HAL
open science

Une meilleure caractérisation des mécanismes d'action toxique à partir de la structure moléculaire

Franklin Bauer

To cite this version:

Franklin Bauer. Une meilleure caractérisation des mécanismes d'action toxique à partir de la structure moléculaire. Autre. Université de Haute Alsace - Mulhouse, 2017. Français. NNT : 2017MULH1360 . tel-01823559

HAL Id: tel-01823559

https://theses.hal.science/tel-01823559v1

Submitted on 26 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE DE HAUTE-ALSACE
UNIVERSITE DE STRASBOURG

Thèse

Présentée pour l'obtention du grade de

Docteur de l'Université de Haute-Alsace

Ecole Doctorale : ED 222 - Sciences Chimiques

Discipline : Chimie, Physique et Matériaux

Présentée et soutenue publiquement

par

Franklin BAUER

Le 30 novembre 2017

Une meilleure caractérisation des mécanismes d'action toxique à partir de la structure moléculaire.

Thèse CIFRE

Sous la Direction de : Directeur de thèse : Prof. Serge NEUNLIST
Co-directeur de thèse : Dr. Samuel FOUCHARD
Encadrée en entreprise par : Dr. Paul THOMAS

Jury : Dr. Karine AUDOUZE, université Paris Diderot (Rapporteur)

Dr. Patrice GONZALEZ, université de Bordeaux (Rapporteur)

Prof. Jean-Philippe GODDARD, université de Haute-Alsace (Président)

Prof. Eric PINELLI, université de Toulouse III

Prof. Serge NEUNLIST, université de Haute-Alsace

Dr. Samuel FOUCHARD, université de Haute-Alsace

Dr. Paul THOMAS, KREATiS

Table des matières

Table des figures 6
Table des tableaux 7
Liste des abréviations 8
Remerciements 10
Résumé de la thèse en anglais – Abstract 11
1) Context and State of the Art 11
2) Our contribution 12
3) Future prospects 14
Chapitre 1 : Introduction 16
1) QSARs et approches par analogie 17
a) Principe de ces méthodes 17
b) Contexte et raisons de leur utilisation 18
2) MechoAs vs MoAs 19
3) Différentes classifications et modèles de prédiction de MechoAs/MoAs 20
a) Classifications et modèles sur les composés organiques en général 20
b) Classifications et modèles sur des familles restreintes 21
4) Objectifs de la thèse 22
Chapitre 2 : Revue bibliographique des mécanismes d'action 24
1) Narcose 25
a) Narcose non-polaire 25
b) Narcose polaire 28
2) Narcose des esters ou pronarcose 29
3) Toxicité par réactivité avec formation d'adduits 30
a) Electrophiles durs 31
b) Electrophiles mous 34
c) Générateurs de radicaux 35
4) Toxicité pro-active 35
a) Métabolisation réduisant la toxicité 36
b) Métabolisation vers un réactif électrophile 39
c) Utilisation d'un cycle RedOx 46
d) Métabolisation en perturbateur indirect d'enzyme 47
e) Métabolisation en perturbateurs directs par docking 48

5) Perturbation de l'environnement enzymatique 49
a) Découplage de la phosphorylation oxydative par transport de protons 49
b) Acidification ou basification des cellules 51
6) Interaction directe dans un site de liaison 51
a) Inhibition de l'AcétylCholinEstérase 51
b) Ligands aux récepteurs de l'AcétylCholine 55
c) Perturbation du transport de la dopamine 56
d) Chélation des métaux 56
e) Inhibition du transport d'électrons du photosystème II 57
f) Modulation des canaux à ions 58
g) Autres mécanismes 61
Chapitre 3 : Organisation des mécanismes d'action et discussion de certains mécanismes 66
1) Une nouvelle classification en mécanismes d'action toxique 67
2) MechoA des amines aliphatiques et ammoniums 69
3) Discussion sur les modèles QSAR MechoA-spécifiques 69
4) Discussion sur l'excès de toxicité observé pour les esters 71
5) Réactivité des γ-dicétones en tant qu'électrophiles durs 74
6) Modélisation QSAR pour les découpleurs de la phosphorylation oxydative 74
7) Mécanisme responsable de la toxicité et corrosivité des acides et bases 75
Chapitre 4 : Développement d'un modèle de prédiction des MechoAs par approche quantique 78
1) Introduction 79
2) Publication 80
A New classification algorithm based on Mechanisms of Action. 80
Abstract 80
1. Introduction 80
1.1. MechoA definitions 82
2. Materials and Methods 84
2.1. Literature data 84
2.2. Computational methods 85
2.3. Classification algorithm 87
3. Results and discussion 90
4. Conclusions 92

Acknowledgments 93
References 93
3) Conclusion 95
Chapitre 5 : Développement d'un modèle de prédiction des MechoAs par alertes structurales. 96
1) Introduction 97
2) Matériels et Méthodes 98
3) Résultats et discussion 101
a) Arbre de décision 101
b) Validation statistique du modèle 105
4) Conclusion 107
Chapitre 6 : Expériences en laboratoire de validation et affinement des modèles. 110
1) Introduction 111
2) Mesures de coefficients de partage octanol-eau 111
a) Introduction 111
b) Protocole utilisé pour les tests 112
c) Substances testées 114
d) Résultats et discussion 116
3) Mesures de solubilités aqueuses 119
a) Introduction 119
b) Protocole utilisé pour les tests 119
c) Substances testées 121
d) Résultats et discussion 121
4) Tests de toxicité aigüe pour les daphnies 122
a) Introduction 122
b) Matériel et méthodes 123
c) Substances testées 123
d) Résultats et discussion 125
5) Conclusion 127
Chapitre 7 : Discussion générale et conclusion. 128
Planche de structures moléculaires 133
Références bibliographiques 135
Annexes 144
A.1) Informations supplémentaires pour la publication « A New classification algorithm based on Mechanisms of Action. » 145

A.2) Informations supplémentaires pour la méthode de prédiction des MechoAs par alertes structurales 179

A.3) lignes directrices OCDE pour les tests en laboratoire..... 210

Table des figures

Figure 1 : superposition de modèles QSARs pour la toxicité environnementale. 26
Figure 2 : bactéries P. fluorescens vues au microscope électronique à transmission (colonne de gauche) et à balayage (colonne de droite) suite à l'exposition (a-b) contrôle sans carvacrol ni 1,8-cinéole, (c-d) avec carvacrol, (e-f) avec 1,8-cinéole. 27
Figure 3 : concentrations léthales médianes (CL50) pour le guppy et le tête de boule pour les narcotiques non polaires (carrés pleins) et polaires (carrés vides) en fonction de a) leur logP et b) leur $logK_{DMPC}$. 29
Figure 4 : hydrolyse enzymatique des esters. 29
Figure 5 : formation d'adduits ou libération d'acide par les électrophiles durs. 31
Figure 6 : voie époxyde-diol pour éliminer une substance de l'organisme. 32
Figure 7 : mécanisme de réaction des électrophiles possédant un bon groupe partant, un état de transition et un groupe partant stabilisés. 33
Figure 8 : formation d'adduits avec les protéines par les électrophiles mous 35
Figure 9 : Voie de métabolisation chez le rat a) du méthacrylate de méthyle, b) de l'acrylate de méthyle 38
Figure 10 : métabolisme du benzo[a]pyrene (B[a]P) qui conduit à la formation d'adduits à l'ADN. 39
Figure 11 : formation de cancer par les anilines et nitrobenzènes. 40
Figure 12 : réticulation des protéines par l'hexan-2,5-dione. 41
Figure 13 : métabolisme des thiophènes et furanes aboutissant à des aldéhydes réactifs. a) métabolisme et réaction des thiophènes, b) des furanes. c) mécanismes de réaction des métabolites du thiophène avec un nucléophile. 42
Figure 14 : métabolisme des alkoxyvinyles en époxydes 43
Figure 15 : métabolisme générant des réactifs du a) tétrachlorure de carbone b) chloroforme c) trichloroéthylène. 44
Figure 16 : mécanisme des alcynes dégradant les CYP450. 45
Figure 17 : métabolisme du 3-méthoxyphénol vers une quinone. 46
Figure 18 : cycle RedOx des quinones. 46
Figure 19 : cycle RedOx du paraquat. 47
Figure 20 : Couplage de l'ATP synthase et de la chaîne de transport d'électrons par le gradient de protons. 49
Figure 21 : transport de protons à travers la membrane mitochondriale interne par les phénols acides. 51
Figure 22 : transmission nerveuse. 52
Figure 23 : mécanisme de l'hydrolyse de l'acétylcholine par l'acétylcholinestérase. 53
Figure 24 : inhibition de l'acétylcholinestérase. 54
Figure 25 : L'acétylcholine et quelques-uns de ses agonistes et antagonistes. 56
Figure 26 : photosystème II. 58
Figure 27 : herbicides inhibant la protéine D1 du photosystème II. 58

Figure 28 : Structure des dérivés de DDT et de pyréthroïdes. 59
Figure 29 : Récepteur à GABA et ses sites de liaison. 60
Figure 30 : insecticides alicycliques chlorés et picrotoxinine. 60
Figure 31 : inhibition de la polymérisation de la tubuline par la trifluraline. 62
Figure 32 : Rôle central de l'acétyl-coenzyme A dans les plantes. 63
Figure 33 : Structure de l'estradiol. 64
Figure 34 : remplissage de la poche de liaison du récepteur aux estrogènes par différentes molécules actives. 65
Figure 35 : comparaison de la narcose non-polaire et polaire pour le poisson en fonction de la solubilité. 70
Figure 36 : toxicité des diesters par rapport aux monoesters pour a) le poisson, b) la daphnie, c) l'algue. 73
Figure 37 : Réactions favorisées par la stabilité de l'état de transition ou par la stabilité du produit final. 74
Figure 38 : toxicité des phénols acides pour le poisson en fonction de la solubilité et du pKa. 75
Figure 39 : acidification des cellules par les acides. 76
Figure 40 : ligands aux récepteurs de l'acétylcholine. A = acétylcholine. B = nicotine. C = muscarine. D = coniine. E = atropine. 99
Figure 41 : Un ester α,β-insaturé dont les électrons ne sont pas délocalisés à cause de l'encombrement stérique. 100
Figure 42 : planche de structures moléculaires 134

Table des tableaux

Tableau 1 : exemple d'approche par analogie 18
Tableau 2 : hydrophobie d'un ester et de ses produits d'hydrolyse. 30
Tableau 3 : groupes partants. 34
Tableau 4 : MechoAs généraux utilisés dans ce travail. 67
Tableau 5 : MechoAs détaillés. (Sous-catégories des MechoAs généraux) 68
Tableau 6 : Arbre de décision pour la prédiction des MechoAs. 104
Tableau 7 : Statistiques de notre arbre de décision. 105
Tableau 8 : comparaison des méthodes de classification. 106
Tableau 9 : cétones α,β-insaturées pour les tests de logP. 115
Tableau 10 : substances variées pour les tests de logP. 116
Tableau 11 : résultats des mesures de logP. 118
Tableau 12 : données de la substance de référence. 117
Tableau 13 : substances pour les tests de solubilité. 121
Tableau 14 : résultats de solubilité. 122
Tableau 15 : substances pour les tests de toxicité aigüe pour les daphnies. 125
Tableau 16 : résultats des mesures de toxicité aïgue pour les daphnies. 126

Liste des abréviations

2,4-D : acide 2,4-dichlorophénoxyacétique.

ACh : AcétylCholine

AChE : AcétylCholinEstérase

AChR (nAChR et mAChR) : Récepteur à l'AcétylCholine (nicotinique et muscarinique)

ADN : Acide DésoxyriboNucléique

AhR : Aryl hydrocarbon Receptor = récepteur des hydrocarbures aromatiques.

AOP : Adverse Outcome Pathway = Voie à conséquences nocives

ATP, ADP : ATP = Adénosine TriPhosphate, ADP = Adénosine DiPhosphate.

CE50 / CL50 : Concentration Effective / Léthale pour 50% des organismes testés

COB : Laboratoire de Chimie Organique et Bioorganique, EA 4566 Université de Haute Alsace

COX : CycloOXygénase

CYP450 : Cytochrome P450

DDT : DichloroDiphénylTrichloroéthane

DMPC : DiMyristoylPhosphatidylCholine

OD : Oxygène Dissous

ECHA : European Chemical Agency

GABA : Acide Gamma-AminoButyrique

GC : chromatographie en phase gazeuse.

GLP : Good Laboratory Practices

GPCR : G Protein-Coupled Receptor = récepteur couplé à une protéine G.

GSH, GSSG, GST, GPx : GSH = glutathion réduit sous sa forme thiol, GSSG = glutathion oxydé sous sa forme de disulfure, GST = Glutathion S-Transférase, GPx = Glutathion Peroxydase

HAP : Hydrocarbure Aromatique Polycyclique

HOMO : Highest Occupied Molecular Orbital = orbitale moléculaire la plus haute occupée

HPLC : Chromatographie Liquide à Haute Performance

KREATiS : Knowledge & Research in Environment And Toxicity *in Silico*, entreprise basée à L'Isle d'Abeau, France, N° SIRET 798 935 656 00024.

logP : logarithme du coefficient de partage octanol-eau

LUMO : Lowest Unoccupied Molecular Orbital = orbitale moléculaire la plus basse vacante

MechoA : Mechanism of Action (mécanisme d'action)

MoA : Mode of Action (mode d'action)

NAD(P)H : Nicotinamide adénine dinucléotide (phosphate). NAD(P)H est la forme réduite, et $NAD(P)^+$ est la forme oxydée.

OCDE : Organisation de Coopération et de Développement Economiques

PCB : PolyChloroBiphényle.

(Q)SAR : (Quantitative) Structure-Activity Relationship = Relation structure-activité (quantitative)

REACH : Registration, Evaluation, Authorisation and restriction of Chemicals. Réglementation européenne pour l'enregistrement, l'évaluation, l'autorisation et la restriction des produits chimiques.

ROS : Reactive Oxygen Species = Espèces réactives de l'oxygène

TPA : Tonnes Par Année

UHA : Université de Haute-Alsace

Remerciements

Avant toute chose, je remercie Dieu qui me dirige dans tous les domaines de ma vie.

Je remercie premièrement le Dr. Paul THOMAS pour m'avoir encadré dans l'entreprise, et avoir permis la mise en place de ce contrat CIFRE au sein de KREATiS. Je remercie également le Professeur Serge NEUNLIST et le Dr. Samuel FOUCHARD pour avoir ouvert les portes de la collaboration avec l'université de Haute-Alsace et pour m'avoir encadré au laboratoire.

Je remercie tout particulièrement ma fiancée Hoong Nee TAN ainsi que ma famille et mes amis pour leur soutien constant dans les moments faciles et les moments difficiles qui se sont succédés durant ces trois années.

Je remercie chaleureusement Nicolas DELPIT du laboratoire LPL pour m'avoir envoyé gratuitement des algues et daphnies de son laboratoire afin de démarrer la culture et l'élevage dans notre laboratoire. Tous mes remerciements aussi à Pascal BICHEREL, Dr. Faizan SAHIGARA, Dr. Nathalie MAYER, Dr. Carole CHARMEAU pour leur soutien scientifique et amical. Un grand merci également à Sarah TERRADOT pour s'occuper des tâches administratives et également pour son amitié. Je suis aussi très reconnaissant envers le Dr. Cécile JOYEUX pour sa patience et son dévouement, pour m'avoir préparé les machines d'analyse au laboratoire, pour m'avoir expliqué leur fonctionnement, pour les avoir maintes fois dépannées et réparées, pour m'avoir conseillé sur les méthodes d'analyse. Je remercie aussi le Professeur Céline TARNUS pour l'accueil dans son laboratoire en tant que directrice, et aussi les doctorants du laboratoire pour m'avoir intégré, montré le fonctionnement de choses diverses. Enfin, je remercie les stagiaires que j'ai encadrés au laboratoire, et qui ont effectué avec moi les tests de logP et de toxicité sur les daphnies : Igor AGAPI, Xu WANG, Geoffrey BERREZ, Maxime PYPEC.

Résumé de la thèse en anglais – Abstract

1) Context and State of the Art

To protect human health and environment, it is important to assess and control the toxicity of chemical substances to which we are exposed. The usual way to do toxicological risk assessment is to perform toxicity studies on animals. Particularly up until REACH (Parlement Européen and Conseil de l'Union Européenne, 2006), the European regulation for chemical substances, either imported of manufactured in Europe, animal testing has been difficult to avoid. However, this new regulation provides opportunities to avoid animal testing by use of alternative methods. ECHA (European Chemicals Agency) is asking the chemical industry to give health and safety information on their products with a tonnage ≥ 1 ton per year. Depending on tonnage (tons of manufactured/imported substance per year) an increasing number of physicochemical and (eco)toxicological endpoints are required. Despite the availability of currently accepted alternatives this set of toxicity tests is expensive and is still posing ethical problems (animal suffering, …). As a consequence, industry is encouraged to use alternative methods, such as *in vitro* approaches, Read-Across, Weight-of-Evidence, and QSARs (Structure-Activity Relationships), which enable the use of existing data on similar substances, to generate a prediction for the target substance.

QSARs are in silico models establishing a relationship between substance properties and a given activity, i.e. a biological effect. The activity, also known as response or output variable, is often a dose of substance leading to a toxicity for organisms, such as an EC50 (the concentration causing the effect for 50% of organisms used in the test). Substance properties which are correlated with activity are named descriptors. The toxicity of most substances being strongly led by their hydrophobicity, the logP (octanol-water partition coefficient) is generally chosen as a descriptor. Conversely, a Read-Across does not aim to establish a mathematical model between input and output variables, but compares two or more substances and evaluates the similarity between their structures, properties and their known activities. If the substances are judged similar enough in all these aspects, the known activity value of one analogue (or the mean of the activities of several analogues) can then be used for the target substance for which activity is unknown.

Mechanisms of Action (MechoAs) describe how substances are toxic for living organisms, and particularly what molecular event will be responsible for their toxicity. When *in silico* methods are used, it is important to ensure that the target substance is behaving with the same mechanism of action as the other substances used to generate the toxicity prediction. Indeed, it has been demonstrated that toxicity to fish requires different regression coefficients, or even different descriptors to be predicted by QSAR modelling when different MechoAs are playing. Hence, if the MechoA of a given substance is not well known, an inappropriate QSAR may be used, resulting in a wrong toxicity prediction. This has led us to create a classification method, sorting substances by categories which may or may not be classed by structure, allowing the choice of the right QSAR, i.e. a classification with well-defined MechoAs. The knowledge of MechoAs predictions based only on the molecular structure can also be useful in medicinal chemistry, in order to develop new hit substances for therapeutic use.

While MechoAs look at the molecular interaction leading to the toxic response, Modes of Action (MoAs) refer to the pathological effects that can be seen at the whole organism level in terms of behaviour or death i.e. at the other end of the Adverse Outcome Pathway (AOP). MechoAs and MoAs

are linked by AOPs. But one MoA can be the result of very different MechoAs. Conversely, the same MechoA but different MoAs can happen mainly when comparing the AOPs across different species.

Several MoA classifications can already be found in the literature, together with predictive models. Relatively few models are built in order to be applicable to the whole organic chemicals family. A review of existing models covering a wide range of chemical groups is presented below:

The most widely known method used in ecotoxicology and to some extent in mammalian toxicology today is the Verhaar scheme, published in 1992 (Verhaar *et al.*, 1992) and last updated in 2008 (Enoch *et al.*, 2008). Today, this method can be used via several interfaces (ToxTree, OECD QSAR Toolbox). This scheme classifies compounds into 4 modes of action: non-polar narcosis or baseline toxicity, polar narcosis or less inert compounds, unspecific reactivity, and specifically-acting compounds. The Verhaar scheme, despite updates, still includes notable errors and a quite restricted applicability domain, with many substances being assigned to the so-called class 5, which corresponds to the unclassified compounds. This method is freely available, and built as a decision tree with simple structural alert rules (topological alerts), which makes it attractive to a broad section of the toxicological community.

Another classification method applied to most organic compounds is that of Russom *et al.* (Russom *et al.*, 1997), which predicts modes of action for fish, solely based on fish toxicity studies. This method classifies compounds into 8 categories: Baseline narcosis or narcosis I, polar narcosis or narcosis II, ester narcosis or narcosis III, oxidative phosphorylation uncoupling, respiratory inhibition, electrophile/proélectrophile reactivity, acetylcholinesterase (AChE) inhibition and central nervous system seizure.

More recently, a more advanced classification into MechoAs was published by Barron *et al.* (Barron *et al.*, 2015), regrouping a wide variety of substances: metals, organometals, pesticides and other organic molecules. Compounds are classified into 6 general MechoAs and 31 specific MechoAs thanks to toxicity studies on fish and aquatic invertebrates, QSARs, and other literature data. The 6 general MechoAs are as follows: narcosis, AChE inhibition, electron transport inhibition, iono/osmoregulatory/circulatory impairment, reactivity, neurotoxicity. A prediction method for MechoAs of the same research group was published 2 years before this classification, using only a part of the above-mentioned classes (Martin *et al.*, 2013). This method is built with all kinds of descriptors (experimental, topological, electronic) and predicts MechoAs for fish toxicity, using statistical methods difficult to follow and reproduce.

The goal of this thesis was to better characterise modes/mechanisms of toxic action than the Verhaar scheme, the most widely used method by ecotoxicologists, and to define a single classification system that can be used by toxicologists and ecotoxicologists alike. We were also aiming to get more accurate classifications than other general prediction methods and to cover the widest range possible of organic substances.

2) Our contribution

The first step for this work was to form a database with information on mechanisms of action. We gathered a dataset of 794 molecules with known biological interactions and/or effects, and classified them into MechoAs organised in 6 general categories and 23 sub-categories, as we wanted to keep a limited number of general categories, but to increase sub-categories number as much as needed. This classification scheme will of course be continued after this thesis to characterise more and more molecules and mechanisms.

The MechoAs defined in this work are organised as follows. General MechoAs are given in the list below, with detailed MechoAs included in brackets:

- Membrane destabilisation (non-polar narcosis, polar narcosis, cationic narcosis)
- Hydrolysis pro-destabilisation (hydrolysis to narcotics, hydrolysis to other MechoAs)
- Reactivity (hard electrophile reactivity, soft electrophile reactivity, homolytic cleavage of a bond generating radicals)
- Pro-activity (metabolisation to less toxic compounds, metabolisation to electrophiles, RedOx cycling, metabolisation to indirect enzyme disruptors, metabolisation to direct docking disruptors)
- Indirect enzyme disruption (oxidative phosphorylation uncoupling, acidification or basification of cells, other indirect mechanisms)
- Direct docking site interaction (AChE inhibition, AChR binding, dopamine transport disruption in synapses, metal chelation, photosystem II electron transport inhibition, ion channel activation/inactivation, other direct mechanisms).

The intention was to completely map the possible Molecular Initiating Events (MIEs) within these six general titles so no chemical substances, regardless of their impact on flora or fauna, would fall under a category "unknown" regardless of its chemical interaction. After gathering a maximum of mechanisms of action for a wide range of chemical families, we started to develop a preliminary prediction method. We chose to use a method based on quantic descriptors, obtained after 3D geometry optimisation, in order to extend the applicability domain of the method further than with structural alerts on two-dimensional structures. Some molecules contain the same moiety which can be a structural alert deemed responsible for a given MechoA. However, these molecules can differ in the molecular environment of this moiety, so that the electronic properties of thes substances can very dramatically, thus leading to a different MechoA in fact. Building a decision tree based on simple structural alerts may thus lead to errors, and a method based on quantic descriptors should not get these errors. In fact, the initial idea was that the error rate (14% as a mean) and the non-classification rate (35% as a mean) obtained using the Verhaar scheme was mostly due to the use of identified structural alerts to classify compounds, implying errors or non-classifications for all the structural alerts that were not identified. This limitation could, in theory, be circumvented thanks to electronic descriptors generated by quantum calculations. This quantic method was developed and has resulted in a publication submitted in the journal *Computational Toxicology*. We obtained with this method 86% of correct classifications for the training set, but only 69% for the validation set. We obtained 13% of classifications slightly different from literature classification (see publication for the definition of slightly different), 17% of misclassifications and 1% outside the applicability domain for the validation set. These results for the validation set have been compared to those of the Verhaar scheme, which obtained 45% of correct classifications, 18% of misclassifications and 37% outside of the applicability domain. So, this method proved to perform better than the Verhaar scheme. However, its success rate still remains inadequate to provide sufficient confidence in predictions, even if calculation means used to obtain these results are relatively high compared to the Verhaar scheme.

