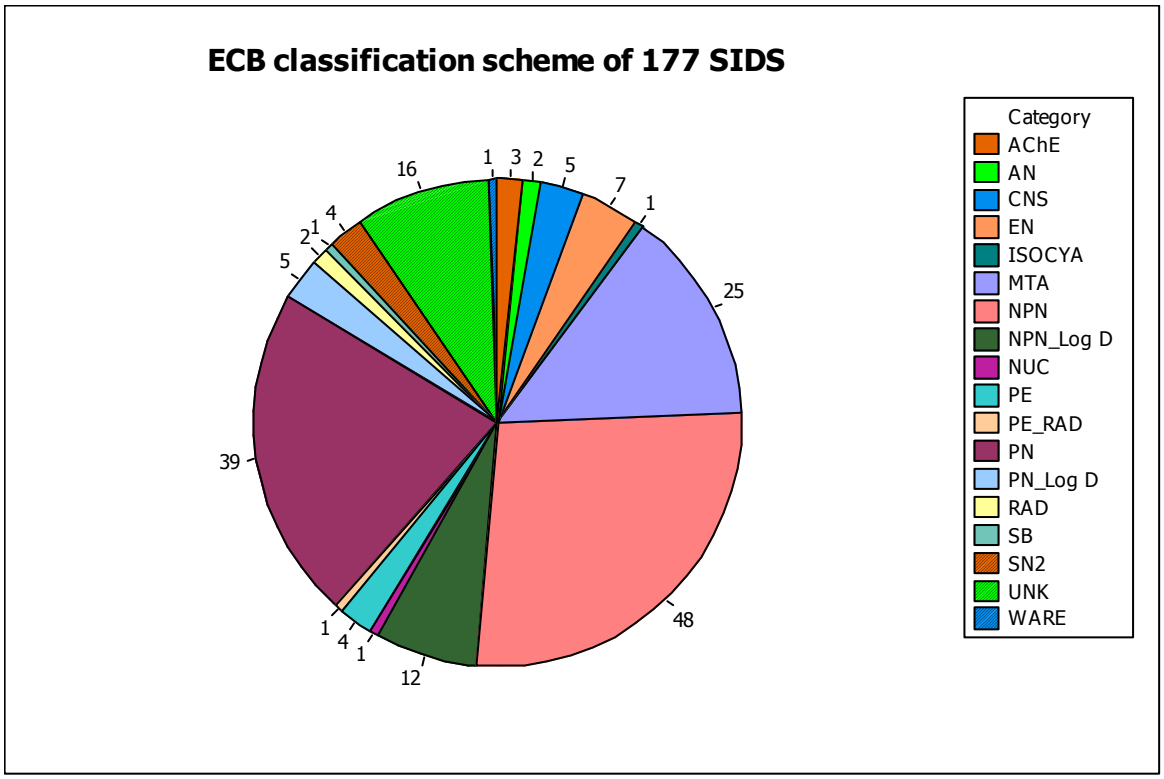


Figure 3 - ECB classification scheme chart.

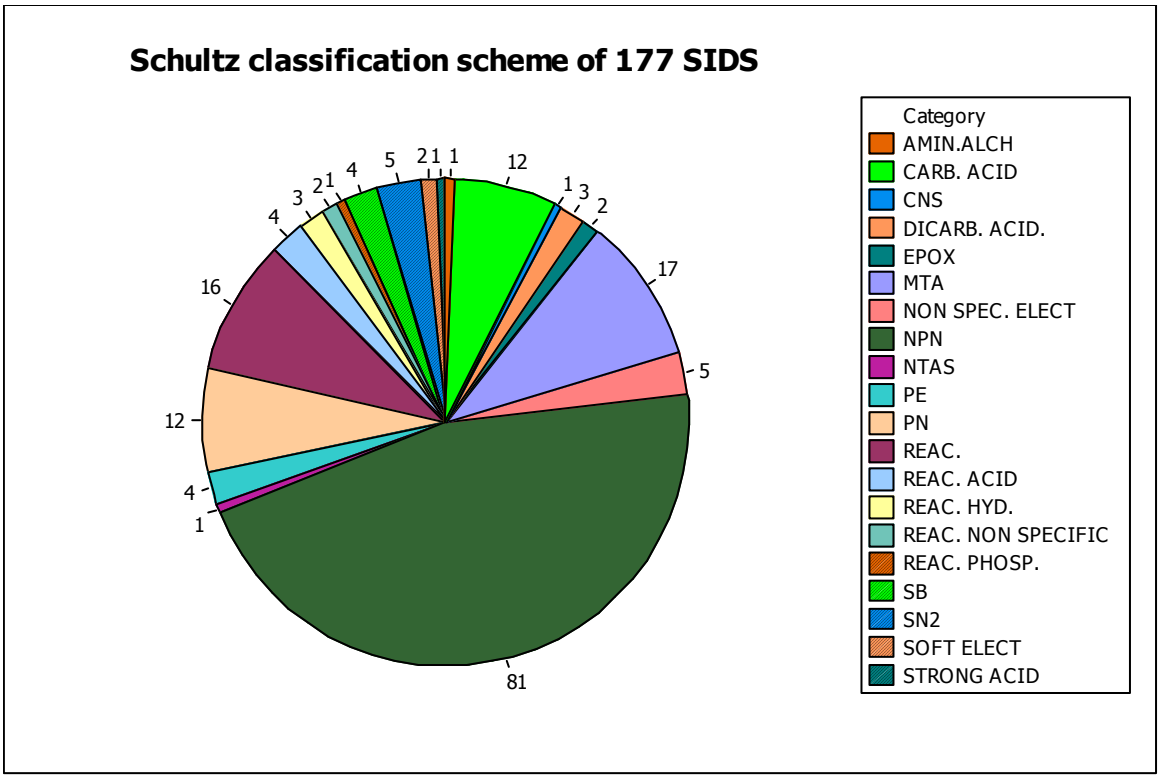


Figure 4 - Schultz classification scheme chart.

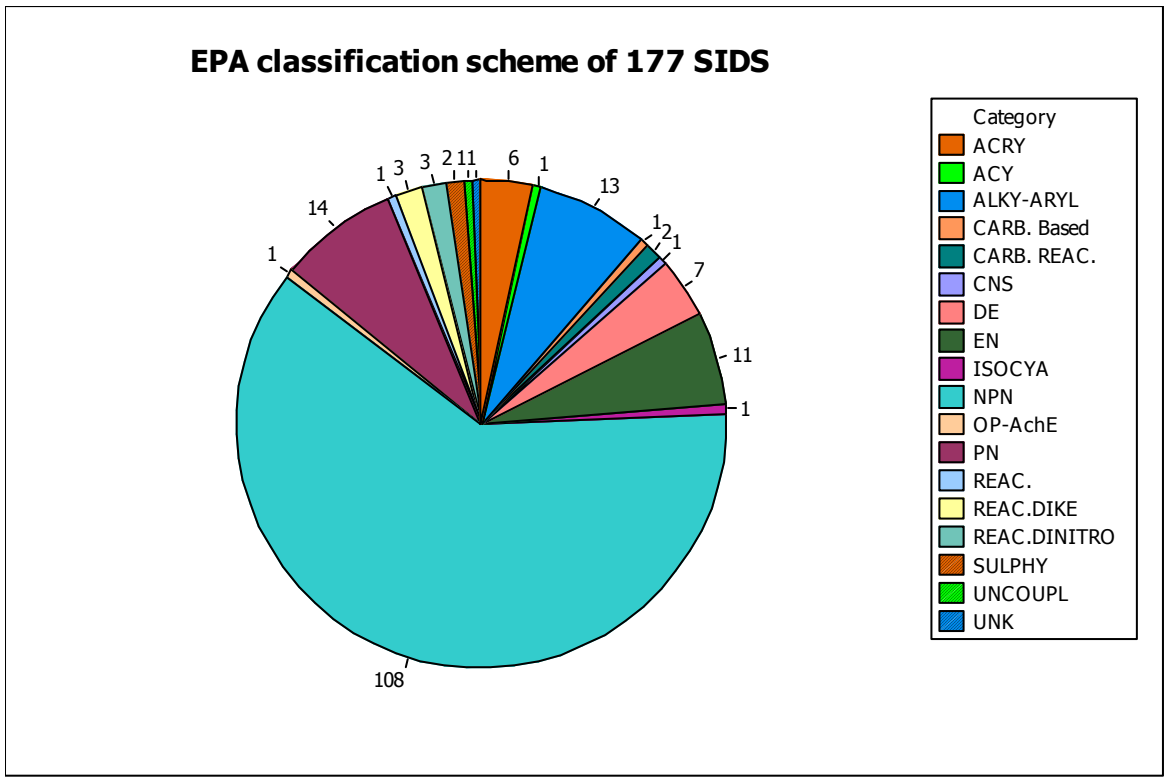


Figure 5 - EPA-ASTER classification scheme chart.

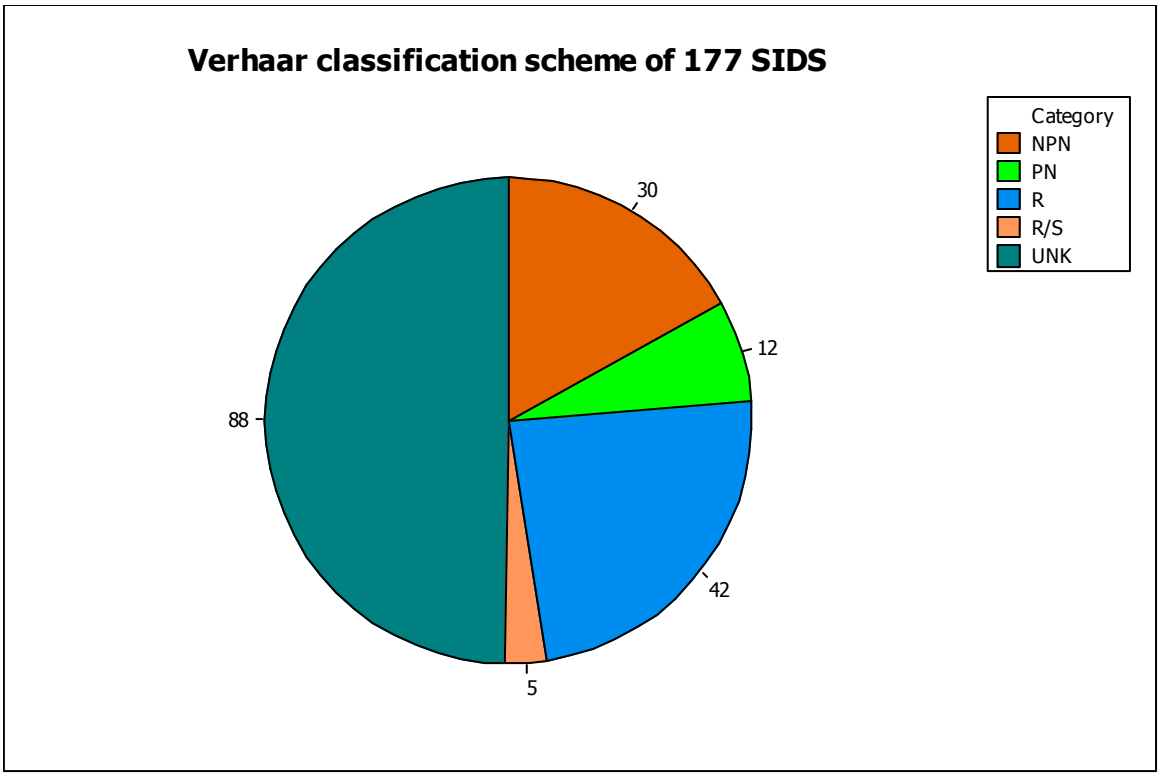


Figure 6 - Verhaar classification scheme chart.

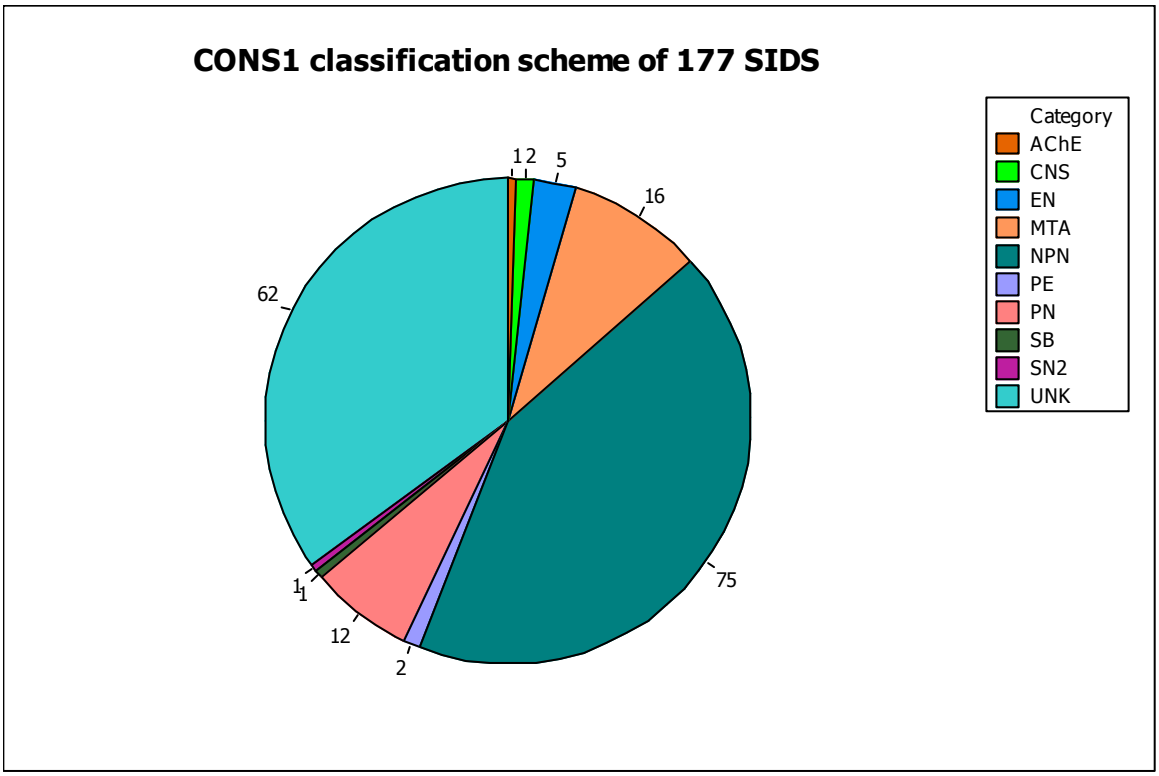


Figure 7 - Consensus1 classification scheme chart.

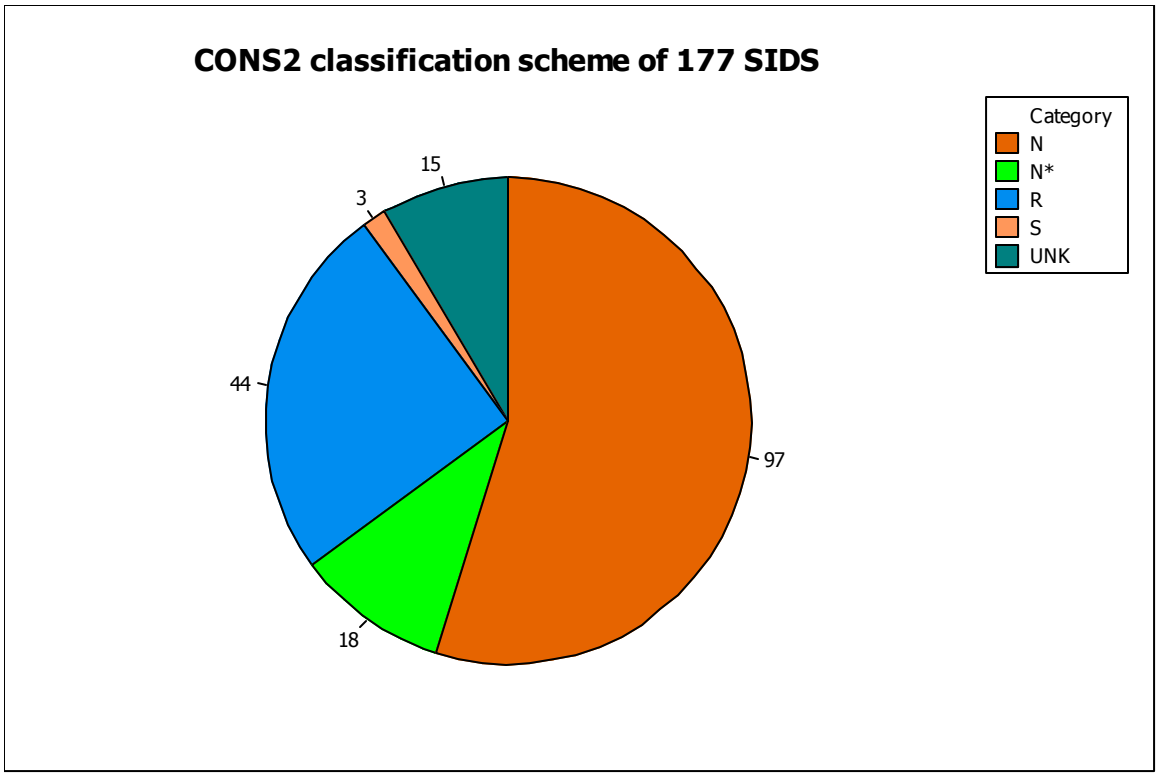


Figure 8 – Consensus2 classification scheme chart.

2.4 SIDS LogKow distribution

The 177 SIDS data have logKow values in the range from -3.89 to 18.08 with 29 chemicals exhibit logKow values lower than 0, while 13 greater than 6.

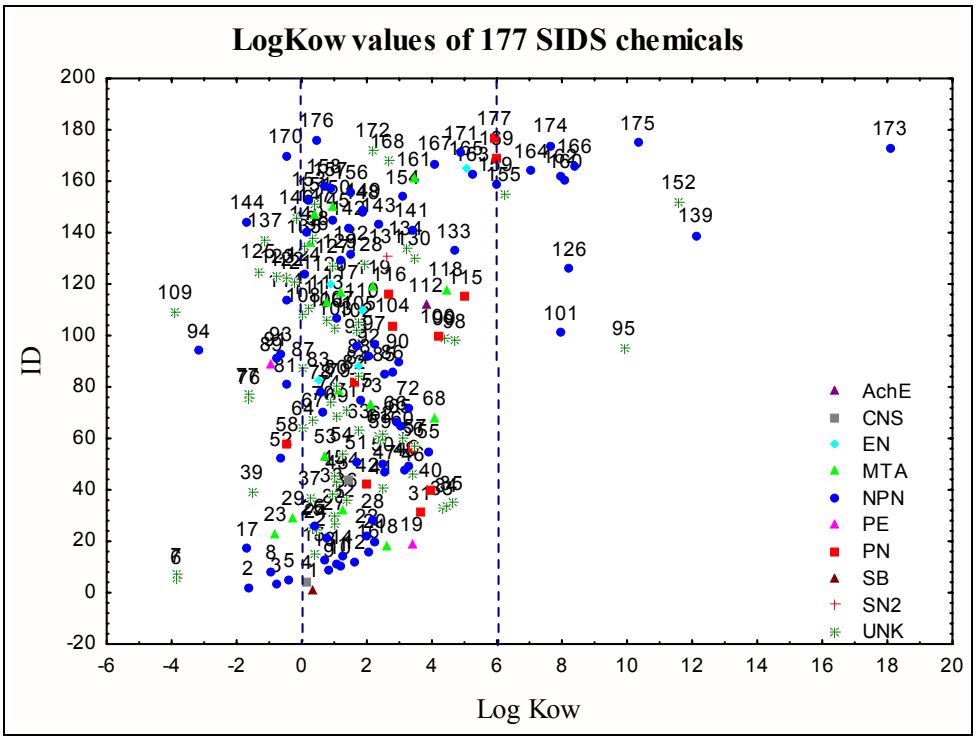


Figure 9 - LogKow values of 177 SIDS chemicals.

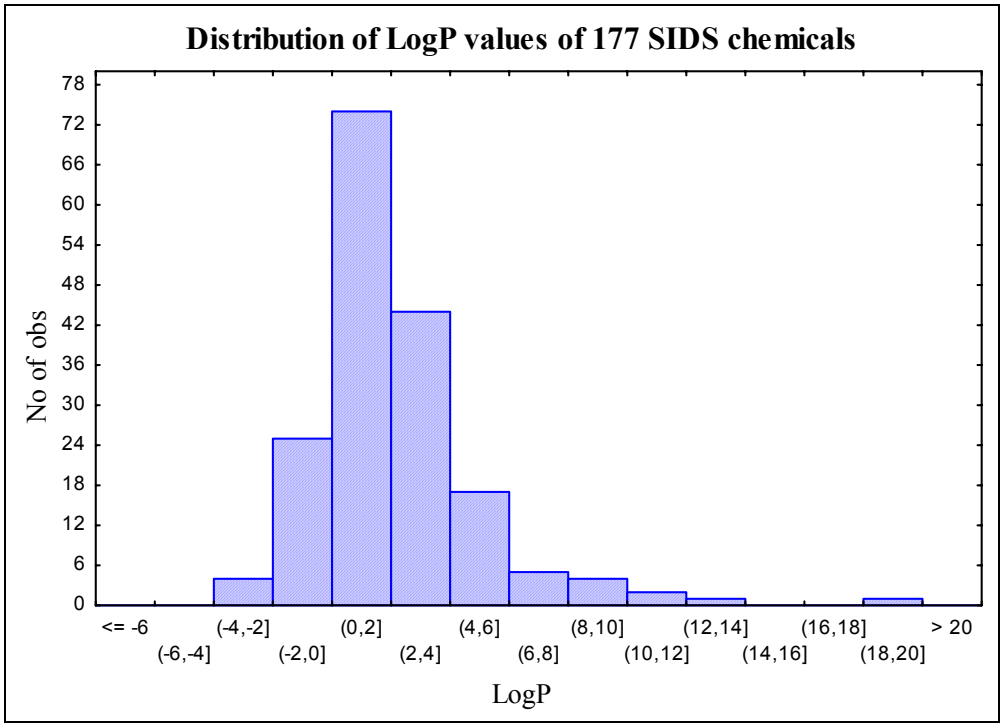


Figure 10 - Distribution of the LogKow values of 177 SIDS chemicals.

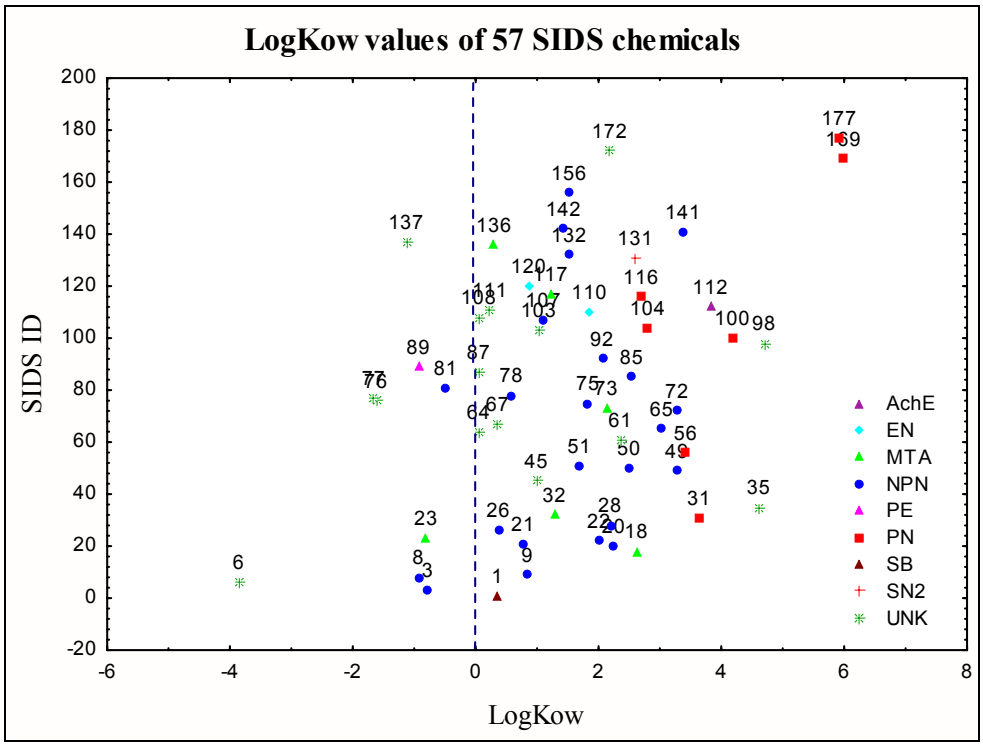


Figure 11 - LogKow values of 57 SIDS chemicals.

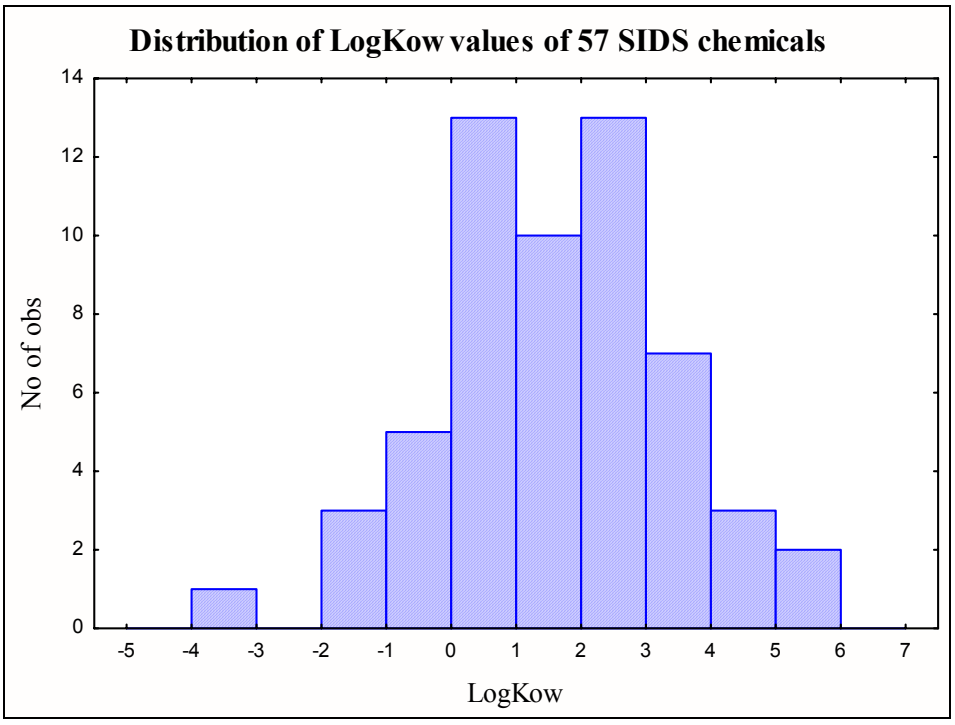


Figure 12 - Distribution of the LogKow values of 32 SIDS chemicals.

A plot of toxicity against logKow for the 57 chemicals shows a baseline effect: within this group 23 compounds were identified as non polar narcotics falling on the baseline, according to their lipophilicity:
1,2-Propanediol (3), Formamide, N,N-dimethyl- (8), 1-Butanol (9), 1,2-dichloro-propane (20), 2-Butanol (21), Ethane, 1,1,2-trichloro- (22), Acetic acid, methyl ester (26), 1,1,2,2-tetrachloro-ethane (28), Benzene, 1,2-dichloro (49), 1,2,3-trichloro-propane (50), 2-Butanone, oxime (51), Benzene, ethyl- (65), 1,4-dichloro-benzene (72), 1,2-dichloro-ethane (75), 2,4-Pentanediol, 2-methyl- (78), 1-methoxy-2 propanol (81), Benzene, methyl- (85), 6-methyl-5-Hepten-2-one (92), Ethanol, 2-phenoxy (107), 2-Propanol, 1-phenoxy (132), 5-methyl-2-(1-methylethyl)-Cyclohexanol (141), Propane, 2-methoxy-2-methyl- (142), 1-Propanol, 2-phenoxy- (156).

The relationship defining the baseline toxic effect is defined by the following model based on the subgroup of 23 compounds:

$$\text{Log}(1/\text{LC50}) = 0.804\ \text{LogKow}\ (\pm 0.073) + 1.317\ (\pm 0.145)$$

$$n = 23\ \ R^2 = 85.22 \qquad Q^2_{LOO} = 82.71\ \ s = 0.428 \qquad F = 121.09$$

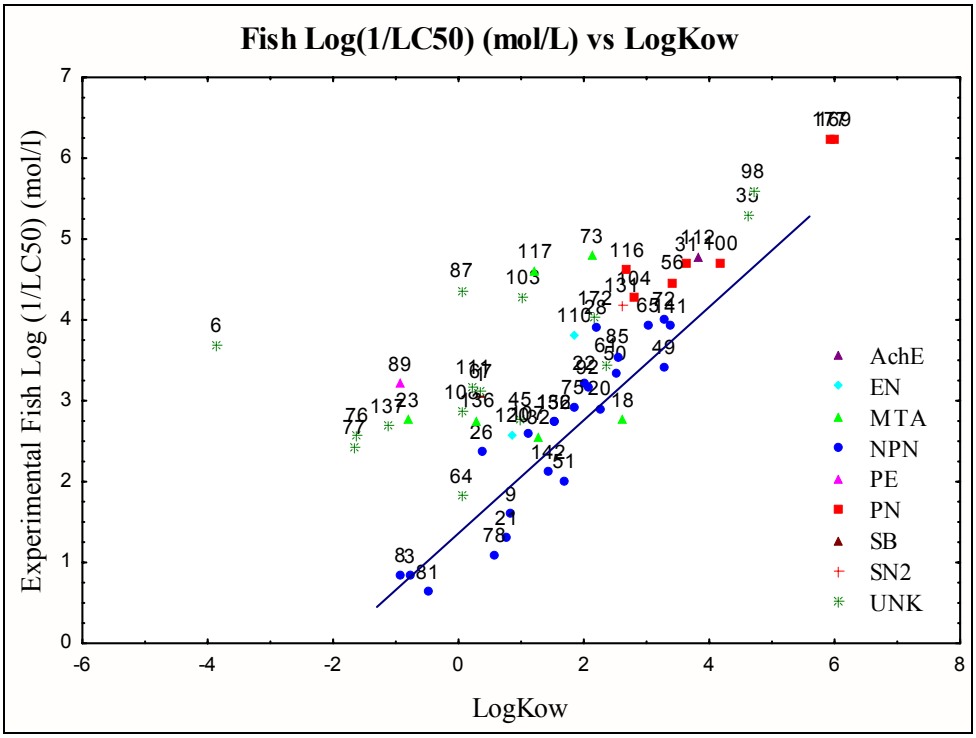


Figure 13 - Log(1/LC50) (mol/l) vs LogKow. Compounds represented by blue points were used to define the baseline (solid line).

A refinement of the training set, excluding the 6 chemicals with a residual in prediction greater than Standard Deviation Error of Prediction (SDEP) provided the following baseline model:

$$\text{Log}(1/\text{LC50}) = 0.810\ \text{LogKow}\ (\pm 0.047) + 1.362\ (\pm 0.095)$$

$$n = 17\ \ R^2 = 95.27 \qquad Q^2_{LOO} = 93.94\ \ s = 0.246 \qquad F = 301.94$$

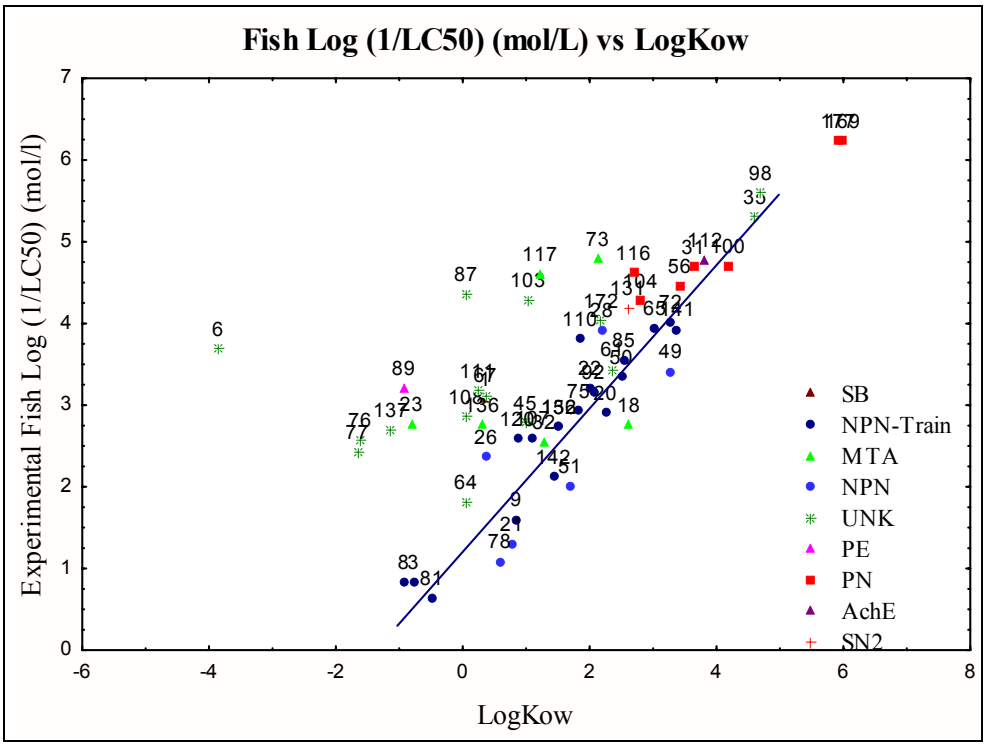


Figure 14 - Log(1/LC50) (mol/l) vs LogKow. Compounds represented by dark blue points were used to define the baseline (solid line).

Within the 57 chemicals 23 compounds were classified as non polar narcotics (NPN), 7 as polar-narcotics (PN), 2 as ester narcotics (EN), 1 as acetylcholinesterase (AChE) inhibitor, 6 as Michael-type reactive (MTA), 1 as electrophile reactor (PE), 1 as Schiff-base reactive, 1 as SN2 reactor, 7 with an unknown mechanism.

CONS1- MOA	Description	N.Chemicals
AChE	acetylcholinesterase (AChE) inhibition	1
EN	Ester narcosis	2
MTA	Michael-type addition	6
NPN	Non polar narcosis	23
PE	Electrophiles and proelectrophile reactivity	1
PN	Polar narcosis	7
SB	Schiff-base formation	1
SN2	SN2 reaction	1
UNK	Unknown mode of action	15

3. SIDS MOLECULAR STRUCTURE FILES

An excel file containing chemical names, CAS numbers and SMILES for 177 SIDS chemicals was kindly provided by Eva Wedebye (DK). The two dimensional structures of SIDS data are collected in Appendix I. The Corina program [Corina software, 2005] was used to create 3D models directly from SMILES strings. Energy optimization of the 3D structures was performed by COSMIC. Molecular geometries were optimized by the Vamp semiempirical molecular orbital package. Total energy, heat of formation, HOMO and LUMO eigenvalues, ionization potential and total dipole were calculated.

Molecular structure files for the SIDS chemicals (Sdf, mol files) were generated for further calculation of both two-dimensional molecular descriptors and three-dimensional descriptors.
The chemical structures of the chemicals were described with more than 1500 molecular descriptors, in order to catch all the structural information. The molecular descriptors were calculated by the DRAGON software [Todeschini *et. al*., 2004] on the basis of the molecular geometry optimization performed by Vamp package [TSAR]. In this study the following sets of molecular descriptors have been used: constitutional descriptors, topological descriptors [Bonchev, 1983; Devillers and Balaban, 2000], WHIM descriptors [Todeschini *et al*., 1994; Todeschini and Gramatica, 1997], GETAWAY descriptors [Consonni *et al*., 2002]. The complete list of the descriptors used together with their symbol, meaning is provided in Appendix II.

3.1 Structure similarity analysis of SIDS data by Principal Component Analysis (PCA)

Structural similarity analysis was performed on 177 chemicals described by 1500 theoretical molecular descriptors. Principal Component Analysis (PCA) was used to identify the orthogonal directions of maximum variance in the original data and to project the data into a lower-dimensionality space formed by a subset of the highest-variance components. The aim of this analysis was to obtain preliminary information on structural similarities and dissimilarities on SIDS test data. The analysis has been performed on subset of molecular descriptors.

The Hotelling T2 control chart was used to evaluate how far away each chemical was from the PC model hyper plane. The Hotelling T2 ellipse was computed with a 0.05 (95% confidence) significance level.

- PCA on 48 constitutional descriptors

  Constitutional descriptors are the most simple and commonly used descriptors, reflecting the molecular composition of a compound without any information about its molecular geometry. The list of the constitutional descriptors, with symbols and meaning is provided in Appendix II.

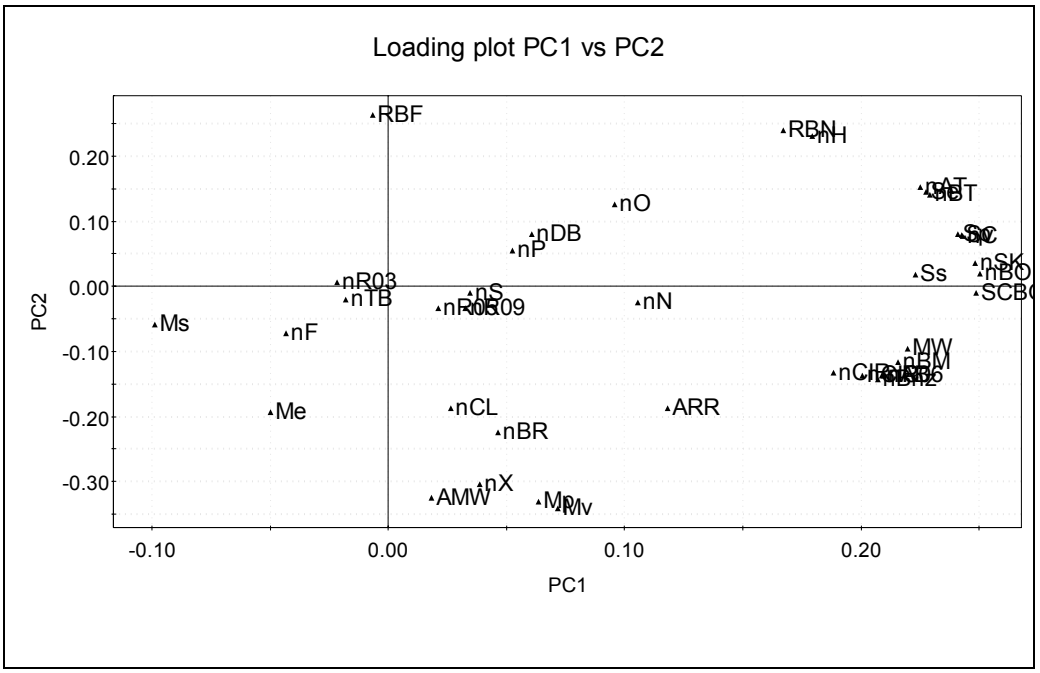


Figure 15 - Loading plot of PC1 *vs* PC2 calculated from constitutional descriptors.

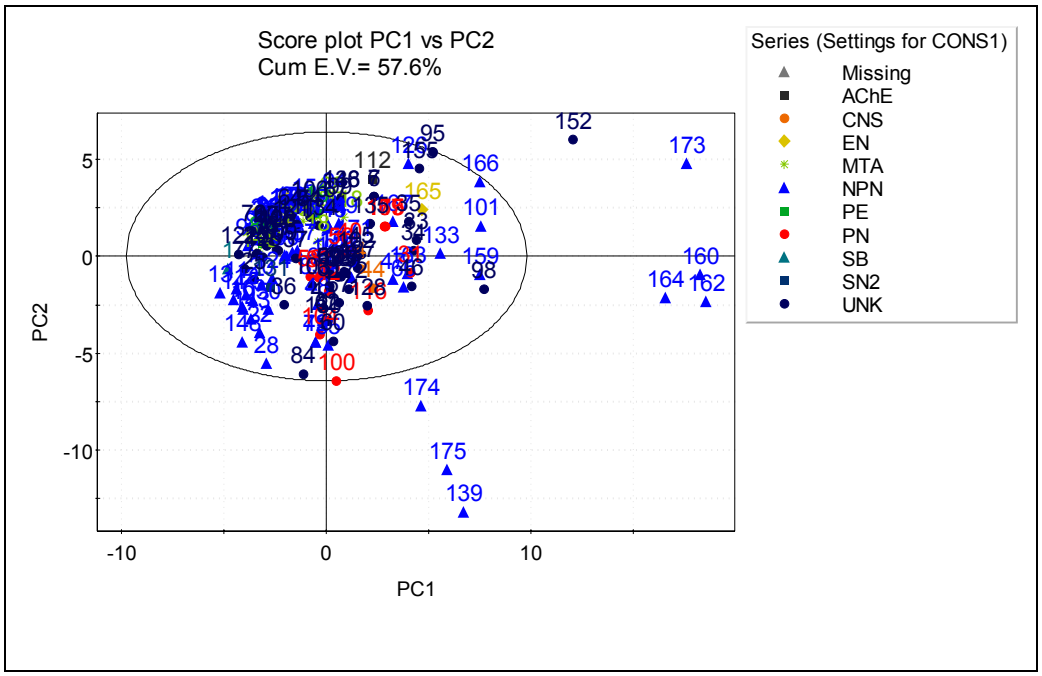


Figure 16 - Score plot of PC1 *vs* PC2 calculated from constitutional descriptors.

The analysis performed highlights a close group of chemicals and a few chemicals which are far apart the others: 1,1'-oxybis[2,3,4,5,6-pentabromo-benzene (139), 1,2,4-Benzenetricarboxylic acid, tris(2-ethylhexyl) ester (152), butanamide, 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[N-(2,4-dimethylphenyl)-3-oxo- (160), butanamide, 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[N-(4-chloro-2,5-dimethoxyphenyl)-3-oxo- (162), butanamide, 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[3-oxo-N-phenyl- (164), 2,4-bis(1,1-

dimethylethyl)- phenol (173), benzene, 1,1'-oxybis-pentabromo deriv (174), benzene, 1,1'-oxybis-octabromo (175). From the loading plot it can be observed that among the SIDS data set, these chemicals are the ones characterized by the highest number of non-hydrogen atoms/bonds, highest sum of conventional bond orders and highest molecular weight.

- PCA on 119 topological descriptors

  Topological descriptors are based on a graph representation of the molecule and quantify the molecular topology obtained by the application of algebraic operators to matrices representing molecular graphs and whose values are independent of vertex numbering or labeling. They can be sensitive to one or more structural features of the molecule such as size, shape, symmetry, branching and cyclicity and can also encode chemical information concerning atom type and bond multiplicity.

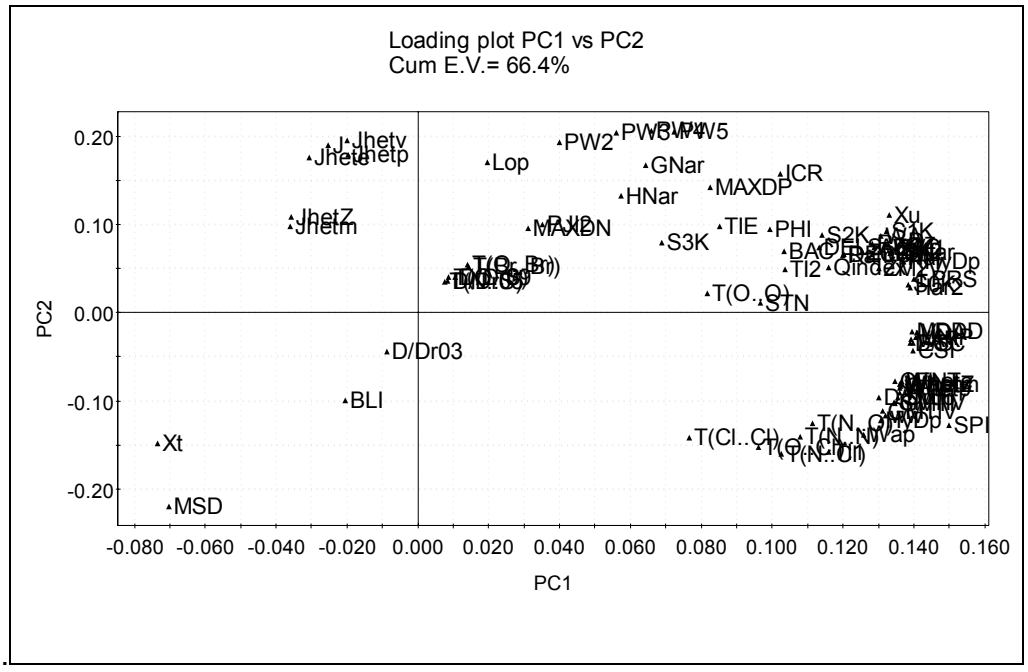


Figure 17 - Loading plot of PC1 *vs* PC2 calculated from topological descriptors.

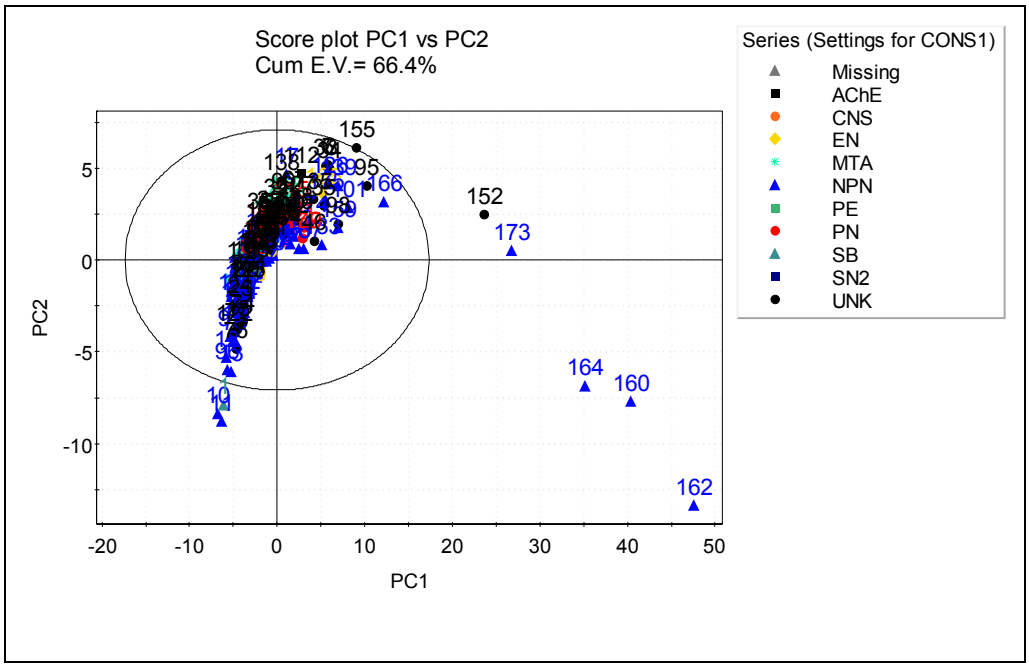


Figure 18 - Score plot of PC1 *vs* PC2 calculated from topological descriptors.

According to the topological representation of the SIDS chemical structure the following chemicals appear different from the others: formaldehyde (1), bromo-methane (10), chloro-methane (11), 1,2,4-benzenetricarboxylic acid, tris(2-ethylhexyl) ester (152) and butanamide, 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[N-(2,4-dimethylphenyl)-3-oxo-butanamide (160), 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[N-(4-chloro-2,5-dimethoxyphenyl)-3-oxo-butanamide (162), 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[3-oxo-N-phenyl- (164), 2,4-bis(1,1-dimethylethyl)-phenol (173).

- <u>PCA on 99 WHIM descriptors and 197 GETAWAY descriptors</u>

  WHIM descriptors (Weighted Holistic Invariant Molecular descriptors) are geometrical descriptors based on statistical indices calculated on the projections of the atoms along principal axes. They are built in such a way as to capture relevant molecular 3D information regarding molecular size, shape, symmetry and atom distribution with respect to invariant reference frames. They are divided into two main classes: directional WHIM descriptors and global WHIM descriptors. The GETAWAY (GEometry, Topology, and Atom-Weights AssemblY) descriptors are chemical structure descriptors encoding the molecule three dimensional information derived from a new representation of molecular structure, the Molecular Influence Matrix (MIM).

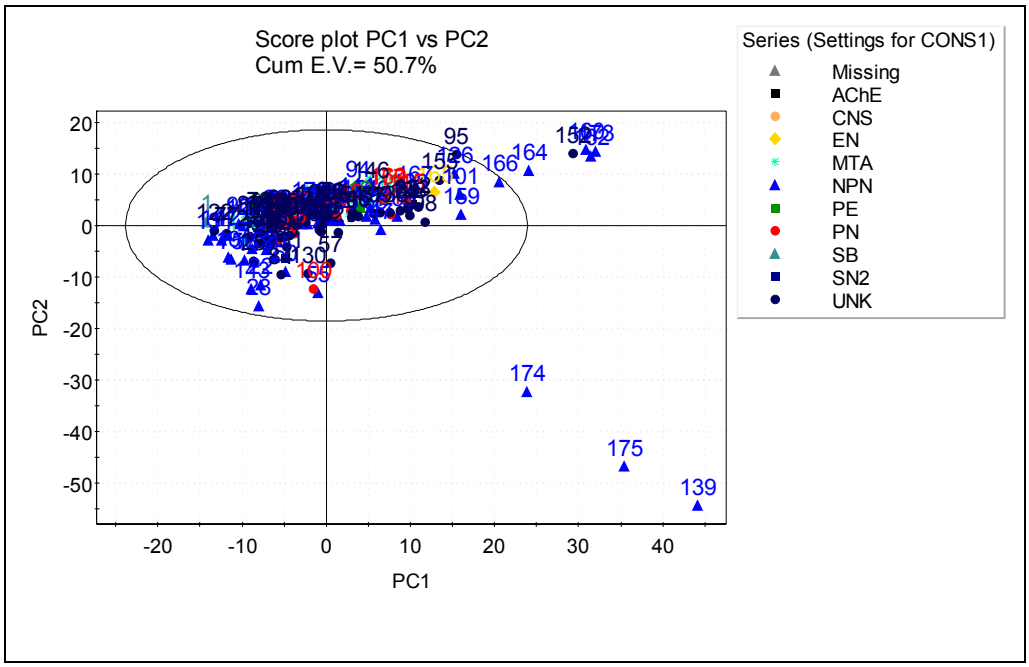


Figure 19 - Score plot of PC1 *vs* PC2 calculated from WHIM and GETAWAY descriptors.

The analysis performed highlights a close group of chemicals and a few chemicals which are far apart the other: 1,1'-oxybis[2,3,4,5,6-pentabromo-benzene (139), 1,2,4-Benzenetricarboxylic acid, tris(2-ethylhexyl) ester (152), butanamide, 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[N-(2,4-dimethylphenyl)-3-oxo- (160), butanamide, 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[N-(4-chloro-2,5-dimethoxyphenyl)-3-oxo- (162), butanamide, 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[3-oxo-N-phenyl- (164), phenol, 2,4-bis(1,1-dimethylethyl)- (173), 1,1'-oxybis-, pentabromobenzene (174), 1,1'-oxybis-, octabromobenzene (175).

- <u>PCA on constitutional, topological, WHIM, GETAWAY</u>

  The principal component analysis developed on 0D-2D-3D descriptors confirms the previous results identifying a close group of chemicals and a few chemicals which are far apart the others: 1,1'-oxybis[2,3,4,5,6-pentabromo-benzene (139), 1,2,4-benzenetricarboxylic acid, tris(2-ethylhexyl) ester (152), 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[N-(2,4-dimethylphenyl)-3-oxo- (160), butanamide, 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[N-(4-chloro-2,5-dimethoxyphenyl)-3-oxo- (162), butanamide, 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[3-oxo-N-phenyl- (164), 1,4-benzenedicarboxylic acid, 2,4-bis(1,1-dimethylethyl)- (173), pentabromobenzene (174), 1,1'-oxybis-, octabromobenzene (175).