Following this first prediction method we tried to develop a second method using structural alerts, i.e. the identification of molecule moieties responsible for a toxic activity, in the hopes that this would improve MechoA predictive hit-rate. Further goals were to obtain faster results, able to be implemented in a software tool that could be run automatically and therefore easily exploitable by anyone. Structure-activity analyses coupled with an expert judgment finally allowed us to develop a decision tree composed of 62 decision rules. This classifies molecules with high accuracy into the 23 detailed MechoAs, with a much reduced calculation time, because geometry optimisation is needed only for a few questions in the scheme, and a basic calculation level appears appropriate enough for this purpose. This geometric optimisation is only used to measure interatomic distances. Comparing the two methods this new method was found not only to be simpler and faster, as expected, but also

obtained much better correlations than the previous one, with 92% of correct classifications, 4% of slightly different classifications, 4% of misclassifications and 0% outside the applicability domain for the validation set. Therefore, we wrote a manuscript to report this method which will soon be submitted to the journal *Computational Toxicology*. Our results were subsequently compared with the Russom method results using MOA classification by the OASIS profiler which is implemented in OECD QSAR Toolbox. In the Russom prediction method, for compounds outside the applicability domain, a baseline mode of action is assigned by default. Consequently, with our validation set, the Russom method obtained 83% correct classifications, 6% of slightly different classifications and 11% of misclassifications. If we separate out the baseline classifications that were assigned because the compound was outside applicability domain, it becomes 70% correct classifications, 6% slightly different, 10% misclassifications and 14% outside the applicability domain. Overall, these are quite good results, better than our quantic method, but not as good as our second structural alerts method.

In addition to the MechoA work outlined above, a shorter experimental part was performed in the scope of this thesis, where we measured partition coefficients (logP), water solubilities and acute toxicity to daphnia for some chosen substances. These experiments enabled a better characterisation of some families of substances for which little or nothing is known in the literature, or with inconsistent data between physico-chemical parameters and aquatic toxicity. For certain substances, we managed to elucidate the experimental issues which prevented accurate experimental values being obtained. For other substances, we were able to obtain new reliable experimental values, which were used to help develop KREATiS models or to validate existing models.

3) Future prospects

The final prediction method is judged satisfying and meets our requirements in terms of completeness accuracy and precision. Furthermore, its simplicity of use enables it to be integrated at least partially in a fully automated software interface. This task is on-going.

Further comparisons with other MechoA prediction methods may be undertaken in the future. This kind of comparative work for other MoA prediction methods has already been done by other groups and published recently (Kienzler *et al.*, 2017). We also think about extending this work to inorganics, metals and organometals. A lot of these compounds indeed exert specific MechoAs. Moreover, further work will be carried out on endocrine disruptors in order to characterise the different active structures and corresponding precise mechanisms.

Chapitre 1 : Introduction

1) QSARs et approches par analogie

a) Principe de ces méthodes

Les QSARs (Relations Quantitatives Structure-Activité) sont des modèles informatiques qui lient des propriétés des substances à une activité, ou un effet au niveau biologique (Madden, 2010). Les propriétés qui sont utilisées pour modéliser l'activité sont appelées descripteurs (ou variable indépendante, ou variable d'entrée). Ces descripteurs peuvent être liés directement à la structure moléculaire (par exemple, le nombre de groupements donneurs de liaison hydrogène), ou bien quantifier les propriétés électroniques des molécules, grâce à des calculs de chimie quantique (par exemple, E(HOMO), l'énergie de l'orbitale moléculaire la plus haute occupée), ou encore des propriétés physicochimiques telles que la solubilité dans l'eau ou le logP (coefficient de partage octanol-eau), qui peuvent être mesurés expérimentalement ou être eux-mêmes prédits par un autre QSAR (Madden, 2010). L'activité, appelée variable dépendante ou variable de réponse ou de sortie, est souvent une concentration qui entraîne une toxicité sur les organismes, comme une CE50, la concentration qui cause l'effet mesuré pour 50% des organismes utilisés dans le test. L'effet mesuré peut être la mortalité, l'inhibition de la croissance, l'immobilisation, l'inhibition d'une activité enzymatique, etc. (Madden, 2010)

Pour construire un modèle QSAR, on s'appuie sur un set de données, renseignant les valeurs des descripteurs et de l'effet qu'on cherche à modéliser pour une série de substances. Ce set de données est classiquement séparé en un set d'entraînement, c'est-à-dire les données qui vont servir à construire le modèle, et un set de validation (ou set de test), c'est-à-dire des données qui seront utilisées pour vérifier que le modèle fonctionne, et qui n'ont de préférence pas été utilisées pour générer le modèle. Une relation est établie entre les valeurs des descripteurs et les valeurs de l'effet. La relation qui lie les variables d'entrée et de sortie peut prendre différentes formes plus ou moins complexes, allant de la simple régression linéaire aux régressions multilinéaires, polynomiales ou non-linéaires (Madden, 2010).

Une approche par analogie ne cherche pas à créer un modèle mathématique entre des variables d'entrée et de sortie, mais à comparer deux substances ou plus, et de juger la similarité entre leurs structures, leurs propriétés, leurs activités biologiques connues. Si les substances sont jugées suffisamment similaires à tous ces égards, on peut alors utiliser la valeur d'activité connue d'une substance (ou la moyenne de l'activité de plusieurs substances éventuellement) pour une autre substance dont on ne connaît pas l'activité (Madden, 2010). Un exemple d'une approche par analogie est donné dans le tableau suivant :

Substance	Masse moléculaire (g/mol)	Coefficient de partage logP	Solubilité aqueuse (mg/L)	Toxicité aigüe pour le poisson (mortalité) : 96h-CL50 (mg/L)	Toxicité aigüe pour l'algue (croissance) : 72h-CE50 (mg/L)
2-propylheptan-1-ol	158.29	4.17[a]	44[c]	1.9[e]	5.01[g]
décan-1-ol	158.29	4.57[b]	39.5[d]	2.4[f]	ND

Tableau 1 : exemple d'approche par analogie

a : base de données de l'ECHA, 1995, ligne directrice EU Method A.8, méthode agitation de flacon, 25°C, score Klimisch K2.
b : ECHA, Hansch *et al.*, 1995, K2 / ECHA, Doucette & Andren, 1988, méthode colonne génératrice, K2 / ECHA, Lide, CRC Handbook, 2008, K2
c : ECHA, 1959, procédure équivalente à la ligne directrice OCDE 105, méthode agitation de flacon, 25°C, K2 (pur à 90%) / ECHA, 2001, procédure équivalente à la ligne directrice OCDE 105, méthode agitation de flacon, 25°C, K2 (pur à 87%)
d : ECHA, Yalkowsky & Valvani, 1980, 25°C, K2
e : ECHA, 1995, ligne directrice OCDE 203, GLP, étude statique, espèce *Danio rerio,* concentration mesurée (moyenne géométrique), aération, durée de 96h, 22°C, K1
f : ECHA, Veith *et al.*, 1983, ligne directrice OCDE 203, étude dynamique, *Pimephales promelas*, concentration mesurée, 96h, 25°C, K2
g : ECHA, 2009, ligne directrice OCDE 201, GLP, étude statique, *Desmodesmus subspicatus*, en flacon fermé, avec espace d'air, concentration mesurée (moyenne géométrique), basée sur le taux de croissance, 72h, 21-23°C, K1

Compte tenu des propriétés similaires entre les deux substances de l'exemple, et sachant que leurs structures sont très proches (l'une est ramifiée et l'autre linéaire, mais ce sont tous les deux des alcools primaires), on peut conclure que le décan-1-ol aura une toxicité sur les algues très proche de celle du 2-propylheptan-1-ol, et ainsi, il est possible d'utiliser la valeur de ce dernier pour le décan-1-ol. Dans le cadre d'une approche par analogie pour une évaluation de risque, on préfère en revanche adopter une approche conservatrice, c'est-à-dire choisir un analogue dont on pense qu'il sera légèrement plus toxique que la substance qu'on cherche à caractériser.

b) Contexte et raisons de leur utilisation

Pour protéger la santé de l'humanité et de notre environnement, il est important de caractériser la toxicité de toutes les substances chimiques que nous produisons, utilisons, ou simplement auxquelles nous sommes exposées, même les substances naturelles. Pour ce faire, on réalise normalement des études de toxicité sur les rongeurs (pour modéliser la toxicité sur les humains), et sur d'autres animaux pour évaluer la toxicité des substances pour l'environnement. Cependant, en particulier depuis la mise en place de REACH (Parlement Européen and Conseil de l'Union Européenne, 2006), la réglementation européenne sur les substances chimiques importées ou fabriquées en Europe, l'utilisation de méthodes alternatives est encouragée afin de réduire au maximum l'expérimentation animale. Dans le cadre de cette réglementation, l'ECHA (European Chemicals Agency) demande que les industriels fournissent des informations sur leurs produits. Selon le tonnage (tonnes de substance par an importée/produite), un nombre croissant d' « endpoints » sont requis. Voici la liste des études requises en fonction du tonnage (TPA = Tonne Par Année) pour la toxicité et l'écotoxicité. Des études pour les propriétés physico-chimiques et de devenir dans l'environnement sont également demandées dans ces annexes, mais nous nous concentrons ici uniquement sur les aspects (éco)toxicologiques.

- De 1 à 10 TPA, annexe VII : toxicité aigüe orale, irritation oculaire et cutanée, sensibilisation cutanée, test de mutagénicité *in vitro* (test AMES), toxicité aigüe pour les invertébrés et pour les algues.
- De 10 à 100 TPA, annexe VIII : toxicité aigüe +1 autre voie (cutanée ou inhalation au choix), toxicité par administration répétée et toxicité pour la reproduction (sauf si les approches par analogie, QSARs ou méthodes *in vitro* indiquent que la substance est toxique pour le

développement) (ces deux endpoints peuvent être complétés en une seule étude), toxicité aigüe sur les poissons, inhibition respiratoire sur boue activée.

- De 100 à 1000 TPA, annexe IX : toxicité subchronique, toxicité sur le développement prénatal, toxicité pour la reproduction, toxicité chronique sur invertébrés et poissons, bioaccumulation dans une espèce aquatique, toxicité aigüe pour les invertébrés terrestres, les micro-organismes du sol et les plantes terrestres.
- Plus de 1000 TPA, annexe X : étude sur cellule somatique *in vivo* si résultat positif d'une étude de génotoxicité *in vitro*, Toxicité pour le développement, toxicité pour la reproduction sur une génération, étude de carcinogénicité, toxicité à long terme pour les invertébrés terrestres et pour les plantes, toxicité à long terme pour les organismes vivant dans les sédiments, toxicité à long terme ou toxicité pour la reproduction chez les oiseaux.

Toute cette batterie de tests de toxicité et d'écotoxicité fait souffrir des animaux vertébrés, protégés sous le règlement REACH, c'est pourquoi, les industriels sont encouragés à utiliser des méthodes alternatives (NC3Rs, n.d.), telles que les méthodes *in vitro*. Cependant, ces alternatives, qui ne remplacent pas directement les études sur animal, coûtent souvent très cher et ceci explique pourquoi les industriels recherchent d'autres méthodes qui n'utilisent pas les modèles animaux ou impliquent des expériences en laboratoire. Des méthodes telles que les approches par analogie et les (Q)SARs, ou les approches par analogie (read-across), qui permettent de se servir des données déjà existantes sur des substances similaires, ou appartenant à une même catégorie, pour générer une prédiction pour la substance d'intérêt.

Dans le cas de l'utilisation des méthodes *in silico*, les (Q)SARs et les approches par analogie, il convient de s'assurer que la substance d'intérêt est toxique par le même processus que la ou les substances qui servent à générer la prédiction de toxicité. En effet, il a été montré que la toxicité sur les poissons par exemple nécessitait différents coefficients de régression, voire différents descripteurs pour prédire la toxicité par un modèle QSAR (Russom *et al.*, 1997). Cela implique que pour une même valeur de logP (coefficient de partage octanol-eau), la valeur de la toxicité est différente pour deux substances qui exercent leur toxicité par un processus biologique différent. Ainsi, si ce processus pour une molécule test est mal connu, un modèle QSAR inapproprié pourra être utilisé, qui donnera finalement une valeur de toxicité erronée.

2) MechoAs vs MoAs

Dans ce travail, nous allons distinguer deux appellations très proches, dont la signification n'est pas toujours différente, ou pas définie de la même manière dans la littérature. Ces deux termes sont mode d'action (MoA, pour « Mode of Action » en anglais) et mécanisme d'action (MechoA pour « Mechanism of Action » en anglais).

Pour expliquer la différence entre ces deux termes, il est nécessaire d'expliquer préalablement le concept d'AOP (European Commission - DG Joint Reseach Centre and U.S. Environmental Protection Agency, n.d.) (« Adverse Outcome Pathway »). Cela désigne une séquence d'évènements au sein de l'organisme qui vont aboutir à un effet nocif pour l'organisme ou la population. On distingue notamment l'évènement moléculaire déclencheur (« molecular initiating event » en anglais) qui est le premier évènement clef qui va entraîner toute la cascade d'évènements aboutissant à l'effet nocif, et qui décrit le mécanisme au niveau moléculaire. Quelques exemples d'évènements moléculaires déclencheurs suivent :

- Accumulation des molécules dans les membranes cellulaires (pour l'AOP narcose)
- Inhibition de l'acétylcholinestérase (pour l'AOP inhibition de l'AChE conduisant à la mortalité aigüe)

- Liaison covalente aux protéines (pour l'AOP sensibilisation cutanée).

Dans le cadre de ces AOPs, les mécanismes d'action se concentrent sur le début de la séquence, alors que les modes d'action se concentrent sur la fin de la séquence. Plus précisément, le mode d'action regroupe les substances en fonction des effets nocifs qu'elles induisent, mais aussi la relation entre la valeur de toxicité et un descripteur de l'hydrophobie de la substance (en général le logP). En effet, certaines substances auront le même type d'effets, mais se positionneront sur une droite (un modèle QSAR) différente lorsqu'on trace leur toxicité en fonction de leur logP. Par exemple, le mode d'action de la narcose non-polaire regroupe les substances qui ont un effet (sur les poissons) d'activité locomotrice réduite, une coloration plus sombre, et une mort rapide en 24h (sauf pour les substances très hydrophobes pour lesquelles l'équilibre thermodynamique est plus long à atteindre), et dont la toxicité est bien décrite par les modèles QSAR de la narcose non-polaire (Russom *et al.*, 1997). En revanche, le mécanisme d'action regroupe les substances en fonction de l'évènement moléculaire déclencheur qu'elles exercent.

C'est donc pour une bonne utilisation des modèles QSARs et approches par analogies que la connaissance des mécanismes ou modes d'action sont importants. Cependant, la connaissance des MechoAs par une prédiction basée uniquement sur la structure moléculaire trouve aussi une application dans une approche de chimie médicinale, afin de développer de nouvelles molécules candidates à visée thérapeutique. En effet, même si dans ce travail nous visons principalement les mécanismes d'action toxique, il faut souligner que ce qui est toxique dans les conditions extrêmes des tests de toxicité, peut être salvateur dans des conditions de maladie ou d'empoisonnement par un mécanisme complémentaire, avec des doses ajustées afin de profiter de l'action thérapeutique sans subir des effets toxiques prononcés. Par exemple, un anticholinergique, un antagoniste de l'acétylcholine (structure en Figure 25) au niveau des récepteurs des terminaisons nerveuses, est un médicament d'urgence pour contrecarrer l'effet des agents innervants, qui inhibent l'acétylcholinestérase. En effet, l'acétylcholinestérase étant bloquée, il y a un excédent d'acétylcholine qui apparait au niveau de la fente synaptique, entraînant une surstimulation des neurones. Un anticholinergique va empêcher cette acétylcholine excédentaire d'agir sur les récepteurs et ainsi bloquer l'effet de l'agent innervant le temps que celui-ci soit éliminé par l'organisme (IPCS Inchem, 2002).

3) Différentes classifications et modèles de prédiction de MechoAs/MoAs

Il existe dans la littérature plusieurs classifications en MoAs ou MechoAs, et des modèles pour les prédire. Relativement peu de modèles sont construits de manière à être applicables à l'ensemble des substances organiques. La plupart des méthodes sont développées pour une famille réduite de composés, comme par exemple celle des phénols. Une revue des modèles existant est présentée ci-après. Cette liste est bien sûr non exhaustive, mais couvre les classifications et modèles les plus connus ainsi que quelques autres exemples.

a) Classifications et modèles sur les composés organiques en général

Tout d'abord, le schéma de Verhaar (Verhaar *et al.*, 1992), est la méthode utilisée par les écotoxicologues de manière générale. Ce modèle a été publié en 1992, puis a subi des mises à jour (Enoch *et al.*, 2008). Aujourd'hui, cette méthode peut être utilisée via plusieurs interfaces (ToxTree (Patlewicz et al., 2014, 2008), OECD QSAR Toolbox (OECD and ECHA, 2017)) soit sous sa forme originale,

soit le schéma mis à jour en 2008 (Enoch *et al.*, 2008). Ce schéma classifie les substances en 4 modes d'action :

- MoA 1 : narcose non-polaire ou toxicité de base
- MoA 2 : narcose polaire ou composés moins inertes
- MoA 3 : réactivité non spécifique
- MoA 4 : composés agissant spécifiquement

Il faut noter que lorsque ce modèle est utilisé via les interfaces des logiciels cités ci-dessus, il y a une cinquième catégorie définie comme suit : MoA 5 : impossible à classer avec ces règles.

Une autre méthode de classification qui s'applique aux composés organiques en général est la méthode de Russom *et al.*, 1997, qui prédit les modes d'action sur le poisson, basé uniquement sur des études sur le poisson tête de boule (*Pimephales promelas*, Rafinesque, 1820). Cette méthode classifie les composés en 8 catégories :

- Narcose de base ou narcose I
- Narcose polaire ou narcose II
- Narcose des ester ou narcose III
- Découplage de la phosphorylation oxydative
- Inhibition de la respiration
- Réactivité électrophile/proélectrophile
- Inhibition de l'AChE
- Crise du système nerveux central

Plus récemment, une classification en MechoAs plus poussée a été publiée (Barron *et al.*, 2015), qui regroupe une grande variété de substances : des métaux, organométalliques, pesticides et autres molécules organiques. Les composés sont classés en 6 MoAs généraux et 31 MoAs spécifiques grâce à des études sur poissons et invertébrés aquatiques, des QSARs, et d'autres données de la littérature.

- Narcose (non-polaire, polaire, ester)
- Inhibition de l'AChE (carbamate, organophosphate)
- Inhibition du transport d'électrons (découplage de la phosphorylation oxydative, inhibition de la phosphorylation oxydative, inhibition de la respiration de l'arsenic)
- Détérioration de l'iono/osmorégulation/circulation (anticoagulation, méthémoglobinémie, détérioration de l'iono/osmorégulation des métaux, autre osmorégulation)
- Réactivité (acrylate, alkylation, carbonyle, chromate, cyanure/nitrile, di/trinitroaromatique, hydrazine, phosphure, autre)
- Neurotoxicité (antagonisme GABA acyclique, antagonisme GABA des pyrazoles, modulation des canaux à sodium du diphényle, modulation des canaux à sodium des pyréthroïdes, agonisme du GABA, agonisme du nAChR, blocage des canaux à sodium, strychnine, autre)

Une méthode de prédiction des MechoAs du même groupe de recherche a été publiée 2 ans avant cette classification (Martin *et al.*, 2013), en utilisant uniquement une partie des MechoAs présentés ci-dessus. Cette méthode est construite à partir de descripteurs de tout type (expérimental, topologique, électronique), et prédit les MechoAs dans le poisson.

b) Classifications et modèles sur des familles restreintes

Karabunarliev *et al.* ont publié en 1996 (Karabunarliev *et al.*, 1996) une méthode de prédiction des MoAs grâce à des descripteurs quantiques. Cette méthode est limitée aux dérivés du benzène, basée sur les MoAs de Russom *et al.* (même équipe, article de Russom *et al.*, 1997 déjà soumis mais non

encore publié), pour deux espèces de poisson, le tête de boule (*Pimephales promelas*) et le guppy (*Poecilia reticulata*, Peters, 1859).

Bearden et Schultz ont publié en 1997(Bearden and Schultz, 1997) une méthode de prédiction des MoAs avec un mélange de descripteurs expérimentaux et quantiques, aussi limitée aux dérivés du benzène, à partir de données sur un poisson (P*imephales promelas*) et un protozoaire cilié (*Tetrahymena pyriformis*, Ehrenberg, 1830). Cette méthode classifie les composés en 6 catégories :

- Narcotiques non-polaires
- Phénols narcotiques polaires
- Electrophiles mous
- Acides faibles découpleurs de la respiration
- Proélectrophiles
- Anilines narcotiques polaires

Une méthode de prédiction des MoAs des phénols a été publiée par Ren en 2003 (Ren, 2003), basée sur des études de toxicité sur le protozoaire *Tetrahymena pyriformis*, et utilisant des descripteurs expérimentaux, topologiques et électroniques.

On peut aussi citer des méthodes de prédiction plus restreintes dans leur domaine d'étude, mais aussi plus poussées dans ce domaine spécifique : celle de Benigni *et al.* (Benigni et al., 2013, 2007) pour prédire le risque de cancérogénicité et mutagénicité (modèle utilisant des informations provenant de tests sur animaux, tests sur cellules de bactérie, et QSARs), celle de Kazius *et al.* (Kazius *et al.*, 2005) pour prédire la mutagénicité (à partir du test d'Ames, effectué sur des cellules de bactérie), celle de LoPachin et Gavin (LoPachin and Gavin, 2014) pour déterminer le MechoA des aldéhydes, celle de Spycher *et al.* (Spycher *et al.*, 2008) pour déterminer si les composés sont des découpleurs de la phosphorylation oxydative.

4) Objectifs de la thèse

Un objectif principal de cette thèse était de développer la première méthode fiable de classification globale qui pourrait être utilisée pour n'importe quelle molécule organique (et plus dans un second temps), aussi bien pour l'écotoxicologie que pour la toxicologie. Nous voulions principalement montrer la valeur ajoutée par rapport au schéma de Verhaar, qui, malgré les mises à jour, présente un taux d'erreur encore élevé et un domaine d'applicabilité assez restreint. En effet, c'est la méthode de prédiction utilisée par la majorité des écotoxicologues, qui est disponible gratuitement, et construite sous forme d'arbre de décision avec des règles de type alertes structurales (topologiques) simples, ce qui le rend attractif pour le plus grand nombre. Nous voulions aussi créer une classification qui organise les substances dans des catégories qui permettent de choisir le modèle QSAR adapté et spécifique à un MechoA (sous-catégories de MechoA). Cette thèse se concentre principalement sur les mécanismes d'action qui induisent des effets significatifs lors des tests de toxicité aigüe (4 jours pour les poissons, 14 jours pour les rats en toxicité orale, 2 jours pour les daphnies, etc.), car ils sont plus facilement observables et attribuables à une exposition donnée que pour les effets chroniques, et surtout les données de toxicité aigüe sont plus abondantes. Quelques données chroniques ont néanmoins été incluses dans ce travail.

Chapitre 2 : Revue bibliographique des mécanismes d'action

Un des travaux les plus importants dans cette thèse fut de collecter des informations sur les mécanismes d'action des substances dans divers organismes de manière la plus complète possible. Il est difficile d'en faire l'état de manière structurée sans en faire une liste fastidieuse à suivre. Nous avons fait le choix ici de classer ces informations en mécanismes d'action généraux regroupant plusieurs mécanismes d'action spécifiques similaires. L'identifications de tous ces mécanismes le plus précisément possible était primordiale pour attribuer une classification correcte aux diverses substances. Les mécanismes bien reconnus qui ont été identifiés dans ce travail seront bien détaillés dans ce chapitre, alors que les mécanismes mal connus seront simplement évoqués.

Les molécules non triviales citées dans ce document sont associées à un numéro (en gras dans le texte), et les structures de ces molécules, référencées par numéro, sont représentées en Figure 42 : planche de structures moléculaires.

1) Narcose

La narcose est un processus non spécifique, non réactif, que chaque molécule organique peut exercer, avec un potentiel différent (Verhaar *et al.*, 1992). Le principe de ce MechoA est une simple accumulation de molécules xénobiotiques dans les membranes cellulaires, qui sont des bicouches lipidiques. Du fait de l'accumulation de ces molécules xénobiotiques dans la membrane, celle-ci est perturbée jusqu'à un point où elle ne remplit plus son rôle (maintien du contenu de la cellule, contrôle du flux de molécules vers et depuis la cellule) correctement. Cela peut se manifester par une perte d'intégrité physique, et donc une rupture de l'enveloppe cellulaire qui maintient le cytosol. Cela peut aussi être le fait que certaines protéines membranaires essentielles (comme les canaux à ions ou les protéines de type GPCR) ne fonctionnent plus correctement, car un environnement perturbé modifie leur structure tridimensionnelle et ainsi leur fonctionnalité (Sikkema *et al.*, 1995). Etant donné que les membranes cellulaires sont des compartiments hydrophobes, la propriété principale qui dirige la toxicité par ce mécanisme est l'hydrophobie des molécules, qui permet d'évaluer la proportion de molécules qui vont rester dans les compartiments aqueux (milieu extérieur pour les organismes aquatiques, liquide extracellulaire, cytosol) et la proportion qui vont s'accumuler dans les compartiments hydrophobes (membranes, tissus adipeux, vésicules lipidiques). L'hydrophobie est quantifiée en général par un coefficient de partage, principalement le coefficient de partage octanol-eau (logP), ou des coefficients de partage membrane-eau ou DMPC-eau. La DMPC **1** (DiMyristoylPhosphatidylCholine) est un des composants majoritaires des membranes cellulaires (Gobas *et al.*, 1988). On peut aussi utiliser la solubilité aqueuse qui est bien corrélée au logP (Thomas *et al.*, 2015).

Le terme de narcose utilisé dans ce contexte est bien le même que lorsqu'on parle des substances narcotiques qui provoquent une anesthésie rapide et réversible. Et en effet, il a été montré que le pouvoir anesthétique des molécules était directement corrélé avec l'hydrophobie des substances, d'après la corrélation de Meyer-Overton (Lugli *et al.*, 2009). Cela semblait indiquer que le mécanisme de cette narcose pourrait être simplement l'accumulation des molécules dans les membranes des cellules du cerveau. Ceci est possible avec les substances volatiles qui sont des petites molécules neutres et légèrement polaires telles que l'éther diéthylique ou le chloroforme. Malgré tout, aujourd'hui plusieurs études sur les mécanismes de l'anesthésie semblent montrer que les narcotiques ciblent au contraire des protéines spécifiques (telles que des canaux à ions), grâce à une poche de liaison peu spécifique (Lugli *et al.*, 2009).

a) Narcose non-polaire

La narcose non-polaire, ou toxicité de base, est le mécanisme d'action qui génère la plus faible toxicité, au moins pour les substances les moins hydrophobes (Verhaar *et al.*, 1992). En effet, à même solubilité,

ou même coefficient de partage logP, c'est le mécanisme qui cause la toxicité la plus faible (i.e. log (1/CE50) le plus faible). C'est ce qu'on observe sur la Figure 1 qui présente différents modèles QSARs superposés, extraite d'une revue des QSARs pour la toxicité environnementale (Dearden, 2002).

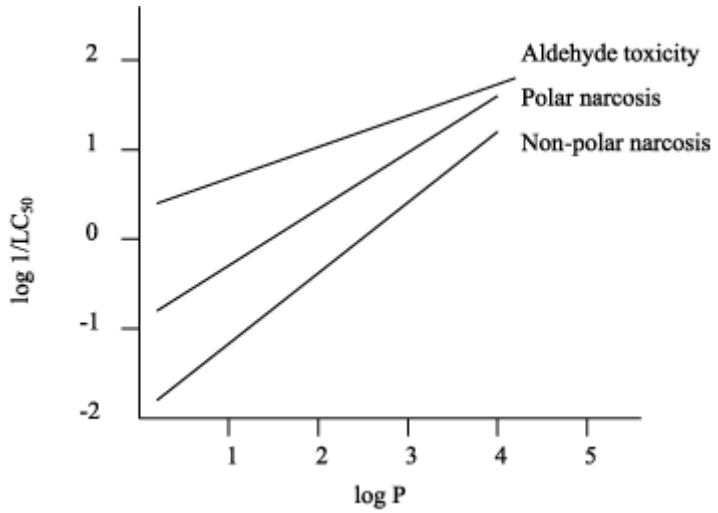


Figure 1 : superposition de modèles QSARs pour la toxicité environnementale. LC50 dans cette figure = CL50 pour le poisson à 96h. Figure tirée de (Dearden, 2002).

Le mécanisme de la narcose non-polaire a pu être compris et caractérisé grâce au développement de modèles QSAR et mis en évidence par quelques expériences. Le principe fondamental qui est lié au fonctionnement de la narcose non-polaire est celui de l'activité chimique. L'activité chimique correspond à la dose réelle qui sera présente au niveau du site d'action (membrane des cellules dans le cas de la narcose non-polaire). En effet, en convertissant la dose d'exposition des organismes en activité chimique, on constate que l'effet toxique dû à ce mécanisme intervient à une même valeur d'activité quelle que soit la substance testée (Ferguson, 1939; Mackay *et al.*, 2009). Du côté plus expérimental, De Sousa *et al.* (de Sousa *et al.*, 2012) ont exposé des bactéries fluorescentes (*Pseudomonas fluorescens*, Migula, 1985) dans des milieux contenant du carvacrol **2** (2-méthyl-5-(propan-2-yl)phénol) et du 1,8-cinéole **3** (1,3,3-triméthyl-2-oxabicyclo[2,2,2]octane) seuls ou en mélange. Ils ont observé les cellules ainsi exposées avec un microscope électronique à transmission et à balayage. Ils ont pu constater que les deux substances perturbaient l'intégrité de la membrane, et des trous apparaissaient, laissant s'échapper le cytosol hors de la cellule. Le 1,8-cinéole est un éther simple, exerçant *a priori* la narcose non-polaire (Enoch *et al.*, 2008), et le carvacrol est un phénol simple, exerçant *a priori* la narcose polaire comme tous les autres phénols qui ne sont substitués que par des chaînes alkyles (Enoch *et al.*, 2008). Ces images de microscopie sont donc vraisemblablement une visualisation de l'effet de la narcose non-polaire et polaire. Elles sont présentées en Figure 2.

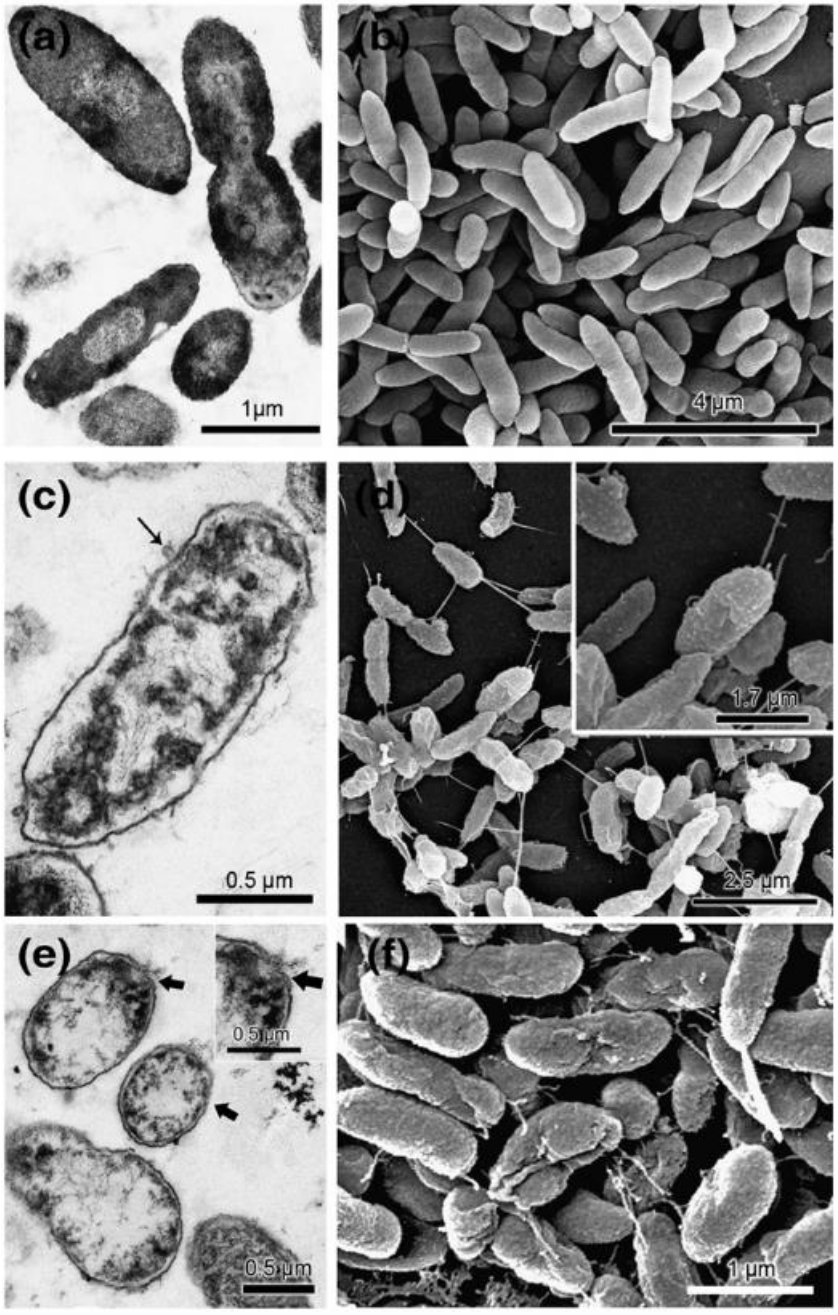


Figure 2 : bactéries P. fluorescens vues au microscope électronique à transmission (colonne de gauche) et à balayage (colonne de droite) suite à l'exposition (a-b) contrôle sans carvacrol ni 1,8-cinéole, (c-d) avec carvacrol, (e-f) avec 1,8-cinéole. Figure tirée de de (de Sousa et al., 2012).

Ainsi, comme la narcose est un mécanisme non spécifique, que toutes les molécules peuvent exercer avec un potentiel différent, ce mécanisme est celui qui dirige la toxicité des substances dites inertes, c'est-à-dire celles qui n'agissent pas par un autre mécanisme d'action toxique. Le terme de substances inertes est un abus de langage car elles sont quand-même toxiques, et on lui préférera le terme narcose ou toxicité de base. C'est le cas pour les familles chimiques suivantes (avec des exceptions) : les hydrocarbures, les cétones, les alcools, les éthers et les versions halogénées de ces types de molécules principalement (Ellison *et al.*, 2015).

D'autres auteurs ont compilé une excellente revue des différentes hypothèses pour expliquer le fonctionnement de la narcose non-polaire ou anesthésie générale (Franks and Lieb, 1982). Selon ces auteurs, l'hypothèse de la perte d'intégrité des membranes par accumulation des molécules

xénobiotiques dans celle-ci n'est pas la plus cohérente en fonction des résultats d'études mécanistiques expérimentales. Ces auteurs expliquent qu'une hypothèse plus prometteuse est que les molécules de narcotique se partitionnent dans des sites de liaison amphiphiles (possédant une partie hydrophobe et une partie hydrophile) de certaines protéines spécifiques, perturbant légèrement leur organisation spatiale, empêchant leur activité normale. Ces sites de liaisons peuvent être les interfaces entre les sous-unités d'une même protéine par exemple, comme il a pu être mis en évidence avec l'hémoglobine. Une protéine dont l'activité est perturbée par des faibles concentrations de substance, et pour une gamme assez diverse de narcotiques est la luciférase (Franks and Lieb, 1982).

b) Narcose polaire

Le mécanisme de la narcose polaire est considéré comme étant très similaire à celui de la narcose non-polaire – d'où le partage du terme narcose – bien qu'une différence de toxicité significative soit observée à même hydrophobie (Figure 1) (Escher and Hermens, 2002; Verhaar et al., 1992). Ce mécanisme est exercé principalement par les phénols et anilines simples, mais aussi certaines amines et des hétéroaromatiques azotés (Ellison *et al.*, 2015).

On peut constater que les effets du carvacrol **2**, narcotique polaire, étaient similaires à ceux du 1,8-cinéole **3**, narcotique non-polaire, sur la Figure 2, si ce n'est qu'une concentration plus faible était nécessaire pour observer ces effets (de Sousa *et al.*, 2012).

La séparation entre narcose non-polaire et polaire reste en fait assez floue. En effet, on peut constater que des familles de molécules polaires sont incluses dans les narcotiques non-polaires, telles que les alcools, les cétones et certaines amines, alors que les phénols et anilines simples constituent la majorité des narcotiques polaires. Certaines pyridines semblent être dans la première catégorie, et d'autres dans la deuxième (Ellison *et al.*, 2015). Certains auteurs considèrent que la séparation entre les deux est liée à la polarisabilité des molécules ou encore à leur capacité à former des liaisons hydrogène fortes (Roberts *et al.*, 2013).

D'autres auteurs (Vaes *et al.*, 1998) ont aussi suggéré que la différence entre les deux catégories était simplement liée au fait qu'on utilise le logP comme modèle pour quantifier l'équilibre de partition des molécules entre les liquides aqueux et les membranes cellulaires. En effet, le logP assimile l'octanol aux lipides membranaires. Des expériences ont été menées en remplaçant l'octanol par la DMPC **1** (phosphatidylcholine de dimyristoyle), un composant majoritaire des membranes cellulaires. En traçant (Figure 3) la toxicité pour les poissons (logCL50) en fonction du coefficient de partage octanol-eau, noté ici $\log K_{ow}$, puis en fonction du coefficient de partage DMPC-eau, $\log K_{DMPC}$, les droites de la narcose non-polaire et de la narcose polaires se sont rejointes en une seule et même droite (Vaes *et al.*, 1998). Enfin, d'autres auteurs (Endo et al., 2013; Escher and Hermens, 2002) ont modélisé les organismes en plusieurs compartiments (lipides de stockage, membranes cellulaires, albumine, protéines musculaires, eau) et expliquent que la différence entre la narcose non-polaire et polaire vient d'une répartition différente dans ces compartiments variés qui n'est pas prise en compte par le logP.

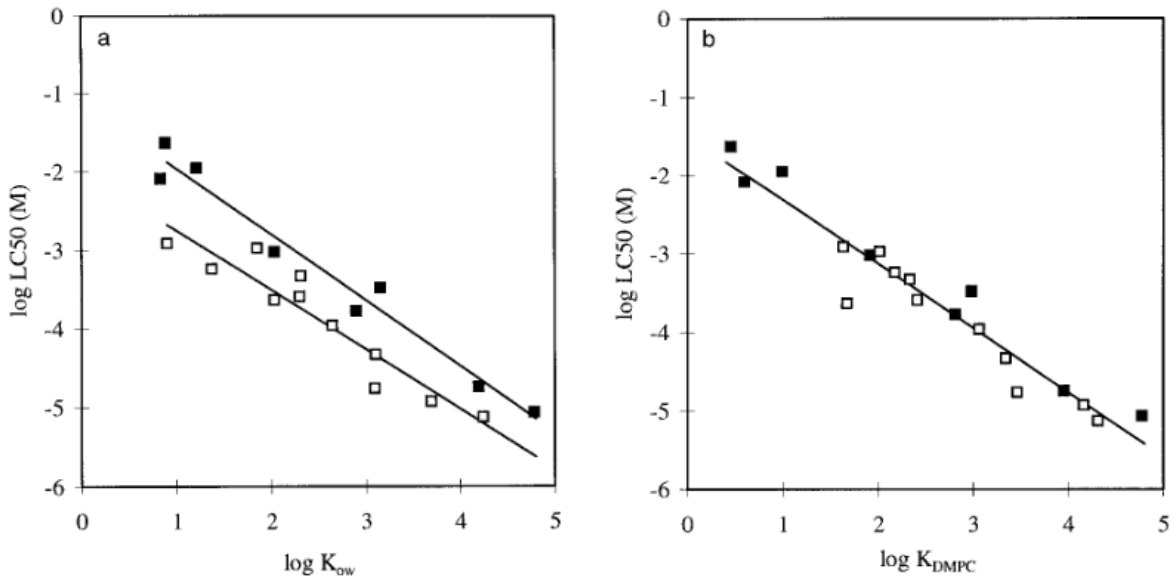


Figure 3 : concentrations léthales médianes (CL50) pour le guppy et le tête de boule pour les narcotiques non polaires (carrés pleins) et polaires (carrés vides) en fonction de a) leur logP et b) leur logK_{DMPC}. Figure tirée de (Vaes et al., 1998).

2) Narcose des esters ou pronarcose

Veith et ses collaborateurs(Veith and Broderius, 1987) ont pu identifier que les esters simples (voir la structure d'un ester à gauche de la Figure 4) provoquaient un symptôme comportemental similaire à la narcose. De plus, ces composés sont peu réactifs en milieu aqueux neutre. Cependant, lorsque leur toxicité a été tracée en fonction de leur logP, ils se sont révélés être plus toxiques que les narcotiques non-polaires et polaires de même logP. Tout cela a conduit ces auteurs à décrire le MoA des esters comme un troisième type de narcose : la narcose des esters. Cette classe a été reprise dans la classification de Russom (Russom *et al.*, 1997).

Par ailleurs, Jaworska *et al.* (Jaworska *et al.*, 1995) ont montré, grâce à des études sur protozoaire (*Tetrahymena pyriformis*) et sur poisson (*Pimephales promelas*) et des analyses QSAR, que les esters se comportaient comme des narcotiques non-polaires pour le protozoaire, alors qu'ils étaient significativement plus toxiques que la narcose de base pour le poisson. Cette observation a été expliquée avec l'hypothèse que ces protozoaires ne métabolisent pas, ou trop lentement, les esters, alors que les poissons ont des estérases bien actives qui vont hydrolyser les esters (Figure 4). Cette hydrolyse déplace l'équilibre de concentrations en ester entre le milieu externe et le poisson, puisque l'ester est consommé au fur et à mesure par l'hydrolyse. En conséquence, une quantité plus importante que si l'ester était inerte va pénétrer dans le poisson, entraînant une plus forte toxicité.

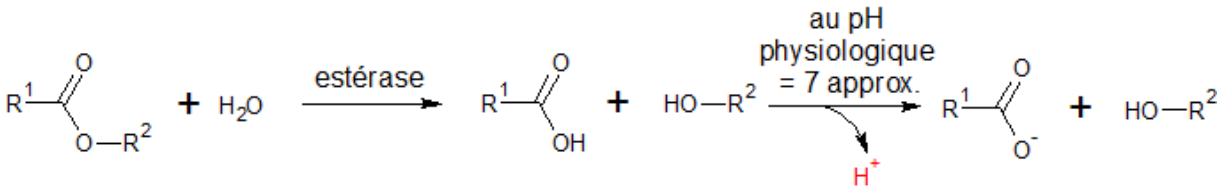


Figure 4 : hydrolyse enzymatique des esters. R^1=H ou Alkyle, R^2=Alkyle.

Il peut être intéressant de préciser que suite à l'hydrolyse, on obtient des composés moins hydrophobes que l'ester parent, tel que mis en évidence avec l'exemple montré dans le Tableau 2 :

	Propanoate d'éthyle	→ acide propanoïque	+ éthanol
logP	1,21	0,33*	-0,31
Référence	EpiSuite, Hansch *et al.*, 1995	ECHA/EpiSuite, Hansch *et al.*, 1995, neutre, K2 / ECHA, GESTIS, 2014, K2	EpiSuite, Hansch *et al.*, 1995 (et plusieurs autres valeurs en accord)

Tableau 2 : hydrophobie d'un ester et de ses produits d'hydrolyse.

* Remarque : la valeur de logP mesurée pour l'acide propanoïque est mesurée pour la forme neutre (acide non dissocié) en laboratoire. Cependant, à pH physiologique, suite à l'hydrolyse enzymatique dans un organisme vivant, l'acide sera déprotoné et on obtiendra l'anion propanoate, qui sera encore nettement moins hydrophobe.

Le mécanisme d'hydrolyse des esters favorise l'élimination du toxique hors de l'organisme, puisque les produits d'hydrolyse sont significativement plus hydrophiles que l'ester parent. Ces produits d'hydrolyse sont aussi moins toxiques que l'ester parent dans la plupart des cas, puisqu'une hydrophobie plus faible est corrélée avec une toxicité plus faible pour les narcotiques. En effet, un alcool (sauf cas particuliers) et un carboxylate (forme déprotonée de l'acide) vont simplement exercer le mécanisme de la narcose non-polaire, mais, étant moins hydrophobes que l'ester parent, seront moins toxique. Cependant, si l'ester est un ester de phénol, par exemple l'acétate de phényle, ce ne sera pas un alcool qui sera relâché, mais un phénol, qui exercera donc un mécanisme plus toxique, au moins la narcose polaire, voire le découplage de la phosphorylation oxydative pour les phénols plus acides, ou encore l'utilisation d'un cycle RedOx si c'est un *ortho*- ou *para*-dihydroxybenzène (voir les sections dédiées à ces mécanismes plus loin). Il convient aussi de mentionner le mécanisme particulier des *ortho*-phtalates. En effet, pour ces diesters aromatiques, après hydrolyse de l'un des deux esters, le monoester produit a des propriétés de perturbateur endocrinien, et induit le PPARα (récepteur activé par les proliférateurs de peroxysomes alpha), qui est un précurseur pour le développement de cancers du foie (Adams *et al.*, 1995). On peut citer également les tartrates, qui après hydrolyse génèrent l'acide tartrique, inhibiteur de la fumarase, une enzyme essentielle du cycle de Krebs (Shaw, 2002). D'autres exemples de toxicité accrue après hydrolyse existent mais ne seront pas détaillés ici.

Selon les espèces, le taux d'hydrolyse enzymatique des esters au cours d'un test de toxicité aigüe est plus ou moins importante. De manière générale, plus les organismes sont complexes (multicellulaires, avec des organes distincts), meilleure est leur capacité métabolique, notamment grâce à la présence du foie, organe spécialisé dans le métabolisme. C'est le constat qu'ont fait Jaworska *et al.* en 1995 (Jaworska *et al.*, 1995), en établissant des relations structure-toxicité des esters pour le poisson (*Pimephales promelas*) et pour un protozoaire (*Tetrahymena pyriformis*). En effet, ils ont pu montrer dans cette étude que les esters avaient la même toxicité que les narcotiques non-polaires pour *T. pyriformis*, alors qu'ils étaient significativement plus toxiques pour *P. promelas*. Cette différence de toxicité a été attribuée à l'hydrolyse enzymatique, que le poisson effectuait largement alors que le protozoaire semblait inactif sur ce point.

3) Toxicité par réactivité avec formation d'adduits

Ce mécanisme d'action concerne les substances qui vont réagir spontanément (c'est-à-dire sans catalyse enzymatique) avec des composés endogènes pour former des adduits aux protéines et/ou à l'ADN. Ceci peut se produire par différents mécanismes moléculaires, et impacter différentes cibles (par exemple les résidus thiols, les résidus aminés), d'où la nécessité de créer des sous-catégories de MechoA.