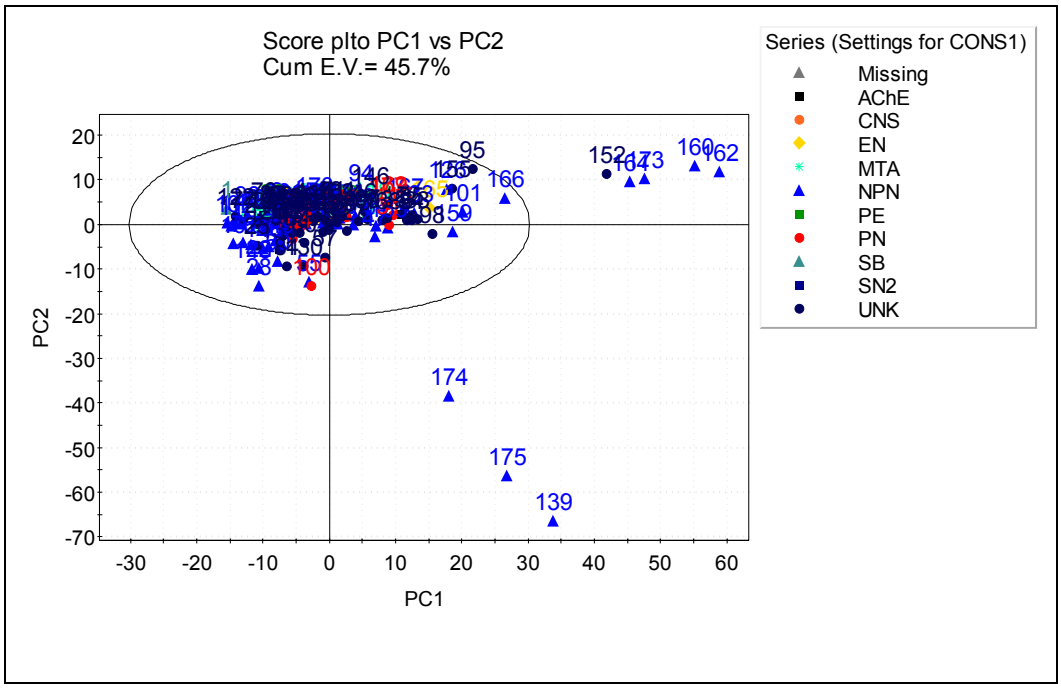


Figure 20 - Score plot of PC1 *vs* PC2 calculated from constitutional, topological, WHIM and GETAWAY descriptors.

4. SELECTION OF LITERATURE-BASED QSAR MODELS FOR NARCOSIS TO PREDICT FISH TOXICITY

The following three QSAR models for narcosis for acute fish toxicity on *Pimephales promelas* were analyzed with respect to their predictive capability on SIDS data set:

- **QSAR 1: non – polar narcosis:** Veith, GD, Call, DJ and Brooke, LT. (1983). Structure-toxicity relationships for the fathead minnow, *Pimephales promelas*: Narcotic industrial chemicals. *Canadian Journal of Fisheries and Aquatic Sciences*. 40, 743-748. Published by the European Commission (European Commission, 1995) and recommended for use in the European Union Technical Guidance Document (European Economic Community 1996).
- **QSAR 2 polar narcosis:** Verhaar, H.J.M., Mulder, W., Hermens, J.L.M. (1995). QSARs for ecotoxicity. In Overview of structure-activity relationships for environmental endpoints, Part I: general outline and procedure. Hermens, J.L.M. (ed), Report in QSAR for Predicting Fate and Effects of Chemicals in the Environment, Final Report of DG XII Contract No. EV5V-CT92-0211 (available at http://ecb.jrc.it/QSAR/).
- **QSAR 3 narcosis model:** developed by ECB by combining the training sets of the two above models.

The first two models represent QSARs for two very well known mechanisms of action: non-polar narcosis (QSAR1) and polar narcosis (QSAR2). The third model developed by ECB is intended to represent the narcosis mechanism of action, including non-polar and polar action.
Each model was analyzed for its correspondence with the OECD principles and for its capability to provide reliable predictions of the fish toxicity of the SIDS chemicals.

5. NON-POLAR NARCOSIS QSAR1 EVALUATION

5.1 Defined endpoint and algorithm

This QSAR developed for predicting acute toxicity of organic chemicals to the fathead minnow is recommended for use in the European Union Technical Guidance Document (European Economic Community 1996).
The model is:

$$\text{LogLC50} = -0.846\ \text{LogKow} - 1.390$$

Where LC50 is the concentration (in moles per litre) causing 50% lethality in *Pimephales promelas*, after an exposure of 96 hours, and Kow is the octanol-water partition coefficient.

The regression coefficients and the intercept of the above equation were not reproducible by OLS. The new OLS equation, recalculated on the molecular descriptors selected by the authors, is:

$$\text{LogLC50} = -0.862\ \text{LogKow} - 1.330$$

5.2 Mechanistic basis

The model was developed for chemicals acting as non-polar narcotics, as defined by Verhaar (Verhaar *et al.*, 1992). The QSAR is based on a single descriptor for hydrophobicity (LogKow), which is relevant to the mechanism of action which consists in accumulation of molecules in biological membranes.

5.3 Domain of applicability

The QSAR model was defined by the developer to be applicable to chemicals with log Kow values in range from -1.24 to 5.13, and exhibiting a non polar narcosis mechanism of action. Thus the structural domain includes aliphatic and aromatic hydrocarbons, halogenated aliphatic and aromatic hydrocarbons, ethers, alcohols.
The domain of applicability has been verified by the leverage approach, which provides a measure of the distance between the descriptor values for a chemical and the mean of descriptor values for all chemicals. A large leverage value indicates that the x-values of a chemical are far from the center of descriptor values for all chemicals. Chemicals with large leverage may exert considerable influence on the fitted value, and thus on the regression model. Thus chemicals with unusual predictor values compared to the rest of the data can be identified by their leverage values. For training set chemicals leverage values fall between 0 and 1. A leverage value greater than $2p/n$ or $3p/n$, where p is the number of predictors plus the constant and n is the number of observations, is considered large and should be examined.

5.4 Model performance

The model quality was evaluated distinguishing between the internal performance of the model (data quality and goodness-of-fit) and the predictivity of the model (external validation).

5.4.1 Internal performance

- Data quality
  The training dataset consists of 58 chemicals listed in Table II. The biological data are considered to be of high quality, provided by a single protocol, measured in the same laboratory.
  The descriptor (Kow) data are both experimental and calculated values. Even if Kow is usually considered a good physicochemical descriptor, there is no evidence that the measurements were made by the same protocol, in the same laboratory. Thus a certain amount of variability could be present.

- Goodness of fit
  The model has been trained by 58 chemicals listed in Table II.

Predictor	*Coeff.*	*SE*
Constant	-1.330	0.088
LogKow	-0.862	0.034

The following fitness regression parameters were calculated for this QSAR:

R^2	R^2_{adj}	*s*	*F*	*LOF*
92.18	92.04	0.411	660.6	0.18

SDEC	*AIC*	*FIT*
0.404	0.18	11.05

R^2 = Coefficient of determination; R^2_{adj} = Coefficient of determination adjusted for the degrees of freedom; *s* = standard error of the estimate; *F* = Fisher function; *LOF* = Friedman modified; *SDEC* = Standard Deviation Error in Calculation; *AIC* = Akaike Information Criterion; *FIT* = Kubinyi function.

- Outliers detection:
  The regression line of the recalculated equation, the Williams and the residual plots are illustrated below: two outliers (Ethanol (48) and 3,3-dimethyl-2-butanone (33)) are present. No highly influential chemicals, with leverage values greater than 3*p/n* (=0.103) are highlighted by the leverage approach.

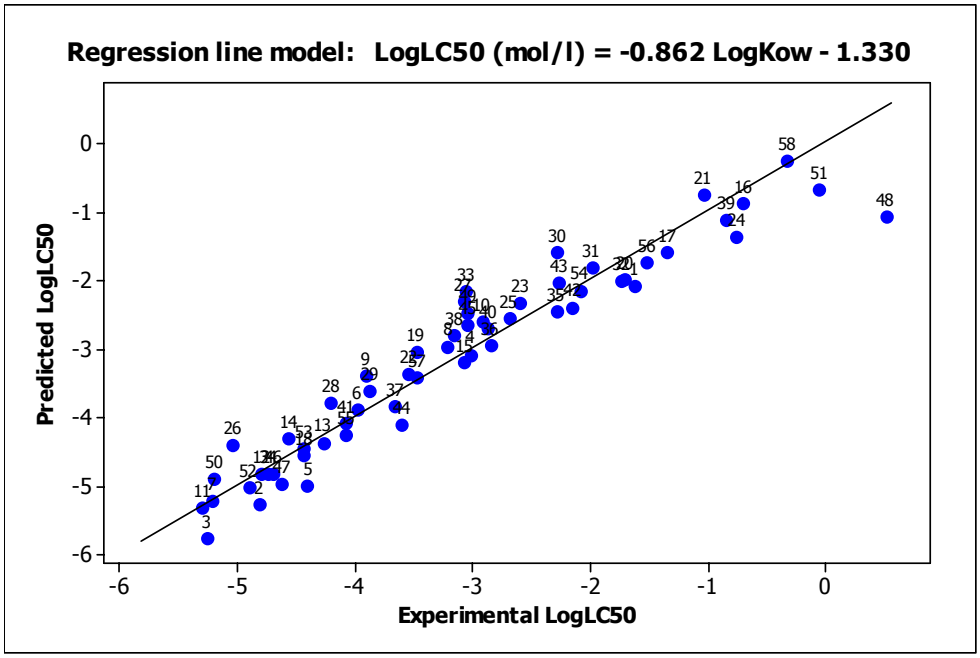


Figure 21 - NPN model regression plot.

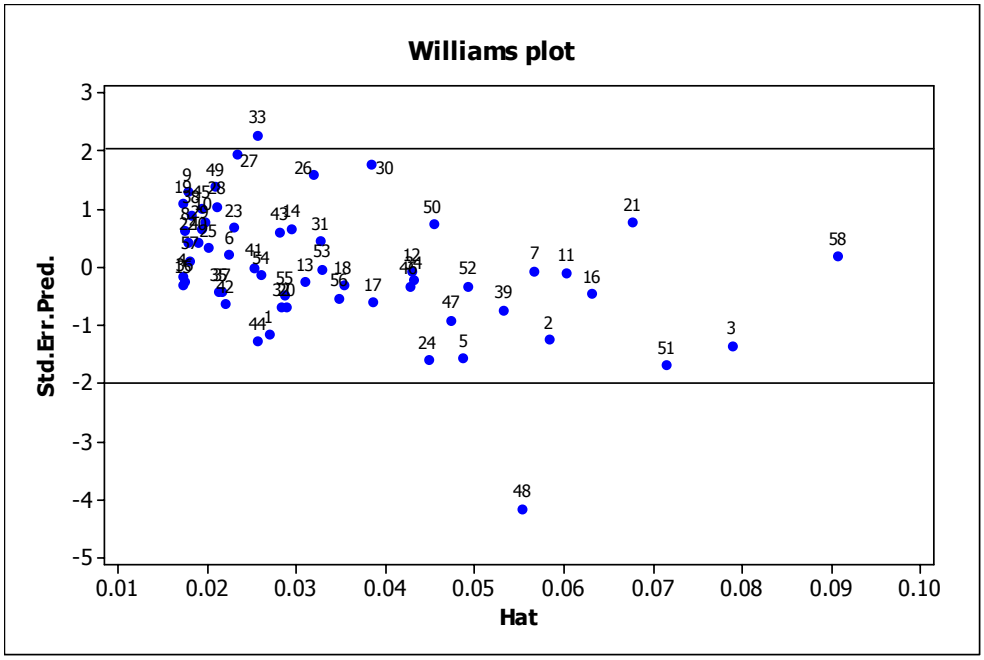


Figure 22 - NPN model Williams plot.

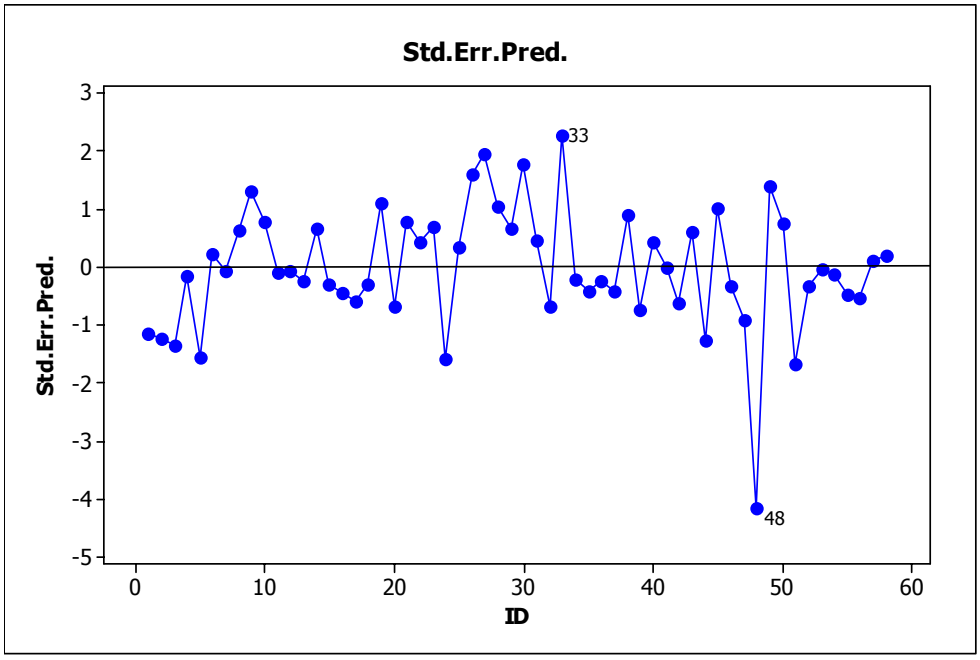


Figure 23 - NPN model residual plot.

The LogKow distribution of the training chemicals was analyzed in order to investigate the distribution of the chemicals in the space of the model descriptor, and to identify anomalous or isolated chemicals: the distribution in this case is essentially homogeneous.

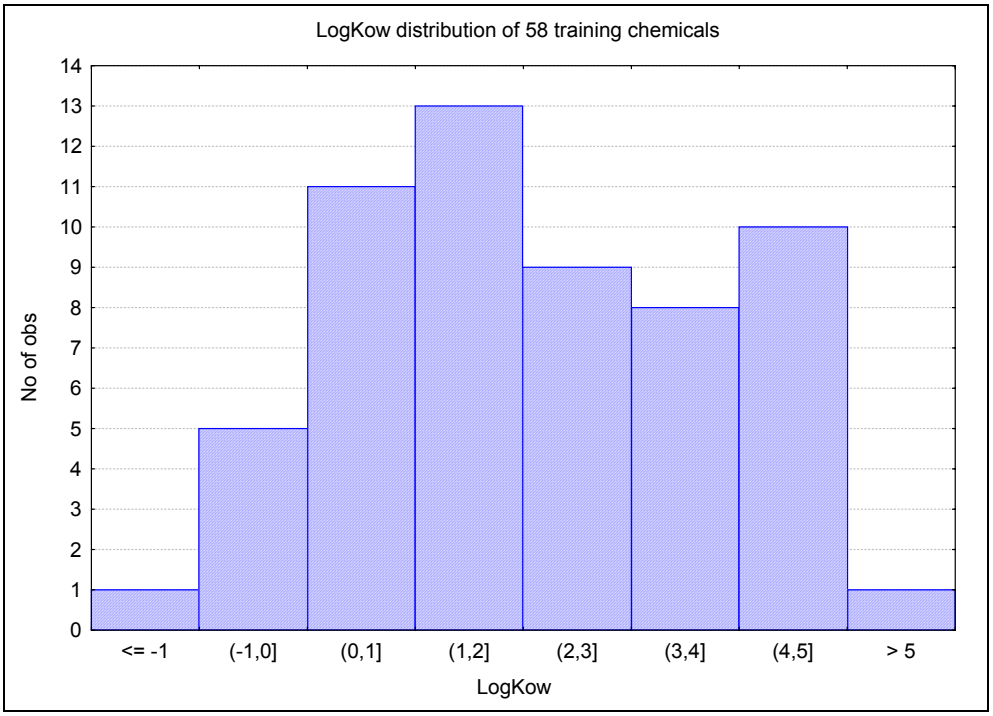


Figure 24 - Histogram of training set LogKow distribution.

- Internal Validation:
  The model evaluated by *leave-one-out* internal cross-validation (Q^2_{LOO}) and bootstrap with 5000 iterations shows a good predictive power. It was also verified by Y-*scrambling* with 300 iterations: the models on randomized response have all extremely low R^2 and Q^2compared with the published models. Thus the model was not obtained by chance correlation.

Q^2_{LOO}	$Q^2_{boostrap}$ *(5000 iterations)*	*SDEP*
91.51	91.66	0.421

Q^2_{LOO} = explained variance in prediction; $Q^2_{boostrap}$ = explained variance in prediction by bootstrapping; *SDEP* = Standard Deviation Error in Prediction.

5.4.2 External validation on SIDS test data

The QSAR model has been used to make predictions of SIDS test data.

- Model descriptor applicability domain
  The simplest method for describing the AD is to consider ranges of individual descriptors. Thus, the domain of applicability with respect to descriptor ranges was evaluated by analyzing the distribution of the SIDS LogKow values with respect to those of the training set.

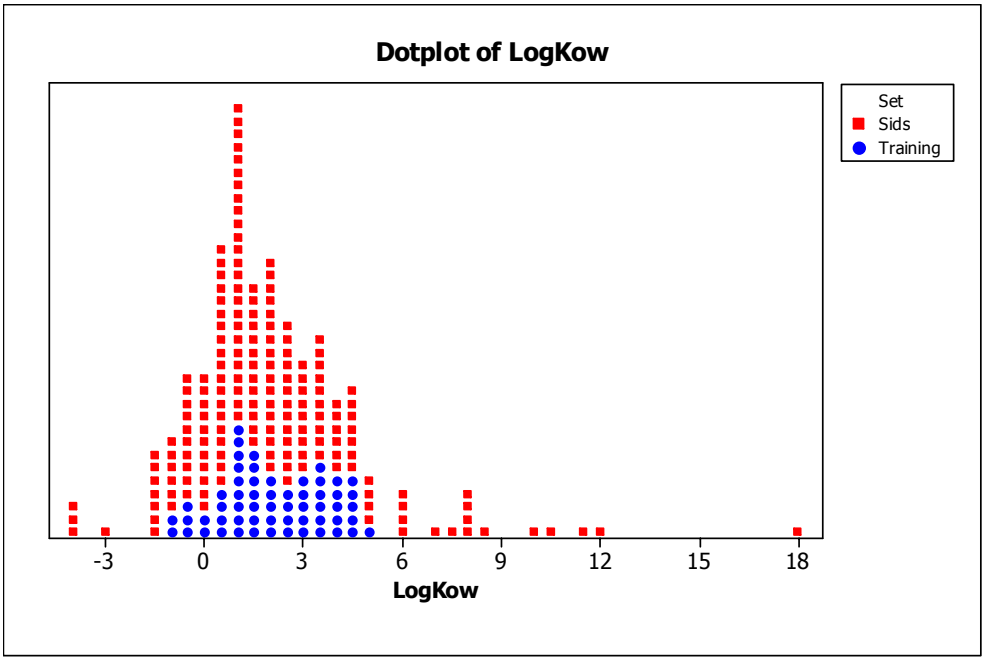


Figure 25 - SIDS and training set LogKow distribution comparison.

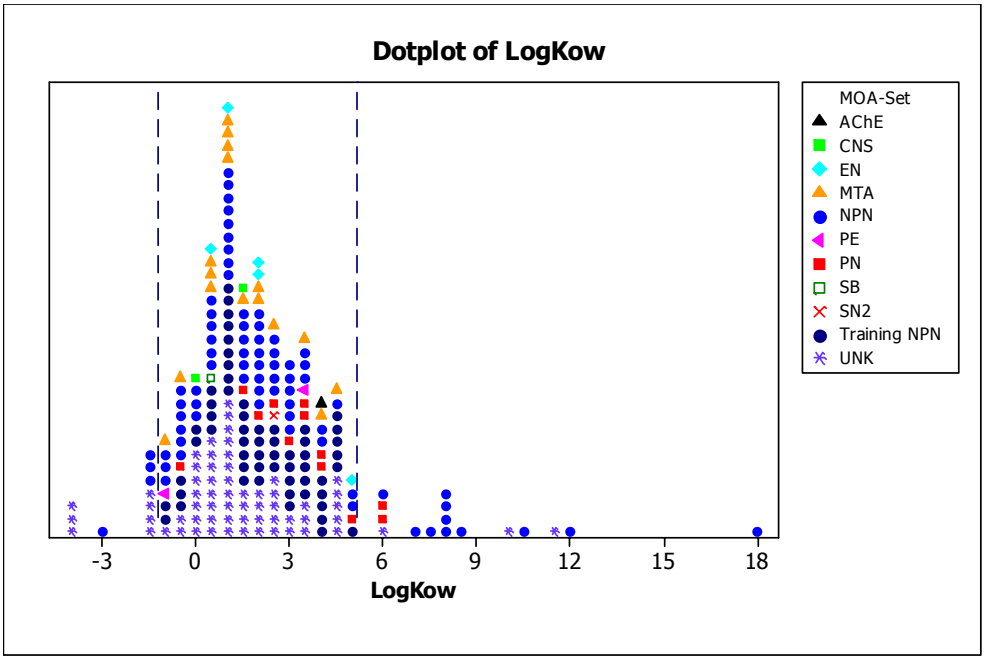


Figure 26 - SIDS and training set LogKow distribution comparison with MOA highlighted.

The LogKow domain of the SIDS test set includes the one of the training set but is much bigger: in fact the range of LogKow values for the SIDS test set is from -3.89 to 18.08. Moreover not all the chemical structures represented by the SIDS test set are consistent with those representing non-polar narcosis.

The non polar narcosis model was evaluated on two subsets of SIDS chemicals: the first set is made of the SIDS chemicals in the descriptor/response domain (XY-domain) and acting as non polar narcotics (MOA domain); the second set is made of the SIDS chemicals in the descriptor domain (XY-domain) without accounting their mechanism of action. The aim of this double evaluation was to verify the opportunity to apply the model only to chemical structures representing non-polar narcosis, and to verify the correctness of the defined mechanism of action of the chemicals under investigation.
The details on the SIDS chemicals disregarded in the two subsets are illustrated in Table III.

- <u>QSAR application on the SIDS subset defined by model domain in descriptor and response space (XY-D) and mode of action domain (MOA-D)</u>

  Predictions was performed only for chemicals with log Kow values in range from -1.24 to 5.13 according to the applicability domain suggested by the authors, and exhibiting a non polar narcosis mechanism of action. Moreover 8 SIDS chemicals (1-Butanol (S9), Ethane, 1,1,2-trichloro (S22), Ethane, 1,1,2,2-tetrachloro- (S28), Benzene, 1,4-dichloro (S72), Ethane, 1,2-dichloro (S75), 5-Hepten-2-one, 6-methyl- (S92), Ethanol, 2-phenoxy- (S107) and Propane, 2-methoxy-2-methyl- (S142)) were in the training set of the model; thus real predictions were performed for a subset of 51 SIDS chemicals.
  The predicted toxicities of the SIDS test set, together with their leverage and standardized residuals in prediction are collected in Table IV.

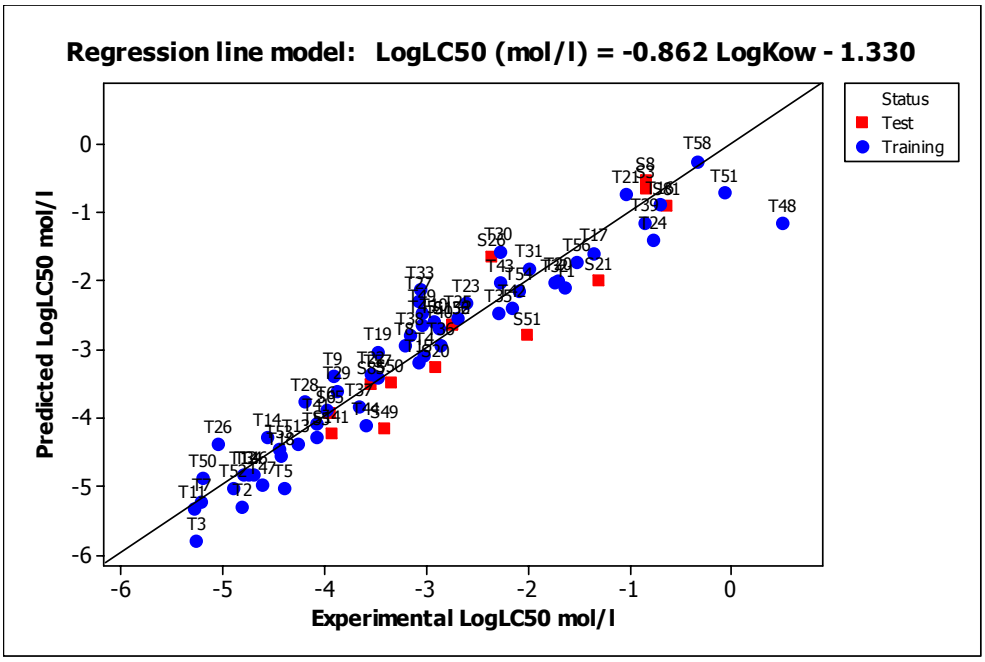


Figure 27 - NPN model regression plot: training and SIDS test data.

The SIDS test set is well predicted: all the chemicals have leverage values lower than the warning leverage ($h^* = 0.103$) meaning that the predicted response is not the result of substantial extrapolation of the model and, therefore, that the predictions are reliable.

Moreover the applicability domain of the model was analyzed by the Williams plot, where the vertical line is $h^* = 0.103$, the warning value for the X descriptor space and the horizontal lines are 2σ the cut off value for Y space. Note that in the Williams plot test chemicals with unknown experimental toxicity values are not represented: even if their leverage values are available, their standardized error in prediction cannot be calculated.

In the Williams plot no SIDS chemical is identified as an outlier: all the SIDS chemicals are into the XY-AD of the model.

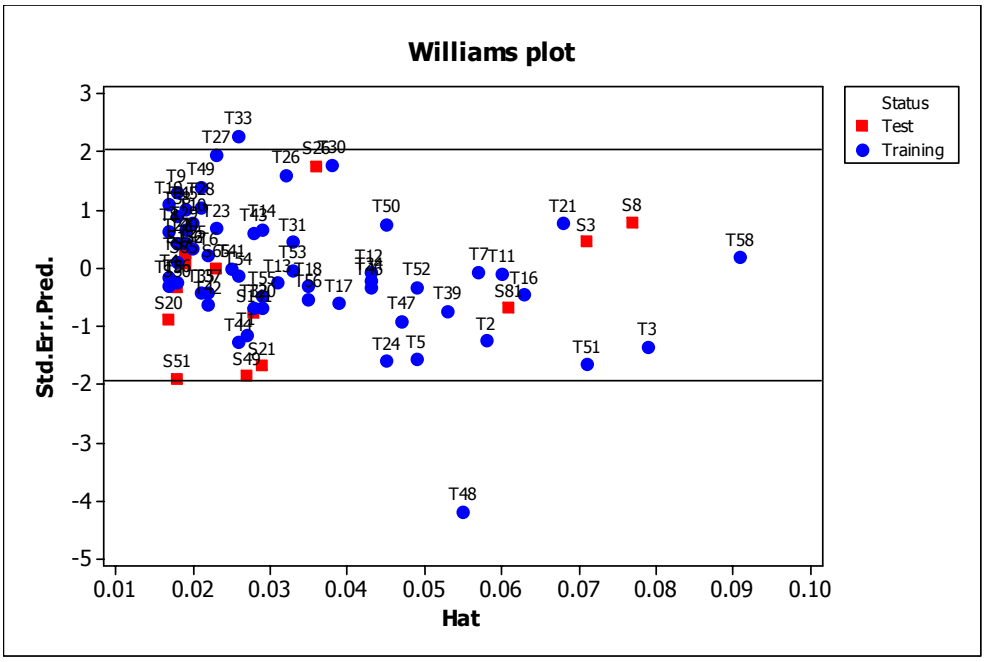


Figure 28 - NPN Williams plot: training and SIDS test data.

Evaluation of predictive performance

The prediction capability of the model evaluated in terms of explained variance (Q^2ext) and external standard deviation error of prediction ($SDEP_{ext}$) shows a pretty high predictive power.

N. ext = 14
Q^2ext = 89.06
$SDEP_{ext}$ = 0.431

- QSAR application on the SIDS subset defined by model domain in descriptor and response space (XY-D)

  To verify the opportunity to apply the model only to chemical structures acting by non-polar narcosis, and to verify the correctness of the defined mechanism of action of the chemicals under investigation, the model was applied to all the SIDS chemicals with log Kow values in range from -1.24 to 5.13 and not already present in the training set.
  The predicted toxicities of the 141 SIDS test chemicals, together with their MOA, leverage and predicted error values are collected in the Table V.

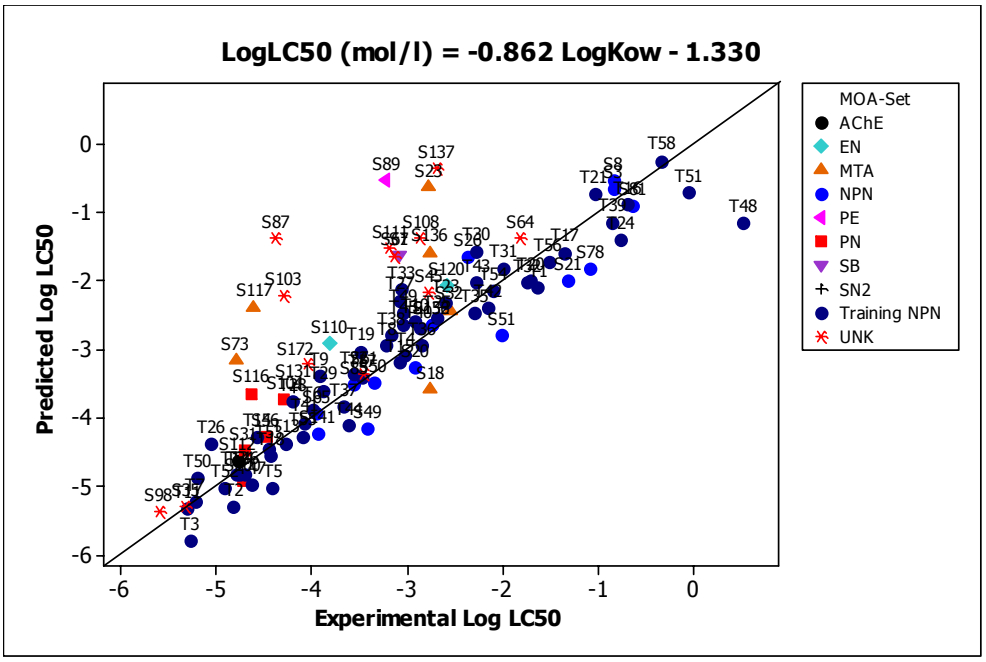


Figure 29 - NPN model regression plot: training and SIDS test data coloured by MOA.

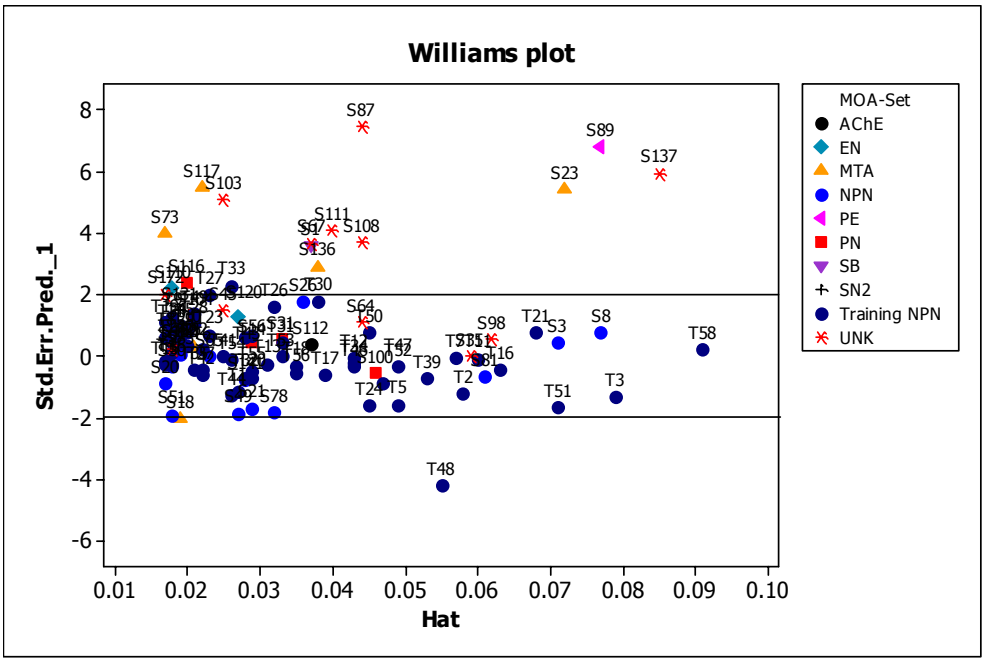


Figure 30 - NPN Williams plot: training and SIDS test data coloured by MOA.

The applicability domain of the model was analyzed by the Williams plot, where the vertical line indicates the warning value for the X space ($h^* = 0.103$) and the horizontal values are 2σ the cut off value for Y space.

Several chemicals are identified as Y-outliers, which are in the X-AD of the model meaning that either their experimental toxicity values are wrong or the model is not accounting some additional features relevant to explain their toxicity. It is important to note that all the outliers

identified by the Williams plot exhibit a diverse mechanism of action which probably needs to be described by other descriptors.
SIDS outliers are collected in the Table VI.

Evaluation of predictive performance
The prediction capability of the model evaluated in terms of explained variance (Q^2ext) and external standard deviation error of prediction ($SDEP_{ext}$) shows that the model is not able to make predictions for all those SIDS chemicals exhibiting a mechanism of action diverse from the non-polar narcotic one.

N. ext = 44
Q^2ext = 13.28
$SDEP_{ext}$ = 1.134

If the sixteen outliers are removed from the explained variance (Q^2ext) and external standard deviation error of prediction ($SDEP_{ext}$) calculation, the model predictive power increases significantly:

N. ext = 28
Q^2ext = 90.86
$SDEP_{ext}$ = 0.417

5.5 Conclusions

In conclusion, having checked the model correspondence with the OECD principles it can be highlighted that, for the investigated QSAR model the OECD principles were completely fulfilled; thus, on the basis of this information, this QSAR model could certainly be regarded as sufficiently well developed to be used for regulatory purposes.
In fact, it should be noted that the model was developed for a clear endpoint defined on a specific experimental system; it shows an unambiguous algorithm which ensures the model algorithm transparency. The applicability domain of the model was defined by the developers and the model exhibits a satisfactory goodness-of–fit, robustness and predictivity.
Finally the model has a mechanistic interpretation being the descriptor used in the model associated to predicted endpoint.
Moreover the exercise pointed out the importance of identifying properly the model applicability domain when it is applied to make predictions on the SIDS test set.
In fact, the applicability domain has to be considered in all three phases of the (Q)SAR life-cycle: in the development to ensure that the domain is defined as broadly as possible, in the model validation, to verified and eventually refined the domain and in the model application.
To apply properly a QSAR model and to identify the subset of reliable predictions provided by the model its domain has to be investigated.
The analysis performed on the non-polar narcosis model confirmed that the model should be applied only to the chemicals in the model descriptor and response space and with a non-polar narcotic mode of action.
A comparison of the model performance on the two subset of SIDS data is given in Table VII.

6. POLAR NARCOSIS QSAR2 EVALUATION

6.1 Defined endpoint and algorithm

This QSAR developed for predicting acute toxicity of organic chemicals to the fathead minnow is recommended for use in the European Union Technical Guidance Document (European Economic Community 1996) for polar narcosis.
The model was developed for predicting the acute toxicity of organic chemicals to the fathead minnow (*Pimephales promelas*) by Verhaar (Verhaar *et al*., 1995):

$$\text{Log(LC50)} = -0.723\ \text{LogKow} - 2.159$$

Where LC_{50} is the concentration (in moles per litre) causing 50% lethality in *Pimephales promelas*, after an exposure of 96 hours, and Kow is the octanol-water partition coefficient.

The regression model is based on a single parameter and it was developed by linear regression.

6.2 Mechanistic basis

The model was developed for chemicals acting as polar narcotics. The QSAR is based on a single descriptor for hydrophobicity (LogKow), which is relevant to the mechanism of action which consists in accumulation of molecules in biological membranes. Polar narcotics are typically defined as aromatic molecules that have a polar group (typically an hydroxyl or amine, but also possibly a nitro group). Further they may have a number of substituents such as alkoxy or alkyl groups and three or less halogens. Such molecules are clearly narcotic since they cause a reversible effect; however, their toxic effects are well in excess of that elicited by non-polar narcosis, and joint binary toxicity studies indicate different mechanisms of action. As concern QSAR modelling, it is commonly considered that there is still a strong relationship between toxicity and hydrophobicity, and QSARs based on log Kow alone should have a lower slope and higher intercept than those for non-polar narcosis.

6.3 Domain of applicability

The applicability domain of the QSAR model was defined by the developers s applicable to chemicals having log K_{ow} values in the range from -1.31 to 6.20: chemicals with a LogKow lower than -1.31 are not considered due to their unrealistic high effect concentrations that will be predicted by a narcosis QSAR. Compounds with a LogKow greater than 6.20 are excluded since they do not normally exhibit acute toxicity being taken up from water too slowly to show acute toxic effect or being too bulky to be uptaken through membranes.
Moreover the model is suitable for chemicals operating by a polar narcosis mechanism of action, i.e. aromatic nitro compounds, anilines and phenols. Aliphatic amines are also included in this class. Although most aliphatic amines are ionized at a pH of 7, they have been included in the model because they perfectly fit the model.
However, the developers highlighted that the uptake of ionized chemicals is complex and therefore, it can not be excluded that aliphatic amines somehow accidentally fit the model.

The domain of applicability was verified by the leverage approach, to analyze the distance of each chemical from the centre of the model space.

6.4 Model performance

The model quality was evaluated according to its internal performance (data quality and goodness-of-fit) and its predictivity on SIDS test data (external validation).

6.4.1 Internal performance

- Data quality
  The training dataset consists of 86 chemicals listed in Table VIII. The biological data are considered to be of high quality, provided by the same source, according to a single protocol, US EPA's Duluth Environmental Research Laboratory's Fathead Minnow database.

- Goodness of fit
  The following regression parameters were calculated for this QSAR:

Predictor	*Coeff.*	*SE*
Constant	-2.159	0.073
LogKow	-0.723	0.026

R^2	R^2_{adj}	*s*	*F*	*LOF*
90.07	89.95	0.332	762.05	0.113

SDEC	*AIC*	*FIT*
0.329	0.116	8.672

R^2 = Coefficient of determination; R^2_{adj} = Coefficient of determination adjusted for the degrees of freedom; *s* = standard error of the estimate; *F* = Fisher function; *LOF* = Friedman modified; *SDEC* = Standard Deviation Error in Calculation; *AIC* = Akaike Information Criterion; *FIT* = Kubinyi function

- Outlier detection:
  The regression line of the equation, the Williams and the residual plots are reported below.
  Several chemicals are identified as Y-outliers, which are into the X-AD of the model meaning that either their toxicity values are wrong or these chemicals have some additional feature not accounted for by the model.
  The Williams plot identifies 3,3-dimethylbutylamine (79) as a strong outlier with a standard deviation error in prediction greater than 3, together with four small outliers: 4-amino-2-nitrophenol (29), 2-chloroaniline (31), 2,5-dichloroaniline (35) and 2,2-dimethyl-1-propylamine (78).
  Moreover, six influential chemicals with leverage values greater than 3*p/n* (=0.070) are identified: 4-nonylphenol (21), 4-decylaniline (49), 2-aminoethanol (82), 1-amino-2-propanol (83), tridecylamine (75) and 2-methoxyethylamine (84).

These chemicals greatly influence the regression line: in fact, the regression line is forced near the observed value and their residuals (observed-predicted values) are small, i.e. they are well predicted.

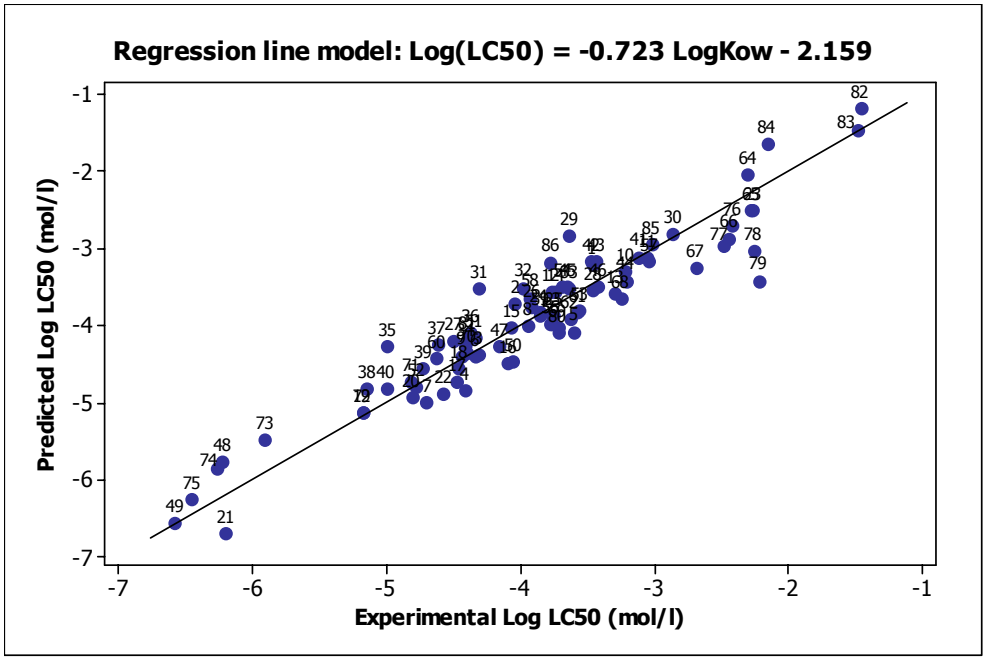


Figure 31 - PN model regression plot.

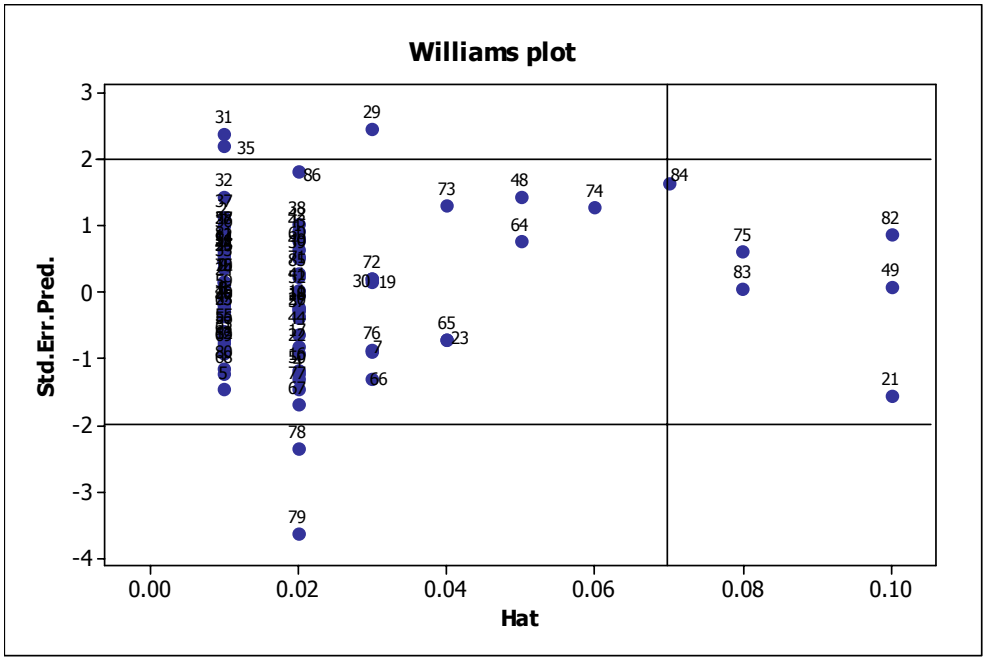


Figure 32 - PN model Williams plot.

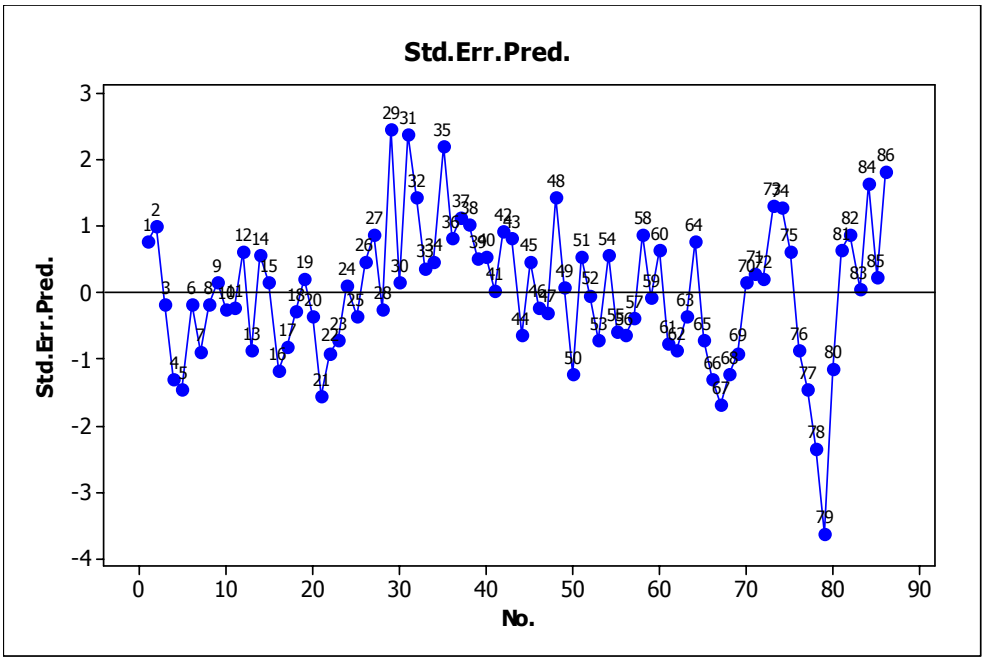


Figure 33 - PN model residual plot.

The LogKow distribution of the training chemicals was analyzed to highlight the distribution of the chemicals in the model descriptor space and to identify anomalous or isolated chemicals: the distribution in this case is essentially homogeneous.

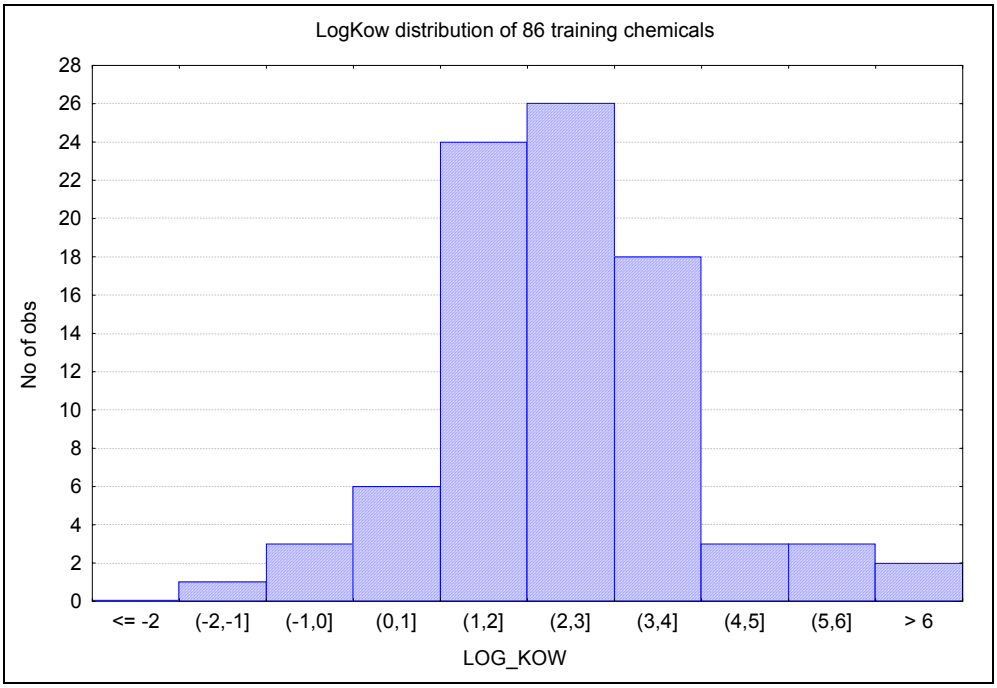


Figure 34 - Histogram of training set LogKow distribution.

- Internal Validation:
  The model evaluated by *leave-one-out* internal cross-validation (Q^2_{LOO}) and by bootstrapping with 5000 iterations shows a good predictive power. It was also verified by Y-*scrambling* with 300 iterations: the models based on randomized responses all have extremely low R^2 and Q^2compared with the published models. Thus the model was not obtained by chance correlation.

Q^2_{LOO}	$Q^2_{boostrap}$ *(5000 iterations)*	*SDEP*
89.59	89.64	0.336

Q^2_{LOO} = explained variance in prediction; $Q^2_{boostrap}$ = explained variance in prediction by bootstrapping; *SDEP* = Standard Deviation Error in Prediction

6.4.2 External validation on SIDS test data

The QSAR model was used to make predictions of SIDS test data.

- Model descriptor applicability domain
  The domain of applicability with respect to descriptor ranges was evaluated by analyzing the distribution of the SIDS LogKow values with respect to the LogKow distribution of the training set.

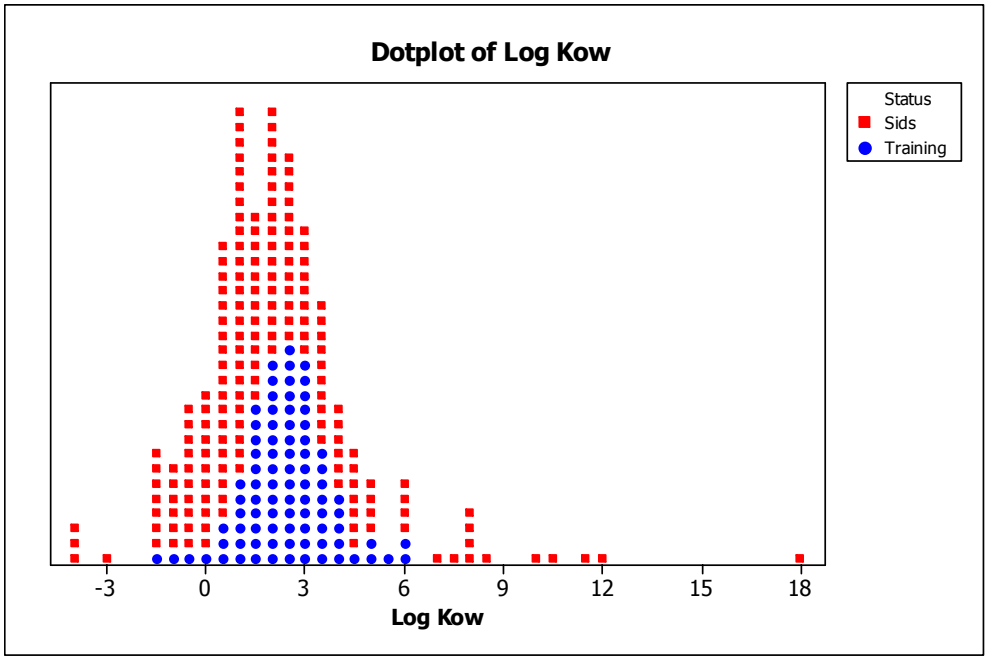


Figure 35 - SIDS and training set LogKow distribution comparison.

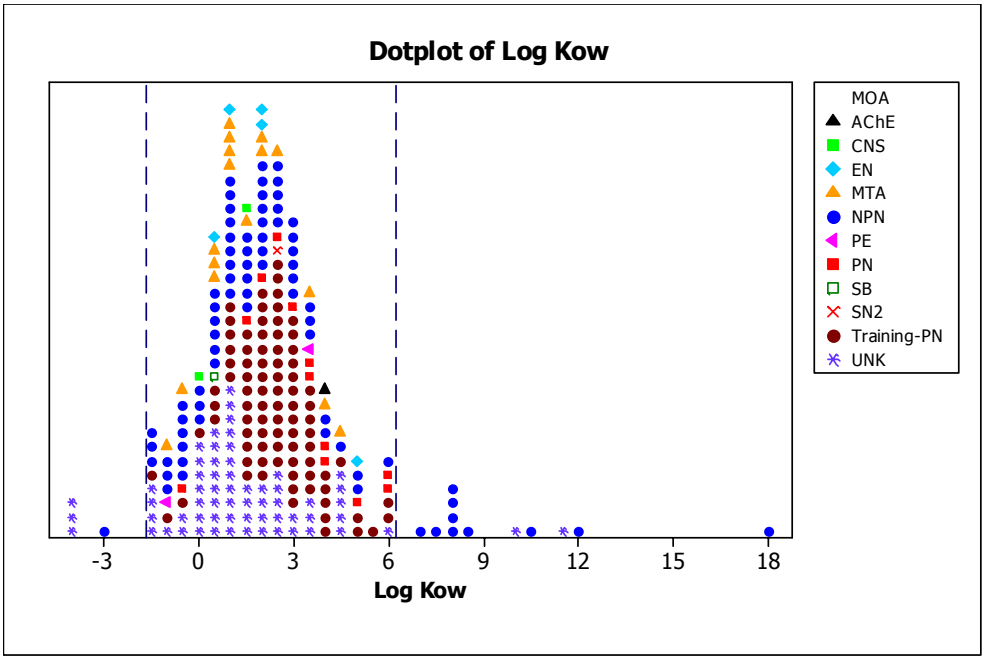


Figure 36 - SIDS and training set LogKow distribution comparison with MOA highlighted.

The LogKow domain of the SIDS test set includes the one of the training set but is much bigger: in fact the range of LogKow values for the SIDS set is from -3.89 to 18.08. Moreover not all the chemical structures represented by the SIDS set are consistent with those representing polar narcosis.

In order to verify the applicability of the model only to chemical structures acting by polar narcosis, and to verify the correctness of the defined mechanism of action of the studied chemicals, the polar narcosis model was evaluated on two subset of SIDS chemicals: the first set containing SIDS chemicals which fall in the descriptor/response domain (XY-domain) and acting as polar narcotics (MOA domain); the second set containing SIDS chemicals which fall in the descriptor domain (XY-domain) without accounting their mechanism of action. Details of the SIDS chemicals disregarded because of their prediction unreliability in the two subsets are given in Table IX.

- <u>QSAR application on the SIDS subset defined by model domain in descriptor and response space (XY-D) and mode of action domain (MOA-D)</u>

  Predictions were considered only for chemicals with log Kow values in range from -1.31 to 6.20 according to the applicability domain of the model, and exhibiting a polar narcosis mechanism of action. Moreover 5 SIDS chemicals (Phenol, 4-(1,1-dimethylethyl) (S56), Benzene, 1-methyl-4-nitro (S61), Benzenamine, 3-methyl- (S82), Phenol, 2,4,6-tribromo (S100), Phenol, 2,4-dichloro (S104)) were already in the model training set, and therefore were not taken into account; thus real predictions were performed for a subset of 8 SIDS chemicals.
  The predicted toxicities of the test set, together with their leverage and predicted error values are collected in Table X.

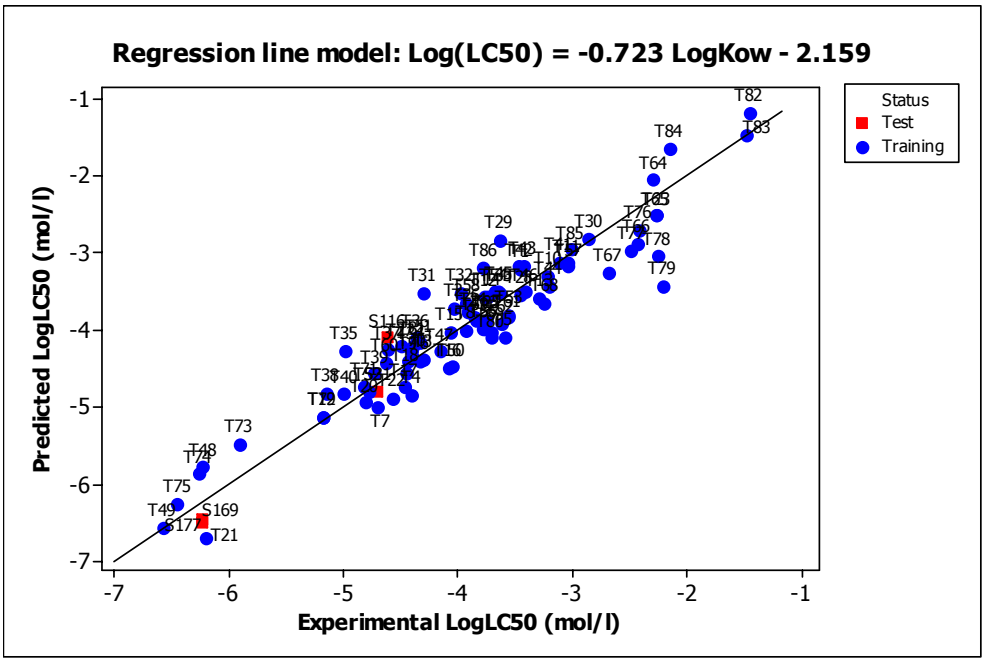


Figure 37 - PN model regression plot: training and SIDS test data.

The applicability domain of the model was analyzed by the Williams plot, where the vertical line indicates the warning value for the X space ($h^* = 0.070$), and the horizontal lines are 2σ the cut off values for Y space. Note that in the Williams plot test chemicals with unknown experimental toxicity values are not represented: even if their leverage values are available, their standardized error in prediction cannot be calculated.

In the Williams plot it is possible to identify two SIDS chemicals with high leverage values, thus being out of the applicability domain of the model: phenol, nonyl- (S169) and phenol, 4-nonyl-, branched (S177). It has to be pointed out that, while the high leverage chemicals in the QSAR model training set reinforce the model itself, the test chemicals with high leverage values have unreliable predicted data, being the result of substantial extrapolation of the model.

Since phenol, nonyl- (S169) and phenol, 4-nonyl-, branched (S177) are outside the model AD, their predictions can be the result of substantial extrapolation of the model and therefore may not be reliable.

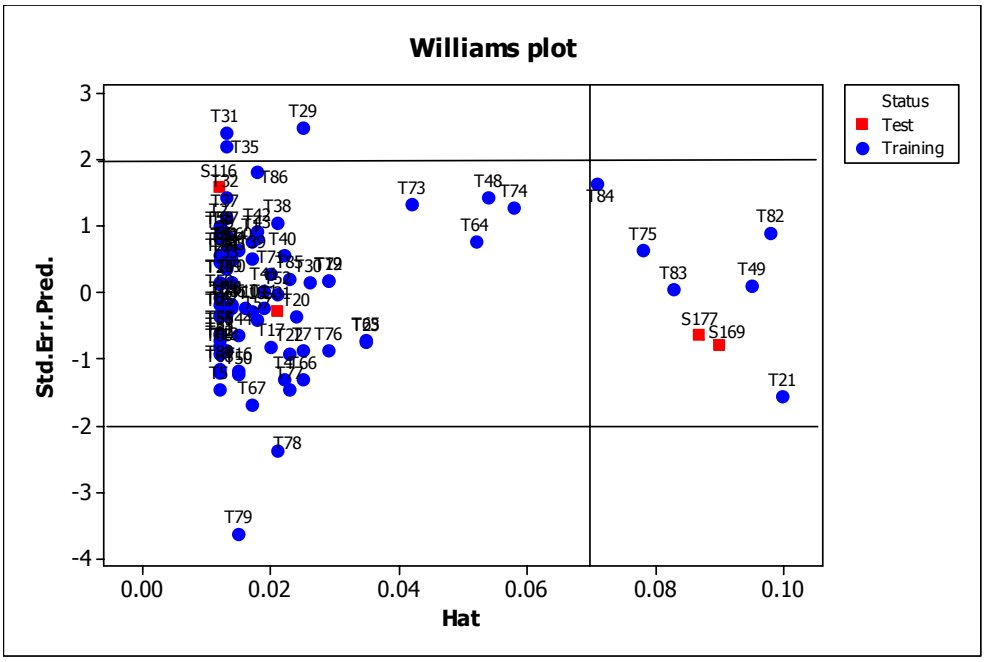


Figure 38 - PN model Williams plot: training and SIDS test data.

Evaluation of predictive performance

Since phenol, nonyl- (S169) and phenol, 4-nonyl-, branched (S177) are outside the model AD, their predictions are considered unreliable and thus the prediction capability of the model in terms of explained variance (Q^2ext) and external standard deviation error of prediction ($SDEP_{ext}$) cannot be evaluated due to the few number of test chemicals.

- QSAR application on the SIDS subset defined by model domain in descriptor and response space (XY-D)

  The applicability of the model only to chemical structures representing polar narcosis was investigated by ignoring the known or expected mechanism of action of the SIDS data and by applying the model to all the SIDS chemicals with log Kow values in range from -1.31 to 6.20, and not already present in the training set.
  The predicted toxicities of the 148 SIDS test chemicals, together with their MOA, leverage and predicted error values are collected in the Table XI.

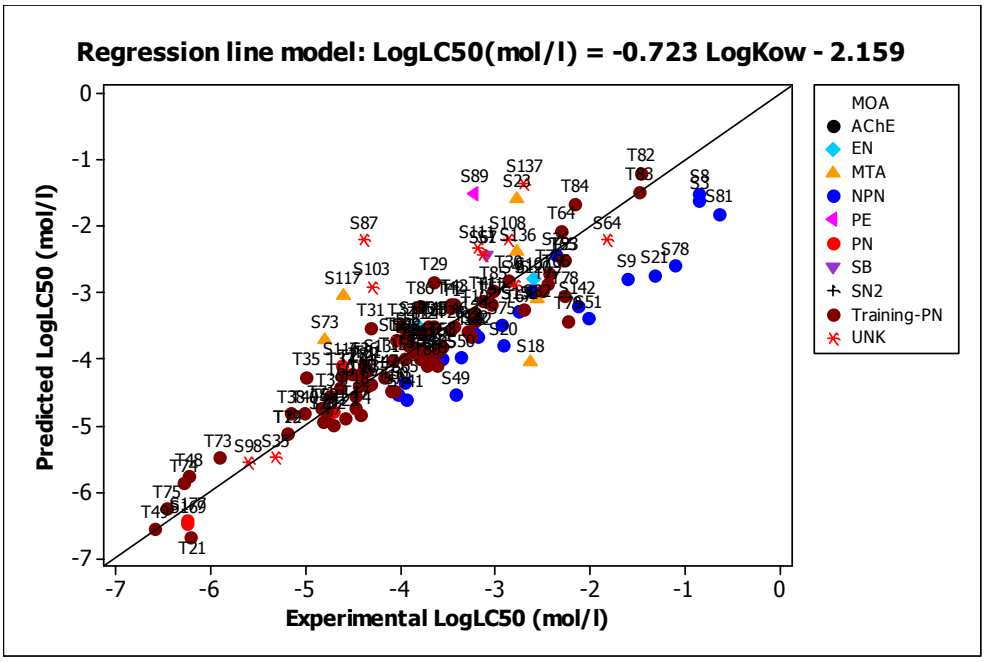


Figure 39 - PN model regression plot: training and SIDS test data colored by MOA.

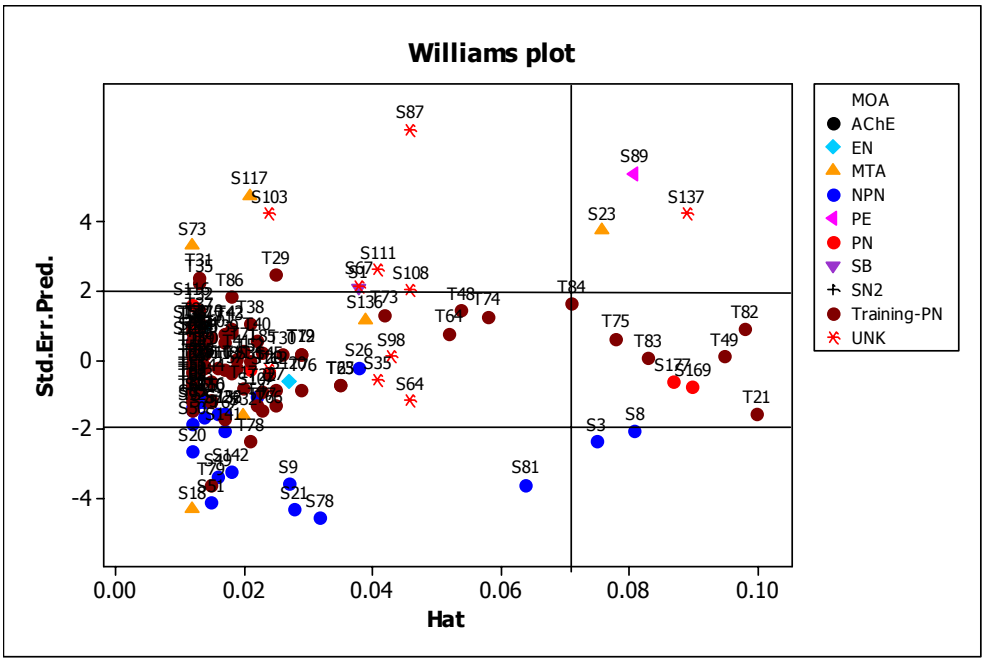


Figure 40 - PN model Williams plot: training and SIDS test data colored by MOA.

The Williams plot highlights that several SIDS chemicals are outside the AD of the model: the five worst predicted chemicals (1,2-Propanediol (S3), Formamide, N,N-dimethyl- (S8), 2-Propenamide (S23), 2-Butyne-1,4-diol (S89) and Phosphonic acid, dimethyl ester (S137)) are both outliers and high leverage chemicals.

Several other chemicals (Formaldehyde (S1), 1-Butanol (S9), 2-Cyclohexen-1-one, 3,5,5-trimethyl (S18), Propane, 1,2-dichloro- (S20), 2-Butanol (S21), Benzene, 1,2-dichloro- (S49), 2-Butanone, oxime (S51), 1,2,3-Propanetriol, triacetate (S67), 2-Propenoic acid, 2-

methylpropyl ester (S73), 2,4-Pentanediol, 2-methyl (S78), 2-Propanol, 1-methoxy- (S81), 2-Butenedioic acid (Z)- (S87), 1,2-Benzenediol (S103), 2,4-Pentanedione (S108), Hexanedioic-acid- (S111), 2-Propenoic acid, ethyl ester (S117), Cyclohexanol, 5-methyl-2-(1-methylethyl)- (S141), Propane, 2-methoxy-2-methyl- (S142)) are only Y-outliers, but they are in the X-AD of the model, confirming that their mechanism of action needs to be described by a diverse model equation and/or by other descriptors able to represent features not accounted for by the polar narcosis model.
The Williams plot identifies two SIDS chemicals as high leverage chemicals, and thus outside the applicability domain of the model: phenol, nonyl- (S169) and phenol, 4-nonyl-, branched (S177). It has to be pointed out that, while high leverage chemicals in the QSAR model training set reinforce the model itself, test chemicals with high leverage values have unreliable predicted data, being the result of substantial extrapolation of the model.
Moreover, other three SIDS chemicals (Piperazine (S91), Butanedioic acid, disodium salt (S123), 2-Benzothiazolesulfenamide, N,N-dicyclohexyl (S159)), not displayed in Williams plot because of lacking experimental toxicity values, are outside the AD of the model according to their leverage values. Their predictions are not reliable.

Evaluation of predictive performance

The prediction capability of the model in terms of explained variance (Q^2ext) and external standard deviation error of prediction ($SDEP_{ext}$), evaluated by including only those SIDS test data with reliable predictions according to the leverage approach, is satisfactory.

N. ext = 43
Q^2ext = 57.68
$SDEP_{ext}$ = 0.840

The SIDS chemicals (1,2-Propanediol (S3), Formamide, N,N-dimethyl- (S8), 2-Propenamide (S23), 2-Butyne-1,4-diol (S89), Phosphonic acid, dimethyl ester (S137), phenol, nonyl- (S169) and phenol, 4-nonyl-, branched (S177)) with leverage values greater than the warning leverage value (h^* = 0.070) were not included in the predictive performance evaluation.

The model predictive power is strongly reduced by the high Y-outliers: Formaldehyde (S1), 1-Butanol (S9), 2-Cyclohexen-1-one, 3,5,5-trimethyl (S18), Propane, 1,2-dichloro- (S20), 2-Butanol (S21), Benzene, 1,2-dichloro- (S49), 2-Butanone, oxime (S51), 1,2,3-Propanetriol, triacetate (S67), 2-Propenoic acid, 2-methylpropyl ester (S73), 2,4-Pentanediol, 2-methyl (S78), 2-Propanol, 1-methoxy- (S81), 2-Butenedioic acid (Z)- (S87), 1,2-Benzenediol (S103), 2,4-Pentanedione (S108), Hexanedioic-acid- (S111), 2-Propenoic acid, ethyl ester (S117), Cyclohexanol, 5-methyl-2-(1-methylethyl)- (S141), Propane, 2-methoxy-2-methyl- (S142)**.** If they are removed from the explained variance (Q^2ext) and external standard deviation Error of Prediction ($SDEP_{ext}$) calculation, because of their suspicious toxicity values or their possession of additional features, the model predictive power increases slightly:

N. ext = 25
Q^2ext = 86.66
$SDEP_{ext}$ = 0.377

6.5 Conclusions

The analysis performed confirmed the model correspondence with the OECD principles: the principles were completely fulfilled, and therefore this QSAR model can certainly be regarded as sufficiently well developed to be used for regulatory purposes.

The model was developed for a clear endpoint defined on a specific experimental system; it's s algorithm is transparent and unambiguous. The applicability domain of the model was defined by the developers and the model exhibits a satisfactory goodness-of–fit, robustness and predictivity.

Finally the model has a mechanistic interpretation being the descriptor used in the model associated to predicted endpoint.

Moreover the exercise confirmed the importance of identifying properly the model applicability domain to apply properly the model and provide reliable predictions on the SIDS test set.

The QSAR polar narcosis model evaluation confirmed that the model should be applied only to the chemicals falling in the model descriptor and response space and with a polar narcotic mode of action.

A comparison of the model performance on the two subset of SIDS data is given in Table XII.

7. NARCOSIS QSAR3 EVALUATION

7.1 Defined endpoint and algorithm

This QSAR was developed by ECB for predicting acute toxicity of organic chemicals to the fathead minnow for chemicals acting by both non-polar and polar narcosis.
Both these two mechanism of actions consist in accumulation of molecules in biological membranes, and thus both can be modeled by a single descriptor for hydrophobicity (LogKow).
The relation between toxicity and logKow values of the chemicals used to train both the non-polar and polar narcosis models is illustrated in the following plot:

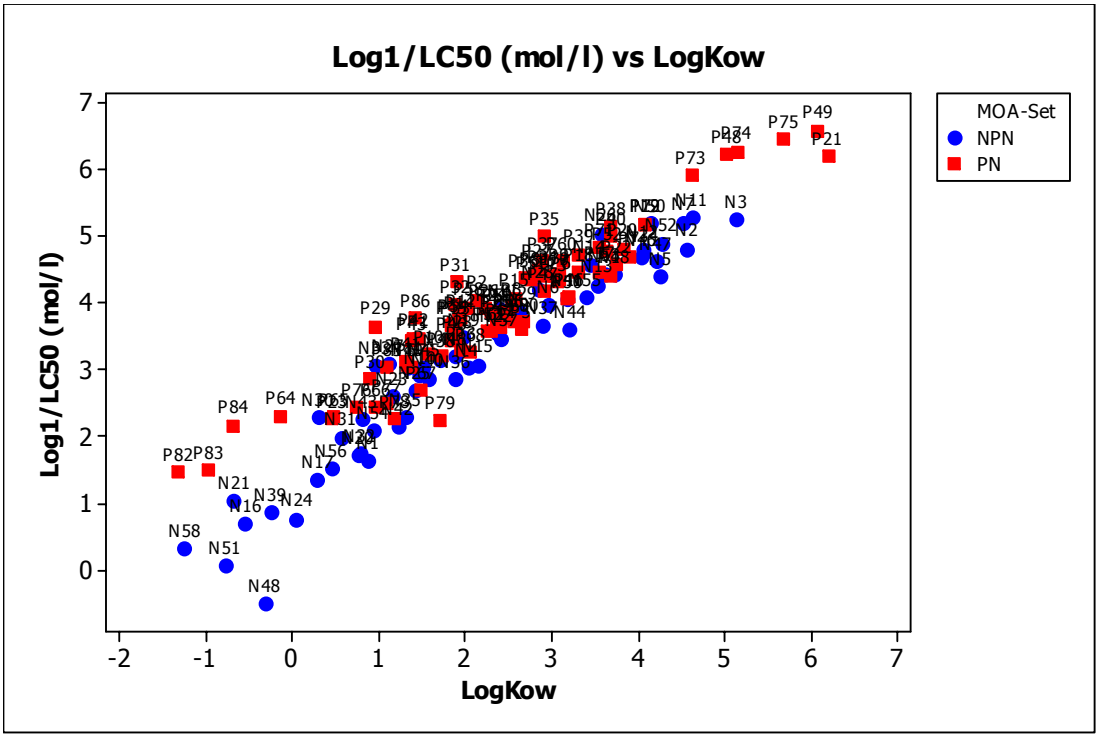


Figure 41 - Correlation between Log(1/LC50) and LogKow in NPN and PN training sets.

Since the two training sets are close enough according to their relation between toxicity and LogKow, they were merged and used to train a new global model for narcosis.

$$\text{Log}(1/\text{LC50})\ \text{mol/l} = 0.810\ \text{LogKow} + 1.744$$

Where LC_{50} is the concentration (in moles per litre) causing 50% lethality in *Pimephales promelas*, after an exposure of 96 hours, and Kow is the octanol-water partition coefficient.
The regression model is based on a single parameter and was developed by linear regression.

A comparison between the narcosis models indicates that the QSAR for polar narcosis based on log Kow alone has a lower slope and higher intercept than that for non-polar narcosis, while the global narcosis model is a compromise between them.

Model	*Equation*
Non polar narcosis	Log(1/LC50) = 0.862 LogKow + 1.330
Polar narcosis	Log(1/LC50) = 0.723 LogKow + 2.159
Global narcosis	Log(1/LC50) = 0.810 LogKow + 1.744

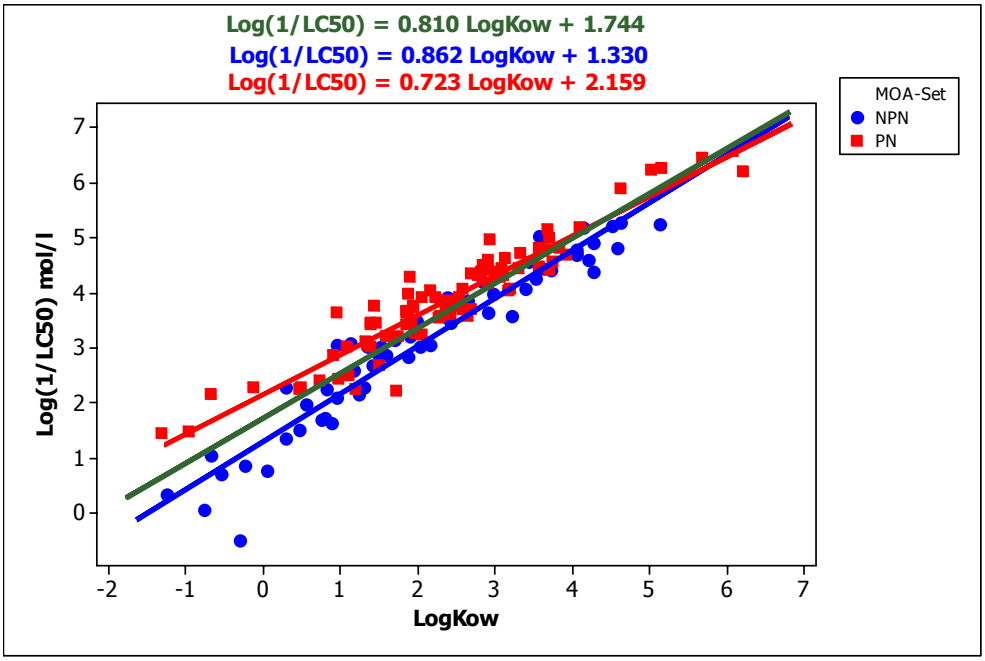


Figure 42 - Comparison regression lines of NPN, PN and N model.

Compounds represented by blue points were used to derive the non-polar narcosis model (blue solid line); compounds represented by red squares were used to derive the polar narcosis model (red solid line); all compounds were used to derive the narcosis model (green solid line);

7.2 Mechanistic basis

The model was developed for chemicals acting as narcotics. The QSAR is based on a single descriptor for hydrophobicity (LogKow), which is relevant to the mechanism of action which consists in accumulation of molecules in biological membranes.

7.3 Domain of applicability

The applicability domain of the QSAR model is limited to chemicals having log K_{ow} values in the range from -1.31 to 6.20: chemicals with a LogKow lower than -1.31 are not considered due to their unrealistic high effect concentrations that will be predicted by a narcosis QSAR. Compounds with a LogKow greater than 6.20 are excluded since they do not normally exhibit acute toxicity, being taken up from water too slowly to show acute toxic effect or being too bulky to pass through membranes.

Moreover the model is suitable for chemicals acting by a narcosis mechanism of action, i.e. aliphatic and aromatic hydrocarbons, halogenated aliphatic and aromatic hydrocarbons, ethers, alcohols, aromatic nitro compounds, anilines and phenols. Aliphatic amines are also included in this class. Although most aliphatic amines are ionized at a pH of 7, they have been included in the model because they perfectly fit the model.

The domain of applicability was verified by the leverage approach.

7.4 Model performance

The model quality was evaluated distinguishing between the internal performance of the model (data quality and goodness-of-fit) and the predictivity of the model (external validation).

7.4.1 Internal performance

- Data quality
  The model has been trained by 144 chemicals listed in Table XIII.

- Goodness of fit

Predictor	*Coeff.*	*SE*
Constant	-1.744	0.070
LogKow	-0.801	0.026

The following fitness regression parameters were calculated for this QSAR:

R^2	R^2_{adj}	*s*	*F*	*LOF*
87.55	87.46	0.455	998.34	0.210

SDEC	*AIC*	*FIT*
0.452	0.213	6.830

R^2 = Coefficient of determination; R^2_{adj} = Coefficient of determination adjusted for the degrees of freedom; *s* = standard error of the estimate; *F* = Fisher function; *LOF* = Friedman modified; *SDEC* = Standard Deviation Error in Calculation; *AIC* = Akaike Information Criterion; *FIT* = Kubinyi function

- Outlier detection:
  The regression line of the equation and the Williams plot are illustrated below.
  Some chemicals are identified as Y-outliers, which are inside the X-AD of the model, meaning that either their toxicity values are wrong or these chemicals have some additional feature not accounted for by the model. The Y-outliers are: 2-propanol N24), ethanol (N48), methanol (N51), 4-amino-2-nitrophenol (P29), 2-chloroaniline (P31), 3,3-dimethylbutylamine (P79), 2-methoxyethylamine (P84).

Moreover five influential chemicals with leverage values greater than 3*p/n* (=0.042) are identified: 4-nonylphenol (P21), 4-decylaniline (P49), 2-aminoethanol (P82), tridecylamine (P75) and Triethylene glycol (N58).

These chemicals greatly influence the regression line: in fact, the regression line is forced near the observed value and their residual (observed-predicted value) are small, i.e. they are well predicted.

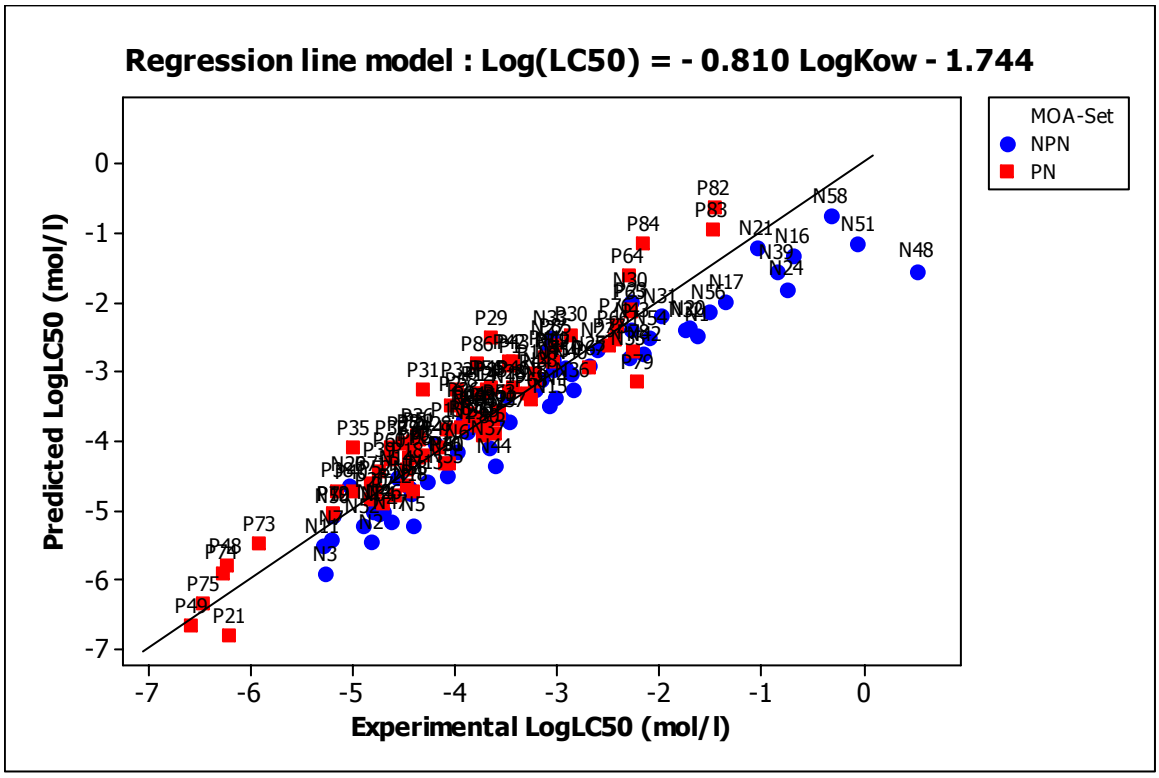


Figure 43 - N model regression plot.

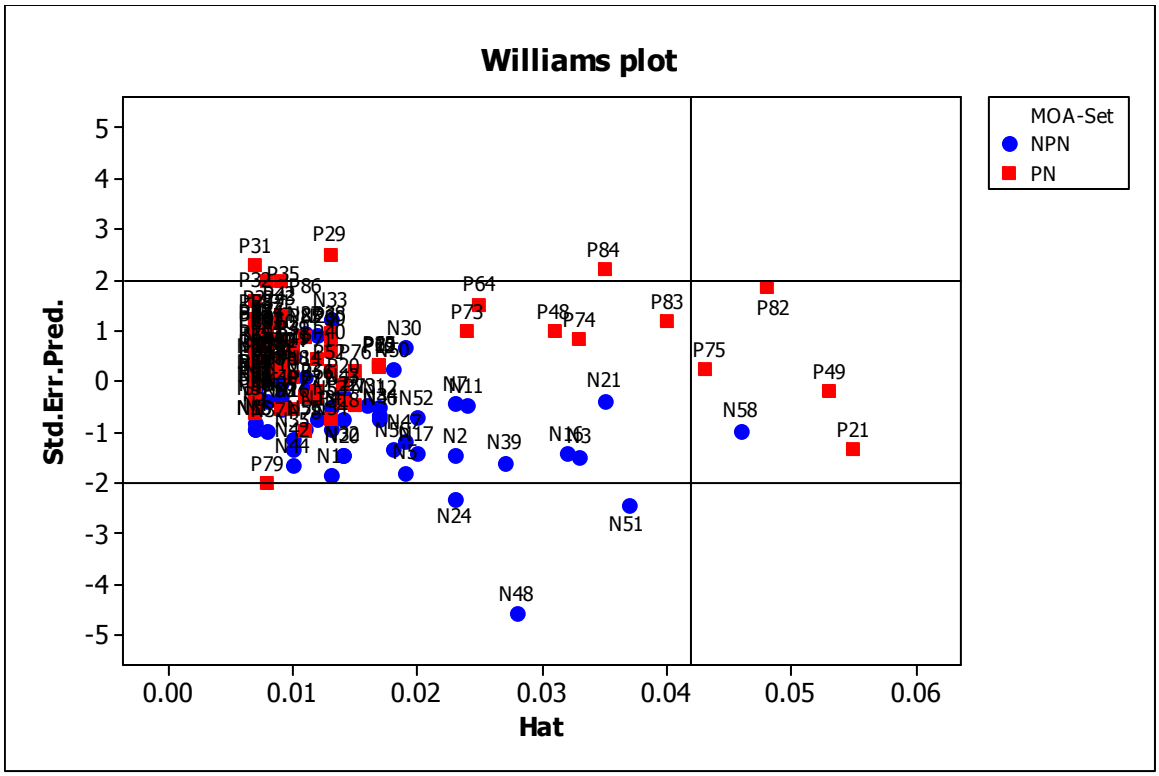


Figure 44 - N model Williams plot.

- Internal validation:

Q^2_{LOO}	$Q^2_{boostrap}$ *(5000 iterations)*	*SDEP*
87.06	87.11	0.461

Q^2_{LOO} = explained variance in prediction; $Q^2_{boostrap}$ = explained variance in prediction by bootstrapping; *SDEP* = Standard Deviation Error in Prediction

The model evaluated by *leave-one-out* internal cross-validation (Q^2_{LOO}) and bootstrap with 5000 iterations shows a good predictive power. It was also verified by Y-*scrambling* with 300 iterations: the models based on randomized responses have all extremely low R^2 and Q^2 compared with the published models, meaning that the model was not obtained by chance correlation.

7.4.2 External validation on SIDS test data

The QSAR model was used to make predictions of SIDS test data. Following the same approach previously applied on the non-polar narcosis and polar narcosis models, the correctness of the hypotised mechanism of action of the SIDS chemicals was evaluated on two subsets of SIDS chemicals: the first set consisting of the SIDS chemicals which fall in the descriptor/response domain (XY-domain) and acting as narcotics (MOA domain); the second set consisting of the SIDS chemicals which fall in the descriptor domain (XY-domain) without accounting for their

mechanism of action. Details of the SIDS chemicals disregarded because of their prediction unreliability in the two subsets are given in Table XIV.

- <u>QSAR application on the SIDS subset defined by model domain in descriptor and response space (XY-D) and mode of action domain (MOA-D)</u>

  Predictions were considered only for chemicals with log Kow values in range from -1.31 to 6.20 according to the applicability domain of the model, and exhibiting a narcosis mechanism of action. Moreover 13 SIDS chemicals (1-Butanol (S9), Ethane, 1,1,2-trichloro (S22), Ethane, 1,1,2,2-tetrachloro- (S28), Phenol, 4-(1,1-dimethylethyl) (S56), Benzene, 1-methyl-4-nitro (S61),Benzene, 1,4-dichloro (S72), Ethane, 1,2-dichloro (S75), Benzenamine, 3-methyl- (S82), 5-Hepten-2-one, 6-methyl- (S92), Phenol, 2,4,6-tribromo (S100), Phenol, 2,4-dichloro (S104), Ethanol, 2-phenoxy- (S107) and Propane, 2-methoxy-2-methyl- (S142),) were already in the model training set and have not been accounted for in the predictions; thus real predictions, were performed for a subset of 61 SIDS chemicals.
  The predicted toxicities of the test set, together with their leverages and predicted error values, are collected in the Table XV.

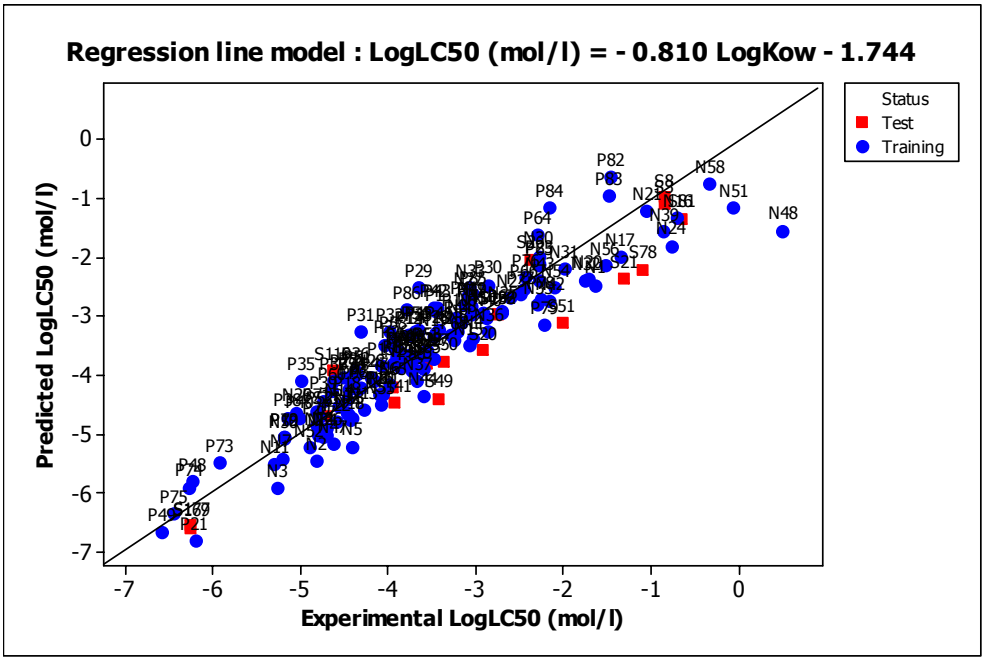


Figure 45 - N model regression plot: training and SIDS test data.

The applicability domain of the model was analyzed by the Williams plot, where the vertical line indicates the warning value for the X space ($h^* = 0.042$), and the horizontal lines are 2σ the cut off value for Y space. Note that in the Williams plot test chemicals with unknown experimental toxicity values are not represented: even if their leverage values are available, their standardized error in prediction cannot be calculated.
In the Williams plot it is possible to identify two SIDS chemicals with high leverage values, and thus outside of the applicability domain of the model: phenol, nonyl- (S169) and phenol, 4-nonyl-, branched (S177). It has to be pointed out that, while the high leverage chemicals in the QSAR model training set reinforce the model itself, test chemicals with high leverage

values greater than the warning value have unreliable predicted data, being the result of substantial extrapolation of the model.
Although not displayed in the Williams plot because of its experimental toxicity value N,N-dicyclohexyl (S159) is outside the applicability domain of the model according to its leverage and thus its prediction is not reliable.
Four SIDS chemicals (2-Butanol (S21), Benzene, 1,2-dichloro (S49), 2-Butanone, oxime (S51), 2,4-Pentanediol, 2-methyl- (S78)) are identified as Y-outliers. These chemicals are outliers only in the Y-response space, since they are inside the X-AD of the model: either their toxicity values are wrong or the model is lacking in some additional feature.

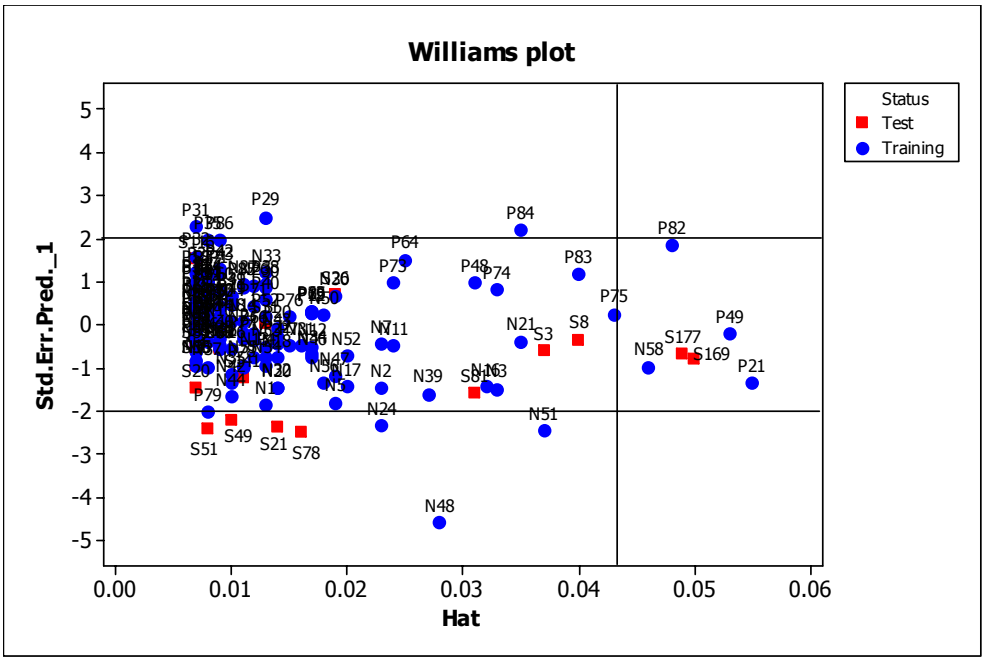


Figure 46 - N model Williams plot: training and SIDS test data.

Evaluation of predictive performance

The prediction capability of the model in terms of explained variance (Q^2ext) and external standard deviation error of prediction ($SDEP_{ext}$), evaluated by including only those SIDS test data with reliable predictions according to the leverage approach, is satisfactory.

N. ext = 17
Q^2ext = 84.31
$SDEP_{ext}$ = 0.637

Since phenol, nonyl- (S169) and phenol, 4-nonyl-, branched (S177) are outside the model AD, their predictions were unreliable and were not accounted for in the predictive performance evaluation.

The model predictive power is strongly reduced down by the four Y-outliers: 2-Butanol (S21), Benzene, 1,2-dichloro (S49), 2-Butanone, oxime (S51), 2,4-Pentanediol, 2-methyl- (S78). If they are removed from the calculation of explained variance (Q^2ext) and External

Standard Deviation Error of Prediction ($SDEP_{ext}$), because of their suspicious toxicity values or their possession of additional features, the model predictive power increases slightly:

N. ext = 13
Q^2ext = 92.18
$SDEP_{ext}$ = 0.425

- <u>QSAR application on the SIDS subset defined by model domain in descriptor and response space (XY-D)</u>

  The model was applied to all the SIDS chemicals with log Kow values in range from -1.31 to 6.20 and not already present in the training set.
  The predicted toxicities of the SIDS test chemicals, together with their MOA, leverage and predicted error values are collected in the Table XVI.

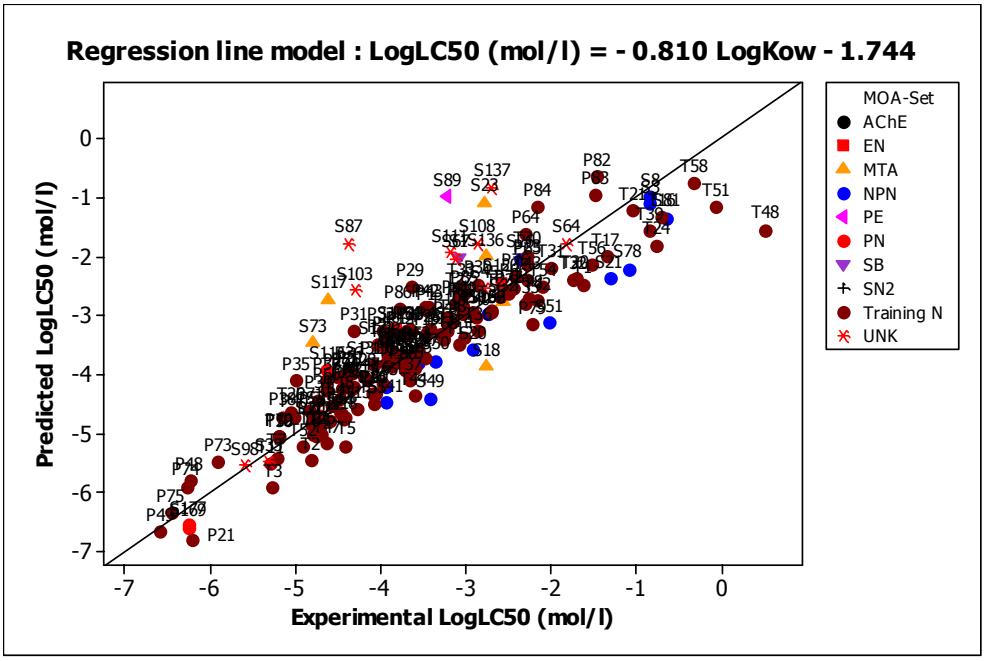


Figure 47 - N model regression plot: training and SIDS test data colored by MOA.

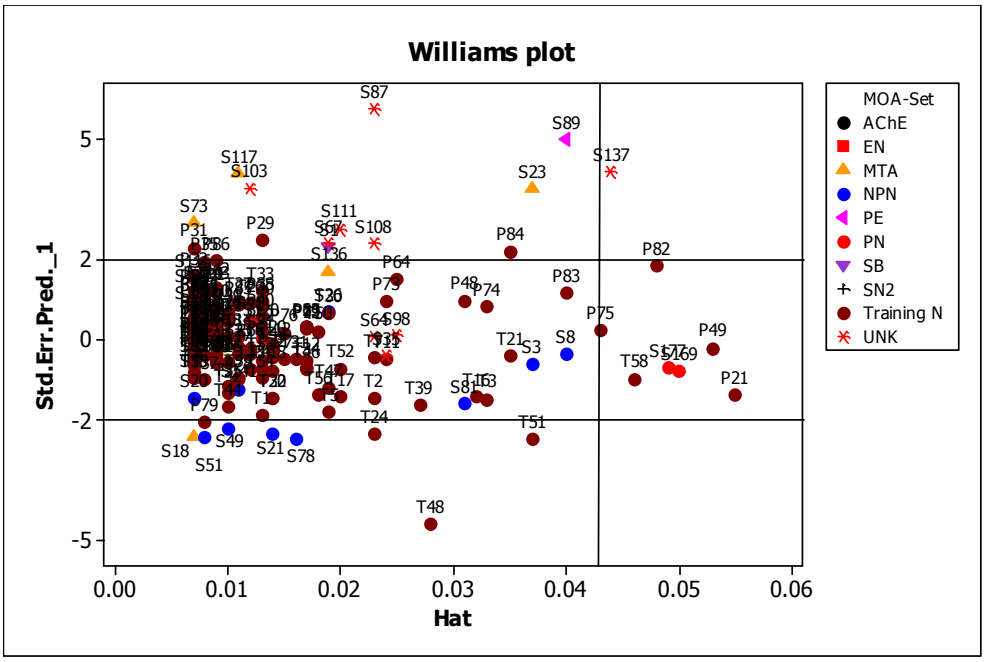


Figure 48 - N model regression plot: training and SIDS test data colored by MOA.

The Williams plot highlights that several SIDS chemicals are outside the AD of the model: the worst predicted chemical (Phosphonic acid, dimethyl ester (S137)) is both an outlier and a high leverage chemical. Several other chemicals (Formaldehyde- (S1), 2-Cyclohexen-1-one, 3,5,5-trimethyl (S18), 2-Butanol (S21), 2-Propenamide (S23), Benzene, 1,2-dichloro- (S49), 2-Butanone, oxime (S51), 1,2,3-Propanetriol, triacetate (S67), 2-Propenoic acid, 2-methylpropyl ester (S73), 2,4-Pentanediol, 2-methyl- (S78), 2-Butyne-1,4-diol (S89), 1,2-Benzenediol (S103), 2,4-Pentanedione (S108), Hexanedioic-acid- (S111), 2-Propenoic acid, ethyl ester (S117)) are only Y-outliers, but they are inside the X-AD of the model, indicating that their mechanism of action needs to be described by a diverse model equation and/or by other descriptors able to represent feature not accounted for by the polar narcosis model. Two SIDS chemicals are outside the applicability domain of the model, due to high leverage values: phenol, nonyl- (S169) and phenol, 4-nonyl-, branched (S177). For these chemicals, predictions can be the result of substantial extrapolation of the model and therefore may not be reliable.
N,N-dicyclohexyl (S159) which is not displayed in Williams plot because its experimental toxicity value is lacking, is outside the applicability domain of the model according to its leverage and thus its prediction is not reliable.

Evaluation of predictive performance

The prediction capability of the model in terms of explained variance (Q^2ext) and external standard deviation error of prediction ($SDEP_{ext}$), evaluated by including only those SIDS test data with reliable predictions, according to the leverage approach is not satisfactory.

N. ext = 39
Q^2ext = 43.48
$SDEP_{ext}$ = 0.976

The three SIDS chemicals (dimethyl ester (S137), phenol, nonyl- (S169), 4-nonyl-, branched (*S177*)) with leverage values greater than the warning value ($h^* = 0.042$) were not included in the predictive performance evaluation).

The model predictive power is thus strongly reduced by some strong Y-outliers: Formaldehyde- (S1), 2-Cyclohexen-1-one, 3,5,5-trimethyl (S18), 2-Butanol (S21), 2-Propenamide (S23), Benzene, 1,2-dichloro- (S49), 2-Butanone, oxime (S51), 1,2,3-Propanetriol, triacetate (S67), 2-Propenoic acid, 2-methylpropyl ester (S73), 2,4-Pentanediol, 2-methyl- (S78), 2-Butyne-1,4-diol (S89), 1,2-Benzenediol (S103), 2,4-Pentanedione (S108), Hexanedioic-acid- (S111), 2-Propenoic acid, ethyl ester (S117). If these outliers are removed from the calculation of explained variance (Q^2ext) and external standard deviation error of prediction ($SDEP_{ext}$), because of their suspicious toxicity values or their possession of additional features, the model predictive power increases slightly:

N. ext = 24
Q^2ext = 91.63
$SDEP_{ext}$ = 0.426

7.5 Conclusions

The global QSAR model developd for narcosis fullfills completely the OECD principles, and therefore it can certainly be regarded as sufficiently well developed to be used for regulatory purposes. The model is well trained and it exhibits a very good goodness-of–fit, robustness and predictivity. Its performance are even higher than the one obtained by the non-polar and polar narcosis models suggesting the opportunity to develop a model predict acute toxicity of organic chemicals to the fathead minnow for chemicals acting by both non-polar and polar narcosis. The model has a mechanistic interpretation since both these two mechanism of actions consisting in accumulation of molecules in biological membranes, are modeled by a single descriptor for hydrophobicity (LogKow).
Moreover the exercise confirmed the importance of identifying properly the model applicability domain to apply properly the model and provide reliable predictions on the SIDS test set.
The QSAR narcosis model evaluation confirmed the model should be applied only to the chemicals in the model descriptor and response space and with a narcotic mode of action.
A comparison of the model performance on the two subset of SIDS data is given in Table XVII.

ACKNOWLEDGEMENTS

The authors would like to thank Eva Bay Wedebye (Danish Institute for Food and Veterinary Research), Terry Schultz (Department of Comparative Medicine, College of Veterinary Medicin at the University of Tennessee), Christine Russom (U.S. EPA, Office of Research and Development, National Health and Environmental Effects Research Laboratory (NHEERL), Mid-Continent Ecology Division (MED), Duluth, MN), Karin Aschberger (European Chemicals Bureau, Joint Research Centre, Italy). Manuela Pavan acknowledges receipt of a post-doctoral grant from the Joint Research Centre (JRC contract 22460-2004-11 P1B30 ISP IT).

REFERENCES

AQUIRE (AQUatic toxicity Information REtrieval), U.S. Environmental Protection Agency. 2002. ECOTOX User Guide: ECOTOXicology Database System. Version 3.0. Available: http:/www.epa.gov/ecotox/

ASTER (ASsessment Tools for the Evaluation of Risk), U.S. Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Mid-Continent Ecology Division (MED), Duluth, Minnesota.