Les organismes complexes ont cependant un bon système de protection contre ce type de substance : en effet, ils possèdent le glutathion **5**, un thiol, et les enzymes glutathion-S-transférase, qui va catalyser

le processus de conjugaison du glutathion avec les molécules réactives, et glutathion peroxydase, qui va réduire les peroxydes. La molécule conjuguée sera ensuite éliminée rapidement car elle n'est plus réactive et assez hydrophile. Ces mécanismes seront détaillés en section 4.a) Métabolisation réduisant la toxicité.

a) Electrophiles durs

Ce mécanisme concerne des molécules réactives qui ont une déficience d'électrons marquée et très localisée, générée par un groupe attracteur d'électrons qui est aussi souvent un bon groupe partant. Les mécanismes inclus dans cette catégorie sont présentés en Figure 5. Dans la réaction c) de la figure, X désigne plus spécifiquement un halogène, et c'est donc un acide fort qui est relâché (acide fluorhydrique, chlorhydrique, etc.). La réaction avec l'eau étant très rapide, la réaction d'acylation b) n'a probablement pas le temps de se passer, car l'eau est omniprésente dans et autour des cellules, et une réaction du type de la c) prédominera.

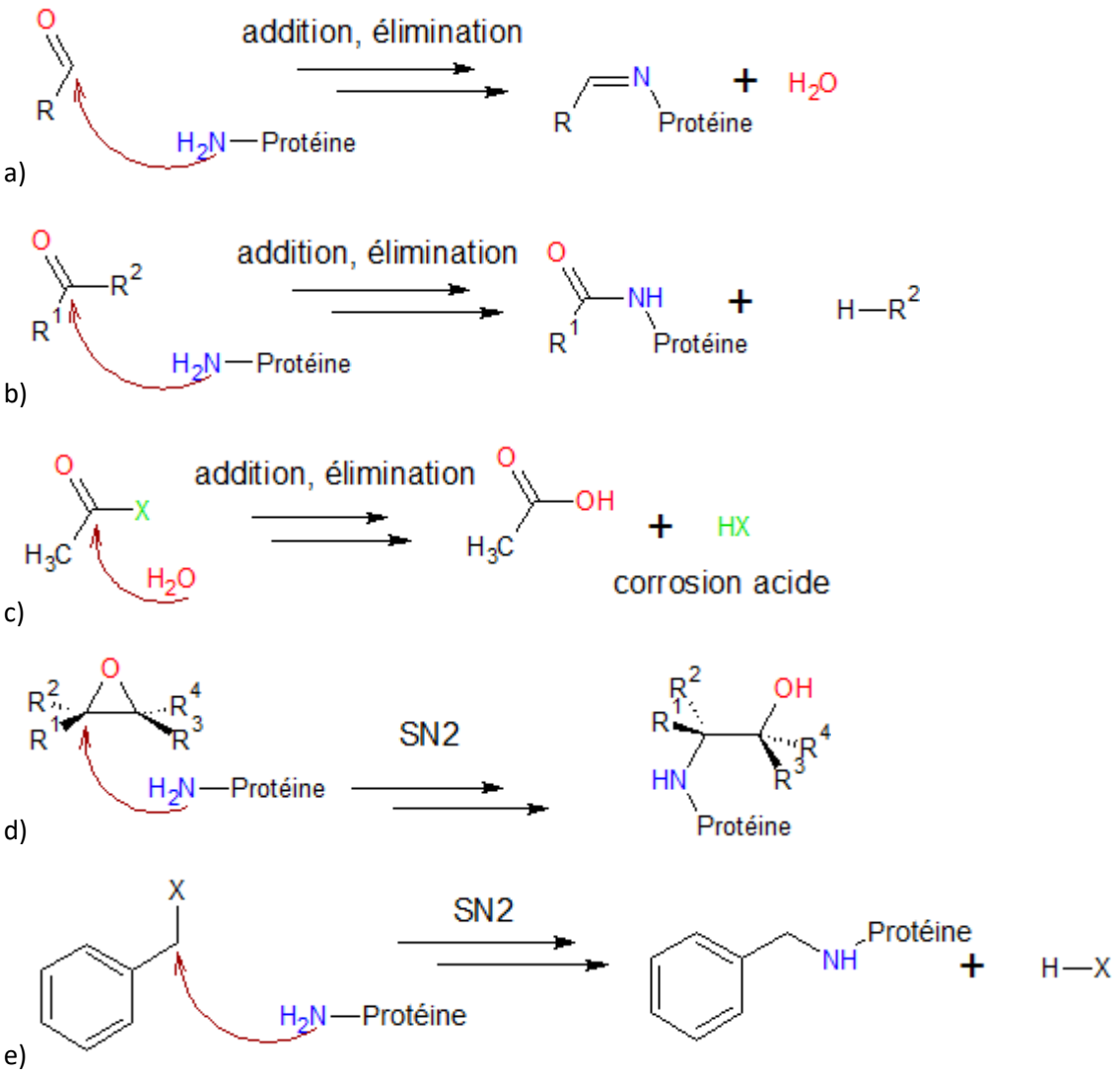


Figure 5 : formation d'adduits ou libération d'acide par les électrophiles durs. X désigne n'importe quel bon groupe partant.

Ces mécanismes et les substances qui y participent sont présentés dans la littérature (Aptula and Roberts, 2006), parmi d'autres mécanismes. Ces substances forment principalement des adduits avec les résidus aminés des protéines et de l'ADN. En effet, d'après la théorie HSAB (Acides et Bases Durs et Mous), un acide (de Lewis) dur réagira préférentiellement avec une base (de Lewis) dure, et non un

dur avec un mou. Cette théorie est extrapolable aux électrophiles (équivalents des acides) et nucléophiles (équivalents des bases). La notion de dur et mou se rapporte à la répartition des électrons. Les électrophiles et nucléophiles durs ont une déficience ou un excès d'électrons très localisé(e) sur la molécule et peu déformable, alors que pour les mous, cette déficience ou cet excès est réparti(e), délocalisé(e) et les électrons peuvent facilement se déplacer sur un plus grand volume (soit sur plusieurs atomes dans la molécule, soit autour d'un atome avec un nuage d'électrons très large et déformable) (Jacobs, 1997). Ainsi, dans les exemples présentés ci-dessus, la molécule de départ est un électrophile dur car la déficience d'électrons est localisée sur un seul carbone (ou deux carbones dans le cas des époxydes), et elle réagit avec l'amine qui est un nucléophile dur, les électrons étant localisés au niveau du doublet non-liant de l'azote.

Il a pu être montré que ce type de substances forment des adduits aux protéines et à l'ADN, notamment les aldéhydes et les époxydes. Pourtant, ces deux fonctions chimiques sont utilisées de manière transitoire dans le fonctionnement normal des organismes vivants. Les aldéhydes ont un degré d'oxydation intermédiaire entre un acide et un alcool primaire. C'est pourquoi ils constituent d'une part des étapes intermédiaires dans des processus d'oxydation ou de réduction importants pour l'organisme (Dunn *et al.*, 2009). D'autre part, ils sont une étape intermédiaire dans les processus de détoxication, en transformant un alcool en acide, composé plus hydrophile, qui pourra en plus être conjugué, et ainsi sera facilement éliminé de l'organisme. Les époxydes sont utilisées de manière transitoire car ils sont un état intermédiaire également entre un hydrocarbure et un diol, afin de rendre les molécules beaucoup plus hydrophiles, pour leur élimination efficace, tel que montré pour la carbamazépine (Tomson *et al.*, 1983) (Figure 6).

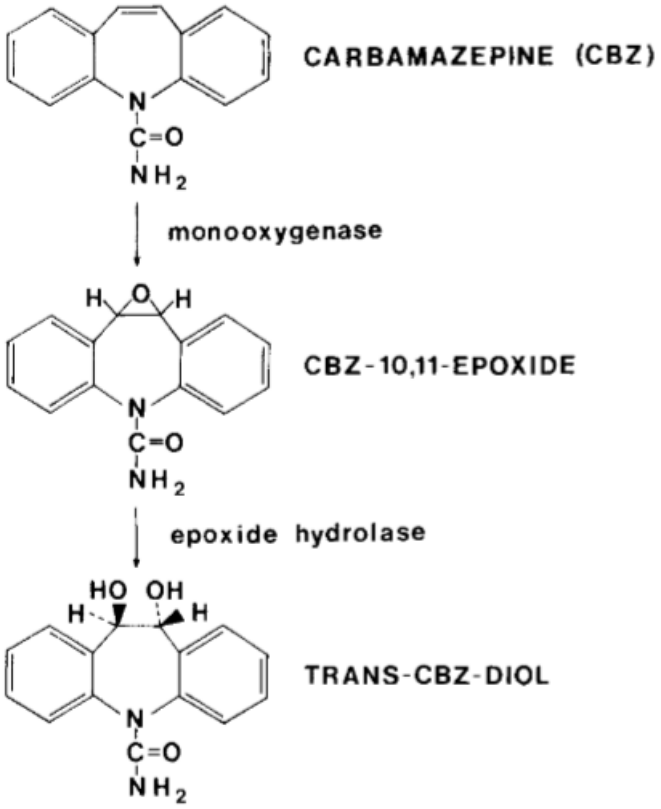


Figure 6 : voie époxyde-diol pour éliminer une substance de l'organisme. Figure tirée de (Tomson et al., 1983)

Ainsi, ces molécules réactives sont générées naturellement dans l'organisme, mais de manière très transitoire. Elles sont transformées rapidement par les enzymes dédiées afin de ne pas exercer leur mécanisme d'action toxique.

Comme montré en Figure 5, un des types d'électrophiles durs sont les molécules qui possèdent un bon groupe partant branché sur une molécule possédant un état de transition favorisé, c'est-à-dire que la charge partielle créée lors du départ du groupe partant est bien stabilisée par une bonne répartition des électrons. Un bon groupe partant désigne un groupe, qui une fois séparé de la molécule de départ, est lui aussi stabilisé. En effet, la rupture de la liaison génère une charge de part et d'autre, qui doivent toutes deux être stabilisées pour que la réaction soit favorable. Ce mécanisme est détaillé en Figure 7. Les molécules qui possèdent un bon potentiel pour cette réactivité sont les molécules possédant un bon groupe partant branché sur un groupe benzyle ou allyle principalement. Une liste non exhaustive des bons groupes partants est donnée dans le Tableau 3.

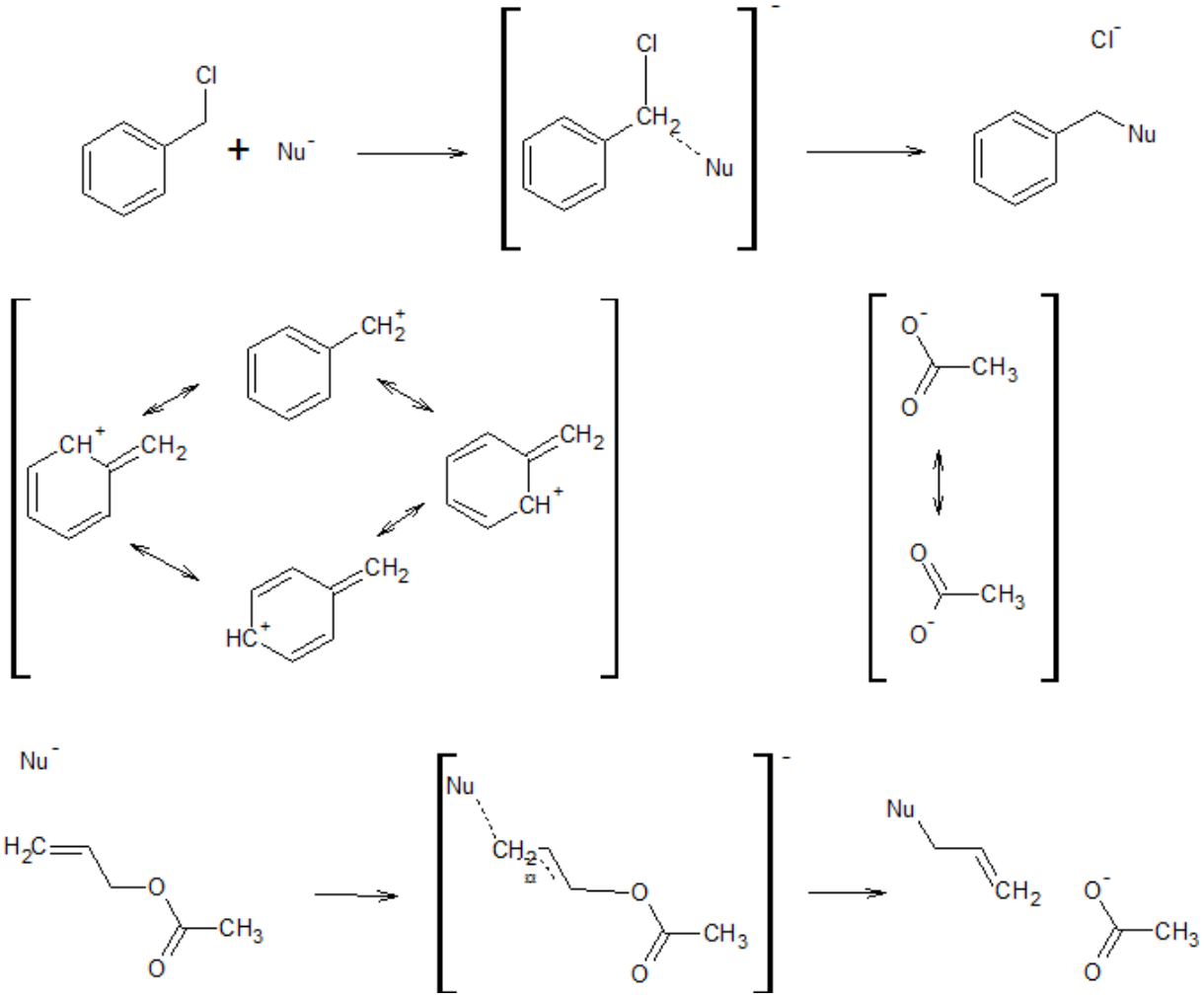


Figure 7 : mécanisme de réaction des électrophiles possédant un bon groupe partant, un état de transition et un groupe partant stabilisés.

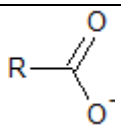
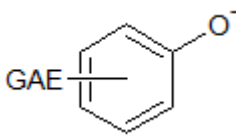
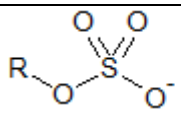
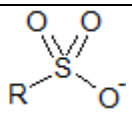
Nom du groupe partant	Structure
Halogénure (chlorure, bromure, iodure) Le fluor n'est pas un bon groupe partant car la liaison carbone-fluor est très stable.	Cl^-, Br^-, I^-
Carboxylate	
Phénolate acide (pKa < 7 : au pH physiologique, il doit être majoritairement anionique, car c'est ainsi qu'il sera éliminé de la molécule lors de la réaction)	 GAE = Groupes Attracteurs d'Electrons
Amine basique (pKa > 8 : protonée sur la molécule de départ, puis éliminée sous la forme neutre)	$R\text{-}NH_2$
Thiocyanate et isothiocyanate	SCN^-
Sulfate, sulfonate	 

Tableau 3 : groupes partants. Les noms et structures sont ceux des groupes une fois éliminés de la molécule initiale.

D'autres électrophiles durs moins évidents existent aussi et doivent être cités. Les γ-dicétones, non gem-disubstituées (c'est-à-dire ne portant pas deux substituants sur le même carbone) peuvent réagir avec des résidus aminés de protéines pour former un pyrrole. Ce mécanisme est présenté en Figure 12. Pour que la formation du pyrrole puisse avoir lieu, il faut donc que les deux cétones soient à la bonne distance, c'est-à-dire que les deux carbonyles soient séparés par deux carbones intermédiaires (ce qui correspond à une γ-dicétone), afin d'obtenir un cycle à 5 maillons lorsque l'amine fera la jonction entre les deux carbones des carbonyles. De même, il faut également qu'aucun des carbones qui constitueront le cycle ne soit disubstitué, car chacun devra perdre un proton lors des étapes de déshydratation afin de générer l'aromaticité du cycle.

b) Electrophiles mous

Comme expliqué dans la section précédente, les électrophiles mous vont donc, par opposition aux électrophiles durs, avoir un nuage d'électrons délocalisé et déformable. Ils vont réagir avec des nucléophiles mous, dont l'exemple majoritaire dans l'organisme est constitué des thiols (chaîne latérale des acides aminés cystéine, ou molécule de glutathion **5**).

Les électrophiles mous les plus fréquents sont les composés possédant un groupe carbonyle α,β-insaturé, dont les esters et aldéhydes principalement, mais aussi les quinones, qui sont des cétones α,β-insaturées de réactivité particulière. D'autres groupes attracteurs conjugués à une liaison multiple, différents d'un carbonyle, peuvent aussi servir à obtenir un électrophile mou. Par exemple un nitrile α,β-insaturé. Ces substances et leur mécanisme de réactivité sont présentées en Figure 8.

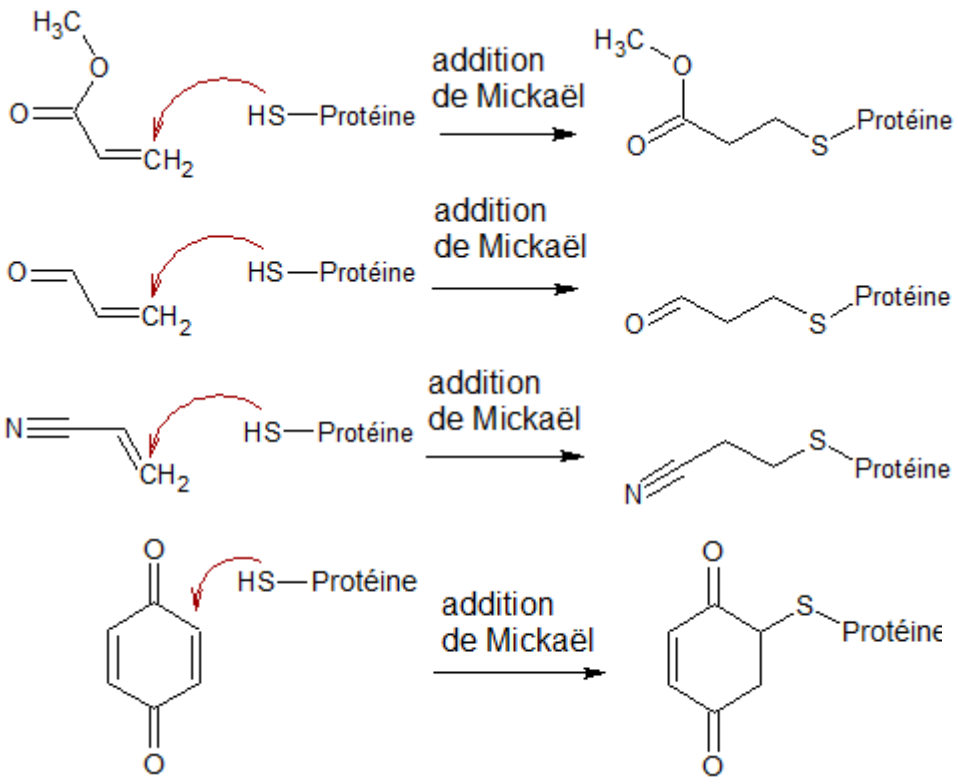


Figure 8 : formation d'adduits avec les protéines par les électrophiles mous

c) Générateurs de radicaux

Les molécules concernées par ce mécanisme d'action sont les molécules possédant une liaison simple oxygène-oxygène ou soufre-soufre, car ce type de liaison a la capacité de se rompre facilement de manière homolytique, c'est-à-dire que la paire d'électrons qui formait la liaison est séparée en deux pour donner deux radicaux libres. Ces radicaux libres pourront ensuite former des adduits directement, ou seront plutôt rapidement piégés par l'oxygène moléculaire, pour former l'anion superoxyde, qui réagira ensuite avec beaucoup de molécules endogènes (lipides, protéines, ADN), en formant des peroxydes (Kazius *et al.*, 2005; Munday, 1989; Wiley-VCH Verlag, 2002a).

4) Toxicité pro-active

Ce mécanisme fait intervenir une métabolisation impliquant une oxydation ou réduction ou une substitution par un groupe détoxifiant.

La métabolisation des substances xénobiotiques par les organismes a pour objectif soit de les utiliser dans les processus de fonctionnement de l'organisme (récupération d'énergie en oxydant les sucres et les lipides, récupération des acides aminés en hydrolysant les protéines, etc.) soit de les éliminer, en les rendant plus hydrophiles et moins toxiques si possible. Les molécules sont rendues plus hydrophiles en étant oxydées, puis conjuguées avec des molécules endogènes hydrophiles. Les molécules plus hydrophiles s'éliminent plus facilement car elles ne vont pas s'accumuler dans les membranes et tissus adipeux, mais vont être éliminées dans les urines. C'est le cas par exemple pour l'alcool benzylique **4**, qui est d'abord oxydé en acide, et qui sera en plus conjugué à la glycine **26** pour être éliminé encore plus facilement (Nair, 2001).

Cependant, ces voies métaboliques peuvent aussi générer des composés plus toxiques que le composé parent. Des sous-catégories ont donc été créées dans cette partie en fonction du type d'activité

exercée par les métabolites. Ainsi, les substances étant métabolisées quantitativement sont rendues plus actives, soit du point de vue de leur réactivité, ou de leur activité inhibitrice d'une enzyme, ou encore du processus d'élimination hors de l'organisme. C'est pourquoi ce mécanisme général a été appelé toxicité pro-active.

Ces mécanismes d'action sont souvent très variables d'une espèce à l'autre, car ils font intervenir le métabolisme, qui peut différer grandement entre les espèces. Une telle différence est par exemple observée pour le 3-methoxyphénol, qui est hydroxylé par les mammifères pour générer un dérivé de catéchol, qui sera toxique par cycle RedOx (Moridani *et al.*, 2003), alors que pour les poissons et protozoaires, il semble simplement exercer la narcose polaire, ce qui laisse penser qu'il n'est alors pas métabolisé de manière quantitative en dérivé d'hydroquinone comme chez les mammifères (Ellison *et al.*, 2015) (détails en 4.b.ix) Toxicité des benzènediols et éthers dérivés).

a) Métabolisation réduisant la toxicité

Comme expliqué dans la section sur les substances réactives, les organismes se dotent de moyens de détoxication, qui fonctionnent bien pour une partie des molécules. Le principal outil de détoxication est le glutathion **5** (GSH), qui peut être oxydé (GSSG) afin de transformer les xénobiotiques toxiques par une réduction (par exemple cétone transformée en alcool), ou bien être conjugué avec les substances, grâce aux enzymes dédiées (Glutathion S-transférase : GST, ou Glutathion peroxydase : GPx). Le glutathion oxydé (GSSG) est ensuite reconverti sous sa forme réduite initiale (GSH) grâce à la glutathion réductase, qui consomme le cofacteur NADPH (Li, 2009; Liska, 1998).

D'autres molécules peuvent être conjuguées aux substances (acides aminés, acide glucuronique **6**, sulfate, acétate, méthyle) pour en réduire la toxicité et/ou augmenter l'hydrophilie. Ces conjugaisons sont connues comme la détoxication de phase II. Elles ne sont possibles que si la substance possède déjà un groupe suffisamment réactif (époxyde, groupe partant, phénol, acide carboxylique, etc.). Les substances qui sont plus inertes auront donc besoin d'être d'abord activées : c'est la détoxication de phase I. Cette activation est effectuée par l'oxydation des substances grâce aux cytochromes P450 (Liska, 1998).

Il faut noter que ces mécanismes de détoxication utilisent des ressources de l'organisme, qui ne sont pas infinies : les réserves de molécules à conjuguer telles que le glutathion, l'acide glucuronique, mais aussi les cofacteurs qui permettent de les restaurer et qui permettent aux cytochromes d'oxyder les substances (NADH, NADPH). Dans le cas d'une exposition à une dose trop forte, ou une exposition chronique ou prolongée, ces ressources peuvent être épuisées, ou pas renouvelées assez vite, et la détoxication ne se fera donc plus.