Bonchev, D. (1983). Information Theoretic Indices for Characterization of Chemical Structures. Research Studies Press: Chichester, UK.

Consonni, V., Todeschini, R. and Pavan, M. (2002). Structure / Response Correlation and Similarity / Diversity Analysis by GETAWAY Descriptors. Part 1. Theory of the Novel 3D Molecular Descriptors. *J.Chem. Comput. Sci*. *42*, 693-705.

CORINA software 2005. 3D Structure Generator CORINA, Generation of High-Quality, Three-Dimensional Molecular Models Molecular Networks GmbH, Computerchemie, Nägelsbachstr. 25 91052 Erlangen,Germany.

Devillers, J. and Balaban, A.T. (2000). Topological Indices and Related Descriptors in QSAR and QSPR. Gordon & Breach: Amsterdam, The Netherlands.

European Commission (1995). QSAR for Predicting Fate and Effects of Chemicals in the Environment, Final Report of DG XII Contract No. EV5V-CT92-0211.

European Economic Community (1996). Technical Guidance Document in Support of Commission Directive 93/67/EEC on Risk Assessment for New Notified Substances and Commission Regulation (EC) No 1488/94 on Risk Assessment for Existing Substances, Luxemburg: European Commission, Office for Official Publications of the European Communities.

OECD ENV/JM/Mono(2004)24. Environment Directorate Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, pesticides and Biotechnology. OECD Series on Testing and assessment. Number 49. The report from the Expert Group on (Quantitative) Structure-Activity Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs.

OECD ENV/JM/TG(2004)26. Comparison of SIDS Test data with (Q)SAR predictions doe acute aquatic toxicity, biodegradability and mutagenicity on organic chemicals discussed at SIAM 11-18.

Russom, C.L., Bradbury, S.P., Broderius, S.J., Hammermeister, D.E., Drummond, R.A. (1997). Predicting modes of action from chemical structure: Acute toxicity in the fathead minnow (Pimephales promelas). *Environmental Toxicology and Chemistry*, 16, 948-967.

TSAR. Quantitative structure-activity relationship (QSAR) package for library design and optimization. Oxford Molecular Ltd., Oxford Science Park, Oxford OX4 4GA, UK

Todeschini, R., Lasagni, M., Marengo, E. (1994). New Molecular Descriptors for 2D- and 3D-Structures. Theory. *J.Chemom*., *8*, 263-273.

Todeschini, R., Gramatica, P. (1997). 3D-Modelling and Prediction by WHIM Descriptors. Part 5. Theory Development and Chemical Meaning of WHIM Descriptors. *Quant.Struct.-Act.Relat*., *16*, 113-119.

Todeschini, R., Consonni, V., Mauri, A., Pavan, M. (2004). DRAGON, rel. 5.2 for Windows; Talete srl: Milano, Italy.

Veith, G.D., Call D.J., and Brooke L.T.. 1983. Structure-toxicity relationships for the fathead minnow, Pimephales promelas: Narcotic industrial chemicals. *Can. J. Fish. Aquat. Sci***.** 40:743-748.

Verhaar, H.J.M., van Leeuwen, C.J. and Hermens, J.L.M. (1992) Classifying environmental pollutants. 1. Structure-activity relationships for prediction of aquatic toxicity. *Chemosphere*, 25, 471-491.

Verhaar, H.J.M., Mulder, W. and Hermens, J.L.M. (1995). QSARs for ecotoxicity. In: Overview of structure-activity relationships for environmental endpoints, Part 1: General outline and procedure. Hermens, J.L.M. (Ed), Report prepared within the framework of the project "QSAR for Prediction of Fate and Effects of Chemicals in the Environment", an international project of the Environmenta; Technologies RTD Programme (DGXII/D-1) of the Europena Commission under contract number EV5V-CT92-0211.

Verhaar, H.J.M., Solbe, J., Speksnijder, J., van Leeuwen, C.J. and Hermens, J.L.M. (2000) 'Classifying environmental pollutants: Part 3. External validation of th classification system. *Chemosphere*, 40, 875-883.

TABLES

Table I – SIDS test data.

ID	*CASN*	*EINECS name*	*ECB-MOA*	*Schultz MOA*	*EPA-MOA*	*CONS1*	*CONS2*	*Verhaar MOA*	*LogLC50 (mol/l)*	*LogP*
1	50-00-0	Formaldehyde-	SB	SB	NPN	SB	R	R	**-3.081**	0.35
2	56-81-5	1,2,3-Propanetriol	NPN	NPN	NPN	NPN	N	NPN		-1.65
3	57-55-6	1,2-Propanediol	NPN	NPN	NPN	NPN	N	NPN	**-0.838**	-0.78
4	58-08-2	1H-Purine-2,6-dione, 3,7-dihydro-1,3,7-trimethyl-	CNS	NPN	CNS	CNS	S			0.16
5	58-55-9	1H-Purine-2,6-dione, 3,7-dihydro-1,3-dimethyl-	CNS	NPN	NPN	NPN	N			-0.39
6	60-00-4	Glycine, N,N'-1,2-ethanediylbis[N-(carboxymethyl)-	NPN_LogD	CARB. ACID	NPN	UNK	N*		**-3.689**	-3.86
7	64-02-8	Glycine, N,N'-1,2-ethanediylbis[N-(carboxymethyl)-, tetrasodium salt	NPN_LogD	CARB. ACID	NPN	UNK	N*	R/S		-3.86
8	68-12-2	Formamide, N,N-dimethyl-	NPN	SB	NPN	NPN	N	R	-0.839	-0.93
9	71-36-3	1-Butanol	NPN	NPN	NPN	NPN	N	NPN	**-1.601**	0.84
10	74-83-9	Methane, bromo-	NPN	NPN	NPN	NPN	N	NPN		1.18
11	74-87-3	Methane, chloro-	NPN	NPN	NPN	NPN	N	NPN		1.09
12	75-01-4	Ethene, chloro-	SN2	NPN	NPN	NPN	N			1.62
13	75-10-5	Methane, difluoro-	NPN	NPN	NPN	NPN	N			0.71
14	75-38-7	Ethene, 1,1-difluoro-	SN2	NPN	NPN	NPN	N			1.24
15	75-56-9	Oxirane, methyl-	RAD	EPOX	ALKY-ARYL	UNK	R	R		0.37
16	75-68-3	Ethane, 1-chloro-1,1-difluoro-	NPN	NPN	NPN	NPN	N	NPN		2.05
17	77-92-9	1,2,3-Propanetricarboxylic acid, 2-hydroxy-	NPN_LogD	CARB. ACID	NPN	NPN	N*			-1.67
18	78-59-1	2-Cyclohexen-1-one, 3,5,5-trimethyl-	MTA	MTA	NPN	MTA	R	R	**-2.762**	2.62

Table I – SIDS test data (continued).

ID	*CASN*	*EINECS name*	*ECB-MOA*	*Schultz MOA*	*EPA-MOA*	*CONS1*	*CONS2*	*Verhaar MOA*	*LogLC50 (mol/l)*	*LogP*
19	78-70-6	1,6-Octadien-3-ol, 3,7-dimethyl-	PE	PE	NPN	PE	R	R		3.38
20	78-87-5	Propane, 1,2-dichloro-	NPN	NPN	NPN	NPN	N	NPN	-2.907	2.25
21	78-92-2	2-Butanol	NPN	NPN	NPN	NPN	N	NPN	-1.305	0.77
22	79-00-5	Ethane, 1,1,2-trichloro-	NPN	NPN	NPN	NPN	N	NPN	**-3.214**	2.01
23	79-06-1	2-Propenamide	MTA	MTA	ALKY-ARYL	MTA	R	R	**-2.767**	-0.81
24	79-10-7	2-Propenoic acid	MTA	REAC. ACID	ALKY-ARYL	UNK	R	R		0.44
25	79-11-8	Acetic acid, chloro-	SN2	REAC. ACID	ALKY-ARYL	UNK	R			0.34
26	79-20-9	Acetic acid, methyl ester	NPN	NPN	EN	NPN	N		-2.365	0.37
27	79-31-2	Propanoic acid, 2-methyl-	NPN_LogD	CARB. ACID	NPN	UNK	N*			1
28	79-34-5	Ethane, 1,1,2,2-tetrachloro-	NPN	NPN	NPN	NPN	N	NPN	-3.917	2.19
29	79-39-0	2-Propenamide, 2-methyl-	MTA	MTA	NPN	MTA	R	R		-0.26
30	79-41-4	2-Propenoic acid, 2-methyl-	MTA	REAC. ACID	NPN	UNK	R	R		0.99
31	80-05-7	Phenol, 4,4'-(1-methylethylidene)bis-	PN	PN	PN	PN	N	PN	-4.696	3.64
32	80-62-6	2-Propenoic acid, 2-methyl-, methyl ester	MTA	MTA	EN	MTA	R	R	**-2.552**	1.28
33	81-14-1	Ethanone, 1-[4-(1,1-dimethylethyl)-2,6-dimethyl-3,5-dinitrophenyl]-	PN	REAC.	REAC.DINITRO	UNK	R			4.31
34	81-15-2	Benzene, 1-(1,1-dimethylethyl)-3,5-dimethyl-2,4,6-trinitro-	PN	REAC.	REAC.DINITRO	UNK	R			4.45
35	84-74-2	1,2-Benzenedicarboxylic acid, dibutyl ester	PN	NPN	DE	UNK	N	R	**-5.306**	4.61
36	87-56-9	2-Butenoic acid, 2,3-dichloro-4-oxo-, (Z)-	MTA	REAC. ACID	ALKY-ARYL	UNK	R	R		1.37

Table I – SIDS test data (continued).

ID	*CASN*	*EINECS name*	*ECB-MOA*	*Schultz MOA*	*EPA-MOA*	*CONS1*	*CONS2*	*Verhaar MOA*	*LogLC50 (mol/l)*	*LogP*
37	88-12-0	2-Pyrrolidinone, 1-ethenyl-	MTA	REAC.	NPN	UNK	R			0.25
38	88-19-7	Benzenesulfonamide, 2-methyl-	UNK	NON SPEC. ELECT	NPN	UNK	UNK			0.92
39	88-44-8	Benzenesulfonic acid, 2-amino-5-methyl-	UNK	STRONG ACID	PN	UNK	UNK	R		-1.53
40	88-60-8	Phenol, 2-(1,1-dimethylethyl)-5-methyl-	PN	PN	PN	PN	N			3.97
41	88-73-3	Benzene, 1-chloro-2-nitro-	PN	SOFT ELECT	NPN	UNK	N	PN		2.46
42	88-74-4	Benzenamine, 2-nitro-	PN	PN	NPN	PN	N			2.02
43	91-15-6	1,2-Benzenedicarbonitrile	UNK	SOFT ELECT	NPN	UNK	UNK	R		1.09
44	91-76-9	1,3,5-Triazine-2,4-diamine, 6-phenyl-	CNS	CNS	PN	CNS	S			1.44
45	93-68-5	Butanamide, N-(2-methylphenyl)-3-oxo-	PN	NON SPEC. ELECT	REAC.DIKE	UNK	R		-2.782	0.99
46	94-36-0	Peroxide, dibenzoyl	UNK	NPN	SULPHY	UNK	UNK	R		3.43
47	95-31-8	2-Benzothiazolesulfenamide, N-(1,1-dimethylethyl)-	CNS	NPN	NPN	NPN	N			2.56
48	95-49-8	Benzene, 1-chloro-2-methyl-	PN	NPN	NPN	NPN	N	NPN		3.18
49	95-50-1	Benzene, 1,2-dichloro-	PN	NPN	NPN	NPN	N	NPN	-3.411	3.28
50	96-18-4	Propane, 1,2,3-trichloro-	NPN	NPN	NPN	NPN	N	NPN	-3.346	2.5
51	96-29-7	2-Butanone, oxime	NPN	NPN	NPN	NPN	N	R	**-2.014**	1.69
52	96-31-1	Urea, N,N'-dimethyl-	NPN	NPN	NPN	NPN	N	R		-0.62
53	96-33-3	2-Propenoic acid, methyl ester	MTA	MTA	ACRY	MTA	R	R		0.73
54	97-72-3	Propanoic acid, 2-methyl-, anhydride	UNK	REAC. HYD.	DE	UNK	R			1.24
55	98-07-7	Benzene, (trichloromethyl)-	PN	NPN	NPN	NPN	N	R		3.9

Table I – SIDS test data (continued).

ID	*CASN*	*EINECS name*	*ECB-MOA*	*Schultz MOA*	*EPA-MOA*	*CONS1*	*CONS2*	*Verhaar MOA*	*LogLC50 (mol/l)*	*LogP*
56	98-54-4	Phenol, 4-(1,1-dimethylethyl)-	PN	PN	PN	PN	N	PN	**-4.466**	3.42
57	98-59-9	Benzenesulfonyl chloride, 4-methyl-	UNK	SN2	SULPHY	UNK	R			3.49
58	98-92-0	3-Pyridinecarboxamide	PN	PN	PN	PN	N			-0.45
59	99-04-7	Benzoic acid, 3-methyl-	PN_Log D	CARB. ACID	NPN	UNK	N*			2.42
60	99-54-7	Benzene, 1,2-dichloro-4-nitro-	PN	SN2	NPN	UNK	N	PN		3.1
61	99-99-0	Benzene, 1-methyl-4-nitro-	PN	NON SPEC. ELECT	NPN	UNK	N	PN	-3.438	2.36
62	100-00-5	Benzene, 1-chloro-4-nitro-	PN	SN2	NPN	UNK	N	PN		2.46
63	100-21-0	1,4-Benzenedicarboxylic acid	PN_Log D	CARB. ACID	NPN	UNK	N*			1.76
64	100-37-8	Ethanol, 2-(diethylamino)-	NUC	AMIN.ALCH	NPN	UNK	UNK		-1.818	0.05
65	100-41-4	Benzene, ethyl-	PN	NPN	NPN	NPN	N	NPN	-3.943	3.03
66	102-06-7	Guanidine, N,N'-diphenyl-	UNK	NPN	NPN	NPN	N	R		2.89
67	102-76-1	1,2,3-Propanetriol, triacetate	EN	NPN	DE	UNK	UNK		-3.121	0.36
68	103-11-7	2-Propenoic acid, 2-ethylhexyl ester	MTA	MTA	ACRY	MTA	R	R		4.09
69	103-84-4	Acetamide, N-phenyl-	PN	NON SPEC. ELECT	NPN	UNK	N	R		1.1
70	105-60-2	2H-Azepin-2-one, hexahydro-	PN	NPN	NPN	NPN	N	R		0.66
71	106-31-0	Butanoic acid, anhydride	UNK	REAC. HYD.	DE	UNK	R	R		1.39
72	106-46-7	Benzene, 1,4-dichloro-	PN	NPN	NPN	NPN	N	NPN	-4.015	3.28
73	106-63-8	2-Propenoic acid, 2-methylpropyl ester	MTA	MTA	ACRY	MTA	R		-4.788	2.13
74	106-88-7	Oxirane, ethyl-	RAD	EPOX	ALKY-ARYL	UNK	R	R		0.86

Table I – SIDS test data (continued).

ID	*CASN*	*EINECS name*	*ECB-MOA*	*Schultz MOA*	*EPA-MOA*	*CONS1*	*CONS2*	*Verhaar MOA*	*LogLC50 (mol/l)*	*LogP*
75	107-06-2	Ethane, 1,2-dichloro-	NPN	NPN	NPN	NPN	N	NPN	-2.931	1.83
76	107-15-3	1,2-Ethanediamine	AN	NPN	REAC.	UNK	UNK		**-2.576**	-1.62
77	107-22-2	Ethanedial-	MTA	SB	CARB. REAC.	UNK	R	R	-2.431	-1.66
78	107-41-5	2,4-Pentanediol, 2-methyl-	NPN	NPN	NPN	NPN	N	NPN	**-1.089**	0.58
79	107-86-8	2-Butenal, 3-methyl-	MTA	MTA	ALKY-ARYL	MTA	R	R		1.15
80	107-92-6	Butanoic-acid-	NPN_log D	CARB. ACID	NPN	UNK	N*			1.07
81	107-98-2	2-Propanol, 1-methoxy-	NPN	NPN	NPN	NPN	N	NPN	-0.637	-0.49
82	108-44-1	Benzenamine, 3-methyl-	PN	PN	PN	PN	N	PN		1.62
83	108-65-6	2-Propanol, 1-methoxy-, acetate	EN	NPN	EN	EN	N			0.52
84	108-77-0	1,3,5-Triazine, 2,4,6-trichloro-	UNK	SN2	NPN	UNK	UNK	R		1.73
85	108-88-3	Benzene, methyl-	PN	NPN	NPN	NPN	N	NPN	-3.549	2.54
86	109-66-0	Pentane-	NPN	NPN	NPN	NPN	N	NPN		2.8
87	110-16-7	2-Butenedioic acid (Z)-	MTA	DICARB. ACID.	ALKY-ARYL	UNK	R		**-4.366**	0.05
88	110-19-0	Acetic acid, 2-methylpropyl ester	EN	NPN	EN	EN	N			1.77
89	110-65-6	2-Butyne-1,4-diol	PE	PE	ALKY-ARYL	PE	R	R	-3.206	-0.93
90	110-83-8	Cyclohexene-	PN	NPN	NPN	NPN	N			2.96
91	110-85-0	Piperazine-	PN	NPN	NPN	NPN	N			-0.8
92	110-93-0	5-Hepten-2-one, 6-methyl-	NPN	NPN	NPN	NPN	N		-3.167	2.06
93	110-98-5	2-Propanol, 1,1'-oxybis-	NPN	NPN	NPN	NPN	N	NPN		-0.64
94	112-57-2	1,2-Ethanediamine, N-(2-aminoethyl)-N'-[2-[(2-aminoethyl)amino]ethyl]-	NPN	NPN	NPN	NPN	N			-3.16
95	112-85-6	Docosanoic-acid-	NPN_log D	CARB. ACID	NPN	UNK	N*			9.91
96	115-07-1	1-Propene	NPN	NPN	NPN	NPN	N	NPN		1.68
97	115-11-7	1-Propene, 2-methyl-	NPN	NPN	NPN	NPN	N	NPN		2.23

Table I – SIDS test data (continued).

ID	*CASN*	*EINECS name*	*ECB-MOA*	*Schultz MOA*	*EPA-MOA*	*CONS1*	*CONS2*	*Verhaar MOA*	*LogLC50 (mol/l)*	*LogP*
98	115-86-6	Phosphoric acid, triphenyl ester	AChE	REAC.	NPN	UNK	UNK		**-5.594**	4.7
99	115-95-7	1,6-Octadien-3-ol, 3,7-dimethyl-, acetate	MTA	REAC.	EN	UNK	R			4.39
100	118-79-6	Phenol, 2,4,6-tribromo-	WARE	PN	PN	PN	N	PN	**-4.705**	4.18
101	119-47-1	Phenol, 2,2'-methylenebis[6-(1,1-dimethylethyl)-4-methyl-	NPN	NPN	PN	NPN	N	PN		7.97
102	120-61-6	1,4-Benzenedicarboxylic acid, dimethyl ester	PN_Log D	NPN	DE	UNK	N*			1.66
103	120-80-9	1,2-Benzenediol	PE_RAD	PE	PN	UNK	R		**-4.288**	1.03
104	120-83-2	Phenol, 2,4-dichloro-	PN	PN	PN	PN	N	PN	**-4.277**	2.8
105	121-91-5	1,3-Benzenedicarboxylic acid	PN_log D	DICARB. ACID.	NPN	UNK	N*			1.76
106	122-52-1	Phosphorous acid, triethyl ester	AChE	REAC.	NPN	UNK	UNK	R/S		0.74
107	122-99-6	Ethanol, 2-phenoxy-	PN	NPN	NPN	NPN	N		**-2.604**	1.1
108	123-54-6	2,4-Pentanedione	UNK	REAC. NON SPECIFIC	REAC.DIKE	UNK	R	NPN	**-2.860**	0.05
109	123-77-3	Diazenedicarboxamide-	MTA	NTAS	NPN	UNK	UNK	R		-3.89
110	123-86-4	Acetic acid, butyl ester	EN	NPN	EN	EN	N		-3.810	1.85
111	124-04-9	Hexanedioic-acid-	NPN_log D	DICARB. ACID.	NPN	UNK	N*		**-3.178**	0.23
112	126-73-8	Phosphoric-acid-tributyl-ester-	AChE	REAC. PHOSP.	OP-AchE	AChE	S	R/S	**-4.774**	3.82
113	126-98-7	2-Propenenitrile, 2-methyl-	MTA	MTA	NPN	MTA	R			0.76
114	127-19-5	Acetamide, N,N-dimethyl-	NPN	NON SPEC. ELECT	NPN	NPN	N	R		-0.49
115	128-37-0	Phenol, 2,6-bis(1,1-dimethylethyl)-4-methyl-	PN	PN	PN	PN	N	PN		5.03
116	135-19-3	2-Naphthalenol	PN	PN	UNCOUPL	PN	N		-4.620	2.69
117	140-88-5	2-Propenoic acid, ethyl ester	MTA	MTA	ACRY	MTA	R	R	**-4.603**	1.22

Table I – SIDS test data (continued).

ID	*CASN*	*EINECS name*	*ECB-MOA*	*Schultz MOA*	*EPA-MOA*	*CONS1*	*CONS2*	*Verhaar MOA*	*LogLC50 (mol/l)*	*LogP*
118	141-10-6	3,5,9-Undecatrien-2-one, 6,10-dimethyl-	MTA	MTA	ALKY-ARYL	MTA	R			4.43
119	141-32-2	2-Propenoic acid, butyl ester	MTA	MTA	ACRY	MTA	R	R		2.2
120	141-78-6	Acetic-acid-ethyl-ester-	EN	MTA	EN	EN	N		-2.583	0.86
121	141-97-9	Butanoic acid, 3-oxo-, ethyl ester	EN	REAC. NON SPECIFIC	REAC.DIKE	UNK	R	R		-0.2
122	144-55-8	Carbonic-acid-monosodium-salt-	PE	CARB. ACID	NPN	UNK	N*			-0.46
123	150-90-3	Butanedioic acid, disodium salt	NPN_log D	CARB. ACID	NPN	UNK	N*			-0.75
124	288-32-4	1H-Imidazole	UNK	NPN	NPN	NPN	N			0.06
125	461-58-5	Guanidine, cyano-	UNK	REAC.	NPN	UNK	UNK	R		-1.34
126	505-32-8	1-Hexadecen-3-ol, 3,7,11,15-tetramethyl-	NPN	PE	NPN	NPN	N			8.23
127	528-44-9	1,2,4-Benzenetricarboxylic acid	PN_log D	CARB. ACID	NPN	UNK	N*			0.95
128	552-30-7	5-Isobenzofurancarboxylic acid, 1,3-dihydro-1,3-dioxo-	UNK	REAC.	ACY	UNK	R	R		1.96
129	556-82-1	2-Buten-1-ol, 3-methyl-	PE	NPN	NPN	NPN	N	R		1.17
130	611-19-8	Benzene, 1-chloro-2-(chloromethyl)-	PN	NPN	ALKY-ARYL	UNK	N			3.44
131	760-23-6	1-Butene, 3,4-dichloro-	SN2	SN2	ALKY-ARYL	SN2	R	R	**-4.184**	2.6
132	770-35-4	2-Propanol, 1-phenoxy-	PN	NPN	NPN	NPN	N		-2.735	1.52
133	793-24-8	1,4-Benzenediamine, N-(1,3-dimethylbutyl)-N'-phenyl-	NPN	NPN	NPN	NPN	N	R/S		4.68
134	822-06-0	Hexane, 1,6-diisocyanato-	ISOCYA	REAC. HYD.	ISOCYA	UNK	R			3.2

Table I – SIDS test data (continued).

ID	*CASN*	*EINECS name*	*ECB-MOA*	*Schultz MOA*	*EPA-MOA*	*CONS1*	*CONS2*	*Verhaar MOA*	*LogLC50 (mol/l)*	*LogP*
135	839-90-7	1,3,5-Triazine-2,4,6(1H,3H,5H)-trione, 1,3,5-tris(2-hydroxyethyl)-	UNK	REAC.	NPN	UNK	UNK			0.07
136	868-77-9	2-Propenoic acid, 2-methyl-, 2-hydroxyethyl ester	MTA	MTA	EN	MTA	R	R	-2.758	0.3
137	868-85-9	Phosphonic acid, dimethyl ester	UNK	REAC.	NPN	UNK	UNK		-2.689	-1.13
138	919-30-2	3-Aminopropyl-triethoxysilane	UNK	REAC.	UNK	UNK	UNK			0.31
139	1163-19-5	Benzene, 1,1'-oxybis[2,3,4,5,6-pentabromo-	NPN	NPN	NPN	NPN	N	R/S		12.11
140	1477-55-0	1,3-Benzenedimethanamine	PN	NPN	NPN	NPN	N			0.15
141	1490-04-6	Cyclohexanol, 5-methyl-2-(1-methylethyl)-	NPN	NPN	NPN	NPN	N	NPN	-3.929	3.38
142	1634-04-4	Propane, 2-methoxy-2-methyl-	NPN	NPN	NPN	NPN	N	NPN	**-2.118**	1.43
143	1717-00-6	HCFC 141b	NPN	NPN	NPN	NPN	N			2.37
144	2216-51-5	Cyclohexanol, 5-methyl-2-(1-methylethyl)-, [1R-(1alpha,2beta,5alpha)]-	NPN	NPN	NPN	NPN	N			-1.67
145	2403-88-5	4-Piperidinol, 2,2,6,6-tetramethyl-	NPN	NPN	NPN	NPN	N			0.94
146	2432-99-7	Undecanoic acid, 11-amino-	NPN_log D	CARB. ACID	NPN	UNK	N*			-0.16
147	2439-35-2	2-Propenoic acid, 2-(dimethylamino)ethyl ester	MTA	MTA	ACRY	MTA	R	R		0.42
148	2837-89-0	Ethane, 2-chloro-1,1,1,2-tetrafluoro-	NPN	NPN	NPN	NPN	N			1.86
149	2855-13-2	Cyclohexanemethanamine, 5-amino-1,3,3-trimethyl-	NPN	NPN	NPN	NPN	N			1.9

Table I – SIDS test data (continued).

ID	*CASN*	*EINECS name*	*ECB-MOA*	*Schultz MOA*	*EPA-MOA*	*CONS1*	*CONS2*	*Verhaar MOA*	*LogLC50 (mol/l)*	*LogP*
150	2867-47-2	2-Propenoic acid, 2-methyl-, 2-(dimethylamino)ethyl ester	MTA	MTA	EN	MTA	R	R		0.97
151	3268-49-3	Propanal, 3-(methylthio)-	NPN	SB	CARB. REAC.	UNK	R	R		0.41
152	3319-31-1	1,2,4-Benzenetricarboxylic acid, tris(2-ethylhexyl) ester	NPN_log D	NPN	DE	UNK	N*			11.59
153	3323-53-3	Hexanedioic acid, compd. with 1,6-hexanediamine (1:1)	NPN_log D	NPN	NPN	NPN	N*			0.23
154	3452-97-9	1-Hexanol, 3,5,5-trimethyl-	NPN	NPN	NPN	NPN	N	NPN		3.11
155	4016-24-4	Hexadecanoic acid, 2-sulfo-, 1-methyl ester, sodium salt	NPN_log D	NPN	EN	UNK	N*			6.21
156	4169-04-4	1-Propanol, 2-phenoxy-	PN	NPN	NPN	NPN	N		-2.735	1.52
157	4454-05-1	2H-Pyran, 3,4-dihydro-2-methoxy-	PN	NPN	NPN	NPN	N	NPN		0.88
158	4457-71-0	1,5-Pentanediol, 3-methyl-	NPN	NPN	NPN	NPN	N			0.69
159	4979-32-2	2-Benzothiazolesulfenamide, N,N-dicyclohexyl-	CNS	NPN	NPN	NPN	N			5.96
160	5102-83-0	Butanamide, 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[N-(2,4-dimethylphenyl)-3-oxo-	NPN	REAC.	NPN	NPN	N			8.11
161	5392-40-5	2,6-Octadienal, 3,7-dimethyl-	MTA	MTA	ALKY-ARYL	MTA	R			3.45

Table I – SIDS test data (continued).

ID	*CASN*	*EINECS name*	*ECB-MOA*	*Schultz MOA*	*EPA-MOA*	*CONS1*	*CONS2*	*Verhaar MOA*	*LogLC50 (mol/l)*	*LogP*
162	5567-15-7	Butanamide, 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[N-(4-chloro-2,5-dimethoxyphenyl)-3-oxo-	NPN	REAC.	NPN	NPN	N			7.94
163	6165-51-1	Benzene, 1,4-dimethyl-2-(1-phenylethyl)-	PN	NPN	NPN	NPN	N			5.24
164	6358-85-6	Butanamide, 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[3-oxo-N-phenyl-	NPN	REAC.	NPN	NPN	N			7.06
165	6386-38-5	Benzenepropanoic acid, 3,5-bis(1,1-dimethylethyl)-4-hydroxy-, methyl ester	EN	NPN	EN	EN	N			5.06
166	6422-86-2	1,4-Benzenedicarboxylic acid, bis(2-ethylhexyl) ester	NPN	NPN	DE	NPN	N			8.39
167	6864-37-5	Cyclohexanamine, 4,4'-methylenebis[2-methyl-	AN	NPN	NPN	NPN	N			4.1
168	11070-44-3	1,3-Isobenzofurandione, tetrahydromethyl-	MTA	REAC.	CARB. Based	UNK	R			2.64
169	25154-52-3	Phenol, nonyl-	PN	PN	PN	PN	N	PN	-6.236	5.99
170	25265-71-8	Propanol, oxybis-	NPN	NPN	NPN	NPN	N	NPN		-0.49
171	25321-09-9	Benzene, bis(1-methylethyl)-	PN	NPN	NPN	NPN	N	NPN		4.9
172	25321-14-6	Benzene, methyldinitro-	PN	REAC.	REAC.DINITRO	UNK	R		-4.030	2.18
173	31570-04-4	Phenol, 2,4-bis(1,1-dimethylethyl)-, phosphite (3:1)	NPN	NPN	NPN	NPN	N			18.08
174	32534-81-9	Benzene, 1,1'-oxybis-, pentabromo deriv.	NPN	NPN	NPN	NPN	N			7.66
175	32536-52-0	Benzene, 1,1'-oxybis-, octabromo deriv.	NPN	NPN	NPN	NPN	N			10.33
176	56539-66-3	1-Butanol, 3-methoxy-3-methyl-	PN	NPN	NPN	NPN	N			0.46
177	84852-15-3	Phenol, 4-nonyl-, branched	PN	PN	PN	PN	N		-6.236	5.92

AQUIRE values are highlighted in bold.

Table II – NPN model training set.

ID	*CASN*	*Chemical*	*LogK_{ow}*	*LogLC50 (mol/l)*		*Hat*	*Err.Calc.*	*Err.Pred.*
				Exp	*Pred.*			
1	71-36-3	1-butanol	0.88	-1.63	-2.10	0.03	-0.46	-0.47
2	112-30-1	1-decanol	4.57	-4.81	-5.30	0.06	-0.46	-0.49
3	112-53-8	1-dodecanol	5.13	-5.26	-5.79	0.08*	-0.49	-0.53
4	111-27-3	1-hexanol	2.03	-3.02	-3.08	0.02	-0.06	-0.06
5	143-08-8	1-nonanol	4.26	-4.40	-5.03	0.05	-0.60	-0.63
6	111-87-5	1-octanol	2.97	-3.98	-3.89	0.02	0.09	0.09
7	112-42-5	1-undecanol	4.52	-5.21	-5.23	0.06	-0.02	-0.02
8	79-00-5	1,1,2-trichloroethane	1.89	-3.21	-2.95	0.02	0.25	0.26
9	79-34-5	1,1,2,2-tetrachloroethane	2.39	-3.91	-3.38	0.02	0.52	0.53
10	107-06-2	1,2-dichloroethane	1.48	-2.92	-2.60	0.02	0.31	0.32
11	79-34-5	1,1,2,2-tetrachloroethane	4.63	-5.29	-5.32	0.06	-0.03	-0.03
12	120-82-1	1,2,4-trichlorobenzene	4.05	-4.79	-4.82	0.04	-0.03	-0.03
13	541-73-1	1,3-dichlorobenzene	3.52	-4.27	-4.37	0.03	-0.09	-0.10
14	106-46-7	1,4-dichlorobenzene	3.44	-4.56	-4.29	0.03	0.26	0.27
15	150-78-7	1,4-dimethoxybenzene	2.15	-3.07	-3.19	0.02	-0.11	-0.12
16	111-90-0	2-(2-ethoxyethoxy)ethanol	-0.54	-0.70	-0.88	0.06	-0.16	-0.18
17	78-93-3	2-butanone	0.29	-1.35	-1.59	0.04	-0.23	-0.24
18	693-54-9	2-decanone	3.73	-4.43	-4.55	0.04	-0.12	-0.12
19	552-41-0	2-hydroxy-4-methoxyacetophenone	1.98	-3.48	-3.03	0.02	0.44	0.45
20	78-83-1	2-methyl-1-propanol	0.76	-1.71	-1.99	0.03	-0.28	-0.28
21	107-41-5	2-methyl-2,4-pentanediol	-0.67	-1.04	-0.73	0.07	0.29	0.31
22	111-13-7	2-octanone	2.37	-3.55	-3.37	0.02	0.18	0.18
23	122-99-6	2-phenoxyethanol	1.16	-2.60	-2.32	0.02	0.27	0.28
24	67-63-0	2-propanol	0.05	-0.76	-1.40	0.05	-0.61	-0.64
25	115-20-8	2,2,2-trichloroethanol	1.42	-2.69	-2.55	0.02	0.14	0.14
26	13608-87-2	2,3,4-trichloroacetophenone	3.57	-5.04	-4.39	0.03	0.63	0.65
27	13909-73-4	2,3,4-trimethoxyacetophenone	1.12	-3.08	-2.28	0.02	0.78	0.80
28	937-20-2	2,4-dichloroacetophenone	2.84	-4.20	-3.77	0.02	0.42	0.43
29	5673-07-4	2,6-dimethoxytoluene	2.64	-3.87	-3.60	0.02	0.26	0.27
30	4412-91-3	3-furanmethanol	0.30	-2.28	-1.56	0.04	0.69	0.72
31	563-80-4	3-methyl-2-butanone	0.56	-1.99	-1.81	0.03	0.18	0.18
32	96-22-0	3-pentanone	0.79	-1.74	-2.02	0.03	-0.27	-0.28
33	75-97-8	3,3-dimethyl-2-butanone	0.96	-3.06	-2.13	0.03	0.90 *	0.93 *
34	95-75-0	3,4-dichlorotoluene	4.06	-4.74	-4.83	0.04	-0.09	-0.09
35	108-10-1	4-methyl-2-pentanone	1.31	-2.29	-2.46	0.02	-0.17	-0.17
36	110-12-3	5-methyl-2-hexanone	1.88	-2.85	-2.95	0.02	-0.10	-0.10
37	502-56-7	5-nonanone	2.90	-3.66	-3.83	0.02	-0.17	-0.17
38	110-93-0	6-methyl-5-hepten-2-one	1.70	-3.16	-2.79	0.02	0.36	0.37
39	67-64-1	Acetone	-0.24	-0.85	-1.14	0.05	-0.27	-0.29
40	98-86-2	Acetophenone	1.58	-2.87	-2.69	0.02	0.18	0.18

Table II – NPN model training set (continued).

ID	*CASN*	*Chemical*	*LogK_{ow}*	*LogLC50 (mol/l)*		*Hat*	*Err.Calc.*	*Err.Pred.*
				Exp	*Pred.*			
41	119-61-9	Benzophenone	3.18	-4.07	-4.07	0.03	0.00	0.00
42	108-93-0	Cyclohexanol	1.23	-2.15	-2.40	0.02	-0.24	-0.25
43	108-94-1	Cyclohexanone	0.81	-2.27	-2.02	0.03	0.24	0.25
44	142-96-1	Dibutyl ether	3.21	-3.60	-4.11	0.03	-0.50	-0.51
45	108-20-3	Diisopropyl ether	1.52	-3.04	-2.63	0.02	0.40	0.41
46	693-65-2	Dipentyl ether	4.04	-4.69	-4.82	0.04	-0.12	-0.13
47	101-84-8	Diphenyl ether	4.21	-4.62	-4.98	0.05	-0.34	-0.36
48	64-17-5	Ethanol	-0.31	0.51	-1.16	0.06	-1.57 **	-1.67 **
49	110-00-9	Furan	1.34	-3.04	-2.47	0.02	0.55	0.57
50	67-72-1	Hexachloroethane	4.14	-5.19	-4.88	0.05	0.29	0.31
51	67-56-1	Methanol	-0.77	-0.06	-0.71	0.07 *	-0.61	-0.65
52	620-88-2	4-nitrophenyl phenylether	4.28	-4.90	-5.03	0.05	-0.12	-0.13
53	76-01-7	Pentachloroethane	3.62	-4.44	-4.45	0.03	-0.01	-0.01
54	1634-04-4	tert-butylmethyl ether	0.94	-2.09	-2.14	0.03	-0.05	-0.05
55	127-18-4	Tetrachloroethene	3.40	-4.08	-4.27	0.03	-0.18	-0.19
56	109-99-9	Tetrahydrofuran	0.46	-1.52	-1.73	0.04	-0.21	-0.21
57	79-01-6	Trichloroethene	2.42	-3.47	-3.42	0.02	0.05	0.05
58	112-27-6	Triethylene glycol	-1.24	-0.33	-0.25	0.09 *	0.07	0.08

* Chemicals with values between 2 times SDEC (or SDEP or critical HAT) and 3 times SDEC (or SDEP or critical HAT). ** Chemicals with values greater than 3 times SDEC (or SDEP or average value of HAT).

Table III – SIDS chemicals not suitable for QSAR 1.

	N.Comp.	SIDS Chemicals							Motivation
	28	2 94 144 164	6 95 152 166	7 101 155 169	17 109 159 173	39 125 160 174	76 126 162 175	77 139 163 177	Out of the X - domain (-1.24≤ LogKow ≤ 5.13)
	8	9 142	22	28	72	75	92	107	In the training set
XY-domain + MOA domain	90	1 25 34 42 56 63 74 88 104 113 121 134 150	4 27 35 43 57 64 79 89 105 115 122 135 151	15 29 36 44 58 67 80 98 106 116 123 136 161	18 30 37 45 59 68 82 99 108 117 127 137 165	19 31 38 46 60 69 83 100 110 118 128 138 168	23 32 40 53 61 71 84 102 111 119 130 146 172	24 33 41 54 62 73 87 103 112 120 131 147	MOA ≠ NPN
XY-domain	16	1 103 137	18 108 172	23 110	67 111	73 116	87 117	89 136	Y-Outliers (cross-validated standardised residual greater than two standard deviation units)

Table IV –QSAR 1 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) and mode of action domain (MOA-D).

ID	*CASN*	*EINECS name*	*LogK$_{ow}$*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
				Exp	*Pred.*		
S3	57556	1,2-Propanediol	-0.78	-0.84	-0.66	0.071	0.45
S5	58559	1H-Purine-2,6-dione, 3,7-dihydro-1,3-dimethyl-	-0.39	-	-0.99	0.057	-
S8	68-12-2	Formamide, N,N-dimethyl-	-0.93	-0.84	-0.53	0.077	0.79
S10	74839	Methane, bromo-	1.18	-	-2.35	0.023	-
S11	74873	Methane, chloro-	1.09	-	-2.27	0.024	-
S12	75014	Ethene, chloro-	1.62	-	-2.73	0.019	-
S13	75105	Methane, difluoro-	0.71	-	-1.94	0.030	-
S14	75387	Ethene, 1,1-difluoro-	1.24	-	-2.40	0.022	-
S16	75683	Ethane, 1-chloro-1,1-difluoro-	2.05	-	-3.10	0.017	-
S20	78875	Propane, 1,2-dichloro-	2.25	-2.91	-3.27	0.017	-0.88
S21	78922	2-Butanol	0.77	-1.31	-1.99	0.029	-1.68
S26	79209	Acetic acid, methyl ester	0.37	-2.36	-1.65	0.036	1.76
S47	95318	2-Benzothiazolesulfenamide, N-(1,1-dimethylethyl)-	2.56	-	-3.54	0.019	-
S48	95498	Benzene, 1-chloro-2-methyl-	3.18	-	-4.07	0.025	-
S49	95501	Benzene, 1,2-dichloro-	3.28	-3.41	-4.16	0.027	-1.85
S50	96184	Propane, 1,2,3-trichloro-	2.50	-3.35	-3.49	0.018	-0.34
S51	96297	2-Butanone, oxime	1.69	-2.01	-2.79	0.018	-1.90
S52	96311	Urea, N,N'-dimethyl-	-0.62	-	-0.80	0.065	-
S55	98077	Benzene, (trichloromethyl)-	3.90	-	-4.69	0.039	-
S65	100414	Benzene, ethyl-	3.03	-3.94	-3.94	0.023	0.00
S66	102067	Guanidine, N,N'-diphenyl-	2.89	-	-3.82	0.021	-
S70	105602	2H-Azepin-2-one, hexahydro-	0.66	-	-1.90	0.031	-
S78	107415	2,4-Pentanediol, 2-methyl-	0.58	-	-1.83	0.032	-
S81	107982	2-Propanol, 1-methoxy-	-0.49	-0.64	-0.91	0.061	-0.68
S85	108883	Benzene, methyl-	2.54	-3.55	-3.52	0.019	0.07
S86	109660	Pentane-	2.80	-	-3.74	0.021	-
S90	110838	Cyclohexene-	2.96	-	-3.88	0.022	-
S91	110850	Piperazine-	-0.80	-	-0.64	0.072	-
S93	110985	2-Propanol, 1,1'-oxybis-	-0.64	-	-0.78	0.066	-
S96	115071	1-Propene	1.68	-	-2.78	0.018	-
S97	115117	1-Propene, 2-methyl-	2.23	-	-3.25	0.017	-
S114	127195	Acetamide, N,N-dimethyl-	-0.49	-	-0.91	0.061	-
S124	288324	1H-Imidazole	0.06	-	-1.38	0.044	-
S129	556821	2-Buten-1-ol, 3-methyl-	1.17	-	-2.34	0.023	-
S132	770354	2-Propanol, 1-phenoxy-	1.52	-2.74	-2.64	0.019	0.25
S133	793248	1,4-Benzenediamine, N-(1,3-dimethylbutyl)-N'-phenyl-	4.68	-	-5.36	0.061	-
S140	1477550	1,3-Benzenedimethanamine	0.15	-	-1.46	0.042	-

Table IV –QSAR 1 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) and mode of action domain (MOA-D) (continued).

ID	*CASN*	*EINECS name*	*LogK$_{ow}$*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
				Exp	*Pred.*		
S141	1490046	Cyclohexanol, 5-methyl-2-(1-methylethyl)-	3.38	-3.93	-4.24	0.028	-0.77
S143	1717006	HCFC 141b	2.37	-	-3.37	0.018	-
S145	2403885	4-Piperidinol, 2,2,6,6-tetramethyl-	0.94	-	-2.14	0.026	-
S148	2837890	Ethane, 2-chloro-1,1,1,2-tetrafluoro-	1.86	-	-2.93	0.018	-
S149	2855132	Cyclohexanemethanamine, 5-amino-1,3,3-trimethyl-	1.90	-	-2.97	0.017	-
S153	3323533	Hexanedioic acid, compd. with 1,6-hexanediamine (1:1)	0.23	-	-1.53	0.040	-
S154	3452979	1-Hexanol, 3,5,5-trimethyl-	3.11	-	-4.01	0.024	-
S156	4169044	1-Propanol, 2-phenoxy-	1.52	-2.74	-2.64	0.019	0.25
S157	4454051	2H-Pyran, 3,4-dihydro-2-methoxy-	0.88	-	-2.09	0.027	-
S158	4457710	1,5-Pentanediol, 3-methyl-	0.69	-	-1.92	0.030	-
S167	6864375	Cyclohexanamine, 4,4'-methylenebis[2-methyl-	4.10	-	-4.86	0.044	-
S170	25265718	Propanol, oxybis-	-0.49	-	-0.91	0.061	-
S171	25321099	Benzene, bis(1-methylethyl)-	4.90	-	-5.55	0.069	-
S176	56539663	1-Butanol, 3-methoxy-3-methyl-	0.46	-	-1.73	0.034	-

Table V –QSAR 1 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D).

ID	*CASN*	*EINECS name*	*MOA*	*LogK$_{ow}$*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
					Exp	*Pred.*		
S1	50000	Formaldehyde-	SB	0.35	-3.08	-1.63	0.037	**3.60**
S3	57556	1,2-Propanediol	NPN	-0.78	-0.84	-0.66	0.071	0.45
S4	58082	1H-Purine-2,6-dione, 3,7-dihydro-1,3,7-trimethyl-	CNS	0.16	-	-1.47	0.041	-
S5	58559	1H-Purine-2,6-dione, 3,7-dihydro-1,3-dimethyl-	NPN	-0.39	-	-0.99	0.057	-
S8	68122	Formamide, N,N-dimethyl-	NPN	-0.93	-0.84	-0.53	0.077	0.79
S10	74839	Methane, bromo-	NPN	1.18	-	-2.35	0.023	-
S11	74873	Methane, chloro-	NPN	1.09	-	-2.27	0.024	-
S12	75014	Ethene, chloro-	NPN	1.62	-	-2.73	0.019	-
S13	75105	Methane, difluoro-	NPN	0.71	-	-1.94	0.030	-
S14	75387	Ethene, 1,1-difluoro-	NPN	1.24	-	-2.40	0.022	-
S15	75569	Oxirane, methyl-	UNK	0.37	-	-1.65	0.036	-
S16	75683	Ethane, 1-chloro-1,1-difluoro-	NPN	2.05	-	-3.10	0.017	-
S18	78591	2-Cyclohexen-1-one, 3,5,5-trimethyl-	MTA	2.62	-2.76	-3.59	0.019	-2.03
S19	78706	1,6-Octadien-3-ol, 3,7-dimethyl-	PE	3.38	-	-4.24	0.028	-
S20	78875	Propane, 1,2-dichloro-	NPN	2.25	-2.91	-3.27	0.017	-0.88
S21	78922	2-Butanol	NPN	0.77	-1.31	-1.99	0.029	-1.68
S23	79061	2-Propenamide	MTA	-0.81	-2.77	-0.63	0.072	**5.40**
S24	79107	2-Propenoic acid	UNK	0.44	-	-1.71	0.035	-
S25	79118	Acetic acid, chloro-	UNK	0.34	-	-1.62	0.037	-
S26	79209	Acetic acid, methyl ester	NPN	0.37	-2.36	-1.65	0.036	1.76
S27	79312	Propanoic acid, 2-methyl-	UNK	1.00	-	-2.19	0.025	-
S29	79390	2-Propenamide, 2-methyl-	MTA	-0.26	-	-1.11	0.053	-
S30	79414	2-Propenoic acid, 2-methyl-	UNK	0.99	-	-2.18	0.025	-
S31	80057	Phenol, 4,4'-(1-methylethylidene)bis-	PN	3.64	-4.70	-4.47	0.033	0.57
S32	80626	2-Propenoic acid, 2-methyl-, methyl ester	MTA	1.28	-2.55	-2.43	0.021	0.30
S33	81141	Ethanone, 1-[4-(1,1-dimethylethyl)-2,6-dimethyl-3,5-dinitrophenyl]-	UNK	4.31	-	-5.05	0.050	-
S34	81152	Benzene, 1-(1,1-dimethylethyl)-3,5-dimethyl-2,4,6-trinitro-	UNK	4.45	-	-5.17	0.054	-
S35	84742	1,2-Benzenedicarboxylic acid, dibutyl ester	UNK	4.61	-5.31	-5.30	0.059	0.01
S36	87569	2-Butenoic acid, 2,3-dichloro-4-oxo-, (Z)-	UNK	1.37	-	-2.51	0.021	-
S37	88120	2-Pyrrolidinone, 1-ethenyl-	UNK	0.25	-	-1.55	0.039	-

Table V –QSAR 1 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) (continued).

ID	*CASN*	*EINECS name*	*MOA*	*LogK_{ow}*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
					Exp	*Pred.*		
S38	88197	Benzenesulfonamide, 2-methyl-	UNK	0.92	-	-2.12	0.026	-
S40	88608	Phenol, 2-(1,1-dimethylethyl)-5-methyl-	PN	3.97	-	-4.75	0.040	-
S41	88733	Benzene, 1-chloro-2-nitro-	UNK	2.46	-	-3.45	0.018	-
S42	88744	Benzenamine, 2-nitro-	PN	2.02	-	-3.07	0.017	-
S43	91156	1,2-Benzenedicarbonitrile	UNK	1.09	-	-2.27	0.024	-
S44	91769	1,3,5-Triazine-2,4-diamine, 6-phenyl-	CNS	1.44	-	-2.57	0.020	-
S45	93685	Butanamide, N-(2-methylphenyl)-3-oxo-	UNK	0.99	-2.78	-2.18	0.025	1.48
S46	94360	Peroxide, dibenzoyl	UNK	3.43	-	-4.29	0.029	-
S47	95318	2-Benzothiazolesulfenamide, N-(1,1-dimethylethyl)-	NPN	2.56	-	-3.54	0.019	-
S48	95498	Benzene, 1-chloro-2-methyl-	NPN	3.18	-	-4.07	0.025	-
S49	95501	Benzene, 1,2-dichloro-	NPN	3.28	-3.41	-4.16	0.027	-1.85
S50	96184	Propane, 1,2,3-trichloro-	NPN	2.50	-3.35	-3.49	0.018	-0.34
S51	96297	2-Butanone, oxime	NPN	1.69	-2.01	-2.79	0.018	-1.90
S52	96311	Urea, N,N'-dimethyl-	NPN	-0.62	-	-0.80	0.065	-
S53	96333	2-Propenoic acid, methyl ester	MTA	0.73	-	-1.96	0.029	-
S54	97723	Propanoic acid, 2-methyl-, anhydride	UNK	1.24	-	-2.40	0.022	-
S55	98077	Benzene, (trichloromethyl)-	NPN	3.90	-	-4.69	0.039	-
S56	98544	Phenol, 4-(1,1-dimethylethyl)-	PN	3.42	-4.47	-4.28	0.029	0.46
S57	98599	Benzenesulfonyl chloride, 4-methyl-	UNK	3.49	-	-4.34	0.030	-
S58	98920	3-Pyridinecarboxamide	PN	-0.45	-	-0.94	0.059	-
S59	99047	Benzoic acid, 3-methyl-	UNK	2.42	-	-3.42	0.018	-
S60	99547	Benzene, 1,2-dichloro-4-nitro-	UNK	3.10	-	-4.00	0.024	-
S61	99990	Benzene, 1-methyl-4-nitro-	UNK	2.36	-3.44	-3.36	0.018	0.20
S62	100005	Benzene, 1-chloro-4-nitro-	UNK	2.46	-	-3.45	0.018	-
S63	100210	1,4-Benzenedicarboxylic acid	UNK	1.76	-	-2.85	0.018	-
S64	100378	Ethanol, 2-(diethylamino)-	UNK	0.05	-1.82	-1.37	0.044	1.12
S65	100414	Benzene, ethyl-	NPN	3.03	-3.94	-3.94	0.023	0.00
S66	102067	Guanidine, N,N'-diphenyl-	NPN	2.89	-	-3.82	0.021	-
S67	102761	1,2,3-Propanetriol, triacetate	UNK	0.36	-3.12	-1.64	0.037	**3.67**
S68	103117	2-Propenoic acid, 2-ethylhexyl ester	MTA	4.09	-	-4.86	0.044	-
S69	103844	Acetamide, N-phenyl-	UNK	1.10	-	-2.28	0.024	-
S70	105602	2H-Azepin-2-one, hexahydro-	NPN	0.66	-	-1.90	0.031	-
S71	106310	Butanoic acid, anhydride	UNK	1.39	-	-2.53	0.020	-
S73	106638	2-Propenoic acid, 2-methylpropyl ester	MTA	2.13	-4.79	-3.17	0.017	**3.98**

Table V –QSAR 1 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) (continued).

ID	*CASN*	*EINECS name*	*MOA*	*LogK_{ow}*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
					Exp	*Pred.*		
S74	106887	Oxirane, ethyl-	UNK	0.86	-	-2.07	0.027	-
S78	107415	2,4-Pentanediol, 2-methyl-	NPN	0.58	-1.09	-1.83	0.032	-1.83
S79	107868	2-Butenal, 3-methyl-	MTA	1.15	-	-2.32	0.023	-
S80	107926	Butanoic-acid-	UNK	1.07	-	-2.25	0.024	-
S81	107982	2-Propanol, 1-methoxy-	NPN	-0.49	-0.64	-0.91	0.061	-0.68
S82	108441	Benzenamine, 3-methyl-	PN	1.62	-	-2.73	0.019	-
S83	108656	2-Propanol, 1-methoxy-, acetate	EN	0.52	-	-1.78	0.033	-
S84	108770	1,3,5-Triazine, 2,4,6-trichloro-	UNK	1.73	-	-2.82	0.018	-
S85	108883	Benzene, methyl-	NPN	2.54	-3.55	-3.52	0.019	0.07
S86	109660	Pentane-	NPN	2.80	-	-3.74	0.021	-
S87	110167	2-Butenedioic acid (Z)-	UNK	0.05	-4.37	-1.37	0.044	**7.45**
S88	110190	Acetic acid, 2-methylpropyl ester	EN	1.77	-	-2.86	0.018	-
S89	110656	2-Butyne-1,4-diol	PE	-0.93	-3.21	-0.53	0.077	**6.79**
S90	110838	Cyclohexene-	NPN	2.96	-	-3.88	0.022	-
S91	110850	Piperazine-	NPN	-0.80	-	-0.64	0.072	-
S93	110985	2-Propanol, 1,1'-oxybis-	NPN	-0.64	-	-0.78	0.066	-
S96	115071	1-Propene	NPN	1.68	-	-2.78	0.018	-
S97	115117	1-Propene, 2-methyl-	NPN	2.23	-	-3.25	0.017	-
S98	115866	Phosphoric acid, triphenyl ester	UNK	4.70	-5.59	-5.38	0.062	0.54
S99	115957	1,6-Octadien-3-ol, 3,7-dimethyl-, acetate	UNK	4.39	-	-5.11	0.052	-
S100	118796	Phenol, 2,4,6-tribromo-	PN	4.18	-4.71	-4.93	0.046	-0.56
S102	120616	1,4-Benzenedicarboxylic acid, dimethyl ester	UNK	1.66	-	-2.76	0.018	-
S103	120809	1,2-Benzenediol	UNK	1.03	-4.29	-2.22	0.025	**5.09**
S104	120832	Phenol, 2,4-dichloro-	PN	2.80	-4.28	-3.74	0.021	1.32
S105	121915	1,3-Benzenedicarboxylic acid	UNK	1.76	-	-2.85	0.018	-
S106	122521	Phosphorous acid, triethyl ester	UNK	0.74	-	-1.97	0.029	-
S108	123546	2,4-Pentanedione	UNK	0.05	-2.86	-1.37	0.044	**3.71**
S110	123864	Acetic acid, butyl ester	EN	1.85	-3.81	-2.92	0.018	**2.19**
S111	124049	Hexanedioic-acid-	UNK	0.23	-3.18	-1.53	0.040	**4.09**
S112	126738	Phosphoric-acid-tributyl-ester-	AChE	3.82	-4.77	-4.62	0.037	0.38
S113	126987	2-Propenenitrile, 2-methyl-	MTA	0.76	-	-1.99	0.029	-
S114	127195	Acetamide, N,N-dimethyl-	NPN	-0.49	-	-0.91	0.061	-
S115	128370	Phenol, 2,6-bis(1,1-dimethylethyl)-4-methyl-	PN	5.03	-	-5.67	0.074	-
S116	135193	2-Naphthalenol	PN	2.69	-4.62	-3.65	0.020	**2.38**
S117	140885	2-Propenoic acid, ethyl ester	MTA	1.22	-4.60	-2.38	0.022	**5.47**
S118	141106	3,5,9-Undecatrien-2-one, 6,10-dimethyl-	MTA	4.43	-	-5.15	0.053	-

Table V –QSAR 1 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) (continued).

ID	*CASN*	*EINECS name*	*MOA*	*LogK_{ow}*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
					Exp	*Pred.*		
S119	141322	2-Propenoic acid, butyl ester	MTA	2.20	-	-3.23	0.017	-
S120	141786	Acetic-acid-ethyl-ester-	EN	0.86	-2.58	-2.07	0.027	1.26
S121	141979	Butanoic acid, 3-oxo-, ethyl ester	UNK	-0.20	-	-1.16	0.051	-
S122	144558	Carbonic-acid-monosodium-salt-	UNK	-0.46	-	-0.93	0.060	-
S123	150903	Butanedioic acid, disodium salt	UNK	-0.75	-	-0.68	0.070	-
S124	288324	1H-Imidazole	NPN	0.06	-	-1.38	0.044	-
S127	528449	1,2,4-Benzenetricarboxylic acid	UNK	0.95	-	-2.15	0.026	-
S128	552307	5-Isobenzofurancarboxylic acid, 1,3-dihydro-1,3-dioxo-	UNK	1.96	-	-3.02	0.017	-
S129	556821	2-Buten-1-ol, 3-methyl-	NPN	1.17	-	-2.34	0.023	-
S130	611198	Benzene, 1-chloro-2-(chloromethyl)-	UNK	3.44	-	-4.30	0.029	-
S131	760236	1-Butene, 3,4-dichloro-	SN2	2.60	-4.18	-3.57	0.019	1.51
S132	770354	2-Propanol, 1-phenoxy-	NPN	1.52	-2.74	-2.64	0.019	0.25
S133	793248	1,4-Benzenediamine, N-(1,3-dimethylbutyl)-N'-phenyl-	NPN	4.68	-	-5.36	0.061	-
S134	822060	Hexane, 1,6-diisocyanato-1,3,5-Triazine-	UNK	3.20	-	-4.09	0.025	-
S135	839907	2,4,6(1H,3H,5H)-trione, 1,3,5-tris(2-hydroxyethyl)-	UNK	0.07	-	-1.39	0.044	-
S136	868779	2-Propenoic acid, 2-methyl-, 2-hydroxyethyl ester	MTA	0.30	-2.76	-1.59	0.038	**2.90**
S137	868859	Phosphonic acid, dimethyl ester	UNK	-1.13	-2.69	-0.36	0.085	**5.93**
S138	919302	3-Aminopropyl-triethoxysilane	UNK	0.31	-	-1.60	0.038	-
S140	1477550	1,3-Benzenedimethanamine	NPN	0.15	-	-1.46	0.042	-
S141	1490046	Cyclohexanol, 5-methyl-2-(1-methylethyl)-	NPN	3.38	-3.93	-4.24	0.028	-0.77
S143	1717006	HCFC 141b	NPN	2.37	-	-3.37	0.018	-
S145	2403885	4-Piperidinol, 2,2,6,6-tetramethyl-	NPN	0.94	-	-2.14	0.026	-
S146	2432997	Undecanoic acid, 11-amino-	UNK	-0.16	-	-1.19	0.050	-
S147	2439352	2-Propenoic acid, 2-(dimethylamino)ethyl ester	MTA	0.42	-	-1.69	0.035	-
S148	2837890	Ethane, 2-chloro-1,1,1,2-tetrafluoro-	NPN	1.86	-	-2.93	0.018	-
S149	2855132	Cyclohexanemethanamine, 5-amino-1,3,3-trimethyl-	NPN	1.90	-	-2.97	0.017	-
S150	2867472	2-Propenoic acid, 2-methyl-, 2-(dimethylamino)ethyl ester	MTA	0.97	-	-2.17	0.025	-
S151	3268493	Propanal, 3-(methylthio)-	UNK	0.41	-	-1.68	0.036	-

Table V –QSAR 1 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) (continued).

ID	*CASN*	*EINECS name*	*MOA*	*LogK_{ow}*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
					Exp	*Pred.*		
S153	3323533	Hexanedioic acid, compd. with 1,6-hexanediamine (1:1)	NPN	0.23	-	-1.53	0.040	-
S154	3452979	1-Hexanol, 3,5,5-trimethyl-	NPN	3.11	-	-4.01	0.024	-
S156	4169044	1-Propanol, 2-phenoxy-	NPN	1.52	-2.74	-2.64	0.019	0.25
S157	4454051	2H-Pyran, 3,4-dihydro-2-methoxy-	NPN	0.88	-	-2.09	0.027	-
S158	4457710	1,5-Pentanediol, 3-methyl-	NPN	0.69	-	-1.92	0.030	-
S161	5392405	2,6-Octadienal, 3,7-dimethyl-	MTA	3.45	-	-4.30	0.029	-
S165	6386385	Benzenepropanoic acid, 3,5-bis(1,1-dimethylethyl)-4-hydroxy-, methyl ester	EN	5.06	-	-5.69	0.075	-
S167	6864375	Cyclohexanamine, 4,4'-methylenebis[2-methyl-	NPN	4.10	-	-4.86	0.044	-
S168	11070443	1,3-Isobenzofurandione, tetrahydromethyl-	UNK	2.64	-	-3.61	0.019	-
S170	25265718	Propanol, oxybis-	NPN	-0.49	-	-0.91	0.061	-
S171	25321099	Benzene, bis(1-methylethyl)-	NPN	4.90	-	-5.55	0.069	-
S172	25321146	Benzene, methyldinitro-	UNK	2.18	-4.03	-3.21	0.017	**2.01**
S176	56539663	1-Butanol, 3-methoxy-3-methyl-	NPN	0.46	-	-1.73	0.034	-

Y outliers are highlighted in bold in the standardized residual in prediction column.

Table VI – outliers predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) in QSAR 1.

ID	*CASN*	*EINECS name*	*MOA*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
				Exp	*Pred.*		
S1	50000	Formaldehyde	SB	-3.08	-1.63	0.037	3.60
S18	78591	2-Cyclohexen-1-one, 3,5,5-trimethyl	MTA	-2.76	-3.59	0.019	-2.03
S23	79061	2-Propenamide	MTA	-2.77	-0.63	0.072	5.40
S67	102761	1,2,3-Propanetriol, triacetate	UNK	-3.12	-1.64	0.037	3.67
S73	106638	2-Propenoic acid, 2-methylpropyl ester	MTA	-4.79	-3.17	0.017	3.98
S87	110167	2-Butenedioic acid (Z)-	UNK	-4.37	-1.37	0.044	7.45
S89	110656	2-Butyne-1,4-diol	PE	-3.21	-0.53	0.077	6.79
S103	120809	1,2-Benzenediol	UNK	-4.29	-2.22	0.025	5.09
S108	123546	2,4-Pentanedione	UNK	-2.86	-1.37	0.044	3.71
S110	123864	Acetic acid, butyl ester	EN	-3.81	-2.92	0.018	2.19
S111	124049	Hexanedioic-acid-	UNK	-3.18	-1.53	0.040	4.09
S116	135193	2-Naphthalenol	PN	-4.62	-3.65	0.020	2.38
S117	140885	2-Propenoic acid, ethyl ester	MTA	-4.60	-2.38	0.022	5.47
S136	868779	2-Propenoic acid, 2-methyl-, 2-hydroxyethyl ester	MTA	-2.76	-1.59	0.038	2.90
S137	868859	Phosphonic acid, dimethyl ester	UNK	-2.69	-0.36	0.085	5.93
S172	25321146	Benzene, methyldinitro-	UNK	-4.03	-3.21	0.017	2.01

Table VII – NPN model performance on the two to subset of SIDS data evaluated.

Test MOA	*N. Chemicals*	*N.Test*	*Unknown SIDS predictions*	*Total SIDS predictions*	Q^2ext
NPN	177 – 28 – 8 – 90 = 51	14	37	51	89.06
Mixed	177 – 28 – 8 – 16= 125	28	97	125	90.86

Test: number of reliable predictions for SIDS data used to evaluate the model quality.
Unknown SIDS predictions: number of reliable predictions for SIDS data lacking the Y response value (experimental LC50).
Total SIDS predictions: number of total reliable predictions provided by the model for SIDS data.

Table VIII – PN model training set.

ID	*CASN*	*Chemical*	*LogK_{ow}*	*Log(LC50) (mol/l)*		*Hat*	*Err.Calc.*	*Err.Pred.*
				Exp	*Pred.*			
1	108-95-2	phenol	1.46	-3.46	-3.21	0.02	0.24	0.25
2	95-57-8	2-chlorophenol	2.15	-4.04	-3.71	0.01	0.33	0.33
3	120-83-2	2,4-dichlorophenol	3.06	-4.31	-4.37	0.01	-0.06	-0.06
4	88-06-2	2,4,6-trichlorophenol	3.69	-4.41	-4.84	0.02	-0.42	-0.43
5	6640-27-3	2-chloro-4-methylphenol	2.65	-3.60	-4.08	0.01	-0.48	-0.48
6	35421-08-0	4-chloro-3-methylphenol	3.10	-4.34	-4.40	0.01	-0.06	-0.06
7	118-79-6	2,4,6-tribromophenol	3.91	-4.70	-4.99	0.03	-0.29	-0.29
8	1745-81-9	2-allylphenol	2.54	-3.94	-4.00	0.01	-0.06	-0.06
9	90-43-7	2-phenylphenol	3.09	-4.44	-4.39	0.01	0.05	0.05
10	150-19-6	3-methoxyphenol	1.58	-3.22	-3.30	0.02	-0.08	-0.08
11	150-76-5	4-methoxyphenol	1.34	-3.05	-3.13	0.02	-0.08	-0.08
12	95-48-7	2-methylphenol	1.95	-3.77	-3.57	0.01	0.20	0.20
13	108-39-4	3-methylphenol	1.96	-3.29	-3.58	0.01	-0.29	-0.29
14	106-44-5	4-methylphenol	1.94	-3.74	-3.56	0.01	0.18	0.18
15	123-07-9	4-ethylphenol	2.58	-4.07	-4.02	0.01	0.04	0.05
16	645-56-7	4-propylphenol	3.20	-4.09	-4.48	0.02	-0.38	-0.39
17	1638-22-8	4-n-butylphenol	3.56	-4.47	-4.74	0.02	-0.26	-0.27
18	27178-34-3	4-tert-butylphenol	3.31	-4.46	-4.55	0.02	-0.09	-0.09
19	14938-35-3	4-n-pentylphenol	4.09	-5.18	-5.12	0.03	0.06	0.06
20	80-46-6	4-tert-pentylphenol	3.83	-4.81	-4.93	0.02	-0.12	-0.12
21	104-40-5	4-nonylphenol	6.20	-6.20	-6.69	0.100 **	-0.44	-0.49
22	831-82-3	4-phenoxyphenol	3.75	-4.58	-4.88	0.02	-0.29	-0.30
23	1687-53-2	4-(N-methoxymethyl)aminophenol	0.47	-2.27	-2.51	0.04	-0.23	-0.24
24	105-67-9	2,4-dimethylphenol	2.30	-3.86	-3.82	0.01	0.04	0.04
25	576-26-1	2,6-dimethylphenol	2.36	-3.75	-3.87	0.01	-0.12	-0.12
26	95-65-8	3,4-dimethylphenol	2.23	-3.92	-3.77	0.01	0.15	0.15
27	90-15-3	1-naphthol	2.84	-4.50	-4.21	0.01	0.29	0.29
28	100-02-7	4-nitrophenol	1.91	-3.46	-3.54	0.01	-0.08	-0.08
29	119-34-6	4-amino-2-nitrophenol	0.96	-3.64	-2.83	0.03	0.79 *	0.81 *
30	62-53-3	aniline	0.90	-2.86	-2.81	0.03	0.05	0.05
31	95-51-2	2-chloroaniline	1.90	-4.31	-3.52	0.01	0.78 *	0.79 *
32	108-42-9	3-chloroaniline	1.88	-3.98	-3.51	0.01	0.46	0.47
33	106-47-8	4-chloroaniline	1.88	-3.64	-3.52	0.01	0.12	0.12
34	554-00-7	2,4-dichloroaniline	2.91	-4.41	-4.26	0.01	0.15	0.15
35	95-82-9	2,5-dichloroaniline	2.92	-4.99	-4.26	0.01	0.72 *	0.73 *
36	95-76-1	3,4-dichloroaniline	2.69	-4.37	-4.10	0.01	0.27	0.27
37	626-43-7	3,5-dichloroaniline	2.90	-4.62	-4.25	0.01	0.36	0.37
38	634-67-3	2,3,4-trichloroaniline	3.68	-5.15	-4.81	0.02	0.33	0.34
39	634-93-5	2,3,6-trichloroaniline	3.32	-4.73	-4.56	0.02	0.17	0.17
40	636-30-6	2,4,5-trichloroaniline	3.69	-5.00	-4.82	0.02	0.17	0.18
41	95-53-4	2-methylaniline	1.32	-3.12	-3.11	0.02	0.01	0.01
42	108-44-1	3-methylaniline	1.40	-3.47	-3.17	0.02	0.30	0.30
43	106-49-0	4-methylaniline	1.39	-3.43	-3.16	0.02	0.27	0.27

Table VIII – PN model training set (continued).

ID	*CASN*	*Chemical*	*LogK_{ow}*	*Log(LC50) (mol/l)*		*Hat*	*Err.Calc.*	*Err.Pred.*
				Exp	*Pred.*			
44	578-54-1	2-ethylaniline	1.74	-3.21	-3.42	0.02	-0.21	-0.21
45	587-02-0	3-ethylaniline	1.85	-3.65	-3.49	0.01	0.15	0.16
46	589-16-2	4-ethylaniline	1.85	-3.42	-3.50	0.01	-0.08	-0.08
47	104-13-2	4-butylaniline	2.91	-4.16	-4.27	0.01	-0.10	-0.11
48	16245-79-7	4-octylaniline	5.02	-6.23	-5.77	0.054 *	0.44	0.46
49	37529-30-9	4-decylaniline	6.08	-6.58	-6.55	0.095 **	0.02	0.03
50	24544-04-5	2,6-diisopropylaniline	3.18	-4.06	-4.47	0.02	-0.40	-0.41
51	536-90-3	3-benzoxyaniline	2.77	-4.34	-4.16	0.01	0.18	0.18
52	39905-57-2	4-hexyloxyaniline	3.64	-4.78	-4.79	0.02	-0.01	-0.01
53	106-40-1	4-bromoaniline	2.26	-3.56	-3.80	0.01	-0.23	-0.24
54	771-60-8	pentafluoroaniline	1.86	-3.69	-3.50	0.01	0.19	0.19
55	452-71-1	4-tetrafluoro-2-methylaniline	2.51	-3.78	-3.98	0.01	-0.19	-0.20
56	443-86-7	4-tetrafluoro-3-methylaniline	2.51	-3.77	-3.98	0.01	-0.20	-0.21
57	100-01-6	4-nitroaniline	1.39	-3.04	-3.17	0.02	-0.12	-0.13
58	121-87-9	2-chloro-4-nitroaniline	2.05	-3.93	-3.64	0.01	0.29	0.29
59	616-86-4	4-ethoxy-2-nitroaniline	2.38	-3.85	-3.88	0.01	-0.03	-0.03
60	618-62-2	3,5-dichloronitrobenzene	3.13	-4.63	-4.42	0.02	0.21	0.21
61	88-72-2	2-nitrotoluene	2.30	-3.57	-3.83	0.01	-0.25	-0.26
62	99-08-1	3-nitrotoluene	2.42	-3.63	-3.91	0.01	-0.28	-0.28
63	99-99-0	4-nitrotoluene	2.37	-3.76	-3.87	0.01	-0.11	-0.11
64	75-04-7	ethylamine	-0.13	-2.30	-2.05	0.052 *	0.23	0.25
65	107-10-8	propylamine	0.48	-2.28	-2.51	0.04	-0.23	-0.23
66	109-73-9	butylamine	0.97	-2.44	-2.87	0.03	-0.42	-0.43
67	110-58-7	amylamine	1.49	-2.69	-3.25	0.02	-0.55	-0.56
68	111-26-2	hexylamine	2.06	-3.25	-3.65	0.01	-0.40	-0.40
69	111-68-2	heptylamine	2.57	-3.72	-4.02	0.01	-0.30	-0.30
70	111-86-4	octylamine	3.03	-4.40	-4.35	0.01	0.05	0.05
71	112-20-9	nonylamine	3.56	-4.82	-4.73	0.02	0.09	0.09
72	2016-57-1	decylamine	4.09	-5.18	-5.12	0.03	0.06	0.06
73	7307-55-3	undecylamine	4.62	-5.91	-5.48	0.04	0.41	0.43
74	124-22-1	dodecylamine	5.15	-6.27	-5.86	0.058 *	0.39	0.41
75	2869-34-3	tridecylamine	5.68	-6.45	-6.25	0.078 **	0.18	0.20
76	13952-84-6	se-butylamine	0.74	-2.42	-2.70	0.03	-0.27	-0.28
77	598-74-3	1,2-dimethylpropylamine	1.10	-2.49	-2.97	0.02	-0.46	-0.48
78	78-81-9	2,2-dimethyl-1-propylamine	1.19	-2.26	-3.04	0.02	-0.76 *	-0.78 *
79	15673-00-4	3,3-dimethylbutylamine	1.72	-2.22	-3.42	0.02	-1.18 **	-1.20 **
80	107-45-9	t-octylamine	2.68	-3.72	-4.10	0.01	-0.38	-0.38
81	693-16-3	1-methylheptylamine	2.81	-4.40	-4.19	0.01	0.21	0.21
82	141-43-5	2-aminoethanol	-1.31	-1.46	-1.18	0.098 **	0.25	0.28
83	78-96-6	1-amino-2-propanol	-0.96	-1.48	-1.46	0.083 **	0.02	0.02
84	109-85-3	2-methoxyethylamine	-0.67	-2.16	-1.64	0.071 **	0.49	0.52
85	100-46-9	benzylamine	1.09	-3.02	-2.95	0.02	0.07	0.07
86	768-94-5	1-adamantanamine	1.43	-3.78	-3.18	0.02	0.59	0.60

* Chemicals with values between 2 times SDEC (or SDEP or critical HAT) and 3 times SDEC (or SDEP or critical HAT). ** Chemicals with values greater than 3 times SDEC (or SDEP or average value of HAT).

Table IX – SIDS chemicals not suitable for QSAR 2:

	N.Comp.	SIDS Chemicals								Motivation
	24	2 94 144 173	6 95 152 174	7 101 155 175	17 109 160	39 125 162	76 126 164	77 139 166		Out of the X - domain (-1.31≤LogKow ≤ 6.20)
	5	82	56	61	100	104				In the training set
XY-domain + MOA domain	140	1 12 21 29 38 49 59 68 78 87 97 108 119 129 137 147 157 170	3 13 22 30 41 50 60 69 79 88 98 110 120 130 138 148 158 171	4 14 23 32 43 51 62 70 80 89 99 111 121 131 140 149 159 172	5 15 24 33 44 52 63 71 81 90 102 112 122 132 141 150 161 176	8 16 25 34 45 53 64 72 83 91 103 113 123 133 142 151 163	9 18 26 35 46 54 65 73 84 92 105 114 124 134 143 153 165	10 19 27 36 47 55 66 74 85 93 106 117 127 135 145 154 167	11 20 28 37 48 57 67 75 86 96 107 118 128 136 146 156 168	MOA ≠ PN
	2	169	177							High leverage chemicals (structurally distant from the training chemicals)
XY-domain	2	169	177							High leverage chemicals
	18	1 67 111	9 73 117	18 78 141	20 81 142	21 87	49 103	51 108		Y-Outliers (cross-validated standardised residual greater than two standard deviation units).
	5	3	8	23	89	137				High leverage chemicals and Y-outliers

Table X –QSAR 2 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) and mode of action domain (MOA-D)

ID	*CASN*	*EINECS name*	*LogK$_{ow}$*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
				Exp	*Pred.*		
S31	80-05-7	Phenol, 4,4'-(1-methylethylidene)bis-	3.64	-4.70	-4.79	0.021	-0.27
S40	88-60-8	Phenol, 2-(1,1-dimethylethyl)-5-methyl-	3.97	-	-5.03	0.026	-
S42	88-74-4	Benzenamine, 2-nitro-	2.02	-	-3.62	0.013	-
S58	98-92-0	3-Pyridinecarboxamide	-0.45	-	-1.83	0.062	-
S115	128-37-0	Phenol, 2,6-bis(1,1-dimethylethyl)-4-methyl-	5.03	-	-5.80	0.053	-
S116	135-19-3	2-Naphthalenol	2.69	-4.62	-4.10	0.012	1.58
S169	25154-52-3	Phenol, nonyl-	5.99	-6.24	-6.49	**0.090**	-0.79
S177	84852-15-3	Phenol, 4-nonyl-, branched	5.92	-6.24	-6.44	**0.087**	-0.63

Unreliable predictions according to the leverage approach are highlighted in bold in the leverage column.

Table XI –QSAR 2 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D).

ID	*CASN*	*EINECS name*	*MOA*	*LogK_{ow}*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
					Exp	*Pred.*		
S1	50-00-0	Formaldehyde-	SB	0.35	-3.08	-2.41	0.038	**2.06**
S3	57-55-6	1,2-Propanediol	NPN	-0.78	-0.84	-1.59	**0.075**	**-2.35**
S4	58-08-2	1H-Purine-2,6-dione, 3,7-dihydro-1,3,7-trimethyl-	CNS	0.16	-	-2.27	0.043	-
S5	58-55-9	1H-Purine-2,6-dione, 3,7-dihydro-1,3-dimethyl-	NPN	-0.39	-	-1.88	0.060	-
S8	68-12-2	Formamide, N,N-dimethyl-	NPN	-0.93	-0.84	-1.49	**0.081**	**-2.04**
S9	71-36-3	1-Butanol	NPN	0.84	-1.60	-2.77	0.027	**-3.57**
S10	74-83-9	Methane, bromo-	NPN	1.18	-	-3.01	0.021	-
S11	74-87-3	Methane, chloro-	NPN	1.09	-	-2.95	0.023	-
S12	75-01-4	Ethene, chloro-	NPN	1.62	-	-3.33	0.016	-
S13	75-10-5	Methane, difluoro-	NPN	0.71	-	-2.67	0.030	-
S14	75-38-7	Ethene, 1,1-difluoro-	NPN	1.24	-	-3.06	0.020	-
S15	75-56-9	Oxirane, methyl-	UNK	0.37	-	-2.43	0.038	-
S16	75-68-3	Ethane, 1-chloro-1,1-difluoro-	NPN	2.05	-	-3.64	0.012	-
S18	78-59-1	2-Cyclohexen-1-one, 3,5,5-trimethyl-	MTA	2.62	-2.62	-4.05	0.012	**-4.34**
S19	78-70-6	1,6-Octadien-3-ol, 3,7-dimethyl-	PE	3.38	-	-4.60	0.017	-
S20	78-87-5	Propane, 1,2-dichloro-	NPN	2.25	-2.91	-3.79	0.012	**-2.67**
S21	78-92-2	2-Butanol	NPN	0.77	-1.31	-2.72	0.028	**-4.31**
S22	79-00-5	Ethane, 1,1,2-trichloro-	NPN	2.01	-3.21	-3.61	0.013	-1.20
S23	79-06-1	2-Propenamide	MTA	-0.81	-2.77	-1.57	**0.076**	**3.75**
S24	79-10-7	2-Propenoic acid	UNK	0.44	-	-2.48	0.036	-
S25	79-11-8	Acetic acid, chloro-	UNK	0.34	-	-2.40	0.038	-
S26	79-20-9	Acetic acid, methyl ester	NPN	0.37	-2.36	-2.43	0.038	-0.21
S27	79-31-2	Propanoic acid, 2-methyl-	UNK	1.00	-	-2.88	0.024	-
S28	79-34-5	Ethane, 1,1,2,2-tetrachloro-	NPN	2.19	-3.92	-3.74	0.012	0.55
S29	79-39-0	2-Propenamide, 2-methyl-	MTA	-0.26	-	-1.97	0.056	-
S30	79-41-4	2-Propenoic acid, 2-methyl-	UNK	0.99	-	-2.88	0.024	-
S31	80-05-7	Phenol, 4,4'-(1-methylethylidene)bis-	PN	3.64	-4.70	-4.79	0.021	-0.27
S32	80-62-6	2-Propenoic acid, 2-methyl-, methyl ester	MTA	1.28	-2.55	-3.08	0.020	-1.61
S33	81-14-1	Ethanone, 1-[4-(1,1-dimethylethyl)-2,6-dimethyl-3,5-dinitrophenyl]-	UNK	4.31	-	-5.28	0.033	-
S34	81-15-2	Benzene, 1-(1,1-dimethylethyl)-3,5-dimethyl-2,4,6-trinitro-	UNK	4.45	-	-5.38	0.037	-
S35	84-74-2	1,2-Benzenedicarboxylic acid, dibutyl ester	UNK	4.61	-5.31	-5.49	0.041	-0.57

Table XI –QSAR 2 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) (continued).

ID	*CASN*	*EINECS name*	*MOA*	*LogK_{ow}*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
					Exp	*Pred.*		
S36	87-56-9	2-Butenoic acid, 2,3-dichloro-4-oxo-, (Z)-	UNK	1.37	-	-3.15	0.018	-
S37	88-12-0	2-Pyrrolidinone, 1-ethenyl-	UNK	0.25	-	-2.34	0.041	-
S38	88-19-7	Benzenesulfonamide, 2-methyl-	UNK	0.92	-	-2.82	0.026	-
S40	88-60-8	Phenol, 2-(1,1-dimethylethyl)-5-methyl-	PN	3.97	-	-5.03	0.026	-
S41	88-73-3	Benzene, 1-chloro-2-nitro-	UNK	2.46	-	-3.94	0.012	-
S42	88-74-4	Benzenamine, 2-nitro-	PN	2.02	-	-3.62	0.013	-
S43	91-15-6	1,2-Benzenedicarbonitrile	UNK	1.09	-	-2.95	0.023	-
S44	91-76-9	1,3,5-Triazine-2,4-diamine, 6-phenyl-	CNS	1.44	-	-3.20	0.018	-
S45	93-68-5	Butanamide, N-(2-methylphenyl)-3-oxo-	UNK	0.99	-2.78	-2.88	0.024	-0.30
S46	94-36-0	Peroxide, dibenzoyl	UNK	3.43	-	-4.64	0.018	-
S47	95-31-8	2-Benzothiazolesulfenamide, N-(1,1-dimethylethyl)-	NPN	2.56	-	-4.01	0.012	-
S48	95-49-8	Benzene, 1-chloro-2-methyl-	NPN	3.18	-	-4.46	0.015	-
S49	95-50-1	Benzene, 1,2-dichloro-	NPN	3.28	-3.41	-4.53	0.016	**-3.40**
S50	96-18-4	Propane, 1,2,3-trichloro-	NPN	2.50	-3.35	-3.97	0.012	-1.88
S51	96-29-7	2-Butanone, oxime	NPN	1.69	-2.01	-3.38	0.015	**-4.14**
S52	96-31-1	Urea, N,N'-dimethyl-	NPN	-0.62	-	-1.71	0.068	-
S53	96-33-3	2-Propenoic acid, methyl ester	MTA	0.73	-	-2.69	0.029	-
S54	97-72-3	Propanoic acid, 2-methyl-, anhydride	UNK	1.24	-	-3.06	0.020	-
S55	98-07-7	Benzene, (trichloromethyl)-	NPN	3.90	-	-4.98	0.025	-
S57	98-59-9	Benzenesulfonyl chloride, 4-methyl-	UNK	3.49	-	-4.68	0.019	-
S58	98-92-0	3-Pyridinecarboxamide	PN	-0.45	-	-1.83	0.062	-
S59	99-04-7	Benzoic acid, 3-methyl-	UNK	2.42	-	-3.91	0.012	-
S60	99-54-7	Benzene, 1,2-dichloro-4-nitro-	UNK	3.10	-	-4.40	0.014	-
S62	100-00-5	Benzene, 1-chloro-4-nitro-	UNK	2.46	-	-3.94	0.012	-
S63	100-21-0	1,4-Benzenedicarboxylic acid	UNK	1.76	-	-3.43	0.014	-
S64	100-37-8	Ethanol, 2-(diethylamino)-	UNK	0.05	-1.82	-2.20	0.046	-1.17
S65	100-41-4	Benzene, ethyl-	NPN	3.03	-3.94	-4.35	0.014	-1.24
S66	102-06-7	Guanidine, N,N'-diphenyl-	NPN	2.89	-	-4.25	0.013	-
S67	102-76-1	1,2,3-Propanetriol, triacetate	UNK	0.36	-3.12	-2.42	0.038	**2.15**
S68	103-11-7	2-Propenoic acid, 2-ethylhexyl ester	MTA	4.09	-	-5.12	0.029	-
S69	103-84-4	Acetamide, N-phenyl-	UNK	1.10	-	-2.95	0.022	-

Table XI –QSAR 2 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) (continued).

ID	*CASN*	*EINECS name*	*MOA*	*LogK_{ow}*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
					Exp	*Pred.*		
S70	105-60-2	2H-Azepin-2-one, hexahydro-	NPN	0.66	-	-2.64	0.031	-
S71	106-31-0	Butanoic acid, anhydride	UNK	1.39	-	-3.16	0.018	-
S72	106-46-7	Benzene, 1,4-dichloro-	NPN	3.28	-4.02	-4.53	0.016	-1.55
S73	106-63-8	2-Propenoic acid, 2-methylpropyl ester	MTA	2.13	-4.79	-3.70	0.012	**3.30**
S74	106-88-7	Oxirane, ethyl-	UNK	0.86	-	-2.78	0.027	-
S75	107-06-2	Ethane, 1,2-dichloro-	NPN	1.83	-2.93	-3.48	0.014	-1.67
S78	107-41-5	2,4-Pentanediol, 2-methyl-	NPN	0.58	-1.09	-2.58	0.032	**-4.57**
S79	107-86-8	2-Butenal, 3-methyl-	MTA	1.15	-	-2.99	0.022	-
S80	107-92-6	Butanoic-acid-	UNK	1.07	-	-2.93	0.023	-
S81	107-98-2	2-Propanol, 1-methoxy-	NPN	-0.49	-0.64	-1.80	0.064	**-3.61**
S83	108-65-6	2-Propanol, 1-methoxy-, acetate	EN	0.52	-	-2.54	0.034	-
S84	108-77-0	1,3,5-Triazine, 2,4,6-trichloro-	UNK	1.73	-	-3.41	0.015	-
S85	108-88-3	Benzene, methyl-	NPN	2.54	-3.55	-4.00	0.012	-1.36
S86	109-66-0	Pentane-	NPN	2.80	-	-4.18	0.012	-
S87	110-16-7	2-Butenedioic acid (Z)-	UNK	0.05	-4.37	-2.20	0.046	**6.68**
S88	110-19-0	Acetic acid, 2-methylpropyl ester	EN	1.77	-	-3.44	0.014	-
S89	110-65-6	2-Butyne-1,4-diol	PE	-0.93	-3.21	-1.49	**0.081**	**5.40**
S90	110-83-8	Cyclohexene-	NPN	2.96	-	-4.30	0.013	-
S91	110-85-0	Piperazine-	NPN	-0.80	-	-1.58	**0.075**	-
S92	110-93-0	5-Hepten-2-one, 6-methyl-	NPN	2.06	-3.17	-3.65	0.012	-1.45
S93	110-98-5	2-Propanol, 1,1'-oxybis-	NPN	-0.64	-	-1.70	0.069	-
S96	115-07-1	1-Propene	NPN	1.68	-	-3.37	0.015	-
S97	115-11-7	1-Propene, 2-methyl-	NPN	2.23	-	-3.77	0.012	-
S98	115-86-6	Phosphoric acid, triphenyl ester	UNK	4.70	-5.59	-5.56	0.043	0.10
S99	115-95-7	1,6-Octadien-3-ol, 3,7-dimethyl-, acetate	UNK	4.39	-	-5.33	0.035	-
S102	120-61-6	1,4-Benzenedicarboxylic acid, dimethyl ester	UNK	1.66	-	-3.36	0.015	-
S103	120-80-9	1,2-Benzenediol	UNK	1.03	-4.29	-2.90	0.024	**4.23**
S105	121-91-5	1,3-Benzenedicarboxylic acid	UNK	1.76	-	-3.43	0.014	-
S106	122-52-1	Phosphorous acid, triethyl ester	UNK	0.74	-	-2.69	0.029	-
S107	122-99-6	Ethanol, 2-phenoxy-	NPN	1.1	-2.60	-2.95	0.022	-1.05
S108	123-54-6	2,4-Pentanedione	UNK	0.05	-2.86	-2.20	0.046	**2.04**
S110	123-86-4	Acetic acid, butyl ester	EN	1.85	-3.81	-3.50	0.014	0.94
S111	124-04-9	Hexanedioic-acid-	UNK	0.23	-3.18	-2.33	0.041	**2.61**

Table XI –QSAR 2 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) (continued).

ID	*CASN*	*EINECS name*	*MOA*	*LogK_{ow}*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
					Exp	*Pred.*		
S112	126-73-8	Phosphoric-acid-tributyl-ester-	AChE	3.82	-4.77	-4.92	0.024	-0.44
S113	126-98-7	2-Propenenitrile, 2-methyl-	MTA	0.76	-	-2.71	0.029	-
S114	127-19-5	Acetamide, N,N-dimethyl-	NPN	-0.49	-	-1.80	0.064	-
S115	128-37-0	Phenol, 2,6-bis(1,1-dimethylethyl)-4-methyl-	PN	5.03	-	-5.80	0.053	-
S116	135-19-3	2-Naphthalenol	PN	2.69	-4.62	-4.10	0.012	1.58
S117	140-88-5	2-Propenoic acid, ethyl ester	MTA	1.22	-4.60	-3.04	0.021	**4.76**
S118	141-10-6	3,5,9-Undecatrien-2-one, 6,10-dimethyl-	MTA	4.43	-	-5.36	0.036	-
S119	141-32-2	2-Propenoic acid, butyl ester	MTA	2.20	-	-3.75	0.012	-
S120	141-78-6	Acetic-acid-ethyl-ester-	EN	0.86	-2.58	-2.78	0.027	-0.61
S121	141-97-9	Butanoic acid, 3-oxo-, ethyl ester	UNK	-0.20	-	-2.01	0.054	-
S122	144-55-8	Carbonic-acid-monosodium-salt-	UNK	-0.46	-	-1.83	0.063	-
S123	150-90-3	Butanedioic acid, disodium salt	UNK	-0.75	-	-1.62	**0.073**	-
S124	288-32-4	1H-Imidazole	NPN	0.06	-	-2.20	0.046	-
S127	528-44-9	1,2,4-Benzenetricarboxylic acid	UNK	0.95	-	-2.85	0.025	-
S128	552-30-7	5-Isobenzofurancarboxylic acid, 1,3-dihydro-1,3-dioxo-	UNK	1.96	-	-3.58	0.013	-
S129	556-82-1	2-Buten-1-ol, 3-methyl-	NPN	1.17	-	-3.01	0.021	-
S130	611-19-8	Benzene, 1-chloro-2-(chloromethyl)-	UNK	3.44	-	-4.65	0.018	-
S131	760-23-6	1-Butene, 3,4-dichloro-	SN2	2.60	-4.18	-4.04	0.012	0.44
S132	770-35-4	2-Propanol, 1-phenoxy-	NPN	1.52	-2.74	-3.26	0.017	-1.58
S133	793-24-8	1,4-Benzenediamine, N-(1,3-dimethylbutyl)-N'-phenyl-	NPN	4.68	-	-5.54	0.043	-
S134	822-06-0	Hexane, 1,6-diisocyanato-	UNK	3.20	-	-4.47	0.015	-
S135	839-90-7	1,3,5-Triazine-2,4,6(1H,3H,5H)-trione, 1,3,5-tris(2-hydroxyethyl)-	UNK	0.07	-	-2.21	0.046	-
S136	868-77-9	2-Propenoic acid, 2-methyl-, 2-hydroxyethyl ester	MTA	0.30	-2.76	-2.38	0.039	1.17
S137	868-85-9	Phosphonic acid, dimethyl ester	UNK	-1.13	-2.69	-1.34	**0.089**	**4.26**
S138	919-30-2	3-Aminopropyl-triethoxysilane	UNK	0.31	-	-2.38	0.039	-
S140	1477-55-0	1,3-Benzenedimethanamine	NPN	0.15	-	-2.27	0.043	-
S141	1490-04-6	Cyclohexanol, 5-methyl-2-(1-methylethyl)-	NPN	3.38	-3.93	-4.60	0.017	**-2.04**

Table XI –QSAR 2 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) (continued).

ID	*CASN*	*EINECS name*	*MOA*	*LogK$_{ow}$*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
					Exp	*Pred.*		
S142	1634-04-4	Propane, 2-methoxy-2-methyl-	NPN	1.43	-2.12	-3.19	0.018	**-3.26**
S143	1717-00-6	HCFC 141b	NPN	2.37	-	-3.87	0.012	-
S145	2403-88-5	4-Piperidinol, 2,2,6,6-tetramethyl-	NPN	0.94	-	-2.84	0.025	-
S146	2432-99-7	Undecanoic acid, 11-amino-	UNK	-0.16	-	-2.04	0.053	-
S147	2439-35-2	2-Propenoic acid, 2-(dimethylamino)ethyl ester	MTA	0.42	-	-2.46	0.036	-
S148	2837-89-0	Ethane, 2-chloro-1,1,1,2-tetrafluoro-	NPN	1.86	-	-3.50	0.014	-
S149	2855-13-2	Cyclohexanemethanamine, 5-amino-1,3,3-trimethyl-	NPN	1.90	-	-3.53	0.013	-
S150	2867-47-2	2-Propenoic acid, 2-methyl-, 2-(dimethylamino)ethyl ester	MTA	0.97	-	-2.86	0.025	-
S151	3268-49-3	Propanal, 3-(methylthio)-	UNK	0.41	-	-2.46	0.037	-
S153	3323-53-3	Hexanedioic acid, compd. with 1,6-hexanediamine (1:1)	NPN	0.23	-	-2.33	0.041	-
S154	3452-97-9	1-Hexanol, 3,5,5-trimethyl-	NPN	3.11	-	-4.41	0.015	-
S156	4169-04-4	1-Propanol, 2-phenoxy-	NPN	1.52	-2.74	-3.26	0.017	-1.58
S157	4454-05-1	2H-Pyran, 3,4-dihydro-2-methoxy-	NPN	0.88	-	-2.80	0.026	-
S158	4457-71-0	1,5-Pentanediol, 3-methyl-	NPN	0.69	-	-2.66	0.030	-
S159	4979-32-2	2-Benzothiazolesulfenamide, N,N-dicyclohexyl-	NPN	5.96	-	-6.47	**0.088**	-
S161	5392-40-5	2,6-Octadienal, 3,7-dimethyl-	MTA	3.45	-	-4.65	0.018	-
S163	6165-51-1	Benzene, 1,4-dimethyl-2-(1-phenylethyl)-	NPN	5.24	-	-5.95	0.060	-
S165	6386-38-5	Benzenepropanoic acid, 3,5-bis(1,1-dimethylethyl)-4-hydroxy-, methyl ester	EN	5.06	-	-5.82	0.054	-
S167	6864-37-5	Cyclohexanamine, 4,4'-methylenebis[2-methyl-	NPN	4.10	-	-5.12	0.029	-
S168	11070-44-3	1,3-Isobenzofurandione, tetrahydromethyl-	UNK	2.64	-	-4.07	0.012	-
S169	25154-52-3	Phenol, nonyl-	PN	5.99	-6.24	-6.49	**0.090**	-0.79
S170	25265-71-8	Propanol, oxybis-	NPN	-0.49	-	-1.80	0.064	-
S171	25321-09-9	Benzene, bis(1-methylethyl)-	NPN	4.90	-	-5.70	0.049	-
S172	25321-14-6	Benzene, methyldinitro-	UNK	2.18	-4.03	-3.74	0.012	0.88
S176	56539-66-3	1-Butanol, 3-methoxy-3-methyl-	NPN	0.46	-	-2.49	0.035	-
S177	84852-15-3	Phenol, 4-nonyl-, branched	PN	5.92	-6.24	-6.44	**0.087**	-0.63

Y outliers are highlighted in bold in the standardized residual in prediction column. Unreliable predictions according to the leverage approach are highlighted in bold in the leverage column.

Table XII – PN model performance on the two to subset of SIDS data evaluated.

Test MOA	*N. Chemicals*	*N.Test*	*Unknown SIDS predictions*	*Total SIDS predictions*	Q^2ext
PN	177 – 24 – 5– 140 - 2= 6	2	4	6	N.A.
Mixed	177 – 24 – 5 – 18 – 5 -2 = 123	25	98	123	86.66

Test: number of reliable predictions for SIDS data used to evaluate the model quality.
Unknown SIDS predictions: number of reliable predictions for SIDS data lacking the Y response value (experimental LC50).
Total SIDS predictions: number of total reliable predictions provided by the model for SIDS data.

Table XIII – N model training set.

ID	*CASN*	*Chemical*	*LogK_{ow}*	*Log(LC50) (mol/l)*		*Hat*	*Err.Calc.*	*Err.Pred.*
				Exp	*Pred.*			
N1	71-36-3	1-butanol	0.88	-1.63	-2.47	0.013	-0.83	-0.84
N2	112-30-1	1-decanol	4.57	-4.81	-5.46	0.023	-0.63	-0.65
N3	112-53-8	1-dodecanol	5.13	-5.26	-5.92	0.033 *	-0.64	-0.66
N4	111-27-3	1-hexanol	2.03	-3.02	-3.39	0.007	-0.37	-0.37
N5	143-08-8	1-nonanol	4.26	-4.40	-5.21	0.019	-0.79	-0.81
N6	111-87-5	1-octanol	2.97	-3.98	-4.15	0.008	-0.17	-0.17
N7	112-42-5	1-undecanol	4.52	-5.21	-5.41	0.023	-0.19	-0.20
N8	79-00-5	1,1,2-trichloroethane	1.89	-3.21	-3.27	0.007	-0.06	-0.06
N9	79-34-5	1,1,2,2-tetrachloroethane	2.39	-3.91	-3.68	0.007	0.23	0.23
N10	107-06-2	1,2-dichloroethane	1.48	-2.92	-2.94	0.009	-0.02	-0.02
N11	79-34-5	1,1,2,2-tetrachloroethane	4.63	-5.29	-5.50	0.024	-0.20	-0.21
N12	120-82-1	1,2,4-trichlorobenzene	4.05	-4.79	-5.03	0.017	-0.23	-0.24
N13	541-73-1	1,3-dichlorobenzene	3.52	-4.27	-4.60	0.012	-0.32	-0.33
N14	106-46-7	1,4-dichlorobenzene	3.44	-4.56	-4.53	0.011	0.03	0.03
N15	150-78-7	1,4-dimethoxybenzene	2.15	-3.07	-3.49	0.007	-0.41	-0.42
N16	111-90-0	2-(2-ethoxyethoxy)ethanol	-0.54	-0.70	-1.33	0.032 *	-0.61	-0.63
N17	78-93-3	2-butanone	0.29	-1.35	-1.99	0.020	-0.63	-0.64
N18	693-54-9	2-decanone	3.73	-4.43	-4.77	0.014	-0.33	-0.34
N19	552-41-0	2-hydroxy-4-methoxyacetophenone	1.98	-3.48	-3.35	0.007	0.13	0.13
N20	78-83-1	2-methyl-1-propanol	0.76	-1.71	-2.37	0.014	-0.65	-0.66
N21	107-41-5	2-methyl-2,4-pentanediol	-0.67	-1.04	-1.21	0.035 *	-0.16	-0.17
N22	111-13-7	2-octanone	2.37	-3.55	-3.66	0.007	-0.11	-0.11
N23	122-99-6	2-phenoxyethanol	1.16	-2.60	-2.68	0.011	-0.08	-0.08
N24	67-63-0	2-propanol	0.05	-0.76	-1.81	0.023	-1.02 *	-1.05 *
N25	115-20-8	2,2,2-trichloroethanol	1.42	-2.69	-2.90	0.009	-0.20	-0.21
N26	13608-87-2	2,3,4-trichloroacetophenone	3.57	-5.04	-4.63	0.012	0.41	0.41
N27	13909-73-4	2,3,4-trimethoxyacetophenone	1.12	-3.08	-2.65	0.011	0.43	0.43
N28	937-20-2	2,4-dichloroacetophenone	2.84	-4.20	-4.04	0.008	0.16	0.16
N29	5673-07-4	2,6-dimethoxytoluene	2.64	-3.87	-3.88	0.007	-0.01	-0.01
N30	4412-91-3	3-furanmethanol	0.30	-2.28	-1.98	0.019	0.29	0.30
N31	563-80-4	3-methyl-2-butanone	0.56	-1.99	-2.20	0.016	-0.21	-0.21
N32	96-22-0	3-pentanone	0.79	-1.74	-2.39	0.014	-0.64	-0.65
N33	75-97-8	3,3-dimethyl-2-butanone	0.96	-3.06	-2.51	0.013	0.54	0.55
N34	95-75-0	3,4-dichlorotoluene	4.06	-4.74	-5.04	0.017	-0.29	-0.30
N35	108-10-1	4-methyl-2-pentanone	1.31	-2.29	-2.81	0.010	-0.51	-0.52
N36	110-12-3	5-methyl-2-hexanone	1.88	-2.85	-3.27	0.007	-0.42	-0.42
N37	502-56-7	5-nonanone	2.90	-3.66	-4.10	0.008	-0.43	-0.44
N38	110-93-0	6-methyl-5-hepten-2-one	1.70	-3.16	-3.12	0.008	0.04	0.04
N39	67-64-1	Acetone	-0.24	-0.85	-1.57	0.027	-0.70	-0.72
N40	98-86-2	Acetophenone	1.58	-2.87	-3.02	0.009	-0.15	-0.15
N41	119-61-9	Benzophenone	3.18	-4.07	-4.32	0.009	-0.25	-0.25
N42	108-93-0	Cyclohexanol	1.23	-2.15	-2.75	0.010	-0.59	-0.60
N43	108-94-1	Cyclohexanone	0.81	-2.27	-2.40	0.014	-0.13	-0.13

Table XIII – N model training set (continued).