Ces mécanismes de détoxication concernent :

- Les électrophiles durs avec un bon groupe partant : le glutathion va jouer le rôle du nucléophile qui vient remplacer le groupe partant, pour empêcher que ce rôle soit joué par un fragment de protéine ou d'ADN.
- Les alcools benzyliques (Nair, 2001) : ils seront oxydés facilement par les alcool déshydrogénases pour générer l'aldéhyde correspondant, puis l'acide correspondant par les aldéhyde déshydrogénases. L'acide pourra ensuite être conjugué. Ce mécanisme intervient aussi pour les alcools aliphatiques, puisqu'il est bien caractérisé pour l'éthanol. Cependant, ces mécanismes ne semblent pas jouer un rôle significatif dans les études de toxicité aigüe pour les alcools aliphatiques, aussi bien pour les mammifères (Koleva *et al.*, 2011) que pour les organismes aquatiques (Russom *et al.*, 1997), où la toxicité des alcools est tout à fait bien modélisée par les QSARs de la narcose non-polaire.

- Les aldéhydes aromatiques riches en électrons (IPCS Inchem, 2015; Wiley-VCH Verlag, 2002b) : ils seront oxydés par les aldéhyde déshydrogénases pour générer l'acide correspondant. On constate que plusieurs aldéhydes aromatiques riches en électrons (non substitués ou substitués avec des groupes donneurs d'électrons) sont très peu toxiques comparés aux aldéhydes aromatiques pauvres en électrons (Ellison *et al.*, 2015). Cela semble indiquer que les aldéhydes aromatiques riches en électrons sont rapidement oxydés pour être éliminés (comme le montrent les études de toxicocinétique), alors que les aldéhydes aromatiques pauvres en électrons sont plus électrophiles, et vont de ce fait très rapidement réagir pour former des adduits avant d'être oxydés (Figure 5a). Cette tendance générale doit être balancée avec les capacités métaboliques de chaque espèce et la cinétique de métabolisation de chaque composé xénobiotique qui dépend aussi de sa diffusion à travers les membranes et des interactions au sein du site actif.
- Les méthacrylates sont métabolisés rapidement pour être transformé en pyruvate, intégré dans le cycle de Krebs (Bonnard *et al.*, 2013a) (Figure 9a), ce qui peut être une des pistes possibles pour expliquer pourquoi ils sont moins toxiques que les acrylates, aussi bien pour les mammifères (Bonnard *et al.*, 2013a, 2013b) que les organismes aquatiques (Schwöbel *et al.*, 2011). Les acrylates ne peuvent être transformés en pyruvates car il leur manque le méthyle indispensable pour obtenir ce métabolite (Figure 9b). Le métabolisme des méthacrylates supprime la double liaison C=C en α,β de l'ester dès la première étape, mais l'obtention du pyruvate ne se fait qu'après plusieurs autres étapes d'oxydation et une étape de déformylation. Cela fournit une explication plausible au fait que la toxicité des méthacrylates pour les poissons et daphnies est parfaitement prédite avec le modèle pour les esters simples (les méthacrylates font partie des points formant le modèle des esters en Figure 36). En effet, ils ont une réactivité similaire dans les expériences de copolymérisation, avec même une meilleure réactivité pour le méthacrylate (Shima and Kotera, 1963), donc l'hypothèse de la réactivité moindre des méthacrylates pour expliquer leur toxicité plus faible ne semble pas pertinente, à moins que les conditions biologiques ne favorisent la réactivité des acrylates par rapport aux méthacrylates.

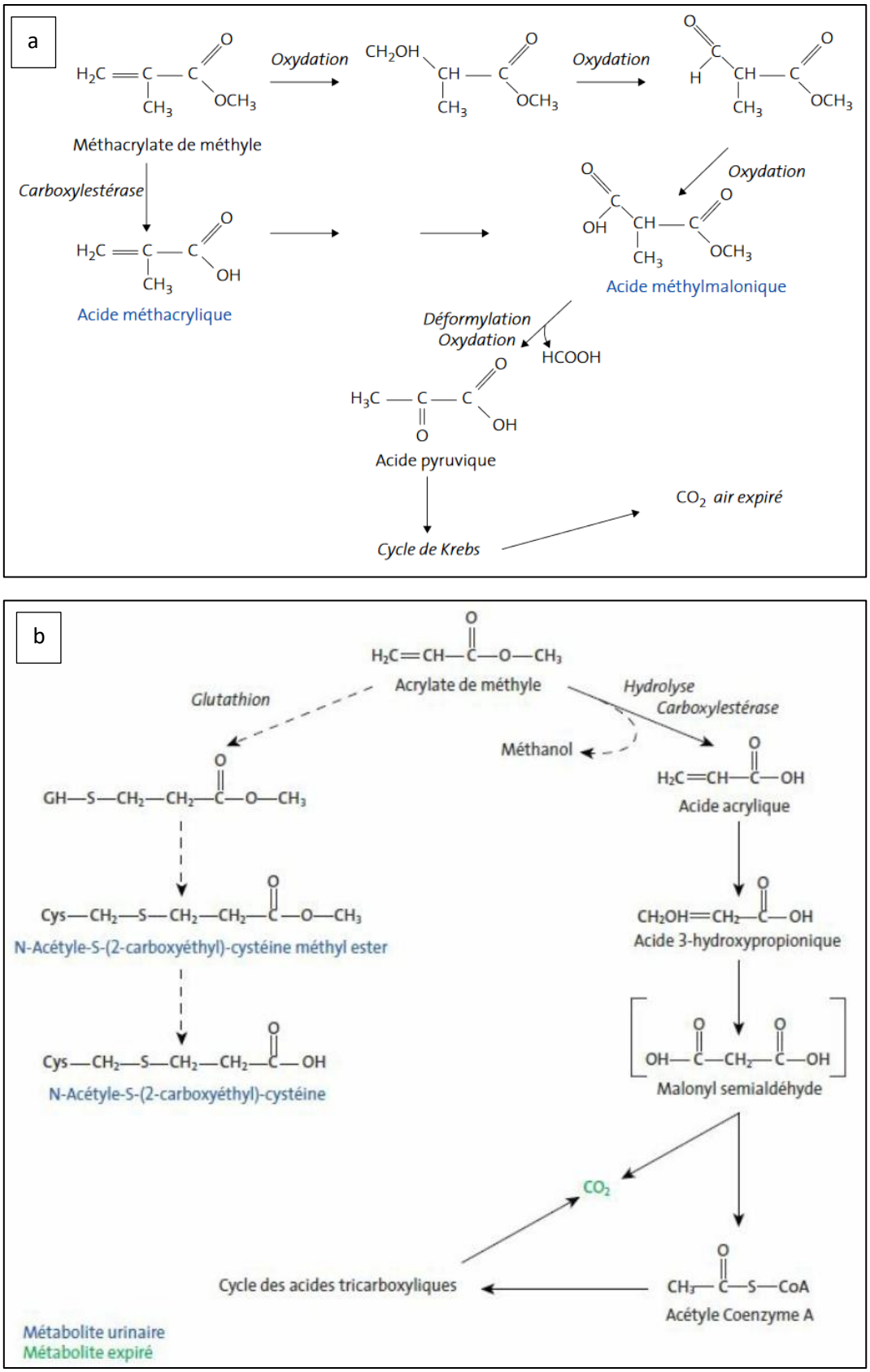


Figure 9 : Voie de métabolisation chez le rat a) du méthacrylate de méthyle (Tiré de (Bonnard et al., 2013a)), b) de l'acrylate de méthyle (Tiré de (Bonnard et al., 2013b)).

- L'acide lactique **7**, et les lactates qui subissent d'abord une étape d'hydrolyse, sont métabolisés de manière quantitative. En effet, en une étape d'oxydation, la lactate déshydrogénase

transforme le lactate en pyruvate, intégré dans le cycle de Krebs, un mécanisme existant dans pratiquement tous les organismes vivants (Adeva-Andany *et al.*, 2014; IPCS Inchem, 1980; Morotomi *et al.*, 1981).

D'autres composés sont bien sûr détoxifiés de manière similaire, de manière plus ou moins quantitative et rapide, mais seuls les principaux cas rencontrés et bien caractérisés ont été listés ici pour illustrer le propos.

b) Métabolisation vers un réactif électrophile

Lorsque les étapes d'oxydation d'une substance passent par un époxyde ou un aldéhyde, et que celui-ci n'est pas converti assez rapidement vers un composé moins toxique, cet état intermédiaire va causer des dommages à l'organisme en formant des adduits aux protéines et à l'ADN. C'est le cas de certains alcools, qui après oxydation en aldéhyde, génèrent un composé spécialement réactif, qui exercera cette réactivité pour former des adduits avant d'être transformé en acide. C'est le cas pour des alcools primaires α,β-insaturés (Koleva and Barzilov, 2010) et pour le 2,2,2-trifluoroéthanol (Airaksinen *et al.*, 1970) par exemple.

i) Métabolisme des hydrocarbures aromatiques

Le passage par un composé intermédiaire toxique, pas assez rapidement éliminé, concerne aussi les HAPs (Hydrocarbures Aromatiques Polycycliques), notamment le benzo[a]pyrène, dont les effets mutagènes sont très bien caractérisés, et ils proviennent d'un époxyde formé au cours des étapes d'oxydation, qui forme des adduits à l'ADN (Figure 10) (Trushin *et al.*, 2012).

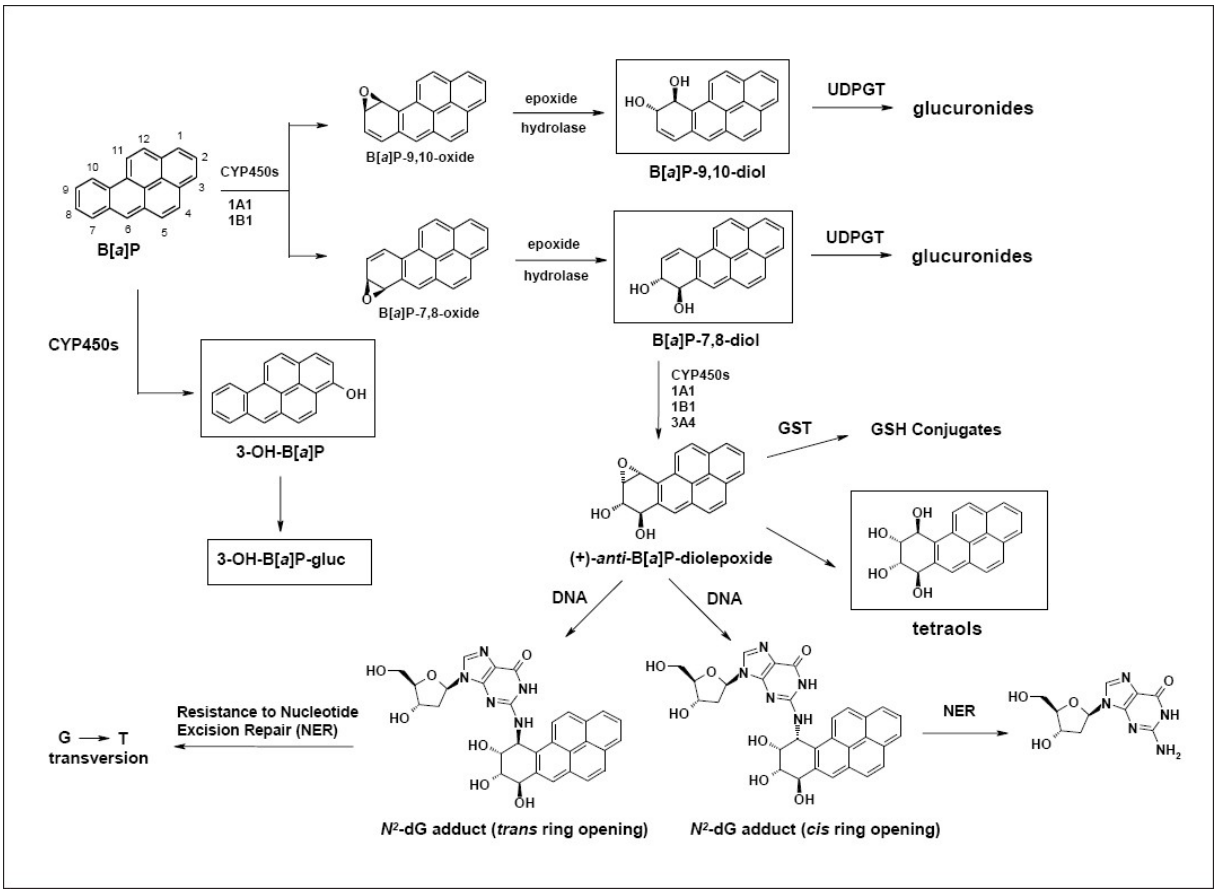


Figure 10 : métabolisme du benzo[a]pyrene (B[a]P) qui conduit à la formation d'adduits à l'ADN. Figure tirée de (Trushin et al., 2012).

En plus de ces époxydes, ces composés peuvent aussi être convertis en quinones. Les quinones vont aussi former des adduits (Buckpitt *et al.*, 2002; Tsuruda *et al.*, 1995), comme montré en Figure 8. C'est ce qui se passe aussi pour le benzène qui est métabolisé par les cytochromes CYP2E1 en hydroquinone, puis oxydé en p-benzoquinone par les myélopéroxydases dans la moëlle épinière. La toxicité de la quinone générée est alors source de leucémie (McHale *et al.*, 2012).

Des mécanismes similaires n'ont pas été identifiés à notre connaissance pour les organismes aquatiques, d'autant plus que leurs effets ne sont visibles qu'à l'échelle de tests de toxicité chronique, peut-être à cause du manque de CYP2E1 dans ces organismes (Ioannides, 1996).

ii) Métabolisme des amines et nitros aromatiques

D'autre part les amines aromatiques (*i.e.* un groupe aminé branché sur un cycle aromatique, principalement les anilines) sont oxydées par le métabolisme au niveau du foie. Si l'amine est primaire, il sera formé une hydroxylamine, puis un nitroso, composé très réactif, qui formera des adduits aux protéines et à l'ADN, qui est la source du caractère mutagène et cancérigène de ces substances, bien que les mécanismes de formation de cancers et d'adduits à l'ADN soient encore flous. On sait que le nitroso réagit avec les groupe thiols pour former des adduits à l'hémoglobine (IARC, 2010). Si l'amine est secondaire ou tertiaire, elle peut être d'abord convertie en amine primaire par une N-déalkylation. De même, les nitroaromatiques sont aussi convertis en nitroso, par une réduction métabolique grâce aux nitroréductases des bactéries intestinales anaérobies, et causent ainsi également des cancers (IARC, 2010; Rafii and Cerniglia, 1995) (Figure 11).

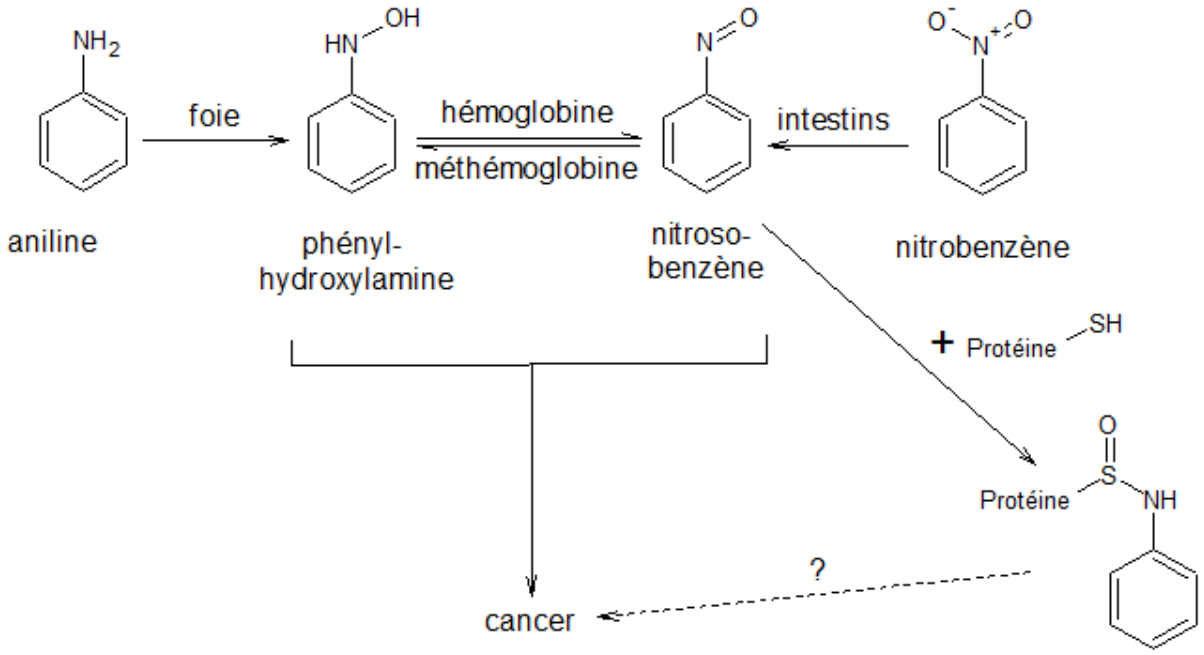


Figure 11 : formation de cancer par les anilines et nitrobenzènes.

iii) Métabolisme de l'hexane

L'hexane est un alcane particulier qui est métabolisé par les mammifères en hexane-2,5-dione (en passant par l'hexan-2-one, qui est donc aussi concernée par ce mécanisme), composé qui va réagir avec les résidus lysine des protéines pour former un pyrrole. Celui-ci va à nouveau être oxydé pour être attaqué par une deuxième lysine, pour former au final une réticulation des protéines (liaison supplémentaire entre 2 parties d'une même protéine, ou entre 2 protéines) (ATSDR, 1999) (Figure 12). Le métabolisme se passe dans le foie, et c'est dans les neurones que la dicétone réagit pour réticuler les protéines constituant les neurofilaments. Cela perturbe en conséquence le transport des neurotransmetteurs le long de l'axone, et c'est la cause de la neuropathie observée pour l'hexane (ATSDR, 1999).

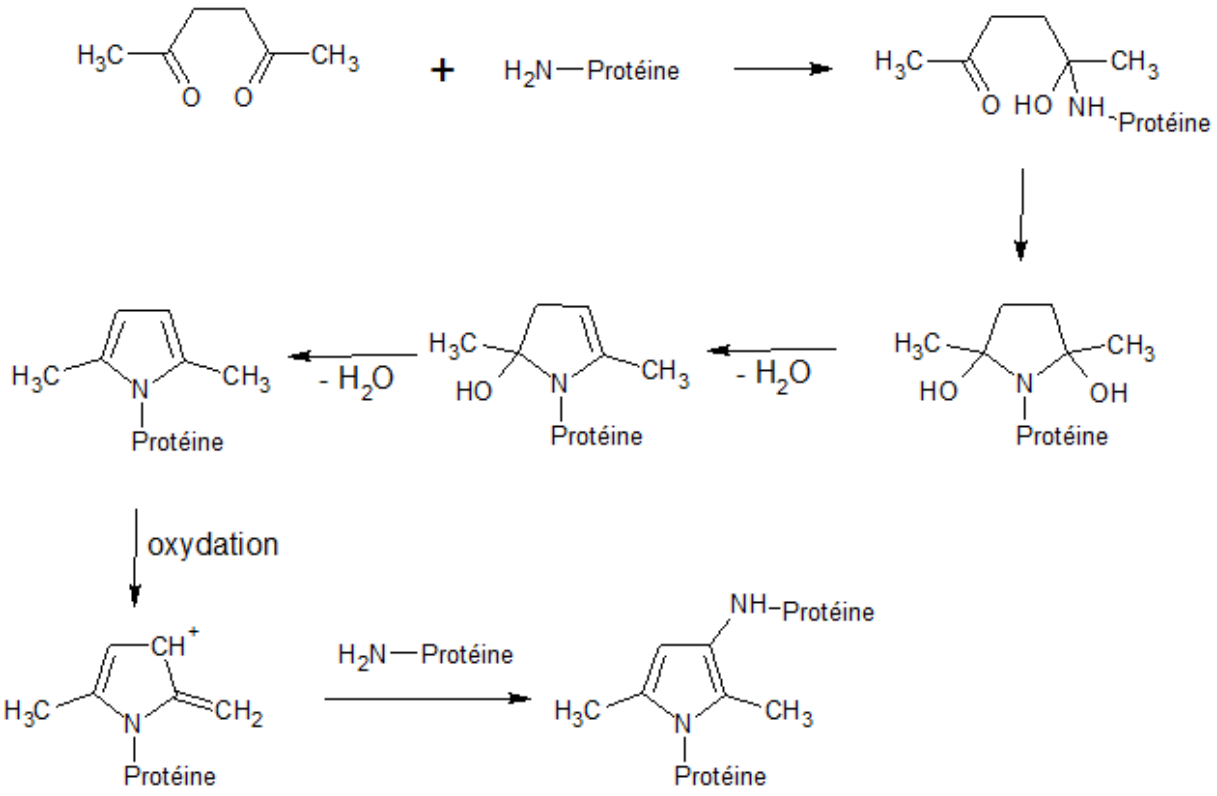


Figure 12 : réticulation des protéines par l'hexan-2,5-dione. Adapté de (ATSDR, 1999).

Par ailleurs, on rappelle que ce mécanisme fonctionne aussi avec d'autres dicétones, si elles répondent à certaines caractéristiques expliquées dans le paragraphe 3.a) Electrophiles durs. Par contre, le métabolisme de l'hexane en hexan-2,5-dione semble unique et la formation de γ-dicétone source de neurotoxicité n'a pas été observée pour d'autres alcanes.

iv) Métabolisme des furanes et thiophènes

Les cycles furanes et thiophènes, non substitués à la fois en position 2 et 5 sont oxydés, passent par un époxyde intermédiaire, puis sont transformés en dialdéhyde α,β-insaturé très réactif (Figure 13), qui formera donc des adduits aux protéines et à l'ADN, et qui sera responsable de la formation de cancer (Food and Drug Administration, 2004; Smith, 2011). On peut constater avec la Figure 13 que dans le cas des thiophènes, la formation d'adduits ne passe pas forcément par l'aldéhyde. Un mode d'action réactif a aussi été identifié en écotoxicologie pour les furanes et thiophènes, ce qui semble indiquer que ce mécanisme se passe aussi chez les organismes aquatiques (Ellison *et al.*, 2015).