ID	*CASN*	*Chemical*	*LogK_{ow}*	*Log(LC50) (mol/l)*		*Hat*	*Err.Calc.*	*Err.Pred.*
				Exp	*Pred.*			
N44	142-96-1	Dibutyl ether	3.21	-3.60	-4.35	0.010	-0.74	-0.75
N45	108-20-3	Diisopropyl ether	1.52	-3.04	-2.97	0.009	0.07	0.07
N46	693-65-2	Dipentyl ether	4.04	-4.69	-5.02	0.017	-0.32	-0.33
N47	101-84-8	Diphenyl ether	4.21	-4.62	-5.16	0.019	-0.53	-0.54
N48	64-17-5	Ethanol	-0.31	0.51	-1.55	0.028 *	-2.00 **	-2.06 **
N49	110-00-9	Furan	1.34	-3.04	-2.83	0.010	0.21	0.21
N50	67-72-1	Hexachloroethane	4.14	-5.19	-5.09	0.018	0.09	0.10
N51	67-56-1	Methanol	-0.77	-0.06	-1.16	0.037 *	-1.06 *	-1.10 *
N52	620-88-2	4-nitrophenyl phenylether	4.28	-4.90	-5.22	0.020	-0.31	-0.32
N53	76-01-7	Pentachloroethane	3.62	-4.44	-4.68	0.013	-0.23	-0.24
N54	1634-04-4	tert-butylmethyl ether	0.94	-2.09	-2.51	0.013	-0.42	-0.42
N55	127-18-4	Tetrachloroethene	3.40	-4.08	-4.50	0.011	-0.42	-0.42
N56	109-99-9	Tetrahydrofuran	0.46	-1.52	-2.13	0.018	-0.60	-0.61
N57	79-01-6	Trichloroethene	2.42	-3.47	-3.71	0.007	-0.23	-0.24
N58	112-27-6	Triethylene glycol	-1.24	-0.33	-0.76	0.046 **	-0.41	-0.43
P1	108-95-2	phenol	1.46	-3.46	-2.92	0.009	0.53	0.54
P2	95-57-8	2-chlorophenol	2.15	-4.04	-3.48	0.007	0.56	0.56
P3	120-83-2	2,4-dichlorophenol	3.06	-4.31	-4.22	0.009	0.09	0.09
P4	88-06-2	2,4,6-trichlorophenol	3.69	-4.41	-4.74	0.013	-0.32	-0.33
P5	6640-27-3	2-chloro-4-methylphenol	2.65	-3.60	-3.89	0.007	-0.29	-0.29
P6	35421-08-0	4-chloro-3-methylphenol	3.10	-4.34	-4.25	0.009	0.09	0.09
P7	118-79-6	2,4,6-tribromophenol	3.91	-4.70	-4.91	0.015	-0.21	-0.21
P8	1745-81-9	2-allylphenol	2.54	-3.94	-3.80	0.007	0.14	0.14
P9	90-43-7	2-phenylphenol	3.09	-4.44	-4.24	0.009	0.19	0.20
P10	150-19-6	3-methoxyphenol	1.58	-3.22	-3.02	0.009	0.20	0.20
P11	150-76-5	4-methoxyphenol	1.34	-3.05	-2.83	0.010	0.22	0.22
P12	95-48-7	2-methylphenol	1.95	-3.77	-3.32	0.007	0.45	0.45
P13	108-39-4	3-methylphenol	1.96	-3.29	-3.33	0.007	-0.04	-0.04
P14	106-44-5	4-methylphenol	1.94	-3.74	-3.31	0.007	0.43	0.43
P15	123-07-9	4-ethylphenol	2.58	-4.07	-3.83	0.007	0.24	0.24
P16	645-56-7	4-propylphenol	3.20	-4.09	-4.34	0.010	-0.24	-0.25
P17	1638-22-8	4-n-butylphenol	3.56	-4.47	-4.63	0.012	-0.16	-0.16
P18	27178-34-3	4-tert-butylphenol	3.31	-4.46	-4.42	0.010	0.04	0.04
P19	14938-35-3	4-n-pentylphenol	4.09	-5.18	-5.05	0.017	0.12	0.13
P20	80-46-6	4-tert-pentylphenol	3.83	-4.81	-4.85	0.014	-0.03	-0.04
P21	104-40-5	4-nonylphenol	6.20	-6.20	-6.80	0.055 **	-0.56	-0.60
P22	831-82-3	4-phenoxyphenol	3.75	-4.58	-4.78	0.014	-0.20	-0.20
P23	1687-53-2	4-(N-methoxymethyl)aminophenol	0.47	-2.27	-2.12	0.017	0.15	0.15
P24	105-67-9	2,4-dimethylphenol	2.30	-3.86	-3.60	0.007	0.25	0.26
P25	576-26-1	2,6-dimethylphenol	2.36	-3.75	-3.65	0.007	0.10	0.10
P26	95-65-8	3,4-dimethylphenol	2.23	-3.92	-3.55	0.007	0.37	0.37
P27	90-15-3	1-naphthol	2.84	-4.50	-4.04	0.008	0.46	0.46
P28	100-02-7	4-nitrophenol	1.91	-3.46	-3.29	0.007	0.17	0.17
P29	119-34-6	4-amino-2-nitrophenol	0.96	-3.64	-2.51	0.013	1.12 *	1.13 *
P30	62-53-3	aniline	0.90	-2.86	-2.47	0.013	0.39	0.39

Table XIII – N model training set (continued).

ID	*CASN*	*Chemical*	*LogK_{ow}*	*Log(LC50) (mol/l)*		*Hat*	*Err.Calc.*	*Err.Pred.*
				Exp	*Pred.*			
P31	95-51-2	2-chloroaniline	1.90	-4.31	-3.27	0.007	1.03 *	1.04 *
P32	108-42-9	3-chloroaniline	1.88	-3.98	-3.26	0.007	0.71	0.72
P33	106-47-8	4-chloroaniline	1.88	-3.64	-3.26	0.007	0.37	0.38
P34	554-00-7	2,4-dichloroaniline	2.91	-4.41	-4.10	0.008	0.31	0.31
P35	95-82-9	2,5-dichloroaniline	2.92	-4.99	-4.10	0.008	0.88	0.89
P36	95-76-1	3,4-dichloroaniline	2.69	-4.37	-3.92	0.007	0.45	0.45
P37	626-43-7	3,5-dichloroaniline	2.90	-4.62	-4.09	0.008	0.53	0.53
P38	634-67-3	2,3,4-trichloroaniline	3.68	-5.15	-4.72	0.013	0.43	0.43
P39	634-93-5	2,3,6-trichloroaniline	3.32	-4.73	-4.43	0.010	0.30	0.30
P40	636-30-6	2,4,5-trichloroaniline	3.69	-5.00	-4.73	0.013	0.27	0.27
P41	95-53-4	2-methylaniline	1.32	-3.12	-2.81	0.010	0.31	0.31
P42	108-44-1	3-methylaniline	1.40	-3.47	-2.87	0.009	0.59	0.60
P43	106-49-0	4-methylaniline	1.39	-3.43	-2.86	0.009	0.56	0.57
P44	578-54-1	2-ethylaniline	1.74	-3.21	-3.15	0.008	0.06	0.06
P45	587-02-0	3-ethylaniline	1.85	-3.65	-3.24	0.008	0.41	0.41
P46	589-16-2	4-ethylaniline	1.85	-3.42	-3.24	0.008	0.18	0.18
P47	104-13-2	4-butylaniline	2.91	-4.16	-4.10	0.008	0.06	0.06
P48	16245-79-7	4-octylaniline	5.02	-6.23	-5.79	0.031 *	0.42	0.44
P49	37529-30-9	4-decylaniline	6.08	-6.58	-6.67	0.053 **	-0.09	-0.09
P50	24544-04-5	2,6-diisopropylaniline	3.18	-4.06	-4.32	0.009	-0.26	-0.26
P51	536-90-3	3-benzoxyaniline	2.77	-4.34	-3.98	0.008	0.35	0.36
P52	39905-57-2	4-hexyloxyaniline	3.64	-4.78	-4.69	0.013	0.09	0.09
P53	106-40-1	4-bromoaniline	2.26	-3.56	-3.57	0.007	-0.01	-0.01
P54	771-60-8	pentafluoroaniline	1.86	-3.69	-3.25	0.008	0.44	0.44
P55	452-71-1	a,a,a,a-4-tetrafluoro-2-methylaniline	2.51	-3.78	-3.78	0.007	0.00	0.00
P56	443-86-7	a,a,a,a-4-tetrafluoro-3-methylaniline	2.51	-3.77	-3.78	0.007	-0.01	-0.01
P57	100-01-6	4-nitroaniline	1.39	-3.04	-2.87	0.009	0.17	0.17
P58	121-87-9	2-chloro-4-nitroaniline	2.05	-3.93	-3.40	0.007	0.53	0.53
P59	616-86-4	4-ethoxy-2-nitroaniline	2.38	-3.85	-3.67	0.007	0.18	0.18
P60	618-62-2	3,5-dichloronitrobenzene	3.13	-4.63	-4.27	0.009	0.35	0.36
P61	88-72-2	2-nitrotoluene	2.30	-3.57	-3.61	0.007	-0.04	-0.04
P62	99-08-1	3-nitrotoluene	2.42	-3.63	-3.70	0.007	-0.07	-0.07
P63	99-99-0	4-nitrotoluene	2.37	-3.76	-3.66	0.007	0.10	0.10
P64	75-04-7	ethylamine	-0.13	-2.30	-1.62	0.025	0.66	0.68
P65	107-10-8	propylamine	0.48	-2.28	-2.13	0.017	0.15	0.15
P66	109-73-9	butylamine	0.97	-2.44	-2.53	0.012	-0.09	-0.09
P67	110-58-7	amylamine	1.49	-2.69	-2.95	0.009	-0.26	-0.26
P68	111-26-2	hexylamine	2.06	-3.25	-3.41	0.007	-0.16	-0.16
P69	111-68-2	heptylamine	2.57	-3.72	-3.83	0.007	-0.10	-0.11
P70	111-86-4	octylamine	3.03	-4.40	-4.20	0.009	0.20	0.20
P71	112-20-9	monylamine	3.56	-4.82	-4.62	0.012	0.19	0.20
P72	2016-57-1	decylamine	4.09	-5.18	-5.05	0.017	0.12	0.13
P73	7307-55-3	undecylamine	4.62	-5.91	-5.47	0.024	0.43	0.44
P74	124-22-1	dodecylamine	5.15	-6.27	-5.90	0.033 *	0.36	0.37

Table XIII – N model training set (continued).

ID	*CASN*	*Chemical*	*LogK_{ow}*	*Log(LC50) (mol/l)*		*Hat*	*Err.Calc.*	*Err.Pred.*
				Exp	*Pred.*			
P75	2869-34-3	tridecylamine	5.68	-6.45	-6.34	0.043 **	0.11	0.11
P76	13952-84-6	se-butylamine	0.74	-2.42	-2.34	0.015	0.08	0.08
P77	598-74-3	1,2-dimethylpropylamine	1.10	-2.49	-2.64	0.011	-0.14	-0.15
P78	78-81-9	2,2-dimethyl-1-propylamine	1.19	-2.26	-2.71	0.011	-0.45	-0.45
P79	15673-00-4	3,3-dimethylbutylamine	1.72	-2.22	-3.14	0.008	-0.92 *	-0.92 *
P80	107-45-9	t-octylamine	2.68	-3.72	-3.92	0.007	-0.19	-0.20
P81	693-16-3	1-methylheptylamine	2.81	-4.40	-4.02	0.008	0.38	0.38
P82	141-43-5	2-aminoethanol	-1.31	-1.46	-0.64	0.048 **	0.78	0.82
P83	78-96-6	1-amino-2-propanol	-0.96	-1.48	-0.95	0.040 *	0.51	0.53
P84	109-85-3	2-methoxyethylamine	-0.67	-2.16	-1.17	0.035 *	0.96 *	0.99 *
P85	100-46-9	benzylamine	1.09	-3.02	-2.62	0.011	0.39	0.40
P86	768-94-5	1-adamantanamine	1.43	-3.78	-2.89	0.009	0.88	0.89

* Chemicals with values between 2 times SDEC (or SDEP or critical HAT) and 3 times SDEC (or SDEP or critical HAT). ** Chemicals with values greater than 3 times SDEC (or SDEP or average value of HAT).

Table XIV – SIDS chemicals not suitable for QSAR 3.

	N.Comp.	SIDS Chemicals							Motivation
	24	2 94 144 173	6 95 152 174	7 101 155 175	17 109 160	39 125 162	76 126 164	77 139 166	Out of the X - domain (-1.31≤LogKow ≤ 6.20)
	13	9 82	22 92	28 100	56 104	61 107	72 142	75	In the training set
XY-domain + MOA domain	79	1 25 35 45 62 73 88 106 118 128 138 168	4 27 36 46 63 74 89 108 119 130 146 172	15 29 37 53 64 79 98 110 120 131 147	18 30 38 54 67 80 99 111 121 134 150	19 32 41 57 68 83 102 112 122 135 151	23 33 43 59 69 84 103 113 123 136 161	24 34 44 60 71 87 105 117 127 137 165	MOA ≠ N
	3	159	169	177					High leverage chemicals (structurally distant from the training chemicals)
	4	21	49	51	78				Y-Outliers (cross-validated standardised residual greater than two standard deviation units)
XY-domain	3	159	169	177					High leverage chemicals (structurally distant from the training chemicals)
	1	137							High leverage chemicals and Y-Outliers
	15	1 73 117	18 78	21 87	23 89	49 103	51 108	67 111	Y-Outliers (cross-validated standardised residual greater than two standard deviation units)

Table XV – QSAR 3 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) and mode of action domain (MOA-D).

ID	*CASN*	*EINECS name*	*LogK$_{ow}$*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
				Exp	*Pred.*		
S3	57-55-6	1,2-Propanediol	-0.78	-0.84	-1.11	0.037	-0.61
S5	58-55-9	1H-Purine-2,6-dione, 3,7-dihydro-1,3-dimethyl-	-0.39	-	-1.43	0.030	-
S8	68-12-2	Formamide, N,N-dimethyl-	-0.93	-0.84	-0.99	0.040	-0.34
S10	74-83-9	Methane, bromo-	1.18	-	-2.70	0.011	-
S11	74-87-3	Methane, chloro-	1.09	-	-2.63	0.011	-
S12	75-01-4	Ethene, chloro-	1.62	-	-3.06	0.008	-
S13	75-10-5	Methane, difluoro-	0.71	-	-2.32	0.015	-
S14	75-38-7	Ethene, 1,1-difluoro-	1.24	-	-2.75	0.010	-
S16	75-68-3	Ethane, 1-chloro-1,1-difluoro-	2.05	-	-3.40	0.007	-
S20	78-87-5	Propane, 1,2-dichloro-	2.25	-2.91	-3.57	0.007	-1.46
S21	78-92-2	2-Butanol	0.77	-1.31	-2.37	0.014	**-2.35**
S26	79-20-9	Acetic acid, methyl ester	0.37	-2.36	-2.04	0.019	0.71
S31	80-05-7	Phenol, 4,4'-(1-methylethylidene)bis-	3.64	-4.70	-4.69	0.013	0.02
S40	88-60-8	Phenol, 2-(1,1-dimethylethyl)-5-methyl-	3.97	-	-4.96	0.016	-
S42	88-74-4	Benzenamine, 2-nitro-	2.02	-	-3.38	0.007	-
S47	95-31-8	2-Benzothiazolesulfenamide, N-(1,1-dimethylethyl)-	2.56	-	-3.82	0.007	-
S48	95-49-8	Benzene, 1-chloro-2-methyl-	3.18	-	-4.32	0.009	-
S49	95-50-1	Benzene, 1,2-dichloro-	3.28	-3.41	-4.40	0.010	**-2.19**
S50	96-18-4	Propane, 1,2,3-trichloro-	2.50	-3.35	-3.77	0.007	-0.93
S51	96-29-7	2-Butanone, oxime	1.69	-2.01	-3.11	0.008	**-2.42**
S52	96-31-1	Urea, N,N'-dimethyl-	-0.62	-	-1.24	0.034	-
S55	98-07-7	Benzene, (trichloromethyl)-	3.90	-	-4.90	0.015	-
S58	98-92-0	3-Pyridinecarboxamide	-0.45	-	-1.38	0.031	-
S65	100-41-4	Benzene, ethyl-	3.03	-3.94	-4.20	0.009	-0.57
S66	102-06-7	Guanidine, N,N'-diphenyl-	2.89	-	-4.08	0.008	-
S70	105-60-2	2H-Azepin-2-one, hexahydro-	0.66	-	-2.28	0.015	-
S78	107-41-5	2,4-Pentanediol, 2-methyl-	0.58	-1.09	-2.21	0.016	**-2.48**
S81	107-98-2	2-Propanol, 1-methoxy-	-0.49	-0.64	-1.35	0.031	-1.59
S85	108-88-3	Benzene, methyl-	2.54	-3.55	-3.80	0.007	-0.55

Table XV – QSAR 3 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) and mode of action domain (MOA-D) (continued).

ID	*CASN*	*EINECS name*	*LogK$_{ow}$*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
				Exp	*Pred.*		
S86	109-66-0	Pentane-	2.80	-	-4.01	0.008	-
S90	110-83-8	Cyclohexene-	2.96	-	-4.14	0.008	-
S91	110-85-0	Piperazine-	-0.80	-	-1.10	0.037	-
S93	110-98-5	2-Propanol, 1,1'-oxybis-	-0.64	-	-1.23	0.034	-
S96	115-07-1	1-Propene	1.68	-	-3.10	0.008	-
S97	115-11-7	1-Propene, 2-methyl-	2.23	-	-3.55	0.007	-
S114	127-19-5	Acetamide, N,N-dimethyl-	-0.49	-	-1.35	0.031	-
S115	128-37-0	Phenol, 2,6-bis(1,1-dimethylethyl)-4-methyl-	5.03	-	-5.82	0.031	-
S116	135-19-3	2-Naphthalenol	2.69	-4.62	-3.92	0.007	1.54
S124	288-32-4	1H-Imidazole	0.06	-	-1.79	0.023	-
S129	556-82-1	2-Buten-1-ol, 3-methyl-	1.17	-	-2.69	0.011	-
S132	770-35-4	2-Propanol, 1-phenoxy-	1.52	-2.74	-2.97	0.009	-0.51
S133	793-24-8	1,4-Benzenediamine, N-(1,3-dimethylbutyl)-N'-phenyl-	4.68	-	-5.53	0.025	-
S140	1477-55-0	1,3-Benzenedimethanamine	0.15	-	-1.87	0.021	-
S141	1490-04-6	Cyclohexanol, 5-methyl-2-(1-methylethyl)-	3.38	-3.93	-4.48	0.011	-1.22
S143	1717-00-6	HCFC 141b	2.37	-	-3.66	0.007	-
S145	2403-88-5	4-Piperidinol, 2,2,6,6-tetramethyl-	0.94	-	-2.51	0.013	-
S148	2837-89-0	Ethane, 2-chloro-1,1,1,2-tetrafluoro-	1.86	-	-3.25	0.008	-
S149	2855-13-2	Cyclohexanemethanamine, 5-amino-1,3,3-trimethyl-	1.90	-	-3.28	0.007	-
S153	3323-53-3	Hexanedioic acid, compd. with 1,6-hexanediamine (1:1)	0.23	-	-1.93	0.020	-
S154	3452-97-9	1-Hexanol, 3,5,5-trimethyl-	3.11	-	-4.26	0.009	-
S156	4169-04-4	1-Propanol, 2-phenoxy-	1.52	-2.74	-2.97	0.009	-0.51
S157	4454-05-1	2H-Pyran, 3,4-dihydro-2-methoxy-	0.88	-	-2.46	0.013	-
S158	4457-71-0	1,5-Pentanediol, 3-methyl-	0.69	-	-2.30	0.015	-
S159	4979-32-2	2-Benzothiazolesulfenamide, N,N-dicyclohexyl-	5.96	-	-6.57	**0.049**	-
S163	6165-51-1	Benzene, 1,4-dimethyl-2-(1-phenylethyl)-	5.24	-	-5.99	0.034	-
S167	6864-37-5	Cyclohexanamine, 4,4'-methylenebis[2-methyl-	4.10	-	-5.06	0.017	-
S169	25154-52-3	Phenol, nonyl-	5.99	-6.24	-6.59	**0.050**	-0.79

Table XV – QSAR 3 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) and mode of action domain (MOA-D) (continued).

ID	*CASN*	*EINECS name*	*LogK$_{ow}$*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
				Exp	*Pred.*		
S170	25265-71-8	Propanol, oxybis-	-0.49	-	-1.35	0.031	-
S171	25321-09-9	Benzene, bis(1-methylethyl)-	4.90	-	-5.71	0.028	-
S176	56539-66-3	1-Butanol, 3-methoxy-3-methyl-	0.46	-	-2.12	0.017	-
S177	84852-15-3	Phenol, 4-nonyl-, branched	5.92	-6.24	-6.54	**0.049**	-0.68

Y outliers are highlighted in bold in the standardized residual in prediction column.
Unreliable predictions according to the leverage approach are highlighted in bold in the leverage column.

Table XVI – QSAR 3 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D).

ID	*CASN*	*EINECS name*	*MOA*	*LogK$_{ow}$*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
					Exp	*Pred.*		
S1	50-00-0	Formaldehyde-	SB	0.35	-3.08	-2.03	0.019	**2.33**
S3	57-55-6	1,2-Propanediol	NPN	-0.78	-0.84	-1.11	0.037	-0.61
S4	58-08-2	1H-Purine-2,6-dione, 3,7-dihydro-1,3,7-trimethyl-	CNS	0.16	-	-1.87	0.021	-
S5	58-55-9	1H-Purine-2,6-dione, 3,7-dihydro-1,3-dimethyl-	NPN	-0.39	-	-1.43	0.030	-
S8	68-12-2	Formamide, N,N-dimethyl-	NPN	-0.93	-0.84	-0.99	0.040	-0.34
S10	74-83-9	Methane, bromo-	NPN	1.18	-	-2.70	0.011	-
S11	74-87-3	Methane, chloro-	NPN	1.09	-	-2.63	0.011	-
S12	75-01-4	Ethene, chloro-	NPN	1.62	-	-3.06	0.008	-
S13	75-10-5	Methane, difluoro-	NPN	0.71	-	-2.32	0.015	-
S14	75-38-7	Ethene, 1,1-difluoro-	NPN	1.24	-	-2.75	0.010	-
S15	75-56-9	Oxirane, methyl-	UNK	0.37	-	-2.04	0.019	-
S16	75-68-3	Ethane, 1-chloro-1,1-difluoro-	NPN	2.05	-	-3.40	0.007	-
S18	78-59-1	2-Cyclohexen-1-one, 3,5,5-trimethyl-	MTA	2.62	-2.76	-3.87	0.007	**-2.44**
S19	78-70-6	1,6-Octadien-3-ol, 3,7-dimethyl-	PE	3.38	-	-4.48	0.011	-
S20	78-87-5	Propane, 1,2-dichloro-	NPN	2.25	-2.91	-3.57	0.007	-1.46
S21	78-92-2	2-Butanol	NPN	0.77	-1.31	-2.37	0.014	**-2.35**
S23	79-06-1	2-Propenamide	MTA	-0.81	-2.77	-1.09	0.037	**3.75**
S24	79-10-7	2-Propenoic acid	UNK	0.44	-	-2.10	0.018	-
S25	79-11-8	Acetic acid, chloro-	UNK	0.34	-	-2.02	0.019	-
S26	79-20-9	Acetic acid, methyl ester	NPN	0.37	-2.36	-2.04	0.019	0.71
S27	79-31-2	Propanoic acid, 2-methyl-	UNK	1.00	-	-2.55	0.012	-
S29	79-39-0	2-Propenamide, 2-methyl-	MTA	-0.26	-	-1.53	0.027	-
S30	79-41-4	2-Propenoic acid, 2-methyl-	UNK	0.99	-	-2.55	0.012	-
S31	80-05-7	Phenol, 4,4'-(1-methylethylidene)bis-	PN	3.64	-4.70	-4.69	0.013	0.02
S32	80-62-6	2-Propenoic acid, 2-methyl-, methyl ester	MTA	1.28	-2.55	-2.78	0.010	-0.50
S33	81-14-1	Ethanone, 1-[4-(1,1-dimethylethyl)-2,6-dimethyl-3,5-dinitrophenyl]-	UNK	4.31	-	-5.23	0.020	-
S34	81-15-2	Benzene, 1-(1,1-dimethylethyl)-3,5-dimethyl-2,4,6-trinitro-	UNK	4.45	-	-5.35	0.022	-
S35	84-74-2	1,2-Benzenedicarboxylic acid, dibutyl ester	UNK	4.61	-5.31	-5.48	0.024	-0.39
S36	87-56-9	2-Butenoic acid, 2,3-dichloro-4-oxo-, (Z)-	UNK	1.37	-	-2.85	0.010	-
S37	88-12-0	2-Pyrrolidinone, 1-ethenyl-	UNK	0.25	-	-1.95	0.020	-

Table XVI – QSAR 3 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) (continued).

ID	*CASN*	*EINECS name*	*MOA*	*LogK$_{ow}$*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
					Exp	*Pred.*		
S38	88-19-7	Benzenesulfonamide, 2-methyl-	UNK	0.92	-	-2.49	0.013	-
S40	88-60-8	Phenol, 2-(1,1-dimethylethyl)-5-methyl-	PN	3.97	-	-4.96	0.016	-
S41	88-73-3	Benzene, 1-chloro-2-nitro-	UNK	2.46	-	-3.74	0.007	-
S42	88-74-4	Benzenamine, 2-nitro-	PN	2.02	-	-3.38	0.007	-
S43	91-15-6	1,2-Benzenedicarbonitrile	UNK	1.09	-	-2.63	0.011	-
S44	91-76-9	1,3,5-Triazine-2,4-diamine, 6-phenyl-	CNS	1.44	-	-2.91	0.009	-
S45	93-68-5	Butanamide, N-(2-methylphenyl)-3-oxo-	UNK	0.99	-2.78	-2.55	0.012	0.51
S46	94-36-0	Peroxide, dibenzoyl	UNK	3.43	-	-4.52	0.011	-
S47	95-31-8	2-Benzothiazolesulfenamide, N-(1,1-dimethylethyl)-	NPN	2.56	-	-3.82	0.007	-
S48	95-49-8	Benzene, 1-chloro-2-methyl-	NPN	3.18	-	-4.32	0.009	-
S49	95-50-1	Benzene, 1,2-dichloro-	NPN	3.28	-3.41	-4.40	0.010	**-2.19**
S50	96-18-4	Propane, 1,2,3-trichloro-	NPN	2.50	-3.35	-3.77	0.007	-0.93
S51	96-29-7	2-Butanone, oxime	NPN	1.69	-2.01	-3.11	0.008	**-2.42**
S52	96-31-1	Urea, N,N'-dimethyl-	NPN	-0.62	-	-1.24	0.034	-
S53	96-33-3	2-Propenoic acid, methyl ester	MTA	0.73	-	-2.34	0.015	-
S54	97-72-3	Propanoic acid, 2-methyl-, anhydride	UNK	1.24	-	-2.75	0.010	-
S55	98-07-7	Benzene, (trichloromethyl)-	NPN	3.90	-	-4.90	0.015	-
S57	98-59-9	Benzenesulfonyl chloride, 4-methyl-	UNK	3.49	-	-4.57	0.012	-
S58	98-92-0	3-Pyridinecarboxamide	PN	-0.45	-	-1.38	0.031	-
S59	99-04-7	Benzoic acid, 3-methyl-	UNK	2.42	-	-3.70	0.007	-
S60	99-54-7	Benzene, 1,2-dichloro-4-nitro-	UNK	3.10	-	-4.25	0.009	-
S62	100-00-5	Benzene, 1-chloro-4-nitro-	UNK	2.46	-	-3.74	0.007	-
S63	100-21-0	1,4-Benzenedicarboxylic acid	UNK	1.76	-	-3.17	0.008	-
S64	100-37-8	Ethanol, 2-(diethylamino)-	UNK	0.05	-1.82	-1.78	0.023	0.09
S65	100-41-4	Benzene, ethyl-	NPN	3.03	-3.94	-4.20	0.009	-0.57
S66	102-06-7	Guanidine, N,N'-diphenyl-	NPN	2.89	-	-4.08	0.008	-
S67	102-76-1	1,2,3-Propanetriol, triacetate	UNK	0.36	-3.12	-2.04	0.019	**2.40**
S68	103-11-7	2-Propenoic acid, 2-ethylhexyl ester	MTA	4.09	-	-5.06	0.017	-
S69	103-84-4	Acetamide, N-phenyl-	UNK	1.10	-	-2.63	0.011	-
S70	105-60-2	2H-Azepin-2-one, hexahydro-	NPN	0.66	-	-2.28	0.015	-
S71	106-31-0	Butanoic acid, anhydride	UNK	1.39	-	-2.87	0.009	-
S73	106-63-8	2-Propenoic acid, 2-methylpropyl ester	MTA	2.13	-4.79	-3.47	0.007	**2.91**

Table XVI – QSAR 3 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) (continued).

ID	*CASN*	*EINECS name*	*MOA*	*LogK$_{ow}$*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
					Exp	*Pred.*		
S74	106-88-7	Oxirane, ethyl-	UNK	0.86	-	-2.44	0.013	-
S78	107-41-5	2,4-Pentanediol, 2-methyl-	NPN	0.58	-1.09	-2.21	0.016	**-2.48**
S79	107-86-8	2-Butenal, 3-methyl-	MTA	1.15	-	-2.68	0.011	-
S80	107-92-6	Butanoic-acid-	UNK	1.07	-	-2.61	0.012	-
S81	107-98-2	2-Propanol, 1-methoxy-	NPN	-0.49	-0.64	-1.35	0.031	-1.59
S83	108-65-6	2-Propanol, 1-methoxy-, acetate	EN	0.52	-	-2.17	0.017	-
S84	108-77-0	1,3,5-Triazine, 2,4,6-trichloro-	UNK	1.73	-	-3.14	0.008	-
S85	108-88-3	Benzene, methyl-	NPN	2.54	-3.55	-3.80	0.007	-0.55
S86	109-66-0	Pentane-	NPN	2.80	-	-4.01	0.008	-
S87	110-16-7	2-Butenedioic acid (Z)-	UNK	0.05	-4.37	-1.78	0.023	**5.75**
S88	110-19-0	Acetic acid, 2-methylpropyl ester	EN	1.77	-	-3.18	0.008	-
S89	110-65-6	2-Butyne-1,4-diol	PE	-0.93	-3.21	-0.99	0.040	**4.98**
S90	110-83-8	Cyclohexene-	NPN	2.96	-	-4.14	0.008	-
S91	110-85-0	Piperazine-	NPN	-0.80	-	-1.10	0.037	-
S93	110-98-5	2-Propanol, 1,1'-oxybis-	NPN	-0.64	-	-1.23	0.034	-
S96	115-07-1	1-Propene	NPN	1.68	-	-3.10	0.008	-
S97	115-11-7	1-Propene, 2-methyl-	NPN	2.23	-	-3.55	0.007	-
S98	115-86-6	Phosphoric acid, triphenyl ester	UNK	4.70	-5.59	-5.55	0.025	0.10
S99	115-95-7	1,6-Octadien-3-ol, 3,7-dimethyl-, acetate	UNK	4.39	-	-5.30	0.021	-
S102	120-61-6	1,4-Benzenedicarboxylic acid, dimethyl ester	UNK	1.66	-	-3.09	0.008	-
S103	120-80-9	1,2-Benzenediol	UNK	1.03	-4.29	-2.58	0.012	**3.78**
S105	121-91-5	1,3-Benzenedicarboxylic acid	UNK	1.76	-	-3.17	0.008	-
S106	122-52-1	Phosphorous acid, triethyl ester	UNK	0.74	-	-2.34	0.014	-
S108	123-54-6	2,4-Pentanedione	UNK	0.05	-2.86	-1.78	0.023	**2.40**
S110	123-86-4	Acetic acid, butyl ester	EN	1.85	-3.81	-3.24	0.008	1.26
S111	124-04-9	Hexanedioic-acid-	UNK	0.23	-3.18	-1.93	0.020	**2.77**
S112	126-73-8	Phosphoric-acid-tributyl-ester-	AChE	3.82	-4.77	-4.84	0.014	-0.15
S113	126-98-7	2-Propenenitrile, 2-methyl-	MTA	0.76	-	-2.36	0.014	-
S114	127-19-5	Acetamide, N,N-dimethyl-	NPN	-0.49	-	-1.35	0.031	-
S115	128-37-0	Phenol, 2,6-bis(1,1-dimethylethyl)-4-methyl-	PN	5.03	-	-5.82	0.031	-
S116	135-19-3	2-Naphthalenol	PN	2.69	-4.62	-3.92	0.007	1.54
S117	140-88-5	2-Propenoic acid, ethyl ester	MTA	1.22	-4.60	-2.73	0.011	**4.14**
S118	141-10-6	3,5,9-Undecatrien-2-one, 6,10-dimethyl-	MTA	4.43	-	-5.33	0.021	-
S119	141-32-2	2-Propenoic acid, butyl ester	MTA	2.20	-	-3.53	0.007	-

Table XVI – QSAR 3 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) (continued).

ID	*CASN*	*EINECS name*	*MOA*	*LogK$_{ow}$*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
					Exp	*Pred.*		
S120	141-78-6	Acetic-acid-ethyl-ester-	EN	0.86	-2.58	-2.44	0.013	0.31
S121	141-97-9	Butanoic acid, 3-oxo-, ethyl ester	UNK	-0.20	-	-1.58	0.026	-
S122	144-55-8	Carbonic-acid-monosodium-salt-	UNK	-0.46	-	-1.37	0.031	-
S123	150-90-3	Butanedioic acid, disodium salt	UNK	-0.75	-	-1.14	0.036	-
S124	288-32-4	1H-Imidazole	NPN	0.06	-	-1.79	0.023	-
S127	528-44-9	1,2,4-Benzenetricarboxylic acid	UNK	0.95	-	-2.51	0.013	-
S128	552-30-7	5-Isobenzofurancarboxylic acid, 1,3-dihydro-1,3-dioxo-	UNK	1.96	-	-3.33	0.007	-
S129	556-82-1	2-Buten-1-ol, 3-methyl-	NPN	1.17	-	-2.69	0.011	-
S130	611-19-8	Benzene, 1-chloro-2-(chloromethyl)-	UNK	3.44	-	-4.53	0.011	-
S131	760-23-6	1-Butene, 3,4-dichloro-	SN2	2.60	-4.18	-3.85	0.007	0.74
S132	770-35-4	2-Propanol, 1-phenoxy-	NPN	1.52	-2.74	-2.97	0.009	-0.51
S133	793-24-8	1,4-Benzenediamine, N-(1,3-dimethylbutyl)-N'-phenyl-	NPN	4.68	-	-5.53	0.025	-
S134	822-06-0	Hexane, 1,6-diisocyanato-	UNK	3.20	-	-4.33	0.010	-
S135	839-90-7	1,3,5-Triazine-2,4,6(1H,3H,5H)-trione, 1,3,5-tris(2-hydroxyethyl)-	UNK	0.07	-	-1.80	0.022	-
S136	868-77-9	2-Propenoic acid, 2-methyl-, 2-hydroxyethyl ester	MTA	0.30	-2.76	-1.99	0.019	1.71
S137	868-85-9	Phosphonic acid, dimethyl ester	UNK	-1.13	-2.69	-0.83	**0.044**	**4.18**
S138	919-30-2	3-Aminopropyl-triethoxysilane	UNK	0.31	-	-2.00	0.019	-
S140	1477-55-0	1,3-Benzenedimethanamine	NPN	0.15	-	-1.87	0.021	-
S141	1490-04-6	Cyclohexanol, 5-methyl-2-(1-methylethyl)-	NPN	3.38	-3.93	-4.48	0.011	-1.22
S143	1717-00-6	HCFC 141b	NPN	2.37	-	-3.66	0.007	-
S145	2403-88-5	4-Piperidinol, 2,2,6,6-tetramethyl-	NPN	0.94	-	-2.51	0.013	-
S146	2432-99-7	Undecanoic acid, 11-amino-	UNK	-0.16	-	-1.61	0.026	-
S147	2439-35-2	2-Propenoic acid, 2-(dimethylamino)ethyl ester	MTA	0.42	-	-2.08	0.018	-
S148	2837-89-0	Ethane, 2-chloro-1,1,1,2-tetrafluoro-	NPN	1.86	-	-3.25	0.008	-
S149	2855-13-2	Cyclohexanemethanamine, 5-amino-1,3,3-trimethyl-	NPN	1.90	-	-3.28	0.007	-
S150	2867-47-2	2-Propenoic acid, 2-methyl-, 2-(dimethylamino)ethyl ester	MTA	0.97	-	-2.53	0.012	-
S151	3268-49-3	Propanal, 3-(methylthio)-	UNK	0.41	-	-2.08	0.018	-

Table XVI – QSAR 3 predictions for the SIDS subset defined by model domain in descriptor and response space (XY-D) (continued).

ID	*CASN*	*EINECS name*	*MOA*	*LogK$_{ow}$*	*Log(LC50) (mol/l)*		*Hat*	*Std. Err.Pred.*
					Exp	*Pred.*		
S153	3323-53-3	Hexanedioic acid, compd. with 1,6-hexanediamine (1:1)	NPN	0.23	-	-1.93	0.020	-
S154	3452-97-9	1-Hexanol, 3,5,5-trimethyl-	NPN	3.11	-	-4.26	0.009	-
S156	4169-04-4	1-Propanol, 2-phenoxy-	NPN	1.52	-2.74	-2.97	0.009	-0.51
S157	4454-05-1	2H-Pyran, 3,4-dihydro-2-methoxy-	NPN	0.88	-	-2.46	0.013	-
S158	4457-71-0	1,5-Pentanediol, 3-methyl-	NPN	0.69	-	-2.30	0.015	-
S159	4979-32-2	2-Benzothiazolesulfenamide, N,N-dicyclohexyl-	NPN	5.96	-	-6.57	**0.049**	-
S161	5392-40-5	2,6-Octadienal, 3,7-dimethyl-	MTA	3.45	-	-4.54	0.011	-
S163	6165-51-1	Benzene, 1,4-dimethyl-2-(1-phenylethyl)-	NPN	5.24	-	-5.99	0.034	-
S165	6386-38-5	Benzenepropanoic acid, 3,5-bis(1,1-dimethylethyl)-4-hydroxy-, methyl ester	EN	5.06	-	-5.84	0.031	-
S167	6864-37-5	Cyclohexanamine, 4,4'-methylenebis[2-methyl-	NPN	4.10	-	-5.06	0.017	-
S168	11070-44-3	1,3-Isobenzofurandione, tetrahydromethyl-	UNK	2.64	-	-3.88	0.007	-
S169	25154-52-3	Phenol, nonyl-	PN	5.99	-6.24	-6.59	**0.050**	-0.79
S170	25265-71-8	Propanol, oxybis-	NPN	-0.49	-	-1.35	0.031	-
S171	25321-09-9	Benzene, bis(1-methylethyl)-	NPN	4.90	-	-5.71	0.028	-
S172	25321-14-6	Benzene, methyldinitro-	UNK	2.18	-4.03	-3.51	0.007	1.15
S176	56539-66-3	1-Butanol, 3-methoxy-3-methyl-	NPN	0.46	-	-2.12	0.017	-
S177	84852-15-3	Phenol, 4-nonyl-, branched	PN	5.92	-6.24	-6.54	**0.049**	-0.68

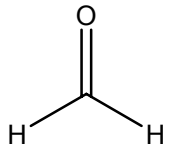
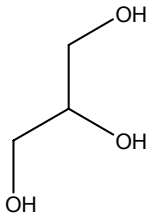
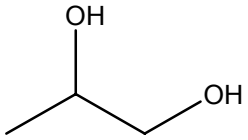
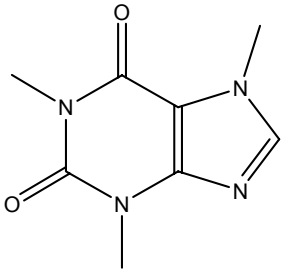
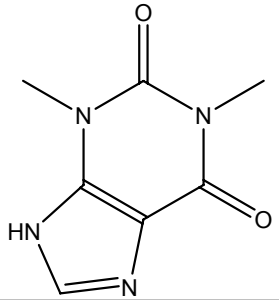
Y outliers are highlighted in bold in the standardized residual in prediction column.
Unreliable predictions according to leverage approach are highlighted in bold in the leverage column.

Table XVII – N model performance on the two to subset of SIDS data evaluated.

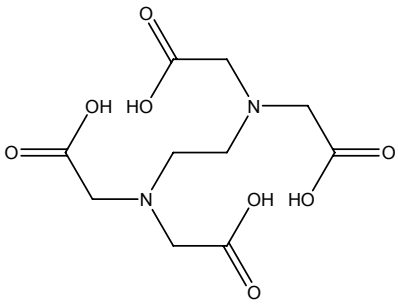
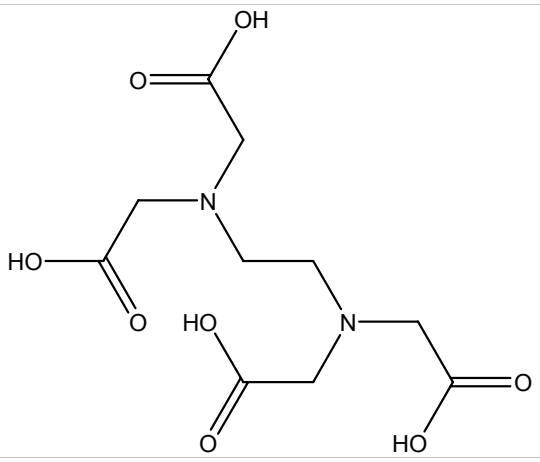
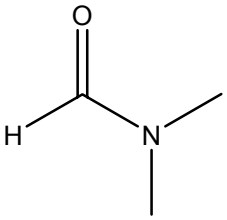
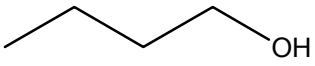
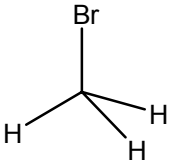
Test MOA	*N. Chemicals*	*N.Test*	*Unknown SIDS predictions*	*Total SIDS predictions*	Q^2ext
N	177 – 24 – 13– 79 – 3 - 4= 54	13	41	54	92.18
Mixed	177 – 24 – 13 – 3 - 1 -15 = 121	24	97	121	91.63

Test: number of reliable predictions for SIDS data used to evaluate the model quality.
Unknown SIDS predictions: number of reliable predictions for SIDS data lacking the Y response value (experimental LC50).
Total SIDS predictions: number of total reliable predictions provided by the model for SIDS data.

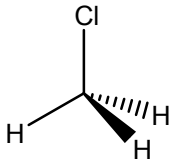
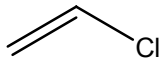
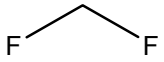
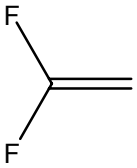
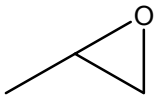
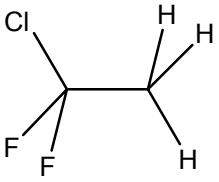
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS.

ID	CASN	EINECS name	2D Structure
1	50000	Formaldehyde O=C	
2	56815	1,2,3-Propanetriol OCC(O)CO	
3	57556	1,2-Propanediol OCC(O)C	
4	58082	1H-Purine-2,6-dione, 3,7-dihydro-1,3,7-trimethyl- N(=CN1C)C(N(C(=O)N2C)C)=C1C2=O	
5	58559	1H-Purine-2,6-dione, 3,7-dihydro-1,3-dimethyl- N(=CN1)C(C(=O)N(C2=O)C)=C1N2C	

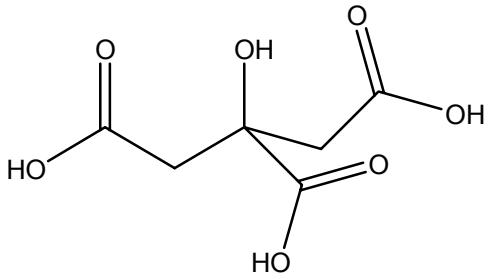
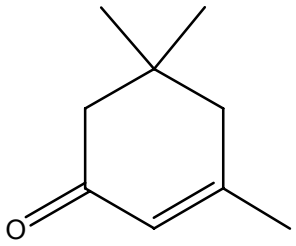
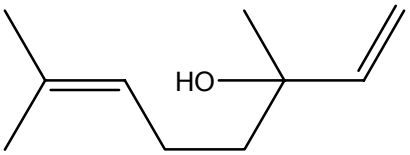
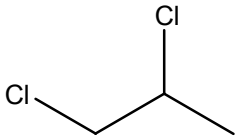
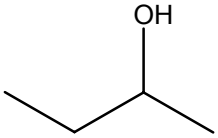
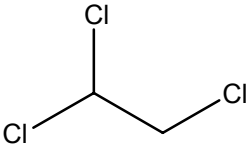
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name	2D Structure
6	60004	Glycine, N,N'-1,2-ethanediylbis[N-(carboxymethyl)- O=C(O)CN(CC(=O)O)CCN(CC(=O)O)CC(=O)O	
7	64028	Glycine, N,N'-1,2-ethanediylbis[N-(carboxymethyl)-, tetrasodium salt OC(=O)CN(CC(=O)O)CCN(CC(=O)O)CC(=O)O	
8	68122	Formamide, N,N-dimethyl- O=CN(C)C	
9	71363	1-Butanol OCCCC	
10	74839	Methane, bromo- BrC	

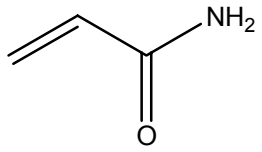
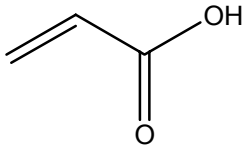
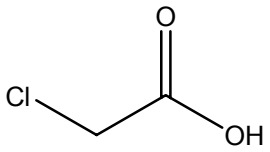
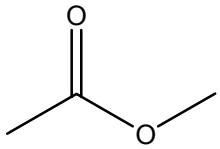
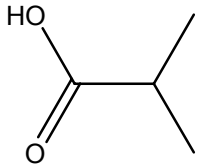
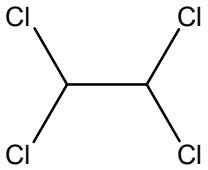
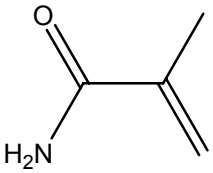
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name	2D Structure
11	74873	Methane, chloro- ClC	
12	75014	Ethene, chloro- ClC=C	
13	75105	Methane, difluoro- FCF	
14	75387	Ethene, 1,1-difluoro- FC(F)=C	
15	75569	Oxirane, methyl- O(C1)C1C	
16	75683	Ethane, 1-chloro-1,1-difluoro- ClC(F)(F)C	

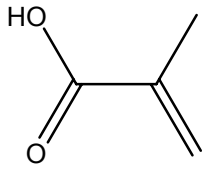
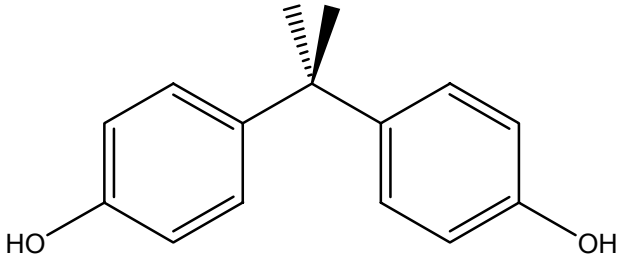
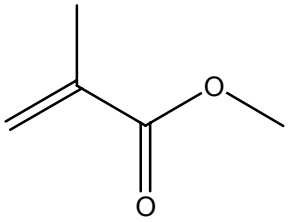
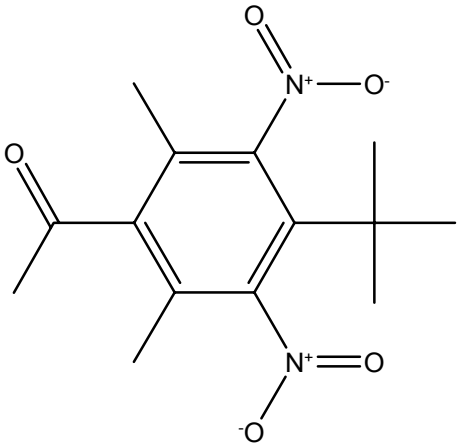
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
17	77929	1,2,3-Propanetricarboxylic acid, 2-hydroxy- O=C(O)CC(O)(C(=O)O)CC(=O)O	
18	78591	2-Cyclohexen-1-one, 3,5,5-trimethyl- O=C(C=C(C1)C)CC1(C)C	
19	78706	1,6-Octadien-3-ol, 3,7-dimethyl- OC(C=C)(C)CCC=C(C)C	
20	78875	Propane, 1,2-dichloro- ClCC(Cl)C	
21	78922	2-Butanol OC(C)CC	
22	79005	Ethane, 1,1,2-trichloro- ClC(Cl)CCl	

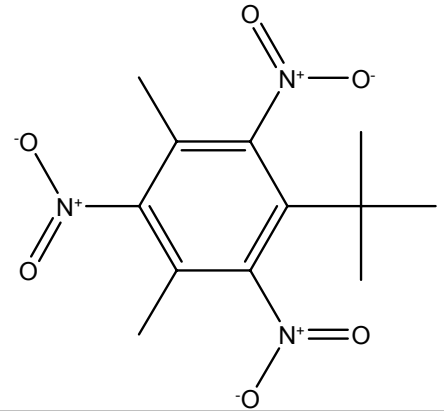
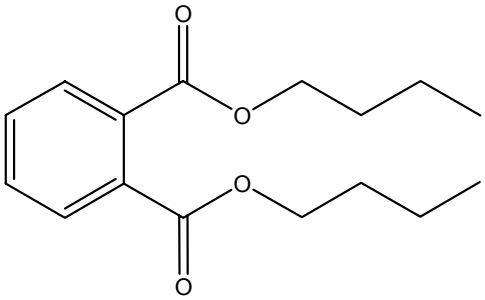
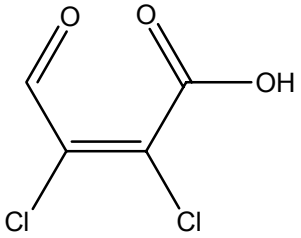
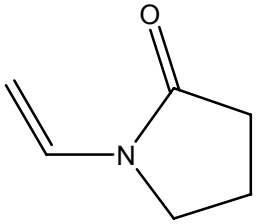
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
23	79061	2-Propenamide O=C(N)C=C	
24	79107	2-Propenoic acid O=C(O)C=C	
25	79118	Acetic acid, chloro- ClCC(=O)O	
26	79209	Acetic acid, methyl ester O=C(OC)C	
27	79312	Propanoic acid, 2-methyl- O=C(O)C(C)C	
28	79345	Ethane, 1,1,2,2-tetrachloro- ClC(Cl)C(Cl)Cl	
29	79390	2-Propenamide, 2-methyl- O=C(N)C(=C)C	

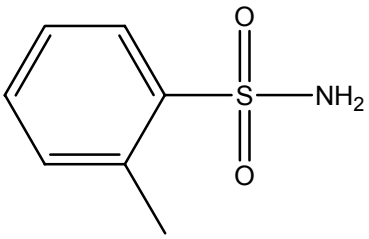
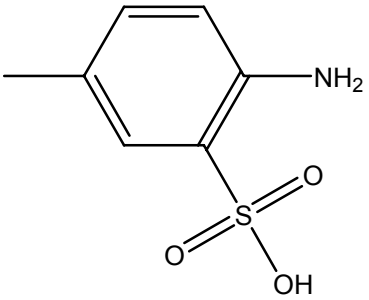
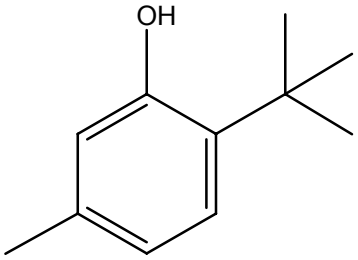
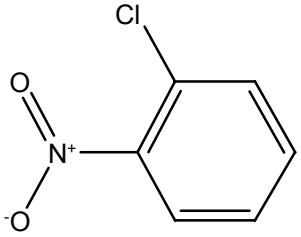
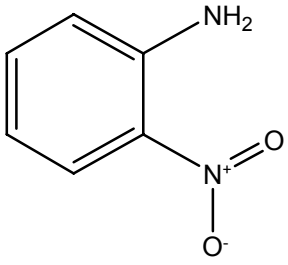
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
30	79414	2-Propenoic acid, 2-methyl- O=C(O)C(=C)C	
31	80057	Phenol, 4,4'-(1-methylethylidene)bis- O-c(ccc1C(-c(ccc2O)cc2)(C)C)cc1	
32	80626	2-Propenoic acid, 2-methyl-, methyl ester O=C(OC)C(=C)C	
33	81141	Ethanone, 1-[4-(1,1-dimethylethyl)-2,6-dimethyl-3,5-dinitrophenyl]- O=N(=O)-c(c(c(c1C)C(=O)C)C)c(c1N(=O)=O)C(C)(C)C	

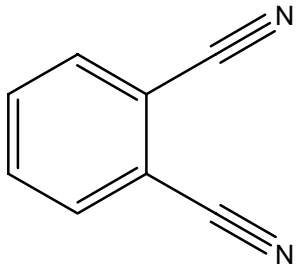
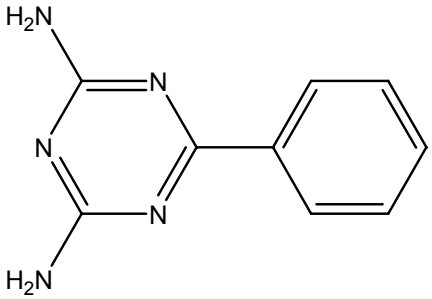
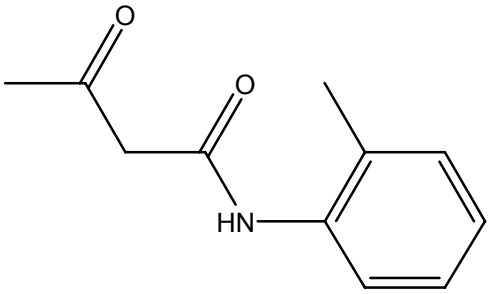
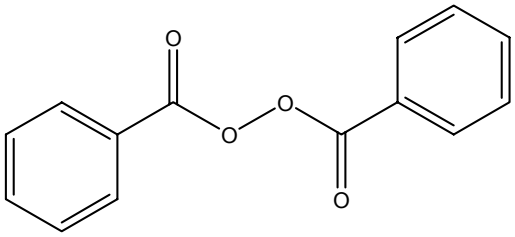
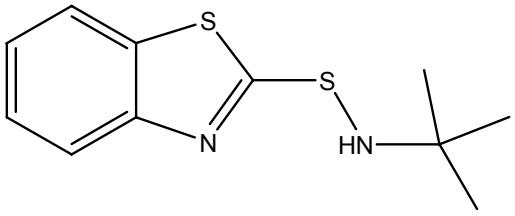
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
34	81152	Benzene, 1-(1,1-dimethylethyl)-3,5-dimethyl-2,4,6-trinitro- O=N(=O)-c(c(c(N(=O)=O)c1C(C)(C)C)C)c(c1N(=O)=O)C	
35	84742	1,2-Benzenedicarboxylic acid, dibutyl ester c(ccc1C(=O)OCCCC)cc1C(=O)OCCCC	
36	87569	2-Butenoic acid, 2,3-dichloro-4-oxo-, (Z)- ClC(=C(Cl)C(=O)O)C=O	
37	88120	2-Pyrrolidinone, 1-ethenyl- O=C(CC1)N(C1)C=C	

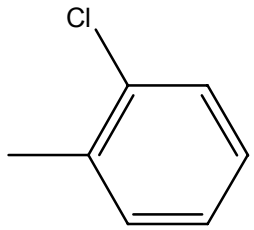
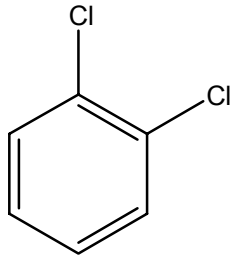
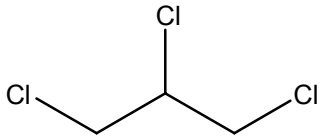
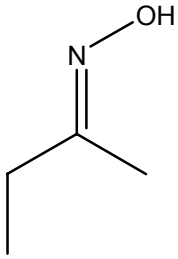
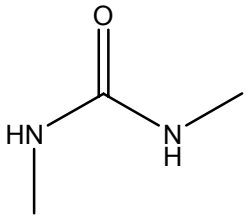
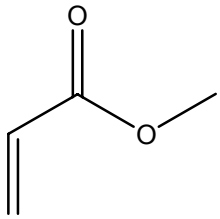
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
38	88197	Benzenesulfonamide, 2-methyl- S(=O)(=O)(N)-c(ccc1)c(c1)C	
39	88448	Benzenesulfonic acid, 2-amino-5-methyl- S(=O)(=O)(O)-c(cc(c1)C)c(c1)N	
40	88608	Phenol, 2-(1,1-dimethylethyl)-5-methyl- O-c(cc(c1)C)c(c1)C(C)(C)C	
41	88733	Benzene, 1-chloro-2-nitro- Cl-c(ccc1)c(c1)N(=O)=O	
42	88744	Benzenamine, 2-nitro- c(ccc1N(=O)=O)cc1N	

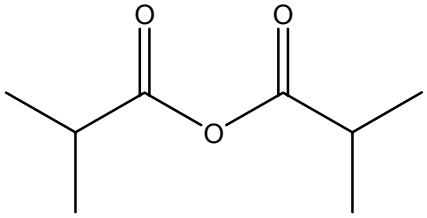
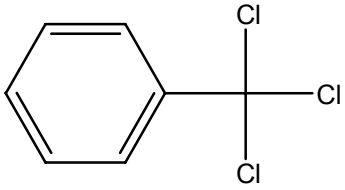
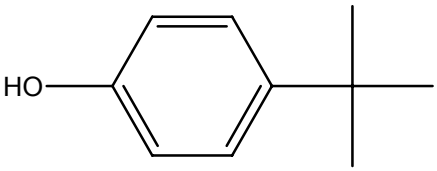
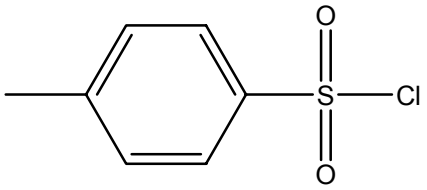
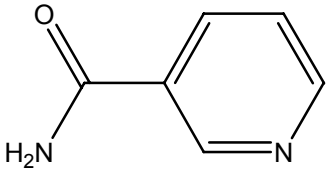
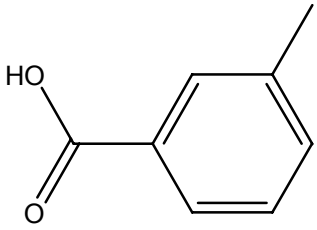
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
43	91156	1,2-Benzenedicarbonitrile c(ccc1C#N)cc1C#N	
44	91769	1,3,5-Triazine-2,4-diamine, 6-phenyl- c(ccc1-c(nc(n2)N)nc2N)cc1	
45	93685	Butanamide, N-(2-methylphenyl)-3-oxo- c(ccc1NC(=O)CC(=O)C)cc1C	
46	94360	Peroxide, dibenzoyl c(ccc1C(=O)OOC(=O)-c(ccc2)cc2)cc1	
47	95318	2-Benzothiazolesulfenamide, N-(1,1-dimethylethyl)- S(-c(ccc1)c2c1)C(SNC(C)(C)C)=N2	

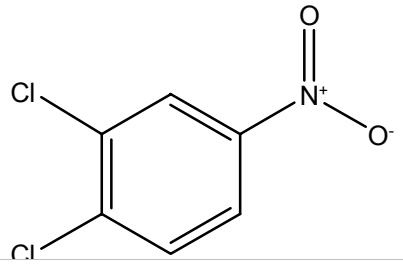
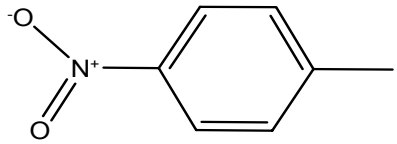
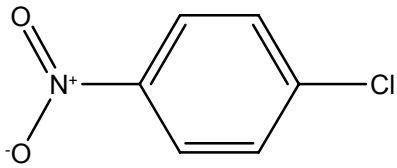
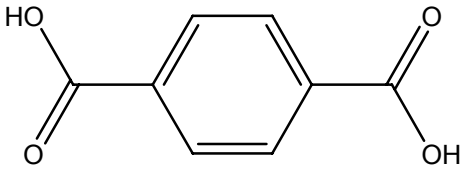
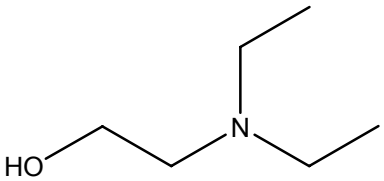
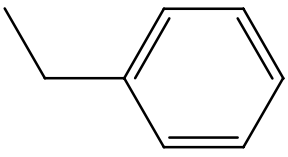
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
48	95498	Benzene, 1-chloro-2-methyl- Cl-c(ccc1)c(c1)C	
49	95501	Benzene, 1,2-dichloro- Cl-c(ccc1)c(Cl)c1	
50	96184	Propane, 1,2,3-trichloro- ClCC(Cl)CCl	
51	96297	2-Butanone, oxime ON=C(C)CC	
52	96311	Urea, N,N'-dimethyl- O=C(NC)NC	
53	96333	2-Propenoic acid, methyl ester O=C(OC)C=C	

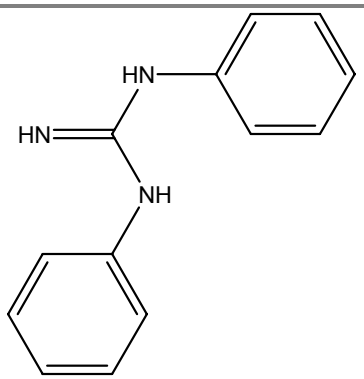
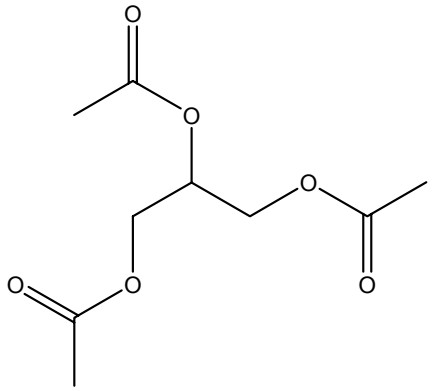
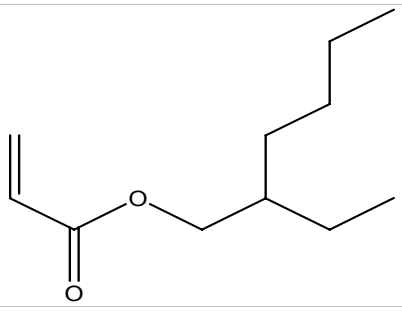
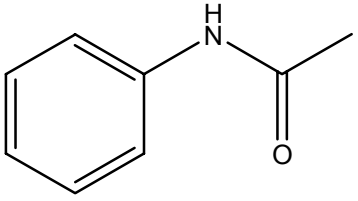
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
54	97723	Propanoic acid, 2-methyl-, anhydride O=C(OC(=O)C(C)C)C(C)C	
55	98077	Benzene, (trichloromethyl)- ClC(Cl)(Cl)-c(ccc1)cc1	
56	98544	Phenol, 4-(1,1-dimethylethyl)- O-c(ccc1C(C)(C)C)cc1	
57	98599	Benzenesulfonyl chloride, 4-methyl- ClS(=O)(=O)-c(ccc1C)cc1	
58	98920	3-Pyridinecarboxamide n(ccc1)cc1C(=O)N	
59	99047	Benzoic acid, 3-methyl- c(cc(c1)C)cc1C(=O)O	

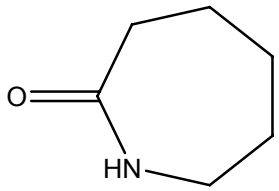
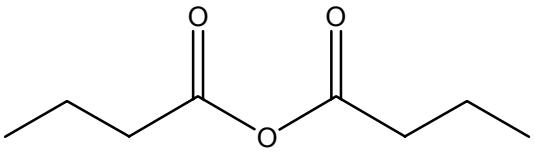
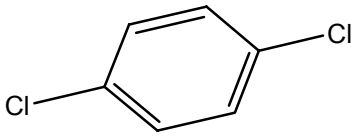
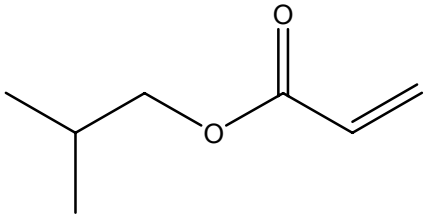
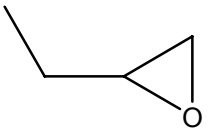
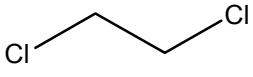
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
60	99547	Benzene, 1,2-dichloro-4-nitro- Cl-c(ccc1N(=O)=O)c(Cl)c1	
61	99990	Benzene, 1-methyl-4-nitro- c(cc(c1)N(=O)=O)c(c1)C	
62	100005	Benzene, 1-chloro-4-nitro- Cl-c(ccc1N(=O)=O)cc1	
63	100210	1,4-Benzenedicarboxylic acid c(cc(c1)C(=O)O)c(c1)C(=O)O	
64	100378	Ethanol, 2-(diethylamino)- OCCN(CC)CC	
65	100414	Benzene, ethyl- c(ccc1CC)cc1	

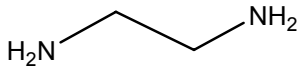
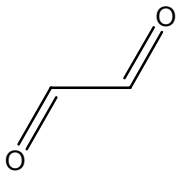
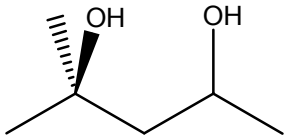
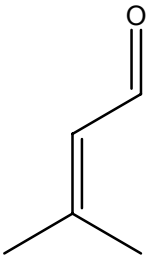
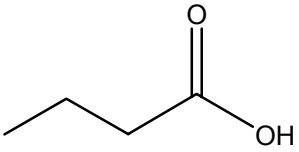
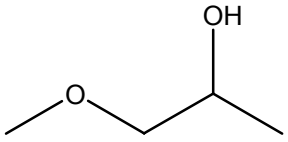
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
66	102067	Guanidine, N,N'-diphenyl- c(ccc1NC(=N)N-c(ccc2)cc2)cc1	
67	102761	1,2,3-Propanetriol, triacetate O=C(OCC(OC(=O)C)COC(=O)C)C	
68	103117	2-Propenoic acid, 2-ethylhexyl ester O=C(OCC(CC)CCCC)C=C	
69	103844	Acetamide, N-phenyl- c(ccc1NC(=O)C)cc1	

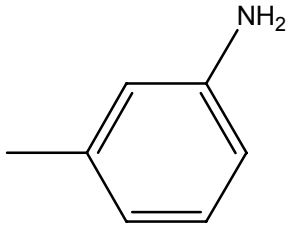
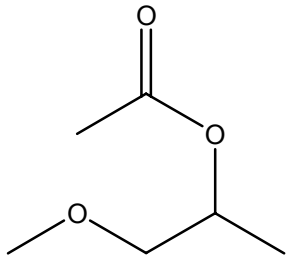
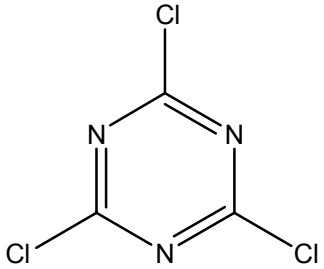
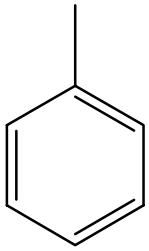
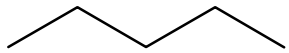
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
70	105602	2H-Azepin-2-one, hexahydro- O=C(NCC1)CCC1	
71	106310	Butanoic acid, anhydride O=C(OC(=O)CCC)CCC	
72	106467	Benzene, 1,4-dichloro- Cl-c(ccc1Cl)cc1	
73	106638	2-Propenoic acid, 2-methylpropyl ester O=C(OCC(C)C)C=C	
74	106887	Oxirane, ethyl- O(C1)C1CC	
75	107062	Ethane, 1,2-dichloro- ClCCCl	

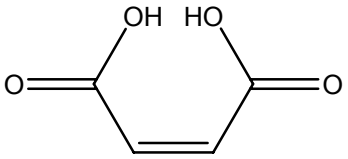
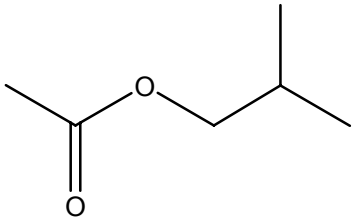
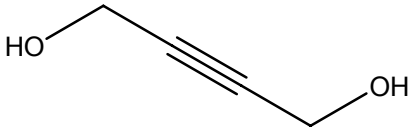
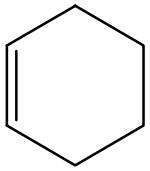
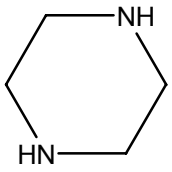
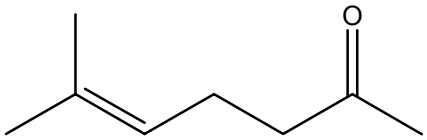
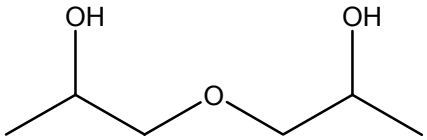
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
76	107153	1,2-Ethanediamine NCCN	
77	107222	Ethanedial- O=CC=O	
78	107415	2,4-Pentanediol, 2-methyl- OC(C)CC(O)(C)C	
79	107868	2-Butenal, 3-methyl- O=CC=C(C)C	
80	107926	Butanoic-acid- O=C(O)CCC	
81	107982	2-Propanol, 1-methoxy- OC(C)COC	

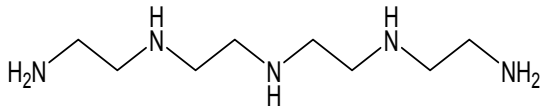
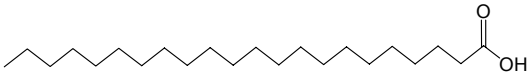
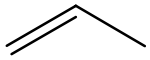
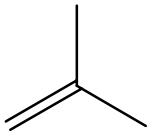
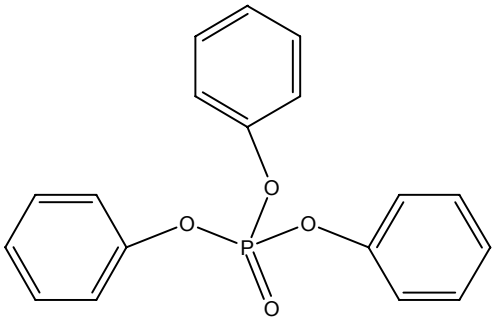
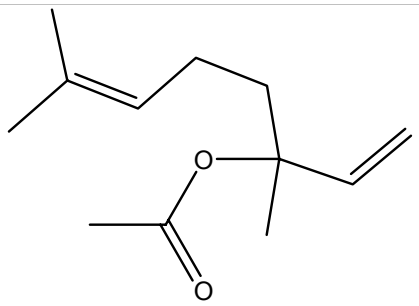
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
82	108441	Benzenamine, 3-methyl- c(cc(c1)N)cc1C	
83	108656	2-Propanol, 1-methoxy-, acetate O=C(OC(C)COC)C	
84	108770	1,3,5-Triazine, 2,4,6-trichloro- Cl-c(nc(Cl)n1)nc1Cl	
85	108883	Benzene, methyl- c(ccc1C)cc1	
86	109660	Pentane- CCCCC	

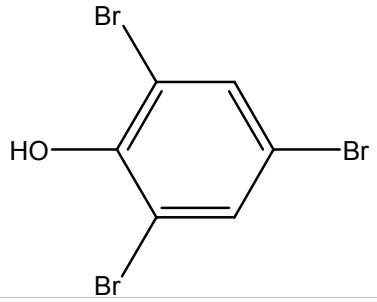
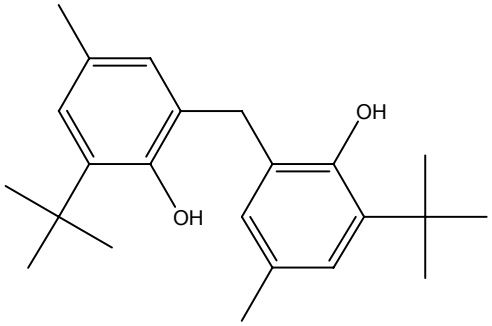
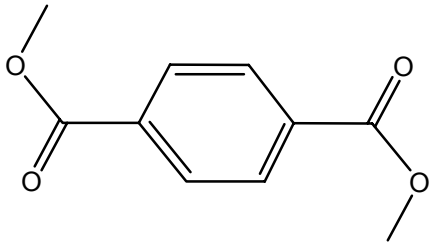
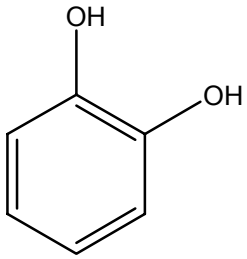
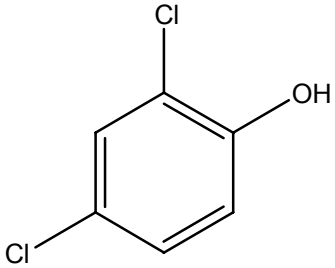
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
87	110167	2-Butenedioic acid (Z)- O=C(O)C=CC(=O)O	
88	110190	Acetic acid, 2-methylpropyl ester O=C(OCC(C)C)C	
89	110656	2-Butyne-1,4-diol OCC#CCO	
90	110838	Cyclohexene- C(=CCC1)CC1	
91	110850	Piperazine- N(CCN1)CC1	
92	110930	5-Hepten-2-one, 6-methyl- O=C(C)CCC=C(C)C	
93	110985	2-Propanol, 1,1'-oxybis- OC(C)COCC(O)C	

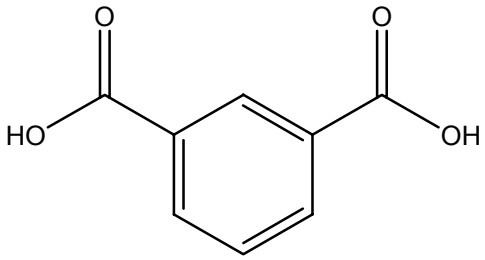
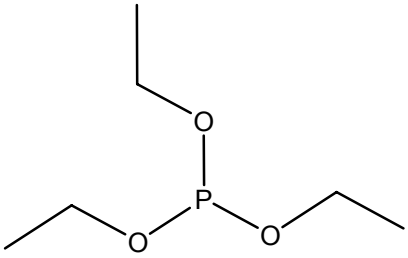
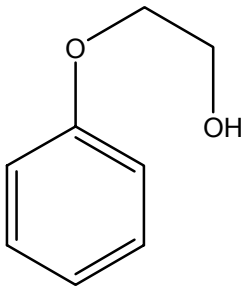
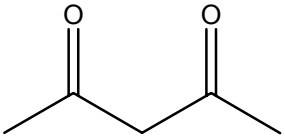
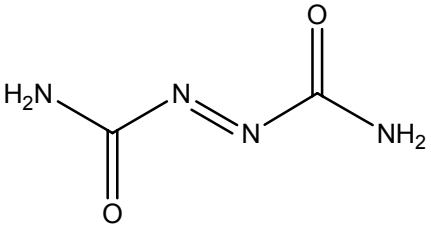
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
94	112572	1,2-Ethanediamine, N-(2-aminoethyl)-N'-[2-[(2-aminoethyl)amino]ethyl]- NCCNCCNCCNCCN	
95	112856	Docosanoic-acid- O=C(O)CCCCCCCCCCCCCCCCCCCCCC	
96	115071	1-Propene C=CC	
97	115117	1-Propene, 2-methyl- C=C(C)C	
98	115866	Phosphoric acid, triphenyl ester c(ccc1)c(c1)OP(=O)(O-c(ccc2)cc2)O-c(ccc3)cc3	
99	115957	1,6-Octadien-3-ol, 3,7-dimethyl-, acetate O=C(OC(C=C)(C)CCC=C(C)C)C	

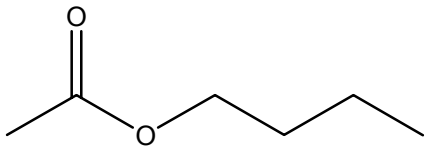
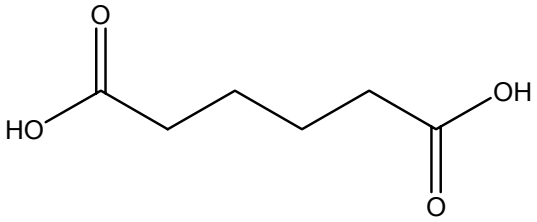
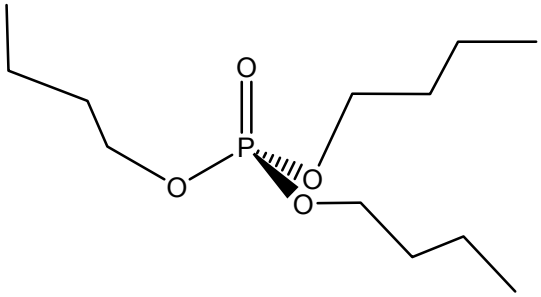
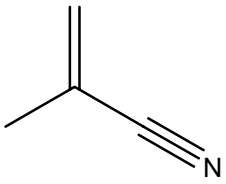
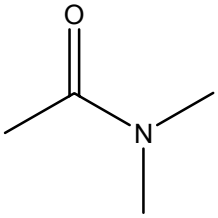
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
100	118796	Phenol, 2,4,6-tribromo- Br-c(cc(Br)c1O)cc1Br	
101	119471	Phenol, 2,2'-methylenebis[6-(1,1-dimethylethyl)-4-methyl- O-c(c(cc1C)C(C)(C)C)c(c1)C-c(cc(c2)C)c(O)c2C(C)(C)C	
102	120616	1,4-Benzenedicarboxylic acid, dimethyl ester c(cc(c1)C(=O)OC)c(c1)C(=O)OC	
103	120809	1,2-Benzenediol c(ccc1O)cc1O	
104	120832	Phenol, 2,4-dichloro- Cl-c(ccc1O)cc1Cl	

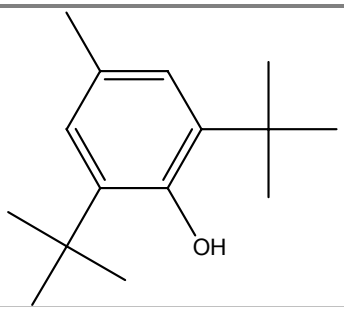
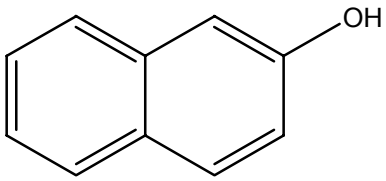
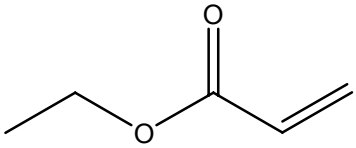
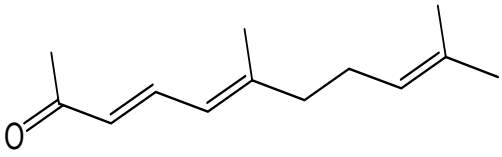
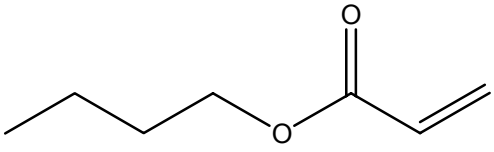
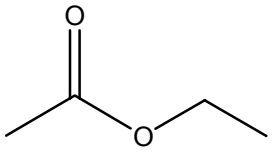
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
105	121915	1,3-Benzenedicarboxylic acid c(cc(c1)C(=O)O)cc1C(=O)O	
106	122521	Phosphorous acid, triethyl ester P(OCC)(OCC)OCC	
107	122996	Ethanol, 2-phenoxy- c(ccc1OCCO)cc1	
108	123546	2,4-Pentanedione O=C(C)CC(=O)C	
109	123773	Diazenedicarboxamide- O=C(N)N=NC(=O)N	

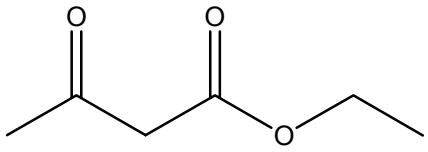
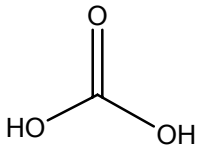
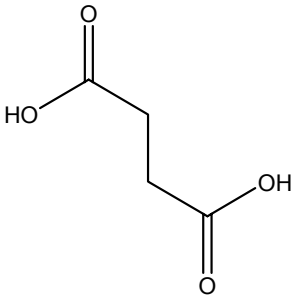
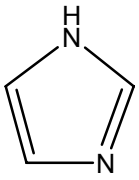
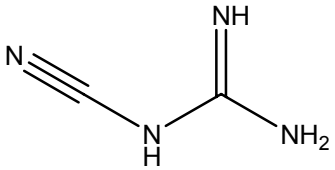
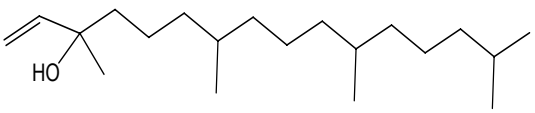
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
110	123864	Acetic acid, butyl ester O=C(OCCCC)C	
111	124049	Hexanedioic-acid- O=C(O)CCCCC(=O)O	
112	126738	Phosphoric-acid-tributyl-ester- P(=O)(OCCCC)(OCCCC)OCCCC	
113	126987	2-Propenenitrile, 2-methyl- N#CC(=C)C	
114	127195	Acetamide, N,N-dimethyl- O=C(N(C)C)C	

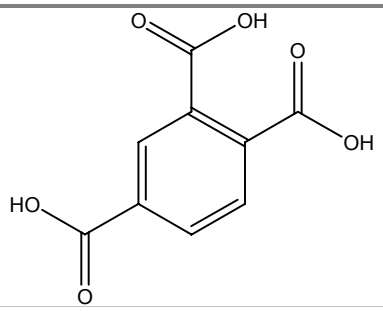
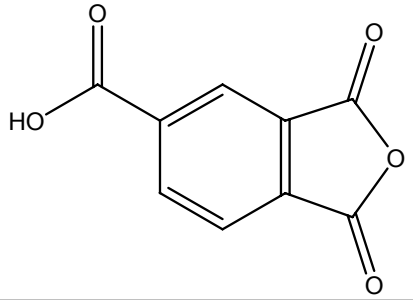
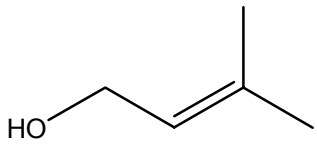
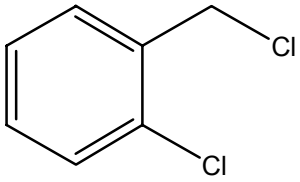
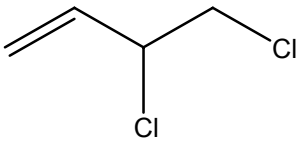
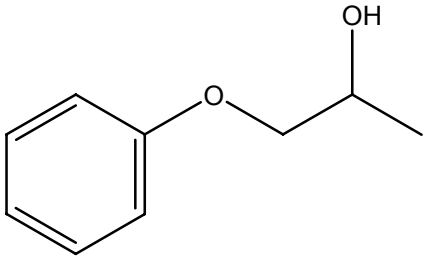
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
115	128370	Phenol, 2,6-bis(1,1-dimethylethyl)-4-methyl- O-c(c(cc1C)C(C)(C)C)c(c1)C(C)(C)C	
116	135193	2-Naphthalenol c(ccc1cc2O)cc1cc2	
117	140885	2-Propenoic acid, ethyl ester O=C(OCC)C=C	
118	141106	3,5,9-Undecatrien-2-one, 6,10-dimethyl- O=C(C=CC=C(C)CCC=C(C)C)C	
119	141322	2-Propenoic acid, butyl ester O=C(OCCCC)C=C	
120	141786	Acetic-acid-ethyl-ester- O=C(OCC)C	

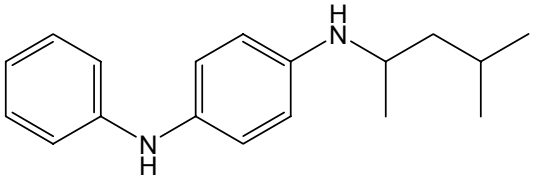
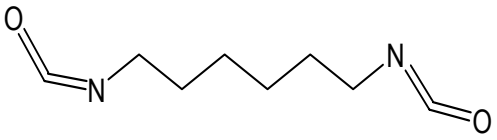
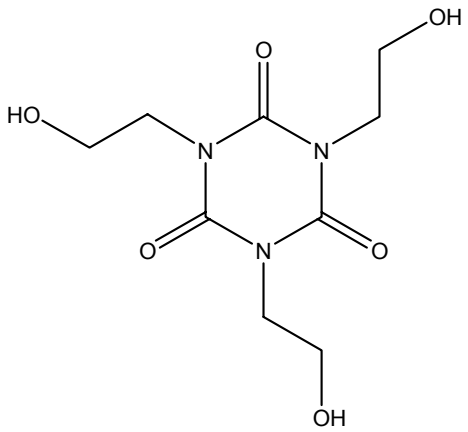
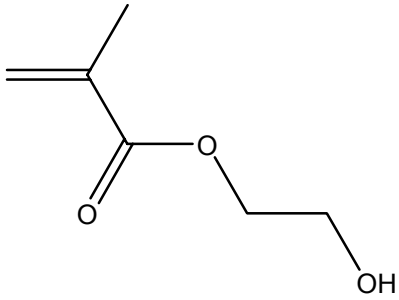
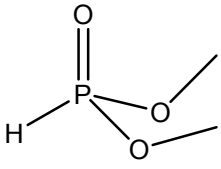
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
121	141979	Butanoic acid, 3-oxo-, ethyl ester O=C(OCC)CC(=O)C	
122	144558	Carbonic-acid-monosodium-salt- OC(=O)O	
123	150903	Butanedioic acid, disodium salt OC(=O)CCC(=O)O	
124	288324	1H-Imidazole N(=CN1)C=C1	
125	461585	Guanidine, cyano- N#CNC(=N)N	
126	505328	1-Hexadecen-3-ol, 3,7,11,15-tetramethyl- OC(C=C)(C)CCCC(C)CCCC(C)CCCC(C)C	

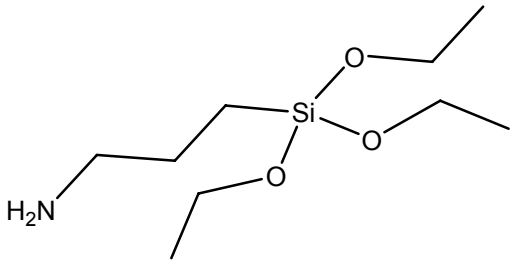
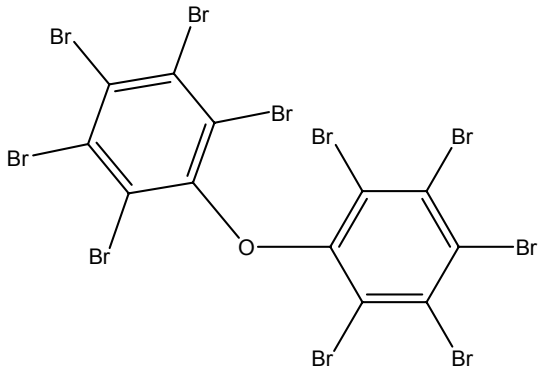
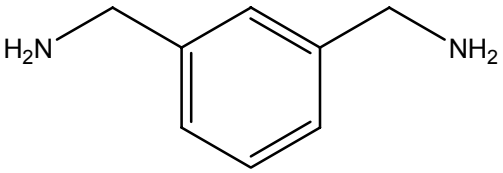
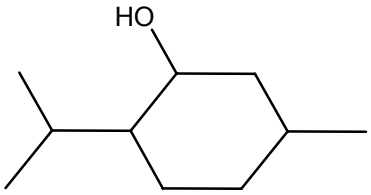
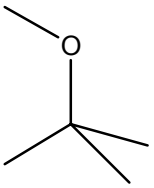
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
127	528449	1,2,4-Benzenetricarboxylic acid c(cc(c1C(=O)O)C(=O)O)c(c1)C(=O)O	
128	552307	5-Isobenzofurancarboxylic acid, 1,3-dihydro-1,3-dioxo- O(C(=O)-c1cc2)C(=O)-c1cc2C(=O)O	
129	556821	2-Buten-1-ol, 3-methyl- OCC=C(C)C	
130	611198	Benzene, 1-chloro-2-(chloromethyl)- Cl-c(ccc1)c(c1)CCl	
131	760236	1-Butene, 3,4-dichloro- ClCC(Cl)C=C	
132	770354	2-Propanol, 1-phenoxy- c(ccc1OCC(O)C)cc1	

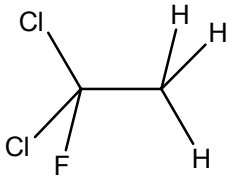
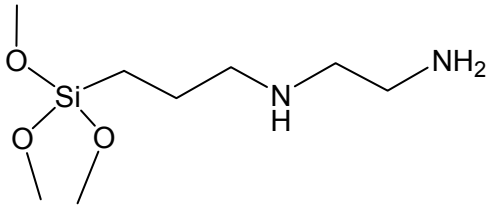
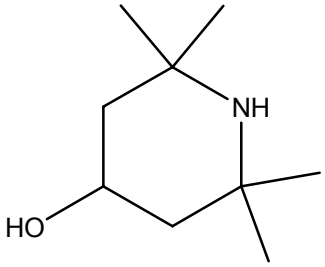
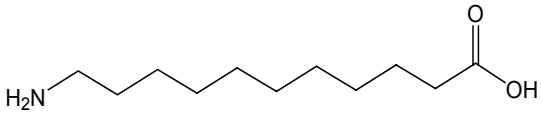
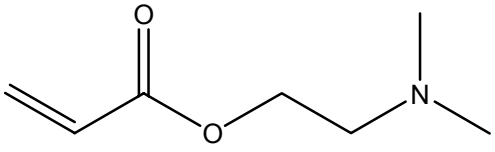
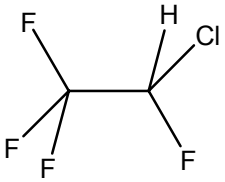
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
133	793248	1,4-Benzenediamine, N-(1,3-dimethylbutyl)-N'-phenyl- c(ccc1N-c(ccc2NC(CC(C)C)C)cc2)cc1	
134	822060	Hexane, 1,6-diisocyanato- O=C=NCCCCCCN=C=O	
135	839907	1,3,5-Triazine-2,4,6(1H,3H,5H)-trione, 1,3,5-tris(2-hydroxyethyl)- O=C(N(C(=O)N1CCO)CCO)N(C1=O)CCO	
136	868779	2-Propenoic acid, 2-methyl-, 2-hydroxyethyl ester O=C(OCCO)C(=C)C	
137	868859	Phosphonic acid, dimethyl ester P(=O)(OC)OC	

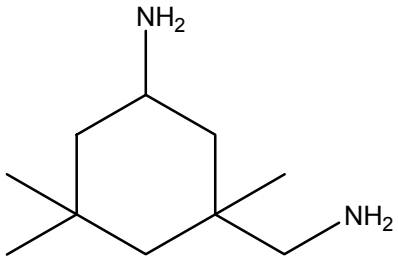
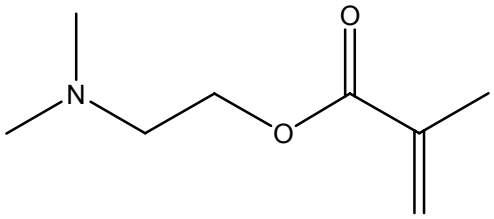
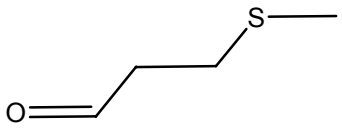
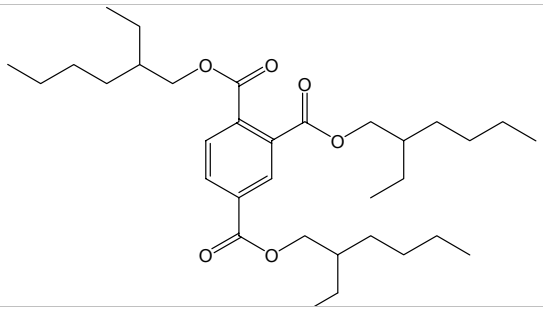
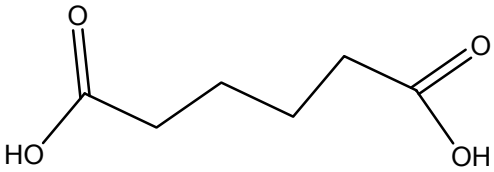
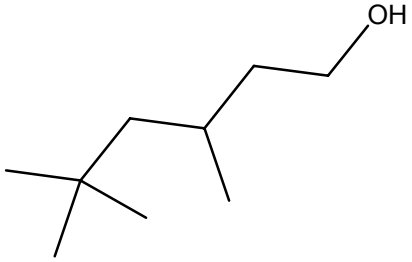
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
138	919302	3-Aminopropyl-triethoxysilane CCO[Si](CCCN)(OCC)OCC	
139	1163195	Benzene, 1,1'-oxybis[2,3,4,5,6-pentabromo- Br-c(c(Br)c(Br)c1O-c(c(Br)c(Br)c2Br)c(Br)c2Br)c(Br)c1Br	
140	1477550	1,3-Benzenedimethanamine c(cc(c1)CN)cc1CN	
141	1490046	Cyclohexanol, 5-methyl-2-(1-methylethyl)- OC(CC(C1)C)C(C1)C(C)C	
142	1634044	Propane, 2-methoxy-2-methyl- O(C)C(C)(C)C	

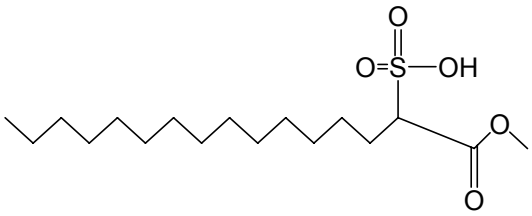
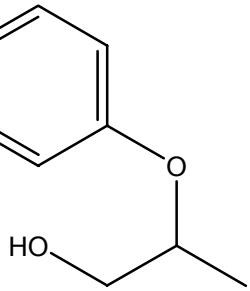
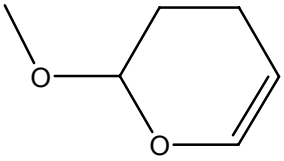
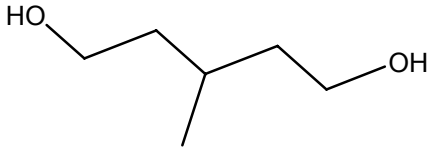
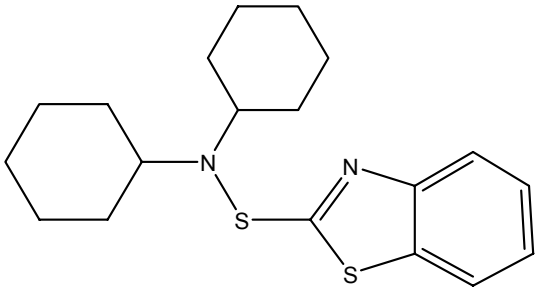
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
143	1717006	HCFC 141b CC(F)(Cl)Cl	
144	1760243	Cyclohexanol, 5-methyl-2-(1-methylethyl)-, [1R-(1alpha,2beta,5alpha)]- OC(CC(C1)C)C(C1)C(C)C	
145	2403885	4-Piperidinol, 2,2,6,6-tetramethyl- OC(CC(N1)(C)C)CC1(C)C	
146	2432997	Undecanoic acid, 11-amino- O=C(O)CCCCCCCCCCN	
147	2439352	2-Propenoic acid, 2-(dimethylamino)ethyl ester O=C(OCCN(C)C)C=C	
148	2837890	Ethane, 2-chloro-1,1,1,2-tetrafluoro- ClC(F)C(F)(F)F	

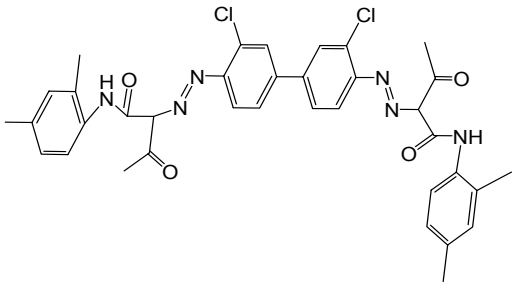
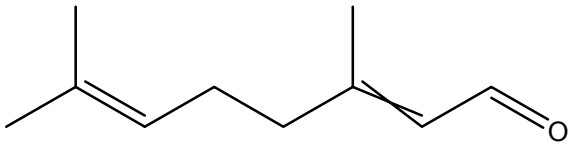
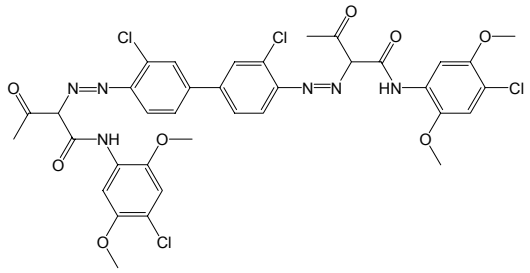
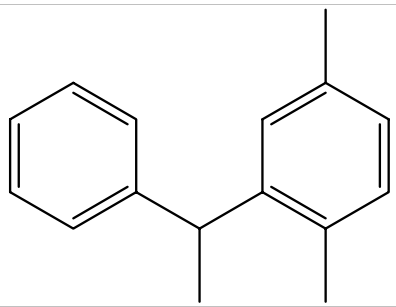
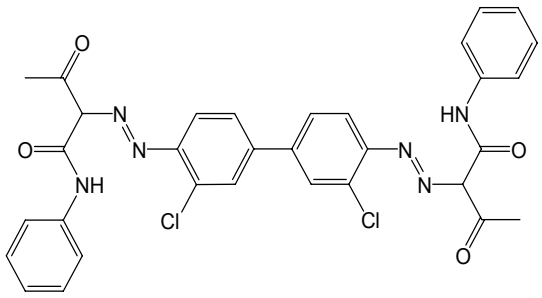
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
149	2855132	Cyclohexanemethanamine, 5-amino-1,3,3-trimethyl- NC(CC(C1)(C)C)CC1(CN)C	
150	2867472	2-Propenoic acid, 2-methyl-, 2-(dimethylamino)ethyl ester O=C(OCCN(C)C)C(=C)C	
151	3268493	Propanal, 3-(methylthio)- S(C)CCC=O	
152	3319311	1,2,4-Benzenetricarboxylic acid, tris(2-ethylhexyl) ester c(cc(c1C(=O)OCC(CC)CCCC)C(=O)OCC(CC)CCCC)c(c1)C(=O)OCC(CC)CCCC	
153	3323533	Hexanedioic acid, compd. with 1,6-hexanediamine (1:1) OC(=O)CCCCC(=O)O	
154	3452979	1-Hexanol, 3,5,5-trimethyl- OCCC(C)CC(C)(C)C	

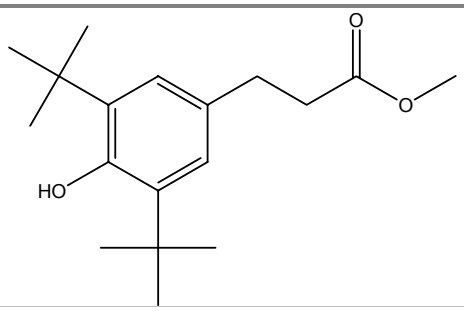
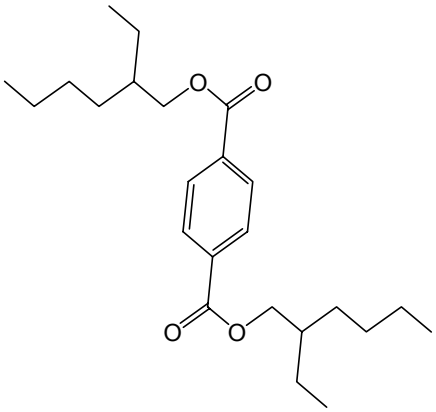
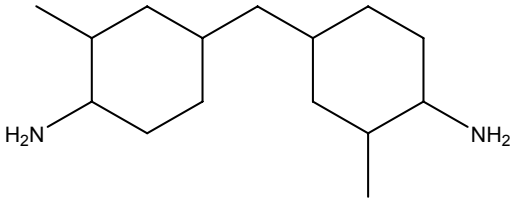
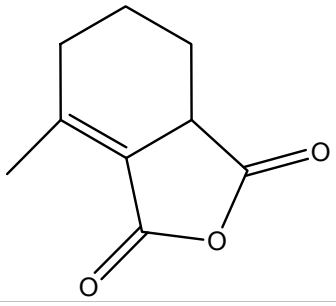
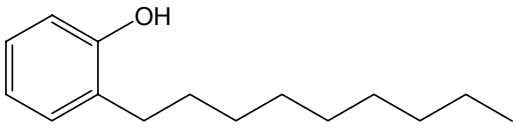
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
155	4016244	Hexadecanoic acid, 2-sulfo-, 1-methyl ester, sodium salt S(=O)(=O)(O)C(CCCCCCCCCCCCCC)C(=O)OC	
156	4169044	1-Propanol, 2-phenoxy- c(ccc1OC(CO)C)cc1	
157	4454051	2H-Pyran, 3,4-dihydro-2-methoxy- O(C=CC1)C(OC)C1	
158	4457710	1,5-Pentanediol, 3-methyl- OCCC(C)CCO	
159	4979322	2-Benzothiazolesulfenamide, N,N-dicyclohexyl- S(-c(ccc1)c2c1)C(=N2)SN(C(CCC3)CC3)C(CCC4)CC4	

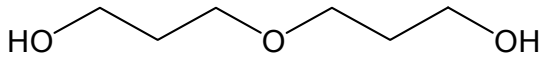
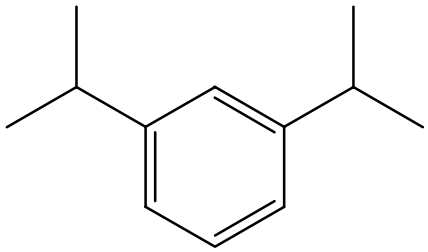
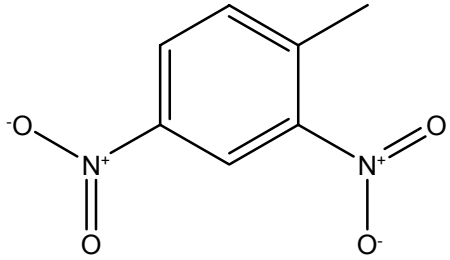
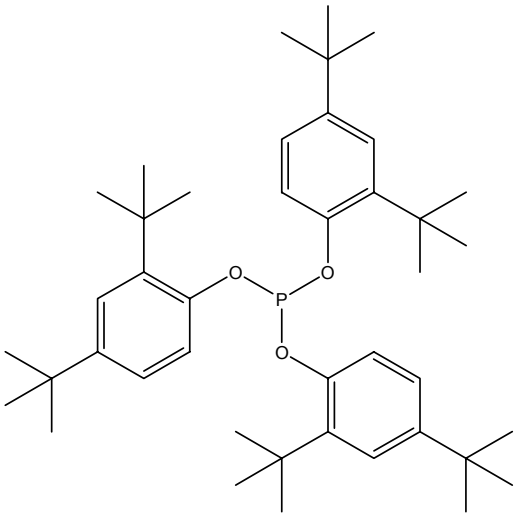
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
160	5102830	Butanamide, 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[N-(2,4-dimethylphenyl)-3-oxo- c(cc(C)c1NC(=O)C(N=N-c(c(Cl)cc2-c(cc(Cl)c3N=NC(C(=O)N-c(c(cc4C)C)cc4)C(=O)C)cc3)cc	
161	5392405	2,6-Octadienal, 3,7-dimethyl- O=CC=C(C)CCC=C(C)C	
162	5567157	Butanamide, 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[N-(4-chloro-2,5-dimethoxyphenyl)-3-oxo- COc1c(Cl)cc(OC)c(NC(=O)C(C(=O)C)N=Nc2c(Cl)cc(-c3cc(Cl)c(N=NC(C(=O)C)C(=O)Nc4ccc(Cl)cc4)cc3)cc2)c1	
163	6165511	Benzene, 1,4-dimethyl-2-(1-phenylethyl)- c(ccc1C(-c(cc(c2)C)c(c2)C)C)cc1	
164	6358856	Butanamide, 2,2'-[(3,3'-dichloro[1,1'-biphenyl]-4,4'-diyl)bis(azo)]bis[3-oxo-N-phenyl- c(ccc1NC(=O)C(N=N-c(c(Cl)cc2-c(cc(Cl)c3N=NC(C(=O)N-c(ccc4)cc4)C(=O)C)cc3)cc2)C(=O)C)cc1	

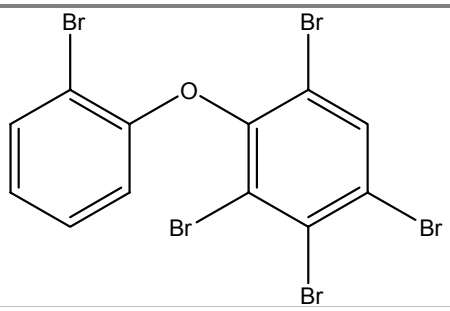
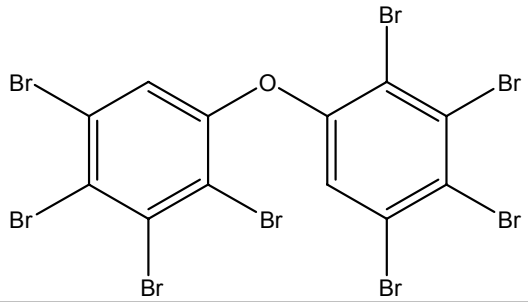
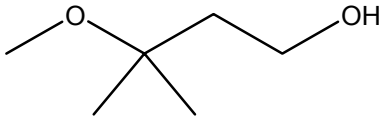
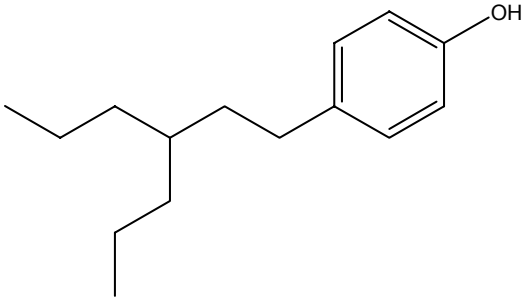
APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
165	6386385	Benzenepropanoic acid, 3,5-bis(1,1-dimethylethyl)-4-hydroxy-, methyl ester O-c(c(cc1CCC(=O)OC)C(C)(C)C)c(c1)C(C)(C)C	
166	6422862	1,4-Benzenedicarboxylic acid, bis(2-ethylhexyl) ester c(cc(c1)C(=O)OCC(CC)CCCC)c(c1)C(=O)OCC(CC)CCCC	
167	6864375	Cyclohexanamine, 4,4'-methylenebis[2-methyl- NC(CCC1CC(CC(C2N)C)CC2)C(C1)C	
168	11070443	1,3-Isobenzofurandione, tetrahydromethyl- O=C1OC(=O)C2C1=C(CCC2)C	
169	25154523	Phenol, nonyl- O-c(ccc1CCCCCCCCC)cc1	

APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
170	25265718	Propanol, oxybis- OCCCOCCCO	
171	25321099	Benzene, bis(1-methylethyl)- c(ccc1C(C)C)cc1C(C)C	
172	25321146	Benzene, methyldinitro- Cc1cccc(N(=O)(=O))c1N(=O)(=O)	
173	31570044	Phenol, 2,4-bis(1,1-dimethylethyl)-, phosphite (3:1) c(cc(c1)C(C)(C)C)c(c1C(C)(C)C)OP(O-c(ccc2C(C)(C)C)c(c2)C(C)(C)C)O-c(ccc3C(C)(C)C)c(c3)C(C)(C)C	

APPENDIX I: 2D STRUCTURES OF 177 SIDS CHEMICALS (continued).