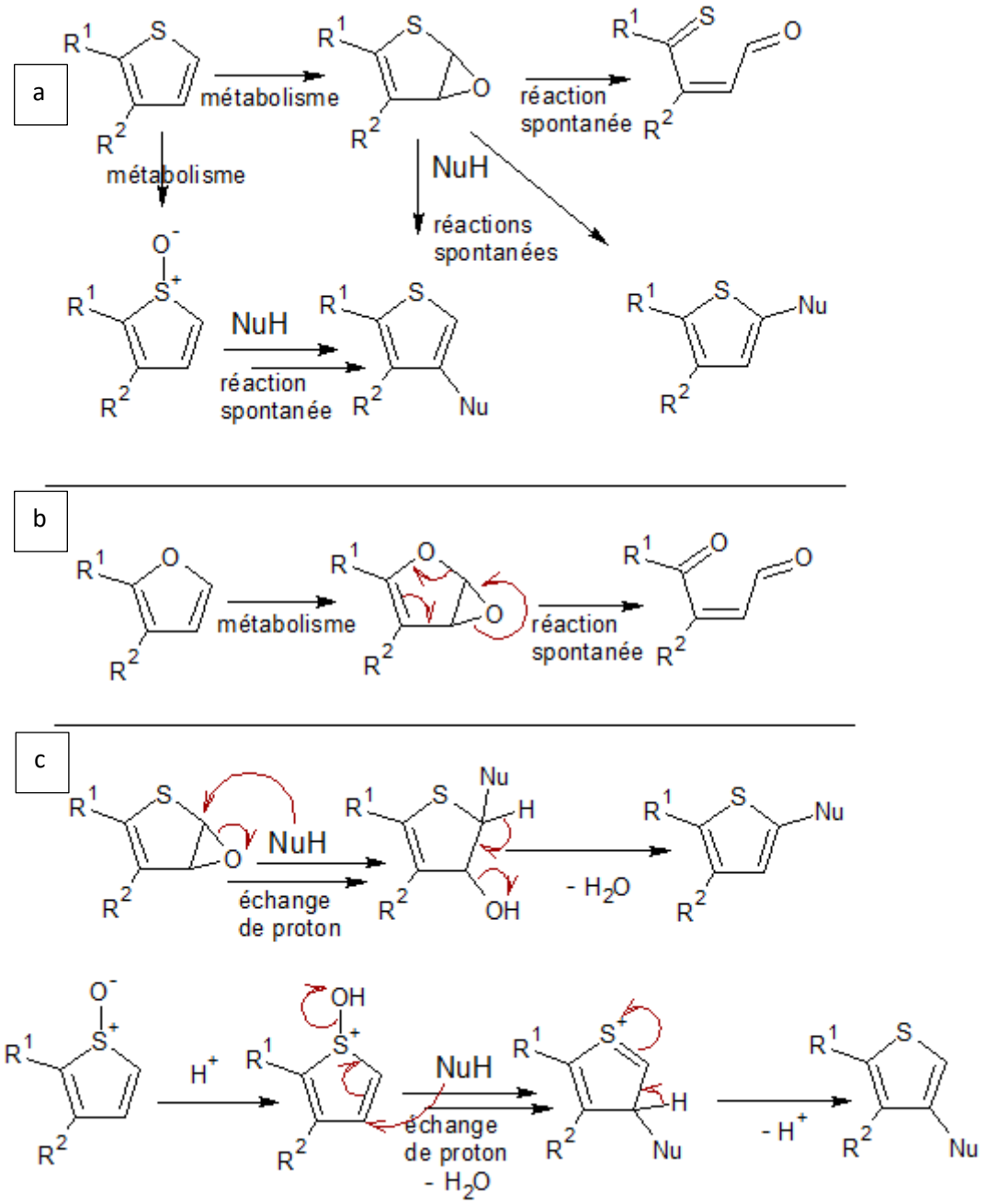


Figure 13 : métabolisme des thiophènes et furanes aboutissant à des aldéhydes réactifs. Figure adaptée de (Smith, 2011). a) métabolisme et réaction des thiophènes, b) des furanes. c) mécanismes de réaction des métabolites du thiophène avec un nucléophile.

v) Métabolisme des alkoxyvinyles

Ce type de structure, une double liaison C=C terminale branchée sur l'oxygène d'un éther ou ester, a été identifié comme une structure cause de mutagénicité, car le métabolisme génère un époxyde qui forme des adduits à l'ADN (Kazius *et al.*, 2005) (Figure 14).

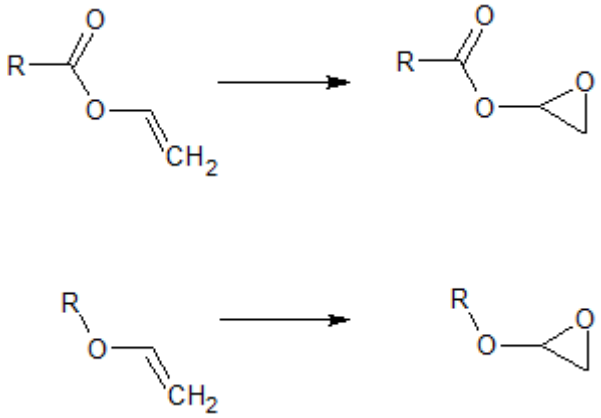


Figure 14 : métabolisme des alkoxyvinyles en époxydes

vi) Métabolisme de certains petits organohalogénés générant des réactifs

Le tétrachlorure de carbone (Recknagel *et al.*, 1989), le chloroforme (ATSDR, 1997a), le trichloroéthylène (Brüning and Bolt, 2000) sont tous métabolisés en composés très réactifs par les mammifères. Nous n'avons pas identifié ce type de mécanisme pour des organohalogénés plus grands, soit à cause d'études moins poussées pour les plus grandes molécules de ce type, soit parce que ces mécanismes sont hautement spécifiques pour les molécules détaillées ici. Il faut cependant noter que ces mécanismes toxiques sont balancés par une élimination rapide, du fait d'une volatilité élevée, et de l'hydrophilie des métabolites générés. Ces mécanismes sont présentés en Figure 15. Cependant, leur écotoxicité aigüe, qui s'aligne sur la valeur attendue pour la narcose non-polaire semble indiquer que ces mécanismes ne sont pas significatifs pour les organismes aquatiques.

Le mécanisme présenté pour le chloroforme est la voie de métabolisation majeure, aérobique. Cependant, la voie anaérobique mineure passe par un radical de manière similaire à ce qui est montré pour le tétrachlorure de carbone. Dans le cas du trichloroéthylène, l'activation par la GST donnant un métabolite réactif est une voie de métabolisation mineure (voie de droite sur le schéma). La voie majeure (voie de gauche) génère une molécule qui active des récepteurs spécifiques et sera discutée dans la section appropriée (4.e) Métabolisation en perturbateurs directs par docking).

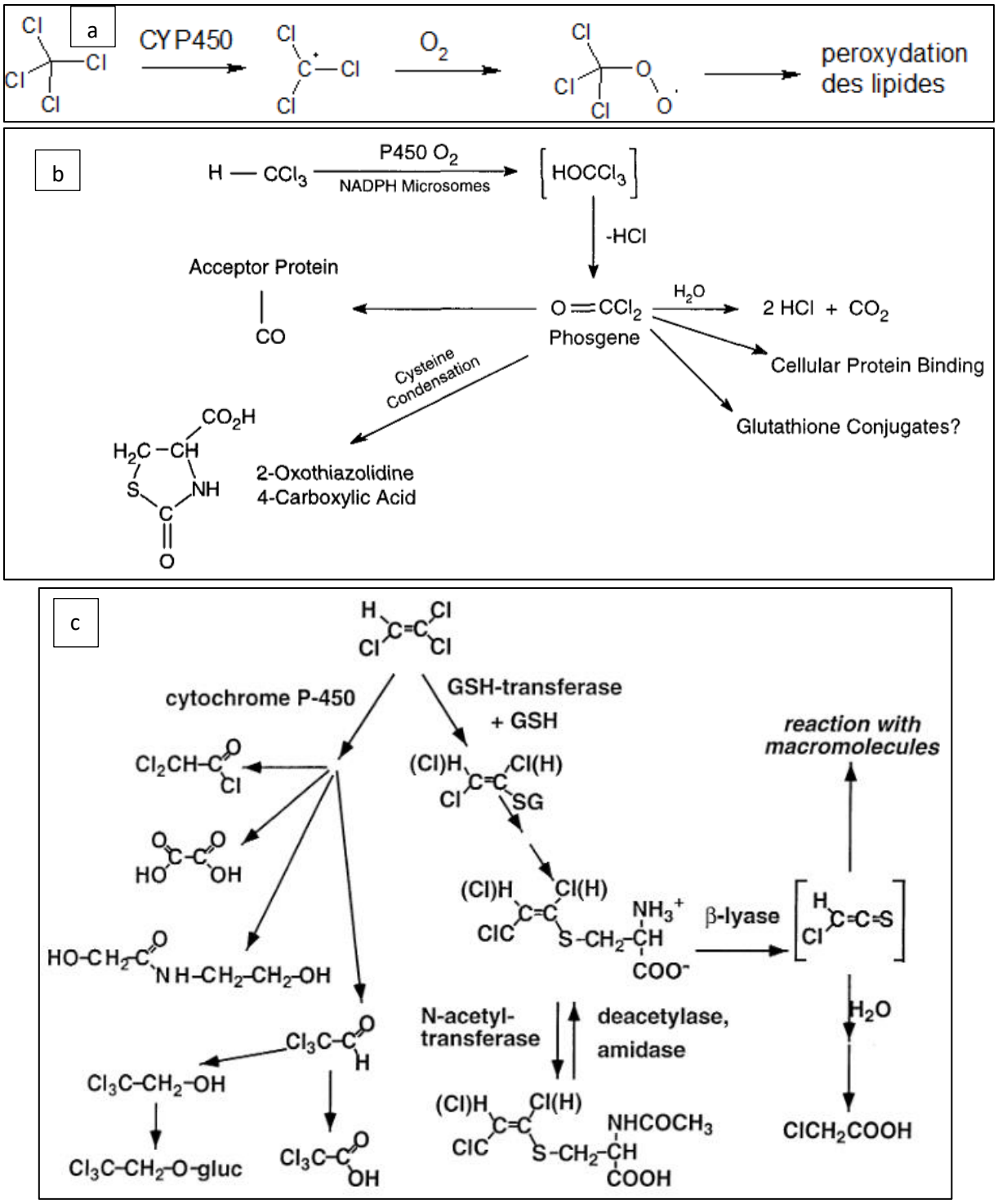


Figure 15 : métabolisme générant des réactifs du a) tétrachlorure de carbone b) chloroforme (Tiré de (ATSDR, 1997a)) c) trichloroéthylène (Tiré de (Brüning and Bolt, 2000)).

vii) Toxicité du but-3-yn-1-ol lors de sa métabolisation

Ce composé a été identifié comme réactif en écotoxicologie (Enoch *et al.*, 2008), et parallèlement, le risque toxique des alcynes en général a été identifié, comme des composés pouvant dégrader l'hème des cytochromes P450, après avoir été oxydés par ceux-ci en composés hautement réactifs (Smith, 2011), selon le mécanisme présenté en Figure 16.

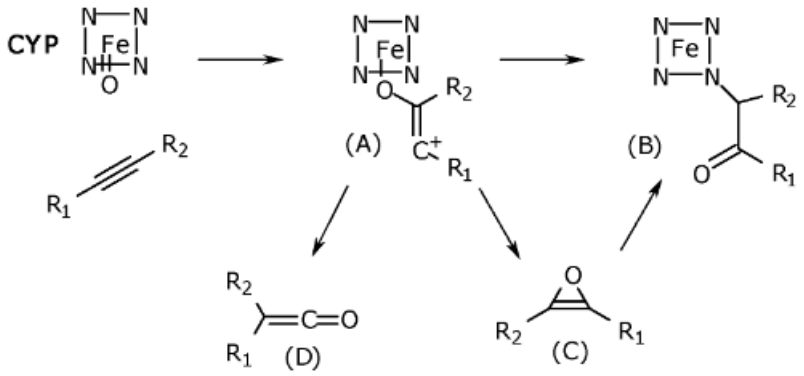


Figure 16 : mécanisme des alcynes dégradant les CYP450. Figure tirée de (Smith, 2011).

Peu d'informations sont disponibles sur la diversité des alcynes pouvant participer à ce mécanisme, et l'importance que celui-ci revêt lors de tests de toxicité aigüe ou chronique, sur les mammifères ou les poissons ou d'autres espèces.

viii) Métabolisme des hydrazines et composés azo

Les hydrazines (composés possédant une liaison simple N-N) sont métabolisées de plusieurs manières possibles selon leur structure, générant de manière générale des radicaux libres, créant un stress oxydatif. Des diazoniums sont aussi générés (ATSDR, 1997b). Ce sont des composés électrophiles très réactifs formant des adduits à l'ADN (Brown and Vito, 1993).

Les composés azo, dont la majorité des représentants sont des colorants, où la liaison azo est branchée sur des cycles aromatiques, sont surtout métabolisés par les bactéries intestinales anaérobies, mais aussi parfois par les enzymes du foie de mammifères, qui vont rompre la liaison N=N, pour générer deux anilines, qui suivront le mécanisme classique des anilines par la suite (voir la section 4.b.ii) Métabolisme des amines et nitros aromatiques). Un autre mécanisme possible est la métabolisation en diazonium (Brown and Vito, 1993).

ix) Toxicité des benzènediols et éthers dérivés

L'hydroquinone (benzène-1,4-diol) et le catéchol (benzène-1,2-diol), ainsi que leurs dérivés substitués, sont oxydés vers les quinones correspondantes (p-benzoquinone et o-benzoquinone respectivement), qui peuvent former des adduits comme expliqué dans le paragraphe 3.b) Electrophiles mous. Ces composés s'engagent également dans un cycle RedOx, mécanisme expliqué dans la section suivante.

Si l'un des hydroxyles est méthylé sur la molécule de départ, pour en faire un éther, mais que l'autre hydroxyle reste intact, les mammifères restent capables de les déméthyler pour retomber sur le diol. Apparemment, les organismes aquatiques n'en sont pas capables, car les alkoxyphénols étudiés dans la littérature ont été identifiés comme narcotiques polaires, basé sur les symptômes comportementaux et les valeurs de LC50 ou EC50 sur les poissons et sur les protozoaires (*T. pyriformis*) (Enoch *et al.*, 2008; Schultz, 1987).

Dans le cas particulier du 3-méthoxyphénol, qui n'est pas simplement un éther d'hydroquinone ou de catéchol (car les deux groupes oxygénés ne sont pas situés en position *ortho* ou *para*, mais en *méta* sur le cycle aromatique), le métabolisme hépatique chez les mammifères parvient quand-même à le

transformer en catéchol substitué, grâce à une hydroxylation en position 6, vraisemblablement effectuée par le CYP 2E1 (Moridani *et al.*, 2003) (Figure 17). Cette enzyme ne semble pas partagée par tous les animaux non mammifères, ce qui peut expliquer la différence de toxicité observée entre les mammifères et les poissons (Ioannides, 1996).

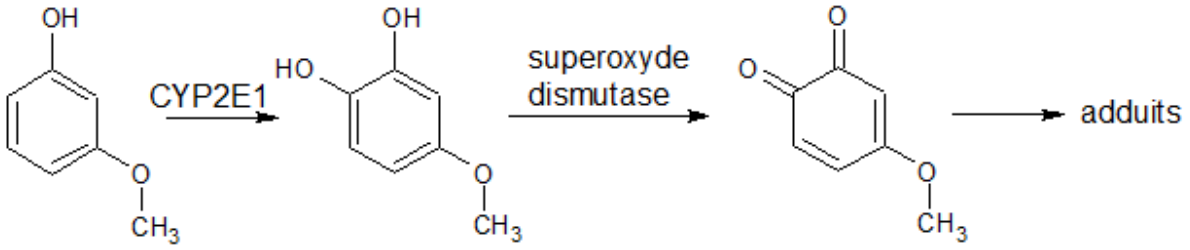


Figure 17 : métabolisme du 3-méthoxyphénol vers une quinone.

c) Utilisation d'un cycle RedOx

Un autre mécanisme, qui est aussi important pour expliquer la toxicité, intervient dans le cadre des benzène-1,2- et -1,4-diols et des quinones. En effet, les diols sont oxydés en deux étapes vers les quinones, en passant par une semi-quinone, un radical source de stress oxydant. Cette oxydation peut se passer lentement en présence de dioxygène de manière spontanée, mais elle est catalysée par la superoxyde dismustase. Cependant, la réaction est aussi effectuée dans l'autre sens par une autre enzyme (NADPH-CYP450 réductase principalement). Ainsi, ces composés peuvent être interconvertis de manière cyclique entre les 3 formes (hydroquinone, semiquinone, quinone), avec une génération d'anion superoxyde à chaque oxydation, et en épuisant les stocks de cofacteur NAD(P)H à chaque réduction (Bolton *et al.*, 2000; Di Francesco *et al.*, 2004).

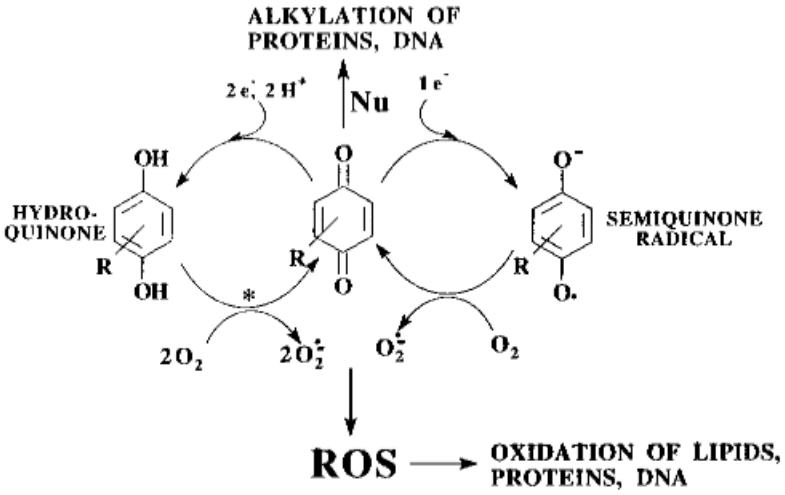


Figure 18 : cycle RedOx des quinones. Figure tirée de (Bolton et al., 2000).

Les diols aromatiques et les quinones ne sont pas le seul type de substance à faire des cycles RedOx. C'est aussi le cas du paraquat (1,1'–diméthyl-4,4'-bipyridinium, Figure 19) et de ses analogues de type bipyridinium ou phénylpyridinium. Ce mécanisme d'action est en effet ce qui leur confère leur propriété d'herbicide, mais qui les rend également toxiques pour les animaux. Leur mécanisme est représenté en Figure 19.

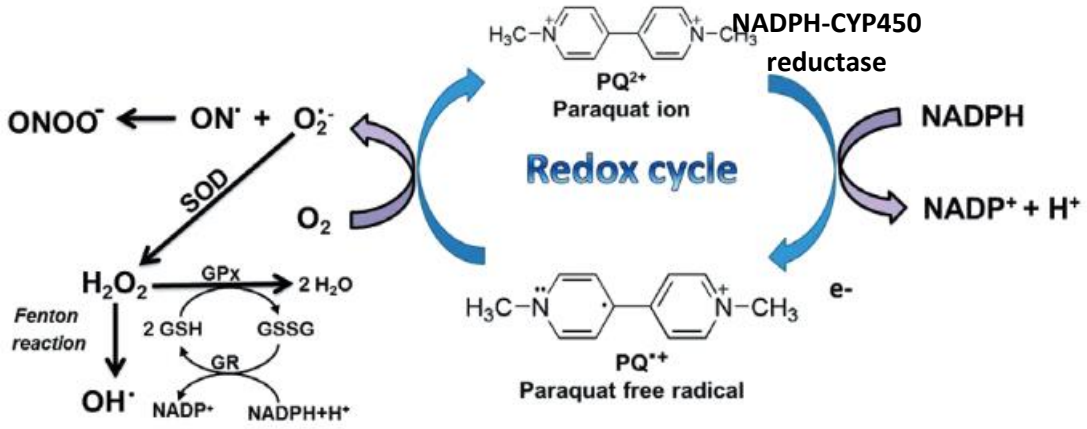


Figure 19 : cycle RedOx du paraquat. Figure tirée de (Blanco-Ayala et al., 2014).

La roténone **8** (1,2,12,12a-tetrahydro-8,9-dimethoxy-2-(1-methylethenyl-(1) benzopyrano (2,4-b) furo (2,3-h) (1)-benzopyran-6 (6H)-one), utilisée en tant que pesticide, utilise aussi un cycle RedOx. Elle vient se fixer au complexe I de la chaîne de transport d'électrons au sein de la membrane mitochondriale interne, et elle subit une réduction de 1 électron. Le radical formé réagira ensuite avec le dioxygène pour reprendre sa forme oxydée initiale, alors que le dioxygène est transformé en anion superoxyde. Ainsi, la roténone, ayant repris sa forme initiale, pourra recommencer cette séquence, et générer un stress oxydatif important à cause de la production continue de superoxyde, en plus du problème de l'inhibition du transport d'électrons au niveau du complexe I, empêchant la production d'ATP, l'énergie des cellules (Sherer *et al.*, 2007).

Enfin, il y a aussi un cycle RedOx qui se met en place entre les thiols et les disulfures, de manière similaire au glutathion **5**. En effet, les thiols, sous leur forme anionique, sont oxydés en radical thiyle et rentrent ensuite dans un cycle RedOx entre le radical thiyle et le disulfure (Munday, 1989). Ce mécanisme dépend des propriétés stéréo-électroniques du thiol. En effet, il a été montré que les thiols tertiaires ne réagissent pas de cette manière à cause de l'encombrement stérique. D'autre part, le branchement du thiol sur un cycle aromatique modifie significativement la réactivité, car la conjugaison des électrons entre le thiol et le cycle aromatique stabilise les états intermédiaires. Si le cycle aromatique porte des groupes attracteurs d'électrons (comme les halogènes, les nitros), celui-ci va stabiliser l'anion initial et ainsi diminuer la réactivité. Au contraire, si le cycle porte des substituants donneurs d'électrons (comme les alkoxys), celui-ci va stabiliser le radical et favoriser ainsi l'oxydation. Pour le cas particulier de ces composés, le glutathion, la GST et la GPx ne sont pas détoxifiants, mais aident au contraire la réactivité et facilitent le cycle RedOx (Munday, 1989).

d) Métabolisation en perturbateur indirect d'enzyme

Ce mécanisme concerne les substances qui vont être transformées par le métabolisme en molécules perturbant des enzymes de manière indirecte. Ce mécanisme de perturbation indirecte d'enzymes est détaillé dans la section 5) Perturbation de l'environnement enzym. Ce mécanisme de métabolisation toxifiante pourrait *a priori* concerner de nombreuses molécules. Cependant, parmi les substances que nous avons étudiées dans le cadre de ce travail, seul l'éthanol passe par ce mécanisme, et ce n'est pas son mécanisme principal (la narcose). Ce qui a été mis en évidence avec l'éthanol qui est très étudié fonctionne probablement aussi pour les autres alcools, avec des degrés variables selon la structure. L'éthanol est oxydé en acétaldéhyde, puis en acide acétique. Les enzymes qui opèrent ces transformations (alcool déshydrogénase et aldéhyde déshydrogénase) consomment du NAD+ pour le réduire en NADH. Cette consommation va dans le même sens que le cycle de Krebs, et la consommation inverse (NADH oxydé en NAD+) est opérée par la phosphorylation oxydative. Par

conséquent, lors d'une exposition importante et répétée à l'éthanol, le taux de glycolyse et d'oxydation des acides gras est diminué par manque de NAD+, et moins d'énergie est produite, alors qu'il y a une compensation de la chaîne de transport d'électrons (qui sert à la phosphorylation oxydative en créant le gradient de protons), qui consomme beaucoup de dioxygène. En conséquence, à terme, il y a une accumulation d'acides gras dans le foie (c'est la stéatose hépatique, autrement appelée syndrôme du foie gras, qui évolue ensuite en cirrhose du foie) et dans le sang, car ils ne sont pas suffisamment utilisés pour alimenter le cycle de Krebs (King, 2015a). Ainsi, on obtient bien une perturbation indirecte de tout un système d'enzymes, car le cofacteur qu'elles utilisent (NAD+ ou NADH) est en trop faible ou trop forte quantité, rompant ainsi l'équilibre nécessaire au bon fonctionnement de l'organisme. De plus, la consommation excessive d'alcool entraîne une production accrue d'enzymes métaboliques dont le CYP2E1, qui arrive également à oxyder l'éthanol en acétaldéhyde. Cependant la concentration anormalement élevée de ce cytochrome génère une quantité non négligeable d'espèces réactives de l'oxygène (ROS), et de peroxydation des lipides. Ceci, ajouté aux adduits formés par l'acétaldéhyde, explique la formation de cancers du foie causés par l'alcoolisme (King, 2015a).

e) Métabolisation en perturbateurs directs par docking

Ce mécanisme concerne les substances qui vont être transformées par le métabolisme en molécules perturbant les enzymes, récepteurs, canaux à ions, en se liant directement dans un site de liaison. En se liant, ces molécules modulent l'action de la protéine liée (inhibition, activation).

Le trichloroéthylène, comme les autres petits organohalogénés, est métabolisé en molécules fortement toxiques chez les mammifères. La voie de métabolisation majoritaire est une oxydation par les cytochromes P450 en chloral (un aldéhyde) puis une réduction en 2,2,2-trichloroéthanol (Brüning and Bolt, 2000) (Figure 15c). Le mécanisme d'action du 2,2,2-trichloroéthanol sera détaillé dans la section 6.f) Modulation des canaux à ions. Cette molécule se lie aux récepteurs à GABA **9** (acide gamma-aminobutyrique) qui contrôlent des canaux à ions chlorure, servant à contrôler le potentiel électrique des neurones.

De même, le dichlorométhane est aussi oxydé par le métabolisme des mammifères, pour générer du monoxyde de carbone (ATSDR, 2014). Le mécanisme du monoxyde de carbone sera détaillé en section 6.g) Autres mécanismes. Le monoxyde de carbone entre en compétition avec l'oxygène pour se lier aux hèmes et empêche ainsi que l'oxygène soit délivré aux différentes cellules de l'organisme en quantité suffisante.