ID	CASN	EINECS name / smiles	2D Structure
174	32534819	Benzene, 1,1'-oxybis-, pentabromo deriv. Brc1cc(c(cc1Oc2c(cc(cc2)Br)Br)Br)Br	
175	32536520	Benzene, 1,1'-oxybis-, octabromo deriv. Brc1c(c(c(c(c1Br)Br)Br)Br)Oc2cc(c(cc2Br)Br)Br	
176	56539663	1-Butanol, 3-methoxy-3-methyl- OCCC(OC)(C)C	
177	84852153	Phenol, 4-nonyl-, branched Oc1ccc(cc1)CCCCCCC(C)C	

APPENDIX II: MOLECULAR DESCRIPTOR LIST

Constitutional descriptors

Symbol	Meaning
MW	molecular weight
AMW	average molecular weight
Sv	sum of atomic van der Waals volumes (scaled on Carbon atom)
Se	sum of atomic Sanderson electronegativities (scaled on Carbon atom)
Sp	sum of atomic polarizabilities (scaled on Carbon atom)
Ss	sum of Kier-Hall electrotopological states
Mv	mean atomic van der Waals volume (scaled on Carbon atom)
Me	mean atomic Sanderson electronegativity (scaled on Carbon atom)
Mp	mean atomic polarizability (scaled on Carbon atom)
Ms	mean electrotopological state
nAT	number of atoms
nSK	number of non-H atoms
nBT	number of bonds
nBO	number of non-H bonds
nBM	number of multiple bonds
SCBO	sum of conventional bond orders (H-depleted)
ARR	aromatic ratio
nCIC	number of rings
nCIR	number of circuits
RBN	number of rotatable bonds
RBF	rotatable bond fraction
nDB	number of double bonds
nTB	number of triple bonds
nAB	number of conjugated bonds
nH	number of Hydrogen atoms
nC	number of Carbon atoms
nN	number of Nitrogen atoms
nO	number of Oxygen atoms
nP	number of Phosphorous atoms
nS	number of Sulfur atoms
nF	number of Fluorine atoms
nCL	number of Chlorine atoms
nBR	number of Bromine atoms
nI	number of Iodine atoms
nB	number of Boron atoms
nHM	number of heavy atoms
nX	number of halogen atoms
nR03	number of 3-membered rings
nR04	number of 4-membered rings
nR05	number of 5-membered rings
nR06	number of 6-membered rings
nR07	number of 7-membered rings
nR08	number of 8-membered rings
nR09	number of 9-membered rings
nR10	number of 10-membered rings

APPENDIX II: MOLECULAR DESCRIPTOR LIST (continued).

Constitutional descriptors

Symbol	Meaning
nR11	number of 11-membered rings
nR12	number of 12-membered rings
nBnz	number of benzene-like rings

Topological descriptors

Symbol	Meaning
ZM1	first Zagreb index M1
ZM1V	first Zagreb index by valence vertex degrees
ZM2	second Zagreb index M2
ZM2V	second Zagreb index by valence vertex degrees
Qindex	Quadratic index
SNar	Narumi simple topological index (log)
HNar	Narumi harmonic topological index
GNar	Narumi geometric topological index
Xt	Total structure connectivity index
Dz	Pogliani index
Ram	ramification index
Pol	polarity number
LPRS	log of product of row sums (PRS)
VDA	average vertex distance degree
MSD	mean square distance index (Balaban)
SMTI	Schultz Molecular Topological Index (MTI)
SMTIV	Schultz MTI by valence vertex degrees
GMTI	Gutman Molecular Topological Index
GMTIV	Gutman MTI by valence vertex degrees
Xu	Xu index
SPI	superpendentic index
W	Wiener W index
WA	mean Wiener index
Har	Harary H index
Har2	square reciprocal distance sum index
QW	quasi-Wiener index (Kirchhoff number)
TI1	first Mohar index TI1
TI2	second Mohar index TI2
STN	spanning tree number (log)
HyDp	hyper-distance-path index
RHyDp	reciprocal hyper-distance-path index
w	detour index
ww	hyper-detour index
Rww	reciprocal hyper-detour index
D/D	distance/detour index
Wap	all-path Wiener index
WhetZ	Wiener-type index from Z weighted distance matrix (Barysz matrix)
Whetm	Wiener-type index from mass weighted distance matrix

APPENDIX II: MOLECULAR DESCRIPTOR LIST (continued).

Topological descriptors

Symbol	Meaning
Whetv	Wiener-type index from van der Waals weighted distance matrix
Whete	Wiener-type index from electronegativity weighted distance matrix
Whetp	Wiener-type index from polarizability weighted distance matrix
J	Balaban distance connectivity index
JhetZ	Balaban-type index from Z weighted distance matrix (Barysz matrix)
Jhetm	Balaban-type index from mass weighted distance matrix
Jhetv	Balaban-type index from van der Waals weighted distance matrix
Jhete	Balaban-type index from electronegativity weighted distance matrix
Jhetp	Balaban-type index from polarizability weighted distance matrix
MAXDN	maximal electrotopological negative variation
MAXDP	maximal electrotopological positive variation
DELS	molecular electrotopological variation
TIE	E-state topological parameter
S0K	Kier symmetry index
S1K	1-path Kier alpha-modified shape index
S2K	2-path Kier alpha-modified shape index
S3K	3-path Kier alpha-modified shape index
PHI	Kier flexibility index
BLI	Kier benzene-likeliness index
PW2	path/walk 2 - Randic shape index
PW3	path/walk 3 - Randic shape index
PW4	path/walk 4 - Randic shape index
PW5	path/walk 5 - Randic shape index
PJI2	2D Petitjean shape index
CSI	eccentric connectivity index
ECC	eccentricity
AECC	average eccentricity
DECC	eccentric
MDDD	mean distance degree deviation
UNIP	unipolarity
CENT	centralization
VAR	variation
BAC	Balaban centric index
Lop	Lopping centric index
ICR	radial centric information index
D/Dr03	distance/detour ring index of order 3
D/Dr04	distance/detour ring index of order 4
D/Dr05	distance/detour ring index of order 5
D/Dr06	distance/detour ring index of order 6
D/Dr07	distance/detour ring index of order 7
D/Dr08	distance/detour ring index of order 8
D/Dr09	distance/detour ring index of order 9
D/Dr10	distance/detour ring index of order 10
D/Dr11	distance/detour ring index of order 11
D/Dr12	distance/detour ring index of order 12
T(N..N)	sum of topological distances between N..N

APPENDIX II: MOLECULAR DESCRIPTOR LIST (continued).

Topological descriptors

Symbol	Meaning
T(N..O)	sum of topological distances between N..O
T(N..S)	sum of topological distances between N..S
T(N..P)	sum of topological distances between N..P
T(N..F)	sum of topological distances between N..F
T(N..Cl)	sum of topological distances between N..Cl
T(N..Br)	sum of topological distances between N..Br
T(N..I)	sum of topological distances between N..I
T(O..O)	sum of topological distances between O..O
T(O..S)	sum of topological distances between O..S
T(O..P)	sum of topological distances between O..P
T(O..F)	sum of topological distances between O..F
T(O..Cl)	sum of topological distances between O..Cl
T(O..Br)	sum of topological distances between O..Br
T(O..I)	sum of topological distances between O..I
T(S..S)	sum of topological distances between S..S
T(S..P)	sum of topological distances between S..P
T(S..F)	sum of topological distances between S..F
T(S..Cl)	sum of topological distances between S..Cl
T(S..Br)	sum of topological distances between S..Br
T(S..I)	sum of topological distances between S..I
T(P..P)	sum of topological distances between P..P
T(P..F)	sum of topological distances between P..F
T(P..Cl)	sum of topological distances between P..Cl
T(P..Br)	sum of topological distances between P..Br
T(P..I)	sum of topological distances between P..I
T(F..F)	sum of topological distances between F..F
T(F..Cl)	sum of topological distances between F..Cl
T(F..Br)	sum of topological distances between F..Br
T(F..I)	sum of topological distances between F..I
T(Cl..Cl)	sum of topological distances between Cl..Cl
T(Cl..Br)	sum of topological distances between Cl..Br
T(Cl..I)	sum of topological distances between Cl..I
T(Br..Br)	sum of topological distances between Br..Br
T(Br..I)	sum of topological distances between Br..I
T(I..I)	sum of topological distances between I..I

WHIM descriptors

Symbol	Meaning
L1u	1st component size directional WHIM index / unweighted
L2u	2nd component size directional WHIM index / unweighted
L3u	3rd component size directional WHIM index / unweighted
P1u	1st component shape directional WHIM index / unweighted
P2u	2nd component shape directional WHIM index / unweighted
G1u	1st component symmetry directional WHIM index / unweighted
G2u	2st component symmetry directional WHIM index / unweighted
G3u	3st component symmetry directional WHIM index / unweighted

APPENDIX II: MOLECULAR DESCRIPTOR LIST (continued).

WHIM descriptors

Symbol	Meaning
E1u	1st component accessibility directional WHIM index / unweighted
E2u	2nd component accessibility directional WHIM index / unweighted
E3u	3rd component accessibility directional WHIM index / unweighted
L1m	1st component size directional WHIM index / weighted by atomic masses
L2m	2nd component size directional WHIM index / weighted by atomic masses
L3m	3rd component size directional WHIM index / weighted by atomic masses
P1m	1st component shape directional WHIM index / weighted by atomic masses
P2m	2nd component shape directional WHIM index / weighted by atomic masses
G1m	1st component symmetry directional WHIM index / weighted by atomic masses
G2m	2st component symmetry directional WHIM index / weighted by atomic masses
G3m	3st component symmetry directional WHIM index / weighted by atomic masses
E1m	1st component accessibility directional WHIM index / weighted by atomic masses
E2m	2nd component accessibility directional WHIM index / weighted by atomic masses
E3m	3rd component accessibility directional WHIM index / weighted by atomic masses
L1v	1st component size directional WHIM index / weighted by atomic van der Waals volumes
L2v	2nd component size directional WHIM index / weighted by atomic van der Waals volumes
L3v	3rd component size directional WHIM index / weighted by atomic van der Waals volumes
P1v	1st component shape directional WHIM index / weighted by atomic van der Waals volumes
P2v	2nd component shape directional WHIM index / weighted by atomic van der Waals volumes
G1v	1st component symmetry directional WHIM index / weighted by atomic van der Waals volumes
G2v	2st component symmetry directional WHIM index / weighted by atomic van der Waals volumes
G3v	3st component symmetry directional WHIM index / weighted by atomic van der Waals volumes
E1v	1st component accessibility directional WHIM index / weighted by atomic van der Waals volumes
E2v	2nd component accessibility directional WHIM index / weighted by atomic van der Waals volumes
E3v	3rd component accessibility directional WHIM index / weighted by atomic van der Waals volumes
L1e	1st component size directional WHIM index / weighted by atomic Sanderson electronegativities
L2e	2nd component size directional WHIM index / weighted by atomic Sanderson electronegativities
L3e	3rd component size directional WHIM index / weighted by atomic Sanderson electronegativities
P1e	1st component shape directional WHIM index / weighted by atomic Sanderson electronegativities
P2e	2nd component shape directional WHIM index / weighted by atomic Sanderson electronegativities

APPENDIX II: MOLECULAR DESCRIPTOR LIST (continued).

WHIM descriptors

Symbol	Meaning
G1e	1st component symmetry directional WHIM index / weighted by atomic Sanderson electronegativities
G2e	2st component symmetry directional WHIM index / weighted by atomic Sanderson electronegativities
G3e	3st component symmetry directional WHIM index / weighted by atomic Sanderson electronegativities
E1e	1st component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities
E2e	2nd component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities
E3e	3rd component accessibility directional WHIM index / weighted by atomic Sanderson electronegativities
L1p	1st component size directional WHIM index / weighted by atomic polarizabilities
L2p	2nd component size directional WHIM index / weighted by atomic polarizabilities
L3p	3rd component size directional WHIM index / weighted by atomic polarizabilities
P1p	1st component shape directional WHIM index / weighted by atomic polarizabilities
P2p	2nd component shape directional WHIM index / weighted by atomic polarizabilities
G1p	1st component symmetry directional WHIM index / weighted by atomic polarizabilities
G2p	2st component symmetry directional WHIM index / weighted by atomic polarizabilities
G3p	3st component symmetry directional WHIM index / weighted by atomic polarizabilities
E1p	1st component accessibility directional WHIM index / weighted by atomic polarizabilities
E2p	2nd component accessibility directional WHIM index / weighted by atomic polarizabilities
E3p	3rd component accessibility directional WHIM index / weighted by atomic polarizabilities
L1s	1st component size directional WHIM index / weighted by atomic electrotopological states
L2s	2nd component size directional WHIM index / weighted by atomic electrotopological states
L3s	3rd component size directional WHIM index / weighted by atomic electrotopological states
P1s	1st component shape directional WHIM index / weighted by atomic electrotopological states
P2s	2nd component shape directional WHIM index / weighted by atomic electrotopological states
G1s	1st component symmetry directional WHIM index / weighted by atomic electrotopological states
G2s	2st component symmetry directional WHIM index / weighted by atomic electrotopological states
G3s	3st component symmetry directional WHIM index / weighted by atomic electrotopological states

APPENDIX II: MOLECULAR DESCRIPTOR LIST (continued).

WHIM descriptors

Symbol	Meaning
E1s	1st component accessibility directional WHIM index / weighted by atomic electrotopological states
E2s	2nd component accessibility directional WHIM index / weighted by atomic electrotopological states
E3s	3rd component accessibility directional WHIM index / weighted by atomic electrotopological states
Tu	T total size index / unweighted
Tm	T total size index / weighted by atomic masses
Tv	T total size index / weighted by atomic van der Waals volumes
Te	T total size index / weighted by atomic Sanderson electronegativities
Tp	T total size index / weighted by atomic polarizabilities
Ts	T total size index / weighted by atomic electrotopological states
Au	A total size index / unweighted
Am	A total size index / weighted by atomic masses
Av	A total size index / weighted by atomic van der Waals volumes
Ae	A total size index / weighted by atomic Sanderson electronegativities
Ap	A total size index / weighted by atomic polarizabilities
As	A total size index / weighted by atomic electrotopological states
Gu	G total symmetry index / unweighted
Gm	G total symmetry index / weighted by atomic masses
Gs	G total symmetry index / weighted by atomic electrotopological states
Ku	K global shape index / unweighted
Km	K global shape index / weighted by atomic masses
Kv	K global shape index / weighted by atomic van der Waals volumes
Ke	K global shape index / weighted by atomic Sanderson electronegativities
Kp	K global shape index / weighted by atomic polarizabilities
Ks	K global shape index / weighted by atomic electrotopological states
Du	D total accessibility index / unweighted
Dm	D total accessibility index / weighted by atomic masses
Dv	D total accessibility index / weighted by atomic van der Waals volumes
De	D total accessibility index / weighted by atomic Sanderson electronegativities
Dp	D total accessibility index / weighted by atomic polarizabilities
Ds	D total accessibility index / weighted by atomic electrotopological states
Vu	V total size index / unweighted
Vm	V total size index / weighted by atomic masses
Vv	V total size index / weighted by atomic van der Waals volumes
Ve	V total size index / weighted by atomic Sanderson electronegativities
Vp	V total size index / weighted by atomic polarizabilities
Vs	V total size index / weighted by atomic electrotopological states

GETAWAY descriptors

Symbol	Meaning
ITH	total information content on the leverage equality
ISH	standardized information content on the leverage equality
HIC	mean information content on the leverage magnitude
HGM	geometric mean on the leverage magnitude

APPENDIX II: MOLECULAR DESCRIPTOR LIST (continued).

GETAWAY descriptors

Symbol	Meaning
H0u	H autocorrelation of lag 0 / unweighted
H1u	H autocorrelation of lag 1 / unweighted
H2u	H autocorrelation of lag 2 / unweighted
H3u	H autocorrelation of lag 3 / unweighted
H4u	H autocorrelation of lag 4 / unweighted
H5u	H autocorrelation of lag 5 / unweighted
H6u	H autocorrelation of lag 6 / unweighted
H7u	H autocorrelation of lag 7 / unweighted
H8u	H autocorrelation of lag 8 / unweighted
HTu	H total index / unweighted
HATS0u	leverage-weighted autocorrelation of lag 0 / unweighted
HATS1u	leverage-weighted autocorrelation of lag 1 / unweighted
HATS2u	leverage-weighted autocorrelation of lag 2 / unweighted
HATS3u	leverage-weighted autocorrelation of lag 3 / unweighted
HATS4u	leverage-weighted autocorrelation of lag 4 / unweighted
HATS5u	leverage-weighted autocorrelation of lag 5 / unweighted
HATS6u	leverage-weighted autocorrelation of lag 6 / unweighted
HATS7u	leverage-weighted autocorrelation of lag 7 / unweighted
HATS8u	leverage-weighted autocorrelation of lag 8 / unweighted
HATSu	leverage-weighted total index / unweighted
H0m	H autocorrelation of lag 0 / weighted by atomic masses
H1m	H autocorrelation of lag 1 / weighted by atomic masses
H2m	H autocorrelation of lag 2 / weighted by atomic masses
H3m	H autocorrelation of lag 3 / weighted by atomic masses
H4m	H autocorrelation of lag 4 / weighted by atomic masses
H5m	H autocorrelation of lag 5 / weighted by atomic masses
H6m	H autocorrelation of lag 6 / weighted by atomic masses
H7m	H autocorrelation of lag 7 / weighted by atomic masses
H8m	H autocorrelation of lag 8 / weighted by atomic masses
HTm	H total index / weighted by atomic masses
HATS0m	leverage-weighted autocorrelation of lag 0 / weighted by atomic masses
HATS1m	leverage-weighted autocorrelation of lag 1 / weighted by atomic masses
HATS2m	leverage-weighted autocorrelation of lag 2 / weighted by atomic masses
HATS3m	leverage-weighted autocorrelation of lag 3 / weighted by atomic masses
HATS4m	leverage-weighted autocorrelation of lag 4 / weighted by atomic masses
HATS5m	leverage-weighted autocorrelation of lag 5 / weighted by atomic masses
HATS6m	leverage-weighted autocorrelation of lag 6 / weighted by atomic masses
HATS7m	leverage-weighted autocorrelation of lag 7 / weighted by atomic masses
HATS8m	leverage-weighted autocorrelation of lag 8 / weighted by atomic masses
HATSm	leverage-weighted total index / weighted by atomic masses
H0v	H autocorrelation of lag 0 / weighted by atomic van der Waals volumes
H1v	H autocorrelation of lag 1 / weighted by atomic van der Waals volumes
H2v	H autocorrelation of lag 2 / weighted by atomic van der Waals volumes
H3v	H autocorrelation of lag 3 / weighted by atomic van der Waals volumes
H4v	H autocorrelation of lag 4 / weighted by atomic van der Waals volumes
H5v	H autocorrelation of lag 5 / weighted by atomic van der Waals volumes

APPENDIX II: MOLECULAR DESCRIPTOR LIST (continued).

GETAWAY descriptors

Symbol	Meaning
H6v	H autocorrelation of lag 6 / weighted by atomic van der Waals volumes
H7v	H autocorrelation of lag 7 / weighted by atomic van der Waals volumes
H8v	H autocorrelation of lag 8 / weighted by atomic van der Waals volumes
HTv	H total index / weighted by atomic van der Waals volumes
HATS0v	leverage-weighted autocorrelation of lag 0 / weighted by atomic van der Waals volumes
HATS1v	leverage-weighted autocorrelation of lag 1 / weighted by atomic van der Waals volumes
HATS2v	leverage-weighted autocorrelation of lag 2 / weighted by atomic van der Waals volumes
HATS3v	leverage-weighted autocorrelation of lag 3 / weighted by atomic van der Waals volumes
HATS4v	leverage-weighted autocorrelation of lag 4 / weighted by atomic van der Waals volumes
HATS5v	leverage-weighted autocorrelation of lag 5 / weighted by atomic van der Waals volumes
HATS6v	leverage-weighted autocorrelation of lag 6 / weighted by atomic van der Waals volumes
HATS7v	leverage-weighted autocorrelation of lag 7 / weighted by atomic van der Waals volumes
HATS8v	leverage-weighted autocorrelation of lag 8 / weighted by atomic van der Waals volumes
HATSv	leverage-weighted total index / weighted by atomic van der Waals volumes
H0e	H autocorrelation of lag 0 / weighted by atomic Sanderson electronegativities
H1e	H autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities
H2e	H autocorrelation of lag 2 / weighted by atomic Sanderson electronegativities
H3e	H autocorrelation of lag 3 / weighted by atomic Sanderson electronegativities
H4e	H autocorrelation of lag 4 / weighted by atomic Sanderson electronegativities
H5e	H autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities
H6e	H autocorrelation of lag 6 / weighted by atomic Sanderson electronegativities
H7e	H autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities
H8e	H autocorrelation of lag 8 / weighted by atomic Sanderson electronegativities
HTe	H total index / weighted by atomic Sanderson electronegativities
HATS0e	leverage-weighted autocorrelation of lag 0 / weighted by atomic Sanderson electronegativities
HATS1e	leverage-weighted autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities
HATS2e	leverage-weighted autocorrelation of lag 2 / weighted by atomic Sanderson electronegativities
HATS3e	leverage-weighted autocorrelation of lag 3 / weighted by atomic Sanderson electronegativities
HATS4e	leverage-weighted autocorrelation of lag 4 / weighted by atomic Sanderson electronegativities
HATS5e	leverage-weighted autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities
HATS6e	leverage-weighted autocorrelation of lag 6 / weighted by atomic Sanderson electronegativities

APPENDIX II: MOLECULAR DESCRIPTOR LIST (continued).

GETAWAY descriptors

Symbol	Meaning
HATS7e	leverage-weighted autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities
HATS8e	leverage-weighted autocorrelation of lag 8 / weighted by atomic Sanderson electronegativities
HATSe	leverage-weighted total index / weighted by atomic Sanderson electronegativities
H0p	H autocorrelation of lag 0 / weighted by atomic polarizabilities
H1p	H autocorrelation of lag 1 / weighted by atomic polarizabilities
H2p	H autocorrelation of lag 2 / weighted by atomic polarizabilities
H3p	H autocorrelation of lag 3 / weighted by atomic polarizabilities
H4p	H autocorrelation of lag 4 / weighted by atomic polarizabilities
H5p	H autocorrelation of lag 5 / weighted by atomic polarizabilities
H6p	H autocorrelation of lag 6 / weighted by atomic polarizabilities
H7p	H autocorrelation of lag 7 / weighted by atomic polarizabilities
H8p	H autocorrelation of lag 8 / weighted by atomic polarizabilities
HTp	H total index / weighted by atomic polarizabilities
HATS0p	leverage-weighted autocorrelation of lag 0 / weighted by atomic polarizabilities
HATS1p	leverage-weighted autocorrelation of lag 1 / weighted by atomic polarizabilities
HATS2p	leverage-weighted autocorrelation of lag 2 / weighted by atomic polarizabilities
HATS3p	leverage-weighted autocorrelation of lag 3 / weighted by atomic polarizabilities
HATS4p	leverage-weighted autocorrelation of lag 4 / weighted by atomic polarizabilities
HATS5p	leverage-weighted autocorrelation of lag 5 / weighted by atomic polarizabilities
HATS6p	leverage-weighted autocorrelation of lag 6 / weighted by atomic polarizabilities
HATS7p	leverage-weighted autocorrelation of lag 7 / weighted by atomic polarizabilities
HATS8p	leverage-weighted autocorrelation of lag 8 / weighted by atomic polarizabilities
HATSp	leverage-weighted total index / weighted by atomic polarizabilities
RCON	Randic-type R matrix connectivity
RARS	R matrix average row sum
REIG	first eigenvalue of the R matrix
R1u	R autocorrelation of lag 1 / unweighted
R2u	R autocorrelation of lag 2 / unweighted
R3u	R autocorrelation of lag 3 / unweighted
R4u	R autocorrelation of lag 4 / unweighted
R5u	R autocorrelation of lag 5 / unweighted
R6u	R autocorrelation of lag 6 / unweighted
R7u	R autocorrelation of lag 7 / unweighted
R8u	R autocorrelation of lag 8 / unweighted
RTu	R total index / unweighted
R1u+	R maximal autocorrelation of lag 1 / unweighted
R2u+	R maximal autocorrelation of lag 2 / unweighted
R3u+	R maximal autocorrelation of lag 3 / unweighted
R4u+	R maximal autocorrelation of lag 4 / unweighted
R5u+	R maximal autocorrelation of lag 5 / unweighted
R6u+	R maximal autocorrelation of lag 6 / unweighted
R7u+	R maximal autocorrelation of lag 7 / unweighted
R8u+	R maximal autocorrelation of lag 8 / unweighted

APPENDIX II: MOLECULAR DESCRIPTOR LIST (continued).

GETAWAY descriptors

Symbol	Meaning
RTu+	R maximal index / unweighted
R1m	R autocorrelation of lag 1 / weighted by atomic masses
R2m	R autocorrelation of lag 2 / weighted by atomic masses
R3m	R autocorrelation of lag 3 / weighted by atomic masses
R4m	R autocorrelation of lag 4 / weighted by atomic masses
R5m	R autocorrelation of lag 5 / weighted by atomic masses
R6m	R autocorrelation of lag 6 / weighted by atomic masses
R7m	R autocorrelation of lag 7 / weighted by atomic masses
R8m	R autocorrelation of lag 8 / weighted by atomic masses
RTm	R total index / weighted by atomic masses
R1m+	R maximal autocorrelation of lag 1 / weighted by atomic masses
R2m+	R maximal autocorrelation of lag 2 / weighted by atomic masses
R3m+	R maximal autocorrelation of lag 3 / weighted by atomic masses
R4m+	R maximal autocorrelation of lag 4 / weighted by atomic masses
R5m+	R maximal autocorrelation of lag 5 / weighted by atomic masses
R6m+	R maximal autocorrelation of lag 6 / weighted by atomic masses
R7m+	R maximal autocorrelation of lag 7 / weighted by atomic masses
R8m+	R maximal autocorrelation of lag 8 / weighted by atomic masses
RTm+	R maximal index / weighted by atomic masses
R1v	R autocorrelation of lag 1 / weighted by atomic van der Waals volumes
R2v	R autocorrelation of lag 2 / weighted by atomic van der Waals volumes
R3v	R autocorrelation of lag 3 / weighted by atomic van der Waals volumes
R4v	R autocorrelation of lag 4 / weighted by atomic van der Waals volumes
R5v	R autocorrelation of lag 5 / weighted by atomic van der Waals volumes
R6v	R autocorrelation of lag 6 / weighted by atomic van der Waals volumes
R7v	R autocorrelation of lag 7 / weighted by atomic van der Waals volumes
R8v	R autocorrelation of lag 8 / weighted by atomic van der Waals volumes
RTv	R total index / weighted by atomic van der Waals volumes
R1v+	R maximal autocorrelation of lag 1 / weighted by atomic van der Waals volumes
R2v+	R maximal autocorrelation of lag 2 / weighted by atomic van der Waals volumes
R3v+	R maximal autocorrelation of lag 3 / weighted by atomic van der Waals volumes
R4v+	R maximal autocorrelation of lag 4 / weighted by atomic van der Waals volumes
R5v+	R maximal autocorrelation of lag 5 / weighted by atomic van der Waals volumes
R6v+	R maximal autocorrelation of lag 6 / weighted by atomic van der Waals volumes
R7v+	R maximal autocorrelation of lag 7 / weighted by atomic van der Waals volumes
R8v+	R maximal autocorrelation of lag 8 / weighted by atomic van der Waals volumes
RTv+	R maximal index / weighted by atomic van der Waals volumes
R1e	R autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities
R2e	R autocorrelation of lag 2 / weighted by atomic Sanderson electronegativities

APPENDIX II: MOLECULAR DESCRIPTOR LIST (continued).

GETAWAY descriptors

Symbol	Meaning
R3e	R autocorrelation of lag 3 / weighted by atomic Sanderson electronegativities
R4e	R autocorrelation of lag 4 / weighted by atomic Sanderson electronegativities
R5e	R autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities
R6e	R autocorrelation of lag 6 / weighted by atomic Sanderson electronegativities
R7e	R autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities
R8e	R autocorrelation of lag 8 / weighted by atomic Sanderson electronegativities
RTe	R total index / weighted by atomic Sanderson electronegativities
R1e+	R maximal autocorrelation of lag 1 / weighted by atomic Sanderson electronegativities
R2e+	R maximal autocorrelation of lag 2 / weighted by atomic Sanderson electronegativities
R3e+	R maximal autocorrelation of lag 3 / weighted by atomic Sanderson electronegativities
R4e+	R maximal autocorrelation of lag 4 / weighted by atomic Sanderson electronegativities
R5e+	R maximal autocorrelation of lag 5 / weighted by atomic Sanderson electronegativities
R6e+	R maximal autocorrelation of lag 6 / weighted by atomic Sanderson electronegativities
R7e+	R maximal autocorrelation of lag 7 / weighted by atomic Sanderson electronegativities
R8e+	R maximal autocorrelation of lag 8 / weighted by atomic Sanderson electronegativities
RTe+	R maximal index / weighted by atomic Sanderson electronegativities
R1p	R autocorrelation of lag 1 / weighted by atomic polarizabilities
R2p	R autocorrelation of lag 2 / weighted by atomic polarizabilities
R3p	R autocorrelation of lag 3 / weighted by atomic polarizabilities
R4p	R autocorrelation of lag 4 / weighted by atomic polarizabilities
R5p	R autocorrelation of lag 5 / weighted by atomic polarizabilities
R6p	R autocorrelation of lag 6 / weighted by atomic polarizabilities
R7p	R autocorrelation of lag 7 / weighted by atomic polarizabilities
R8p	R autocorrelation of lag 8 / weighted by atomic polarizabilities
RTp	R total index / weighted by atomic polarizabilities
R1p+	R maximal autocorrelation of lag 1 / weighted by atomic polarizabilities
R2p+	R maximal autocorrelation of lag 2 / weighted by atomic polarizabilities
R3p+	R maximal autocorrelation of lag 3 / weighted by atomic polarizabilities
R4p+	R maximal autocorrelation of lag 4 / weighted by atomic polarizabilities
R5p+	R maximal autocorrelation of lag 5 / weighted by atomic polarizabilities
R6p+	R maximal autocorrelation of lag 6 / weighted by atomic polarizabilities
R7p+	R maximal autocorrelation of lag 7 / weighted by atomic polarizabilities
R8p+	R maximal autocorrelation of lag 8 / weighted by atomic polarizabilities
RTp+	R maximal index / weighted by atomic polarizabilities

APPENDIX III: TERMINOLOGY AND STATISTICAL BACKGROUND

Boot-strapping

By this validation technique, the original size of the data set (*n*) is preserved for the training set, by the selection of *n* objects with repetition; in this way the training set usually consists of repeated objects and the evaluation set of the objects left out [Efron, B. 1982; 1987]. The model is calculated on the training set and responses are predicted on the evaluation set. All the squared differences between the true response and the predicted response of the objects of the evaluation set are collected in *PRESS*. This procedure of building training sets and evaluation sets is repeated thousands of time, *PRESS* are summed up and the average predictive power is calculated and the average predictive power is calculated (Q^2_{Boot}). Thus, the validation is performed by randomly generating training sets with sample repetitions and then evaluating the predicted responses of the samples not included in the training set.

Chemical Domain of Model applicability

The chemical domain of a model applicability has been recently [Netzeva *et al*., 2005] defined as: "The applicability domain of a (Q)SAR model is the response and chemical structure space in which the model makes predictions with a given reliability. "
Where the chemical structure can be expressed by physicochemical and/or fragmental information, and response can be any physicochemical, biological or environmental effect that is being predicted. The relationship between chemical structure and the response can be developed by a variety of SARs and QSARs. Thus, the chemical domain of applicability is a theoretical region in the space defined by the modeled response and the descriptors of the model, for which a given QSAR should make reliable predictions. This region is defined by the nature of the chemicals in the training set, and can be characterized in various ways: in this work the leverage approach has been used.
Williams plot or Ordinary Least Squares (OLS) Outlier and Leverage Plot is the plot of jackknifed residuals versus leverages (hat diagonals). In this plot the horizontal and vertical straight lines indicate the limits of normal values: the first for the outliers and the second for influential chemicals.
The jackknifed residuals, also called standardized residual in prediction, referred to as Std Error, is calculated as the ordinary residual in prediction divided by the residual standard deviation:

$$\hat{e}_{i/s} = \frac{\hat{e}_{i/i}}{s \cdot \sqrt{1 - h_{ii}}}$$

where $\hat{e}_{i/i}$ is the ordinary residual in prediction of the *i-th* object, *s* is the standard error of the estimate and h_{ii} is the leverage value of the *i-th* object.
It can be noted that, while the outliers for the response can be highlighted only for chemicals with known responses, the possibility of a chemical to be out of the structural applicability domain of a model, and thus the reliability of its predictions, can be verified for every new chemicals, the only knowledge needed being the molecular structure. The Williams plot of the regression allows a graphical detection of both the outliers for the response and the structurally influential chemicals in a model.

External validation

The external validation technique makes use of a test set retained to perform a further check on the predictive capabilities of a model obtained from a training set and with predictive power optimized by an evaluation set. By using the selected model the values of the response for the test objects are calculated and the quality of these predictions is defined in terms of Q^2_{ext}, which is defined as:

$$Q^2_{ext} = 1 - \frac{\sum_{1=1}^{n_{ext}} (y_i - \hat{y}_i)^2}{\sum_{1=1}^{n_{ext}} (y_i - \bar{y})^2}$$

where the sum runs over the test set objects (n_{ext}) and $\bar{y}$ is the average value of the training set responses.

Fitness regression parameters

The performance of the QSAR model can be evaluated by several regression parameters. A first group of them are devoted to evaluate the goodness of fit, i.e. the model capability to fit the data of the training set, providing a measure of how well the regression model accounts for the variance of the response variable.
Some of the ones more used and proposed for comparison or selection of the best subset of models are the following:

- ***Residual Sum of Squares***, *RSS* (*: error sum of squares*). The sum of squared differences between the observed (y) and estimated response ($\hat{y}$):

  $$RSS = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

  being n the number of training objects.
  This quantity is minimized by the least square estimator.

- ***Model Sum of Squares***, *MSS* defined as the sum of the squared differences between the estimated responses and the average response:

  $$MSS = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

  This is a part of the total variance explained by the regression model as opposed to the residual sum of squares *RSS*.

- ***Total Sum of Squares, TSS, defined as the sum of the squared differences between the*** experimental responses and the average response

  $$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

This is the total variance that a regression model has to explain and is used as a no-model reference quantity to calculate standard quality parameters such as the coefficient of determination.

- ***Coefficient of determination,*** R^2. The squared multiple correlation coefficient that is the total variance of the response explained by a regression model. It can be calculated from the model sum of squares MSS or from the residual sum of squares RSS:

$$R^2 = \frac{MSS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where *TSS* is the total sum of squares around the mean. A value of one indicates perfect fit, i.e. a model with zero error term.

- ***Residual Mean Square,*** RMS or s2 (: mean square error, expected squared error). The estimate s2 of the error variance σ 2, defined as:

$$s^2 = \frac{RSS}{df_E}$$

where *RSS* is the residual sum of squares and df_E is the error degrees of freedom, i.e. to $n - p'$, where n is the number of objects (samples), p' the number of model parameters (for example, $n - p - 1$ for a regression model with p variables and the intercept). The standard error of the estimate s is the square root of the residual mean square.

- ***Standard Deviation Error in Calculation,*** SDEC also known as standard error in calculation, SEC. A function of the residual sum of squares, defined as:

$$SDEC = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} = \sqrt{\frac{RSS}{n}}$$

- ***F Fisher function.*** Among the most known statistical tests, it is defined as the ratio between the model sum of squares MSS and the residual sum of squares RSS:

$$F = \frac{MSS / df_M}{RSS / df_E}$$

where df_M and df_E refer to the degrees of freedom of the model and error, respectively. The calculated value is compared with the critical value F_{crit} for the corresponding degrees of freedom. It is a comparison between the model explained variance and the residual variance: high values of the F-ratio test indicate reliable models.

- ***Adjusted R2.*** A fitness parameter adjusted for the degrees of freedom, so that it can be used for comparing models with different numbers of predictor variables:

$$R^2_{adj} = 1 - \frac{RSS / df_E}{TSS / df_T} = 1 - (1 - R^2)\left(\frac{n-1}{n-p'}\right)$$

where *RSS* and *TSS* are the residual sum of squares and the total sum of squares, respectively; df_T refers to the total degrees of freedom; R^2 is the coefficient of determination.

- ***FIT Kubinyi function*** [Kubinyi, H. 1994]:

$$FIT = \frac{R^2}{(1-R^2)} \cdot \frac{(n-p')}{(n+p)^2}$$

where R^2 is the coefficient of determination.

- ***Akaike Information Criterion***, *AIC*. A model selection criterion for choosing between models with different parameters and defined as:

$$AIC = RSS \cdot \frac{(n+p')}{(n-p')^2}$$

Hotelling ellipse

The Hotelling's T^2 statistic is the multivariate equivalent of the Student's *t* statistic, and provides a check for observations adhering to multivariate normality. The Hotelling T2 for observation *i*, based on *p* components is defined as:

$$T_i^2 = \sum_{j=1}^{p} \frac{t_{ij}^2}{s_{tj}^2}$$

s^2_{tj} = variance of t_j

For a given observation, *i*, the Hotelling T2 is a combination of all the X-scores (*t*) in all *p* components. The Hotelling T2 control chart yields a summary of all the process variables and all model dimensions, displaying how far away from the center (target) the process is along the PC model hyper plane.

The significance level to compute the Hotelling T2 ellipse and the critical distance to the model is often by default 0.05 (95% confidence).

Leverage

The leverage of a chemical provides a measure of the distance of the chemical from the centroid of X. Chemicals close to the centroid are less influential in model building than extreme points. The leverages of all chemicals in the data set are generated by manipulating X to give the so-called Influence Matrix or Hat Matrix (**H**), a symmetric matrix defined as:

$$\mathbf{H} = \mathbf{X} \cdot (\mathbf{X}^\mathbf{T}\mathbf{X})^{-1} \cdot \mathbf{X}^\mathbf{T}$$

where **X** is the descriptor matrix, $\mathbf{X}^\mathbf{T}$ is the transpose of **X**, and $(\mathbf{A})^{-1}$ is the inverse of matrix **A**. The leverages or hat values (h_i) of the chemicals (*i*) in the descriptor space are the diagonal elements of **H**, and can be computed as:

$$h_{ii} = x_{i}^{T} \cdot (\mathbf{X^T X})^{-1} \cdot x_i$$

where x_i is the descriptor row-vector of the query chemical.
The leverage matrix is related to the response vector **y** by the following relationship:

$$\hat{\mathbf{y}} = \mathbf{Hy}$$

where $\hat{\mathbf{y}}$ is the calculated response vector from the model.
A "warning leverage" (h^*) is generally fixed at 3p/n, where n is the number of training chemicals, and p the number of model variables plus one. A chemical with high leverage in the training set greatly influences the regression line: the fitted regression line is forced near the observed value and its residual (observed-predicted value) is small, so the chemical does not appear to be an outlier, even though it may actually be outside the AD. In contrast, if a chemical in the test set has a hat value greater than the warning leverage h^*, this means that the prediction is the result of substantial extrapolation and therefore may not be reliable.

Leave-one-out cross-validation

The simplest and most general cross-validation procedure is the leave-one-out technique (LOO technique), where each object is taken away, one at a time. In this case, given n objects, n reduced models have to be calculated.
For each reduced data set, the model is calculated and responses for the deleted object are predicted from the model. The squared differences between the true response and the predicted response for the object left out are added to *PRESS* (*predictive residual sum of squares*). From the final *PRESS*, the Q^2 (or R^2_{CV}) and *SDEP* (*standard deviation error of prediction*) values are usually calculated.
This technique is particularly important as this deletion scheme is unique and the predictive ability of the different models can be compared accurately. However, in several cases, the predictive ability obtained is too optimistic, particularly when the number of objects is quite large. This is due to a too small perturbation of the data when only one object is left out.

Multidimensional scaling (MDS)

Multidimensional scaling (MDS) is a largely used multivariate technique for explorative data analysis, which can be considered to be an alternative to factor analysis, typically used as an exploratory technique to visualize objects in a low dimensional space. In general, the analysis allows detecting meaningful underlying dimensions for similarities or dissimilarities (distances) between the investigated chemicals. In factor analysis, the similarities between objects (e.g., variables) are expressed in the correlation matrix. With MDS it is possible to analyze not only correlation matrices but also any kind of similarity or dissimilarity matrix. The Non-metric multidimensional scaling is works on the distance matrix **D** obtained from the original multidimensional data matrix **X**, using the Euclidean distance; starting from a scaling of the objects in full-dimensional space it attempts to obtain a representation in a Cartesian coordinate system of a set of objects whose relationships are measured by a dissimilarity coefficient, i.e. the selected distance. The principal coordinates are functions of the original variables, mediated through the similarity or distance function used and explaining the largest percentage of the total variance.

Outlier

An object that is atypical (different from the average) of the rest of the objects in a data set is deemed an outlier. A chemical may be an outlier with respect to the response variable (Y) and/or with respect to the independent variables. Thus, to make a decision regarding the inclusion of a particular chemical in an model two aspects have to be accounted for: whether or not that chemical is an outlier and the influence or weight that the chemical has on the results.
Regarding the first aspect, since the assumption of normality of the residuals is a given with any regression equation, the two (or three) times standard deviation rule can be used to identify a potential outlier, simply finding the Standard Deviation Error in Calculation (SDEC) and multiplying it by 2 (or 3) in order to get the bounds within which all of residuals should lie. Therefore, if a particular residual lies outside of these bounds, it is deemed to be an outlier. If it is close to three standard deviations, these chemicals should be examined further. Once an observation has been established as an outlier, another decision must be made, that is whether or not it can be retained in the equation. If the outlier is due to miscoding, the user simply make the correction and proceeds from there. However, if the observation is atypical, the concept of leverage and/or influence enters the analysis.

Predictive regression parameters

This group of regression parameters are devoted to evaluate the goodness of prediction, i.e. the model capability to estimate future (test) data, providing a measure of how well the regression model estimates the response variable given a set of values for predictor variables. These quantities are obtained using validation techniques and are also used as criteria for model selection.
The most important regression parameters are listed below:

- ***Predictive Residual Sum of Squares, PRESS.*** The sum of squared differences between the observed and estimated response by validation techniques:

$$PRESS = \sum_{i=1}^{n} (y_i - \hat{y}_{i/i})^2$$

  where $\hat{y}_{i/i}$ denotes the response of the *i-th* object estimated by using a model obtained without using the *i-th* object. Using validation techniques minimizes this quantity.

- ***Cross-validated R^2***, R^2_{cv} (or Q^2). The explained variance in prediction:

$$R^2_{cv} = Q^2 = 1 - \frac{PRESS}{TSS} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_{i/i})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

  where *PRESS* is the predictive error sum of squares and *TSS* the total sum of squares.

- ***External Q^2***. The explained variance in prediction:

$$Q^2_{ext} = 1 - \frac{\sum_{1=1}^{n_{ext}} (y_i - \hat{y}_i)^2}{\sum_{1=1}^{n_{ext}} (y_i - \bar{y})^2}$$

where the sum runs over the test set objects (n_{ext}) and $\bar{y}$ is the average value of the training set responses.

- ***Standard Deviation Error of Prediction***, *SDEP* also known as *standard error in prediction* SEP or PSE. A function of the predictive residual sum of squares, defined as:

$$SDEP = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_{i/i})^2}{n}} = \sqrt{\frac{PRESS}{n}}$$

- ***External Standard Deviation Error of Prediction***, $SDEP_{ext}$. A function of the predictive residual sum of squares, defined as:

$$SDEP_{ext} = 1 - \frac{\sum_{1=1}^{n_{ext}}(y_i - \hat{y}_i)^2}{n_{ext}}$$

where the sum runs over the test set objects (n_{ext}).

Principal component analysis (PCA)

Principal component analysis is a statistical technique for exploratory data analysis, modelling the p variables in the data matrix $\mathbf{X}$ (n x p), where n is the number of objects, as linear combinations of the common factors $\mathbf{T}$ (n x M), called principal components $\mathbf{t_m}$:

$\mathbf{X} = \mathbf{T} \cdot \mathbf{L}^{\mathbf{T}}$

where $\mathbf{T}$ is the score matrix, $\mathbf{L}$ (p x M) is the loading matrix and M is the number of significant principal components ($M \leq p$). The columns of the loading matrix are the eigenvectors $\mathbf{l}_m$; the eigenvector coefficients ℓ_{jm}, called *loadings*, represent the importance of each original variable in the considered eigenvector. The components are calculated according to the maximum variance criterion, i.e. each successive component is an orthogonal linear combination of the original variables such that it covers the maximum of the variance not accounted for by the previous components. The eigenvalue λ_m associated with each *m-th* component represents the variance explained by the considered component.

The principal components can also be viewed as linear combinations of the p original variables.

The main advantages of principal components are that:

1) each component is orthogonal to all the remaining components, i.e. the information carried by this component is unique;
2) each component represents a *macrovariable* of the data;
3) components associated with the lowest eigenvalues do not usually contain useful information (noise, spurious information, etc.).

When PCA is performed on a set of compounds characterized by molecular descriptors (physico-chemical properties, structural variables, etc.) the significant principal components are called principal properties PP because they summarize the main information of the original molecular descriptors:

Principal component analysis is often used to identify groups of inter-related variables, reduce the number of variables, as well as discover extreme cases on one variable, or a combination of variables, which have a strong influence on the calculation of statistics (outlier detection).

Validation techniques

Validation techniques constitute a fundamental tool for the assessment of the validity of models obtained by multivariate regression methods. Validation techniques are used to check the prediction power of the models, i.e. to give a measure of their capability to perform reliable predictions of the modelled response for new cases where the response is unknown.
A necessary condition for the validity of a regression model is that the multiple correlation coefficient R^2 is as close as possible to one and the standard error of the estimate s small. However, this condition (fitting ability) is not sufficient for model validity as the models give a closer fit (smaller s and larger R^2) the larger the number of parameters and variables in the models. Moreover, unfortunately, these parameters are not related to the capability of the model to make reliable predictions on future data.
Other problems for the validity of the models arise when models, often with only few variables, are obtained by using procedures based on variable selection [Allen, D.M. 1971]. In fact, when a set with a large number of descriptors to select from is available, simple models can be found with apparently good fitting properties due to chance correlation, i.e. collinearity without predictive ability [Topliss, J.G. and Edwards, R.P. 1979; Wold S, *et al.* 1983; Clark M and Cramer IRD 1993].
To avoid models with chance correlation, a check with different validation procedures must be adopted.
The more common statistical techniques proposed to simulate the predictive ability of a model are the following:

- leave-one-out
- bootstrap
- Y-scrambling
- external validation

Y-Scrambling

This validation technique is adopted to check models with chance correlation, i.e. models where the independent variables are randomly correlated to the response variables. The test is performed by calculating the quality of the model (usually R2 or, better, Q2) randomly modifying the sequence of the response vector y, i.e. by assigning to each object a response randomly selected from the true responses. Each scrambling is characterised in terms of the correlation of the scrambled response with the unperturbed data ($R^2_{yy'}$). If the original model has no chance correlation, there is a significant difference in the quality of the original model and that associated with a model obtained with random responses. The procedure is repeated several hundred of times.
Once the model validation has been performed the Y-scrambling parameters (a(R2) and a(Q2)) are calculated as the intercepts of the equations:

$$R^2 = b_o + bR^2(yy') \qquad b_0 = a(R^2)$$
$$Q^2 = b_o + bQ^2(yy') \qquad b_0 = a(Q^2)$$

Models which are unstable (that is, which change greatly with small changes in underlying response values) are characterized by high intercept value. Stable models (that is, which change proportionally with small changes in underlying data) have low intercept value.

References

Allen, D.M. (1971). Mean square error of prediction as a criterion for selecting variables. *Technometrics* 13, 469-475.

Clark M and Cramer IRD (1993). The probability of chance correlation using partial least squares (PLS). *Quant Struct-Act Relat* 12**:** 137-145.

Efron, B. (1982). The Jackknife: the Bootstrap and Other Resampling Planes, Society for Industrial and Applied Mathematics, Philadelphia (PA).

Efron, B. (1987).Better bootstrap confidence intervals. (with discussion). *Journal of American Statistical Association*, 82, 171-200.

Kubinyi, H. (1994). Variable selection in QSAR studies. I. An evolutionary algorithm. *Quant Struct-Act Relat*. 13, 285-294.

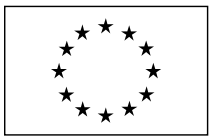
Netzeva,T., Worth, A., Aldenberg,T., Benigni, R., Cronin, M., Gramatica,P., Jaworska, J., Kahn, S., Klopman,G., Marchant, C., Myatt, G., Nikolova-Jeliazkova,N., Patlewicz, G., Perkins, R., Roberts, D., Schultz,T.W., Stanton, D., van de Sandt, J., Tong, W., Veith, G., Yang, C. (2005). Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure–Activity Relationships. ECVAM Workshop 52, *ATLA* 33, 1–19.

Topliss, J. G. and Edwards, R. P. (1979). Chance Factors in Studies of Quantitative Structure-Activity Relationships, *J. Med. Chem.* 22, 1238-1244.

Wold S, Albano C, Dunn WJ III, Esbensen K, Hellberg S, Johansson E. (1983). Pattern recognition: finding and using regularities in multivariate data. In: *Food Research and Data Analysis* (Martens H, Russwurm H, eds). Essex, UK: Applied Science Publishers, 147-188.

Mission of the JRC

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.



EUROPEAN COMMISSION
DIRECTORATE GENERAL
JOINT RESEARCH CENTRE