D'autre part, les nitriles dont le carbone en alpha est lié à au moins un hydrogène sont hydroxylés par les CYP450 au niveau de ce carbone alpha, pour générer une cyanohydrine **10**. Le composé ainsi obtenu est réactif et libère spontanément l'anion cyanure d'une part, et un aldéhyde ou une cétone d'autre part (Grogan *et al.*, 1992). L'anion cyanure inhibe la cytochrome c oxydase, le 4e complexe de la chaîne de transport d'électrons des mitochondries, découplant ainsi cette chaîne de transport d'électrons de la synthèse d'ATP, supprimant l'énergie de la cellule, tout en générant un stress oxydatif (Way *et al.*, 2007).

Enfin, il y a le cas particulier de la diphénylamine, qui est métabolisée en petites proportions par certains organismes (des mammifères) en indophénol **11** (IPCS Inchem, 1998). Ce dernier est potentiellement un inhibiteur du transport d'électrons dans la mitochondrie. En effet, en écotoxicologie, il a été identifié un mode d'action de découpleur de la phosphorylation oxydative pour la diphénylamine (Enoch *et al.*, 2008), bien que la structure du parent ne permette pas d'expliquer un tel effet. Il n'a pas été identifié de substances similaires suivant le même mécanisme, à notre connaissance.

5) Perturbation de l'environnement enzymatique

Nous arrivons maintenant aux catégories de substances qui perturbent le fonctionnement d'une enzyme (ou autre protéine active, comme les récepteurs ou les canaux à ions), ou d'un système d'enzymes en particulier, par une interaction spécifique à ces enzymes. Dans cette section, on traitera des mécanismes de perturbation indirecte. C'est-à-dire que les substances toxiques ne vont pas se lier dans un site de liaison de la protéine affectée, mais vont plutôt modifier son environnement. Cette modification de l'environnement de la protéine modifiera le fonctionnement de l'enzyme. Les conditions de fonctionnement des enzymes qui peuvent ainsi être perturbées sont le pH, le potentiel électrique, la présence de cofacteurs sous leur forme oxydée ou réduite, etc.

a) Découplage de la phosphorylation oxydative par transport de protons

Ce mécanisme s'attaque au système d'enzymes intégrées dans la membrane interne de la mitochondrie. Il existe un gradient de protons à travers cette membrane, créé par un transport de protons (H^+) par les complexes de la chaîne de transport d'électrons depuis la matrice interne vers l'espace entre la membrane interne et la membrane externe. Ce gradient de protons est indispensable pour le fonctionnement de l'ATP-synthase, qui catalyse la réaction ADP + Phosphate → ATP, soit la fixation d'un phosphate supplémentaire à l'adénosine diphosphate, en transportant un proton vers la matrice interne. Cette liaison entre le deuxième et le troisième phosphate attachés à l'adénosine est très haute en énergie (environ 12 kcal/mol). Lorsque cette liaison sera rompue, elle libérera une quantité importante d'énergie, qui pourra être utilisée pour effectuer un autre processus non spontané. L'ATP est donc un cofacteur pour un bon nombre d'enzymes, qui apporte l'énergie nécessaire à la catalyse de la réaction gérée par l'enzyme. C'est une molécule essentielle pour le fonctionnement de tout organisme. Si la production de l'ATP est stoppée, l'organisme ne peut plus survivre. La chaîne de transport d'électrons, qui va oxyder l'oxygène moléculaire en eau, et la synthèse d'ATP sont couplées, c'est-à-dire que la synthèse de l'ATP ne fonctionne pas sans cette chaîne de transport d'électrons, qui produit le gradient de protons (Berg *et al.*, 2002) (Figure 20).

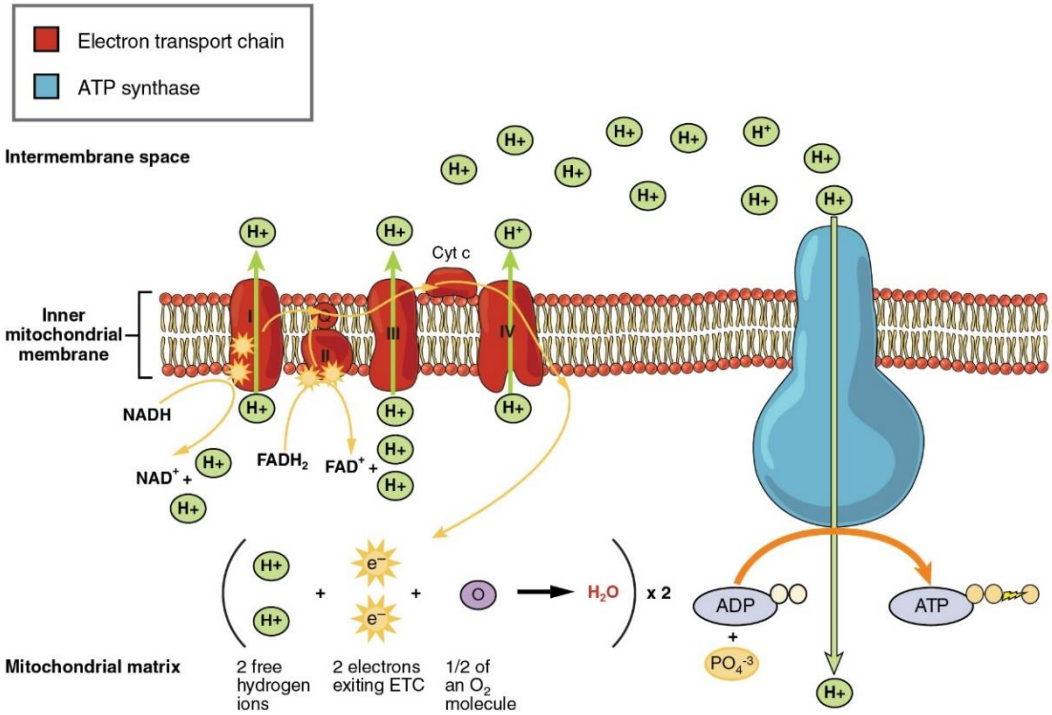


Figure 20 : Couplage de l'ATP synthase et de la chaîne de transport d'électrons par le gradient de protons. Tiré de Anatomy & Physiology (OpenStax, 2014).

Cependant, certaines substances ont la capacité de transporter les protons librement à travers la membrane mitochondriale interne, car ces substances la traversent aussi bien sous forme protonée (forme acide de la substance) dans un sens que sous forme non protonée (forme basique) dans l'autre sens. Cela a pour effet de supprimer le gradient de protons, et donc d'arrêter la synthèse d'ATP. Ce processus lutte alors contre la chaîne de transport d'électrons qui continue à pomper les protons hors de la matrice interne. Cette dernière fonctionne donc toujours, mais plus l'ATP-synthase. C'est ce qu'on appelle le découplage de la phosphorylation oxydative (Escher et al., 1999). Les substances capables de faire cette navette à protons sont les phénols suffisamment acides. Un phénol est acide lorsqu'il porte des groupes attracteurs d'électrons, appauvrissant l'oxygène qui libérera donc plus facilement son proton, en formant un phénolate, dont la charge négative est justement stabilisée par ces groupes attracteurs d'électrons.

Dans des conditions normales, le pH de la matrice interne de la mitochondrie est d'environ 8, alors que l'espace intermembranaire est de 6,6 (Berg *et al.*, 2002; Casey *et al.*, 2010). Ainsi, il paraît raisonnable de penser que le pKa optimal pour un phénol qui va traverser la membrane sous sa forme protonée et non protonée serait dans cette même gamme, pour que les deux formes coexistent en proportions suffisamment importantes. D'autres auteurs affirment plutôt que ce mécanisme fonctionne efficacement si le pKa est inférieur à 6,5 (Schultz, 1987). Le passage de la membrane sous la forme anionique, donc non protonée, est le passage difficile, à cause de la charge négative qui ne favorise pas l'entrée de la molécule dans la couche lipidique. Cependant, pour certaines substances, il a pu être mis en évidence qu'un dimère était formé entre l'anion et la molécule neutre, afin de répartir la charge négative, permettant ainsi un passage efficace de l'anion par diffusion passive (Figure 21). Ce dimère a tendance à être formé par les phénols non substitués en position *ortho*, alors que les phénols substitués en *ortho* passeront plus facilement la membrane sans former de dimère, car le substituant en *ortho* cache partiellement la charge négative de l'oxygène (Escher *et al.*, 1999). Finalement, pour obtenir un potentiel de découplage significatif, plusieurs facteurs interviennent :

- l'hydrophobie (pour le passage dans la membrane de toutes les formes),
- le pKa (pour la facilité de captage et relargage de proton à ces pH),
- l'encombrement stérique (pour cacher la charge de l'anion, ou à défaut faciliter la formation de dimère).

L'ensemble de ces facteurs implique qu'il est très difficile de quantifier l'effet découpleur des phénols, et donc leur toxicité.

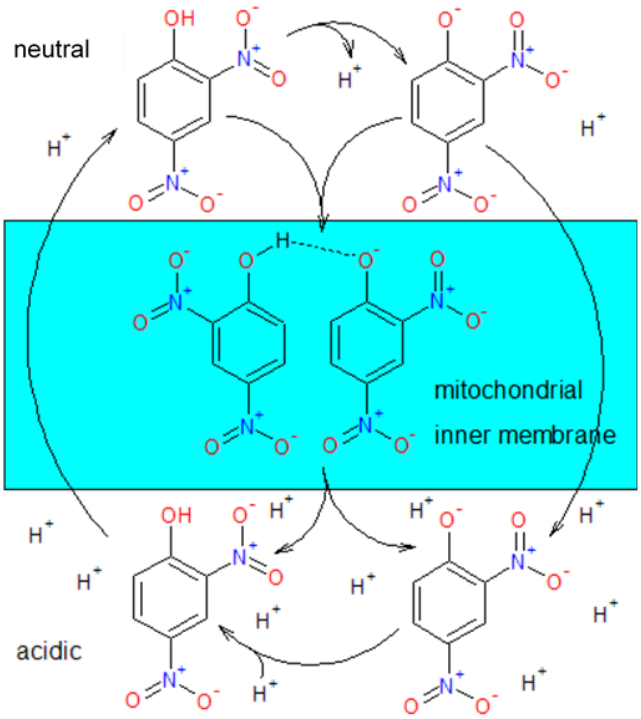


Figure 21 : transport de protons à travers la membrane mitochondriale interne par les phénols acides.

Certaines substances sont reconnues comme interagissant directement dans un site de liaison de l'un des complexes de la chaîne de transport d'électrons, ce qui a également pour effet de perturber le processus de phosphorylation oxydative. Cependant, dans ce cas, c'est une interaction directe, et cela sera donc traité dans la partie suivante 6) Interaction directe dans un site de liaison.

b) Acidification ou basification des cellules

Ce mécanisme est celui qui intervient avec les acides et les bases suffisamment forts, c'est-à-dire un acide avec un pKa < 5 ou une base avec le pKa de l'acide conjugué > 9. Leur mécanisme est discuté dans le chapitre suivant.

6) Interaction directe dans un site de liaison

Ce dernier mécanisme général regroupe certainement le plus grand nombre de sous-groupes. En effet, cela regroupe tous les mécanismes d'interaction directe avec une enzyme, un récepteur, un canal à ions, qui seront regroupés sous le terme de protéines actives. Parmi ces mécanismes, les mieux caractérisés, et ceux pour lesquels on connaît le plus grand nombre de molécules actives sont les suivants : Inhibition de l'acétylcholinestérase, liaison aux récepteurs nicotiniques ou muscariniques, liaison aux canaux à ions (calcium, sodium, chlorure, etc.), liaison aux récepteurs opioïdes, liaison aux hèmes, liaison aux récepteurs d'hormone (estrogène, androgène, thyroïde), etc. La liaison de la molécule toxique peut se faire dans le site actif d'une enzyme (cas de l'acétylcholinestérase), ou un site de liaison allostérique pour une enzyme, un canal à ions ou un récepteur, ou encore au niveau du site de liaison principal d'un récepteur.

a) Inhibition de l'AcétylCholinEstérase

L'acétylcholinestérase (AChE) est une enzyme essentielle pour la transmission nerveuse. Cette enzyme hydrolyse l'acétylcholine (ACh, Figure 25) en choline + acétate.

L'acétylcholine est un transmetteur nerveux, qui sert à conduire le signal entre un premier neurone (présynaptique) et un second neurone ou une autre cellule, musculaire par exemple (postsynaptique),

au niveau du synapse, c'est-à-dire la jonction entre les deux cellules. Pour cela, l'ACh est relâchée au niveau de la terminaison nerveuse, puis se lie aux récepteurs de l'acétylcholine (AChR) de la membrane postsynaptique. Ceci va déclencher l'ouverture de canaux à ions (le récepteur lui-même dans le cas des récepteurs nicotiniques, ou un canal à ion distinct du récepteur dans le cas des récepteurs muscariniques), qui vont changer la polarisation de la cellule postsynaptique et initier ainsi la conduction du signal électrique à travers celle-ci. Une fois que cette transmission du signal électrique a été effectuée, l'action de l'ACh doit être stoppée, afin que les canaux à ions soient refermés et que la cellule postsynaptique repasse à l'état de repos. C'est là qu'intervient l'enzyme AChE, qui dégrade le transmetteur nerveux. La choline, produite par l'hydrolyse, est ensuite recyclée en étant transportée à nouveau dans le premier neurone, pour y être retransformée en acétylcholine pour la prochaine transmission nerveuse (King, 2015b). Ces mécanismes sont représentés en Figure 22.

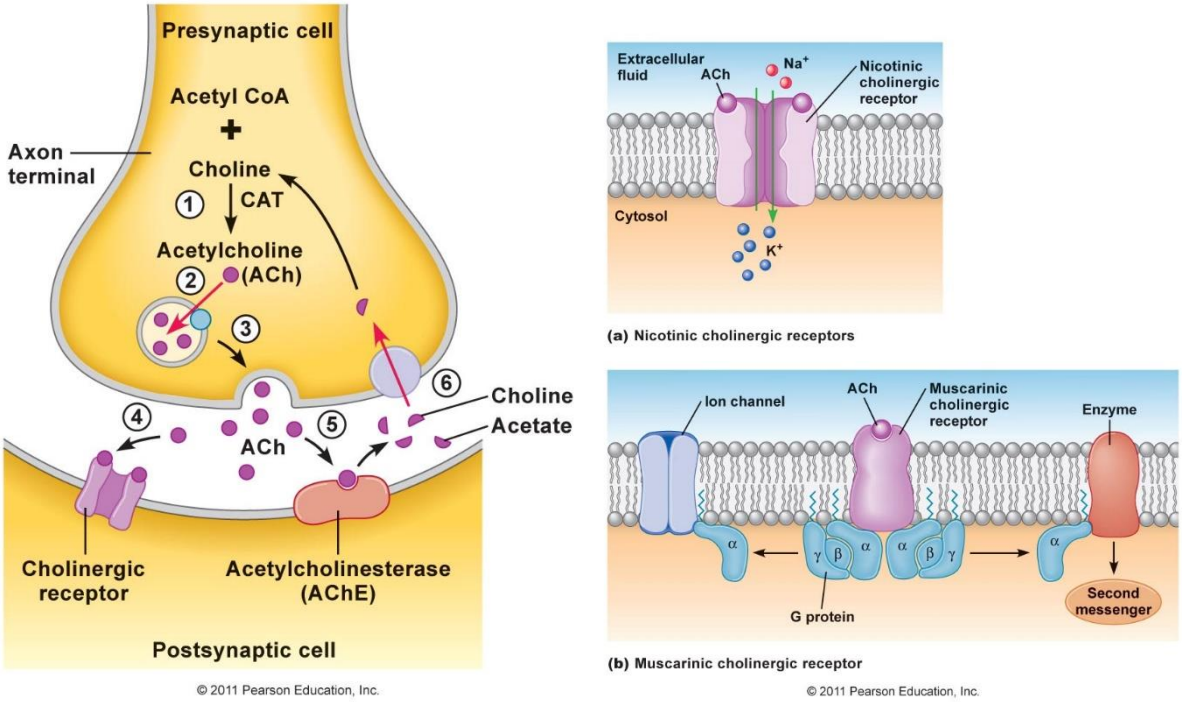


Figure 22 : transmission nerveuse. Figure tirée de (Droual, 2011).

Ainsi, l'AChE est une enzyme essentielle, et son inhibition, entraîne des effets mortels rapides. C'est ce mécanisme qui est utilisé par les agents innervants utilisés comme arme chimique, tels le VX, le sarin, le tabun. C'est aussi ce mécanisme qui est utilisé par de nombreux insecticides, comme l'azinphos-méthyl **12**, le méthomyl **13**, l'aldicarbe **14**, le malathion **15**. La mort est provoquée par l'arrêt du fonctionnement des muscles respiratoires (Elersek and Filipic, 2011; Fukuto, 1990).

Au niveau moléculaire, l'AChE hydrolyse l'ACh en plusieurs étapes : liaison de l'ACh à la sérine du site actif de l'AChE ; élimination de la choline ; élimination de l'acétate. Ces étapes sont représentées en Figure 23 (Fukuto, 1990).

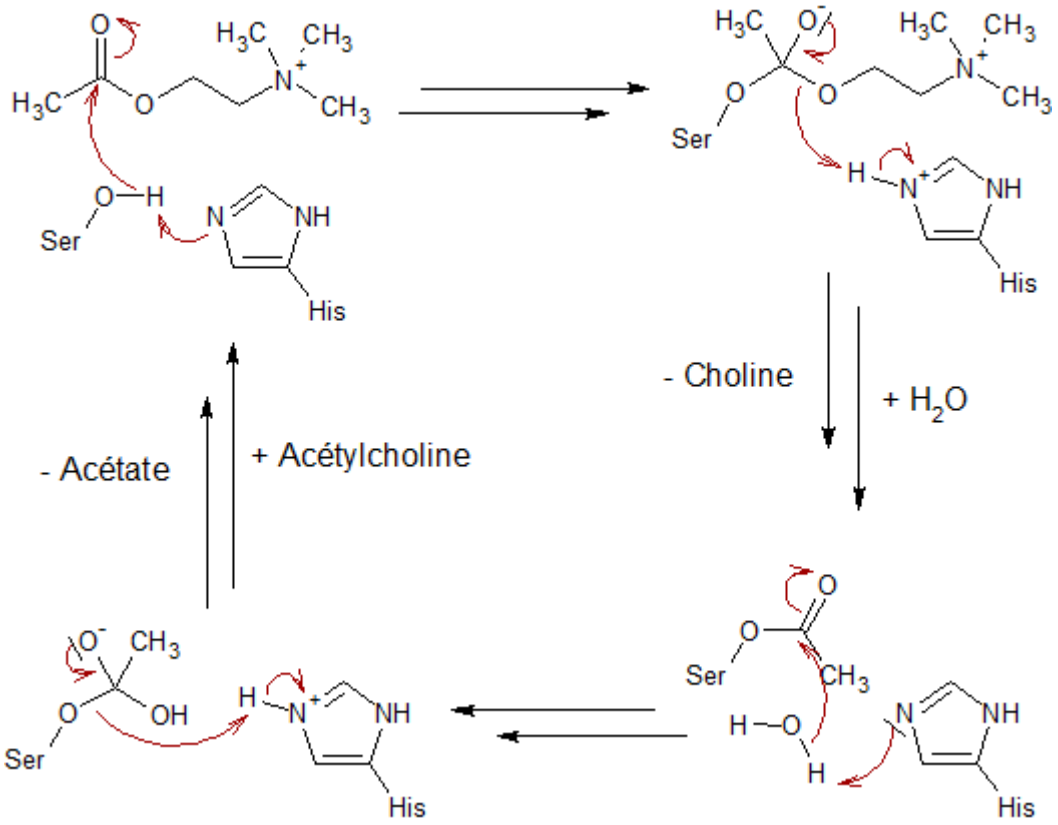


Figure 23 : mécanisme de l'hydrolyse de l'acétylcholine par l'acétylcholinestérase. Les doubles flèches de réaction indiquent que la réaction se passe en plusieurs étapes. Les étapes d'échange de proton ont été groupées avec les autres étapes pour simplifier le schéma.

L'inhibition de cette enzyme se fait en imitant le ligand naturel, l'acétylcholine, pour se lier au groupement sérine (acide aminé avec un groupement alcool sur sa chaîne latérale), comme dans la première étape de l'hydrolyse. Un groupe partant est éliminé comme dans la 2e étape. Cependant, l'élimination du groupe restant attaché à la sérine se passe très lentement, voire pas du tout, contrairement au cas de l'acétylcholine (Elersek and Filipic, 2011) (Figure 24).

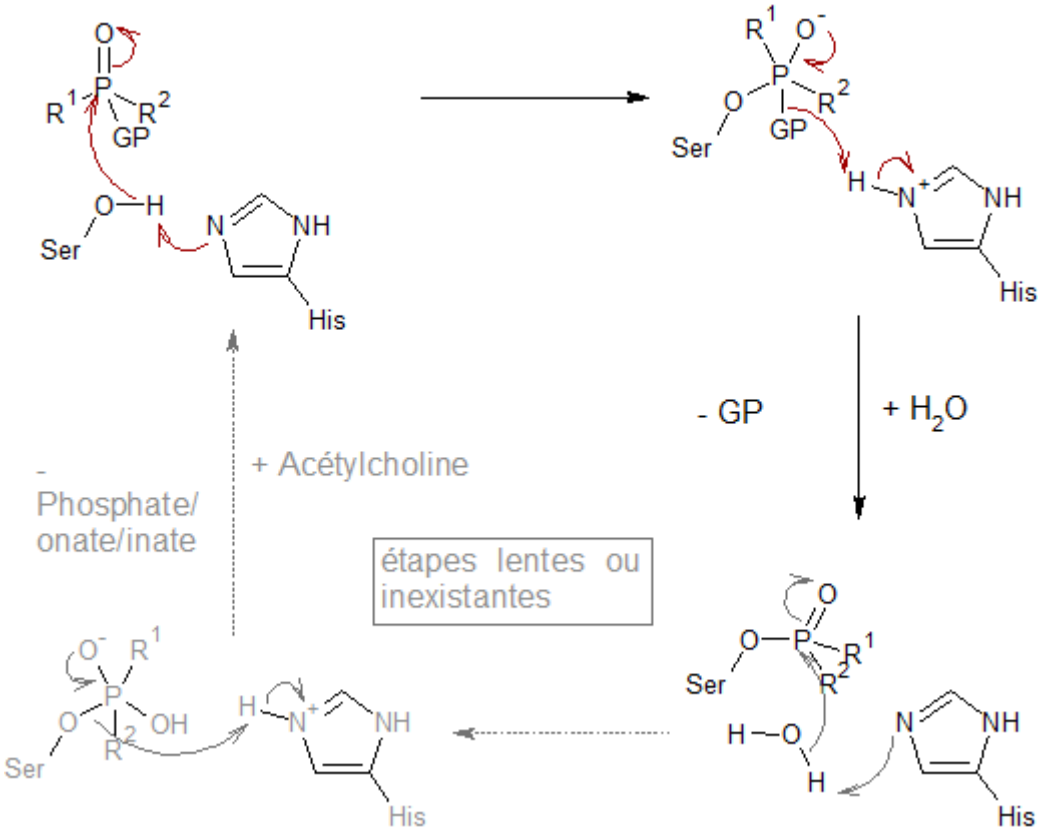


Figure 24 : inhibition de l'acétylcholinestérase. GP = groupe partant ; R1, R2 = groupes alkoxy, alkylamino, alkyle.

L'effet d'inhibition de l'AChE est connu et fort pour les organophosphorés avec le type de structure montré en Figure 24, mais les effets d'autres types d'organophosphorés sont moins connus. La revue d'Elersek et Filipic (Elersek and Filipic, 2011) renseigne des effets variés et plus ou moins forts pour différents organophosphorés : inhibition de l'AChE, des carboxylases, indirectement des DAG-lipases, des protéines kinases PKC (probablement indirectement par génération de ROS), des récepteurs androgènes, ou encore activation de la protéine CREB, *etc*. Cela invite à une évaluation de risques prudente pour les organophosphorés en général, et en particulier ceux qui sont des esters centrés sur le phosphore.

Une autre famille de substances qui sont des bons inhibiteurs de l'AChE sont les carbamates (Fukuto, 1990). Leur mécanisme d'inhibition est le même : ils se lient à la sérine en éliminant un groupe partant, puis ne sont que lentement éliminés pour rétablir la sérine. L'élimination des carbamates est cependant nettement plus rapide que pour les organophosphorés.

Quelques relations structure-activité ont été établies pour ce type de substances : Pour les organophosphorés, il a été démontré que l'inhibition de l'AChE était dirigée par la réactivité de l'ester. Plus l'ester est réactif, plus l'adduit sera formé facilement. La réactivité de l'ester est liée à la charge partielle de l'atome de phosphore. Plus celle-ci est élevée (déficient en électrons), plus l'attaque par le groupe hydroxyle de la sérine (charge partielle négative, riche en électrons) sera favorisée. La force du groupe partant joue un rôle décisif également, car l'élimination du groupe partant est en compétition avec l'élimination de la sérine pour retourner à l'état initial. Une liste de groupes partants a été donnée dans le Tableau 3. Cependant, il faut préciser ici que le fluor est un bon groupe partant lorsqu'il est attaché au phosphore, car la liaison phosphore-fluor n'a pas la stabilité particulière de la liaison carbone-fluor. En revanche, la réactivité n'est pas le seul paramètre à prendre en compte pour quantifier l'inhibition. En effet, les substituants trop encombrants réduisent l'activité, pour une même

réactivité de l'ester. Les carbamates doivent également posséder un groupe partant suffisamment bon (phénoxy ou oxime), mais le plus important est leur structure tridimensionnelle qui doit correspondre à la forme et à la polarité du site actif (Fukuto, 1990).

Ce mécanisme n'intervient que pour les organismes qui possèdent des neurones dont le neurotransmetteur est l'ACh, et donc qui possèdent aussi l'AChE, ce qui constitue le règne animal (à part quelques exceptions). Cependant, les substances exerçant ce mécanisme ne sont pas forcément inertes pour les autres organismes, car elles peuvent inhiber d'autres estérases, particulièrement pour les organophosphorés.

b) Ligands aux récepteurs de l'AcétylCholine

Une autre manière de perturber la transmission nerveuse est d'agir non au niveau de l'AChE, mais au niveau des récepteurs à l'ACh (les AChR). Deux types de perturbation peuvent avoir lieu au niveau de ces récepteurs : l'agonisme ou l'antagonisme. L'agonisme est lorsqu'une substance produit le même effet que le substrat naturel (ici l'ACh, effet d'activation du récepteur), l'antagonisme est lorsqu'une substance empêche l'effet normal de se produire, en bloquant le récepteur sans l'activer.

Les agonistes de l'ACh au niveau de ses récepteurs entraînent un effet similaire à celui des inhibiteurs de l'AChE. En effet, tout comme l'ACh, ils vont activer les récepteurs, mais ne vont pas être dégradés ensuite par l'AChE, et l'effet n'est donc pas stoppé, et la cellule post-synaptique ne peut pas retourner en état de repos. Les agonistes les plus connus sont évidemment la nicotine au niveau des récepteurs nicotiniques (nAChR) et la muscarine au niveau des récepteurs muscariniques (mAChR) (King, 2015b).

Les antagonistes administrés seuls sont toxiques également en empêchant la transmission nerveuse. En revanche, ce sont des antidotes pour contrer les effets des inhibiteurs de l'AChE et des agonistes. La molécule la plus connue comme antidote est l'atropine (et aussi quelques dérivés) (IPCS Inchem, 2002), qui bloque les récepteurs muscariniques. Il y a également la coniine (Figure 25) (poison Hemlock) (National Center for Biotechnology Information, 2017; Vetter, 2004), qui bloque les récepteurs nicotiniques. Les molécules similaires à la coniine, malgré une structure très simple, peuvent aussi provoquer ces effets, dont la force dépend de l'affinité pour le site de liaison (Committee on Acute Exposure Guideline Levels *et al.*, 2012).

Pour toutes ces molécules, on peut reconnaître la similarité avec l'ACh (voir Figure 25), qui possède une terminaison encombrée et chargée positivement d'un côté (en bleu), et, à quelques Angström, un groupement riche en électrons et accepteur de liaison hydrogène (en rouge), c'est-à-dire un ester dans le cas de l'ACh, ou bien la pyridine dans le cas de la nicotine. En effet, cet aromatique est riche en électrons et l'azote est un bon accepteur de liaison hydrogène, des caractéristiques qui rendent la nicotine suffisamment similaire à l'acétylcholine vis-à-vis de l'interaction au sein de la poche de liaison. La tubocurarine (King, 2015b) (Figure 25) est aussi un antagoniste de l'ACh au niveau des récepteurs nicotiniques. Cette molécule contient 2 équivalents d'ACh. En effet, au niveau du récepteur, il y a 2 sites de liaison pour l'ACh qui doivent être occupés pour ouvrir le canal à ions (Droual, 2011). En revanche, il reste à comprendre quelles sont les propriétés des molécules qui font que celles-ci seront plutôt un agoniste ou plutôt un antagoniste.

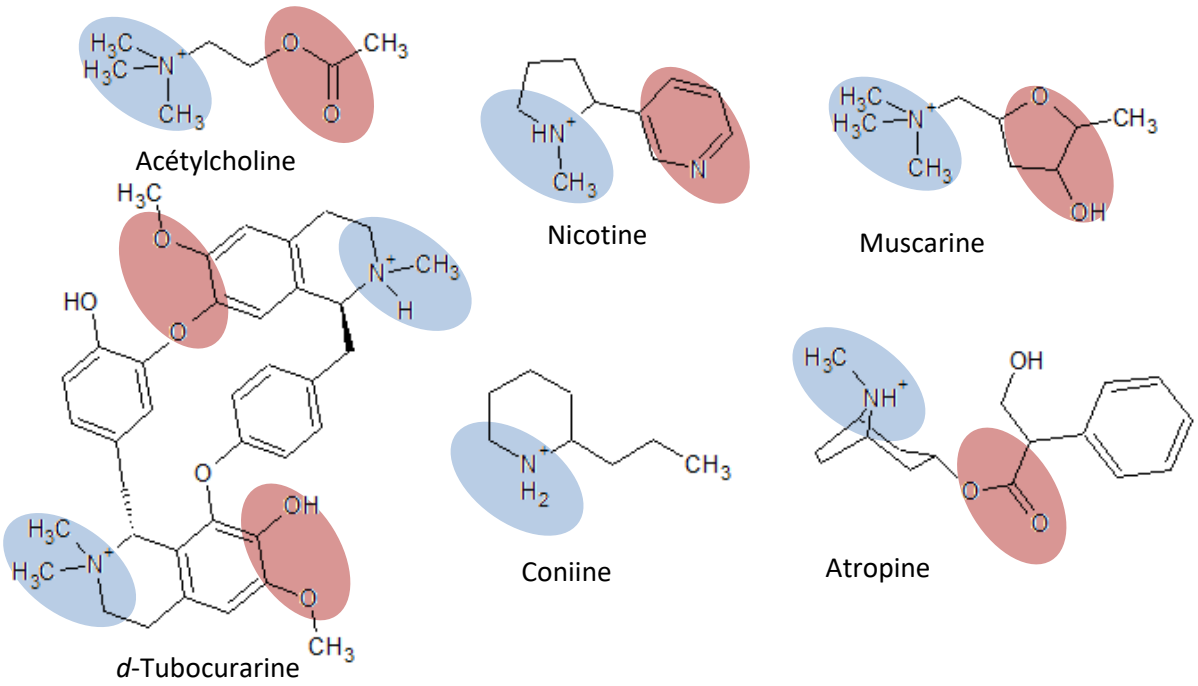


Figure 25 : L'acétylcholine et quelques-uns de ses agonistes et antagonistes.

Ce mécanisme n'intervient que pour les organismes qui possèdent des neurones dont le neurotransmetteur est l'ACh, et donc qui possèdent aussi les récepteurs à l'ACh, ce qui constitue le règne animal (à part quelques exceptions).

c) Perturbation du transport de la dopamine

La dopamine **16** est un autre neurotransmetteur, qui intervient dans d'autres connexions neuronales que l'acétylcholine (contrôlant les fonctions exécutives, le contrôle moteur et aussi en partie la respiration et le rythme cardiaque par exemple). Le fonctionnement de la transmission nerveuse est identique, avec le relargage du neurotransmetteur, la dopamine, au niveau de la fente synaptique pour venir activer des récepteurs de type GPCR (comme les récepteurs à l'acétylcholine muscariniques). Puis la dopamine est recyclée en étant transportée à nouveau dans le neurone présynaptique (Calipari and Ferris, 2013; King, 2015b).

Ce sont ces mécanismes qui sont modifiés par l'amphétamine **18** et les molécules similaires. En effet, l'amphétamine force la libération de dopamine dans la fente synaptique, et en même temps inhibe le transport de la dopamine dans la cellule, ce qui conduit à une forte surdose de dopamine dans l'espace intercellulaire, et donc une excitation continue du neurone postsynaptique (Calipari and Ferris, 2013).

Ce mécanisme n'intervient que pour les organismes qui possèdent des neurones dont le neurotransmetteur est la dopamine, et donc qui possèdent aussi les mécanismes de transport de la dopamine, ce qui constitue le règne animal (à part de rares exceptions) (Barron *et al.*, 2010). La dopamine a aussi été détectée dans les plantes, mais avec un rôle différent (Kulma and Szopa, 2007).

d) Chélation des métaux

Dans les organismes vivants, beaucoup d'enzymes utilisent des ions métalliques au sein de leur site actif afin de catalyser la réaction. Par exemple les cytochromes P450 ont un ion de fer au centre de l'hème, ou encore les métalloprotéases, qui coupent les liaisons peptidiques en se servant par exemple d'un ion de magnésium ou de zinc pour catalyser la réaction d'hydrolyse. De plus les ions métalliques

jouent aussi parfois le rôle de messager (particulièrement le calcium), ou encore servent à établir un courant électrique (comme les ions de sodium et de potassium dans les neurones).

La chélation des métaux est le processus par lequel des molécules forment un complexe assez stable avec l'ion métallique. Les ions métalliques sont chélatés au sein du site actif de certaines enzymes. Cependant, certaines substances xénobiotiques ont aussi des propriétés chélatantes, et entrent ainsi en compétition avec les enzymes pour capter le métal. Le chélatant peut soit se fixer dans le site actif et partager la chélation du métal avec l'enzyme (dans ce cas, le site actif est bloqué et l'enzyme est de toute façon inactivée), soit extraire l'ion métallique de l'enzyme pour former un chélate indépendant (ce qui inactive l'enzyme). Il faut noter que les organismes sont également dotés de protéines telles que les métallothionéines qui servent à chélater les ions métalliques afin d'en contrôler la concentration libre. Mais l'expression de ces protéines est finement régulée, et l'exposition à une forte dose de chélatants xénobiotiques n'aura pas le même effet positif (Petering and Fowler, 1986).

Pour obtenir des bonnes propriétés de chélation, il faut des charges partielles négatives à bonne distance qui correspond au diamètre de l'ion métallique visé. Par exemple les *beta*-dicarbonyles, à condition que les deux carbonyles puissent se positionner de manière parallèle, sont des bons chélatants du fer et dégradent les cytochromes P450 (O'Donoghue, 2001). D'autre part, l'oxalate **17** chélate le calcium, et les complexes formés font des cristaux, qui vont donc former des dépôts dans la circulation sanguine, bloquant le flux normal (Wiley-VCH Verlag, 2002c).

e) Inhibition du transport d'électrons du photosystème II

Ce mécanisme d'action est similaire à celui de l'utilisation de cycle RedOx car la substance va bloquer une chaîne de transport d'électrons. Cependant, au lieu d'impacter la chaîne de transport d'électrons de la mitochondrie, c'est le transport d'électrons du photosystème, qui sert à la photosynthèse dans les plantes, qui est perturbé. Plusieurs substances, dérivées de l'urée, inhibent la protéine D1 du photosystème II (van Rensen, 1982) empêchant la transmission des électrons, issus de la réduction de l'eau en oxygène moléculaire, aux prochains complexes de la chaîne qui vont réduire le NADP+ en NADPH, tout en alimentant le gradient de protons servant à la production d'ATP. Les électrons ne diffusant plus, la production d'oxygène est aussi stoppée. La Figure 26 représente la structure du photosystème II et son fonctionnement. Les substances identifiées dans ce travail comme inhibant la protéine D1 sont présentées en Figure 27. L'atrazine fait partie du groupe d'herbicides des triazines, et est très utilisée, mais son utilisation est de plus en plus restreinte du fait de ses effets à long terme sur les humains.

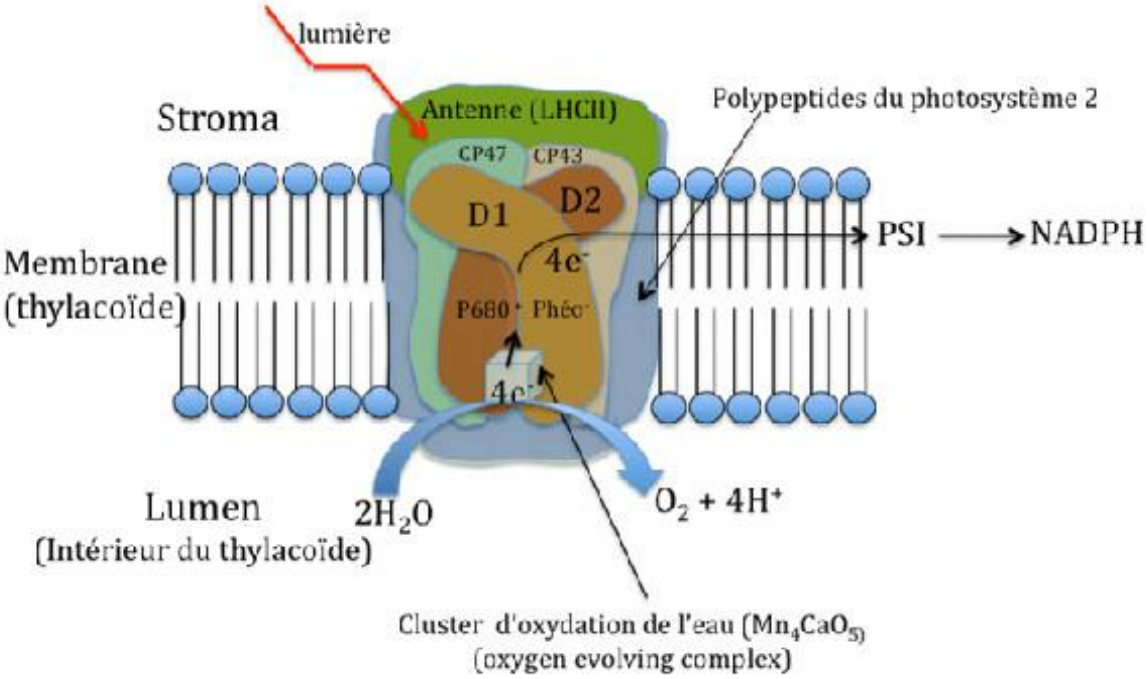


Figure 26 : photosystème II. Figure tirée de (Mache, 2015).

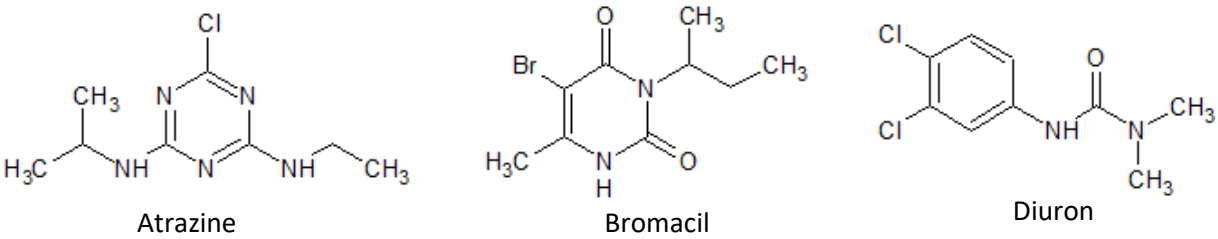


Figure 27 : herbicides inhibant la protéine D1 du photosystème II.

Aucun effet significatif pour les animaux n'est connu pour le bromacil et le diuron. En revanche, l'atrazine a été identifiée comme cancérigène et toxique pour la reproduction et le développement chez les mammifères. Plusieurs mécanismes d'action sont postulés à ce jour pour expliquer tous ces effets, mais dont l'importance relative et l'application aux humains ne sont pas encore certaines (APVMA, 2010).

f) Modulation des canaux à ions

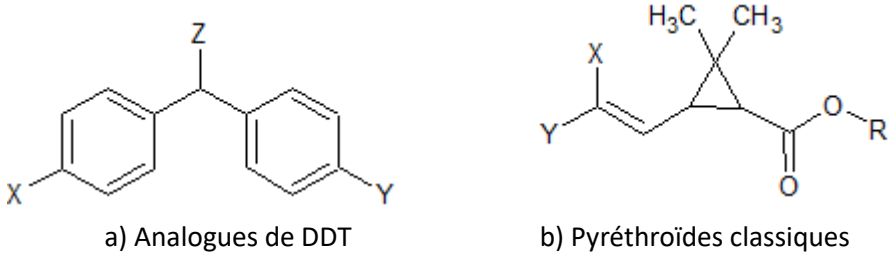
Les canaux à ions spécifiques que sont les récepteurs à l'ACh nicotiniques ont déjà été traités dans la section dédiée. De manière assez similaire, certains neurones fonctionnement avec l'acide gamma-aminobutyrique **9** (GABA) comme neurotransmetteur, qui va se lier sur deux types de récepteurs. Les récepteurs $GABA_A$, comme les récepteurs à l'ACh nicotiniques, sont des canaux à ions dont l'ouverture est gérée par le neurotransmetteur. Les récepteurs $GABA_B$, comme les récepteurs à l'ACh muscariniques, sont des GPCRs qui vont déclencher l'ouverture d'un canal à ion indépendant. Les canaux à ions gérés par l'action du GABA sont des canaux à ions chlorures. L'activation des neurones GABAergiques a des effets anti-dépresseurs, qui se traduisent par des effets anticonvulsifs, anxiolytiques, sédatifs et hypnotiques (Olsen and DeLorey, 1999).

Le 2,2,2-trichloroéthanol, déjà évoqué plus haut, rend les récepteurs plus sensibles au GABA en s'y liant, de sorte qu'une plus faible concentration de GABA suffit à déclencher l'ouverture du canal à chlorures (Peoples and Weight, 1994).

Les neurones possèdent aussi des récepteurs aux opioïdes, qui, lorsqu'ils sont activés, vont activer l'efflux de potassium par les canaux à potassium tout en activant l'influx de calcium par les canaux à calcium, réduisant ainsi le potentiel d'action et empêchant la propagation du signal nerveux. L'opioïde le plus connu agissant au niveau de ce type de récepteurs est la morphine (Chahl, 1996).

Certaines substances empêchent la fermeture des canaux à sodium des neurones, empêchant ceux-ci de retourner à l'état de repos. C'est le cas du DDT **19** (dichlorodiphényltrichloroéthane) et de ses analogues ainsi que des pyréthroïdes, tels la resméthrine **20**, la cyperméthrine **21** ou la téfluthrine **22**. En 1990, Coats a établi des relations structure-activité pour ces substances, en fonction des substituants (Coats, 1990).

Les analogues du DDT qui exercent ce mécanisme d'action ont la structure présentée en Figure 28a), avec X et Y étant des substituants pas ou peu polaires (comme des alkoxy, des chloro, des alkyles, mais pas des hydroxyles) et assez petits (rarement plus grand que des butoxy), et Z ne devant pas être trop polaire non plus. Pour les pyréthroïdes (Figure 28b)), les plus actifs sont ceux avec le carbone du cyclopropane le plus proche de l'ester en stéréoconfiguration R, et avec le substituant sur le deuxième carbone du même côté du cycle que l'ester (*cis*). Le substituant R est variable, mais possède en général un cycle, souvent aromatique, et les substituants X et Y sont en général des halogénés ou des substituants comportant des halogènes. Il faut néanmoins noter que d'autres dérivés déviant de ce squelette existent aussi. D'autre part, l'encombrement stérique et la lipophilie jouent des rôles importants pour l'activité.



a) Analogues de DDT b) Pyréthroïdes classiques

Figure 28 : Structure des dérivés de DDT et de pyréthroïdes.

En plus de leur effet sur les canaux à sodium, certains pyréthroïdes (type II) inhibent aussi l'ouverture des canaux à chlorure GABAergiques et les pyréthroïdes de type I inhibent la Ca-ATPase alors que les type II inhibent la Ca-Mg-ATPase, affectant dans tous les cas la régulation de la concentration en calcium dans le neurone (Coats, 1990).

De plus, une variété de substances peut moduler les canaux à calcium GABAergiques, qui possèdent plusieurs sites de liaison différents, permettant une grande variété de modulateurs de ces canaux (Figure 29). Le site de liaison utilisé par le 2,2,2-trichloroéthanol et l'éthanol est incertain (Olsen and DeLorey, 1999). Les pyréthroïdes semblent se lier au site de la picrotoxine (Coats, 1990). Les insecticides alicycliques chlorés se lient aussi dans ce site de la picrotoxine, et empêchent son ouverture, empêchant le neurone de revenir à l'état de repos, entraînant la mort par arrêt de la respiration (Coats, 1990). Ces insecticides ont des structures variées ayant en commun d'avoir au moins un cycle non aromatique (= alicyclique), et plusieurs substituants chloro (au moins 4 concentrés sur une partie de la molécule composée de moins de 5 carbones). Quelques structures sont dessinées en Figure 30, avec la picrotoxinine.

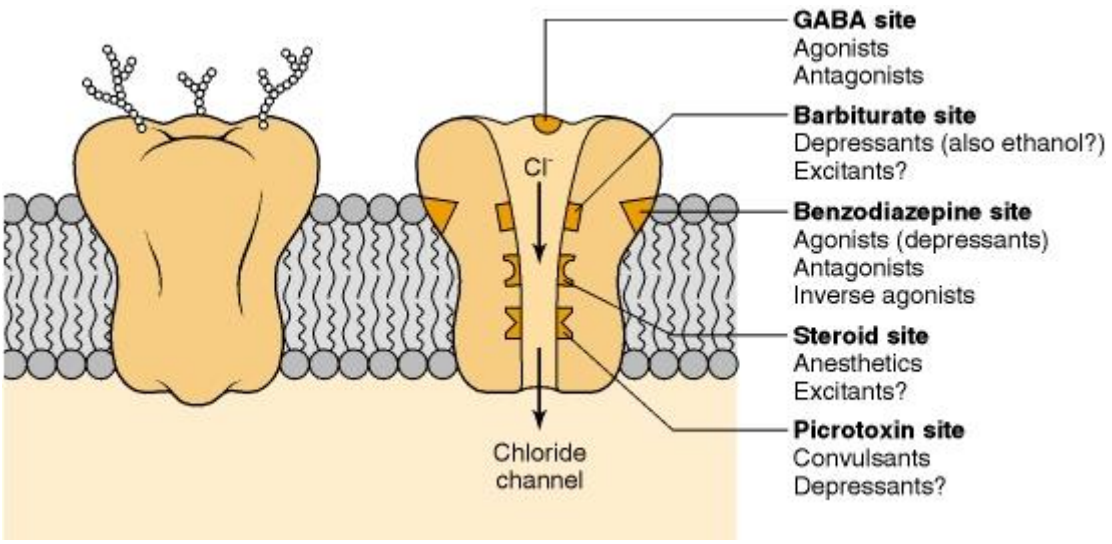


Figure 29 : Récepteur à GABA et ses sites de liaison. Figure tirée de (Olsen and DeLorey, 1999).

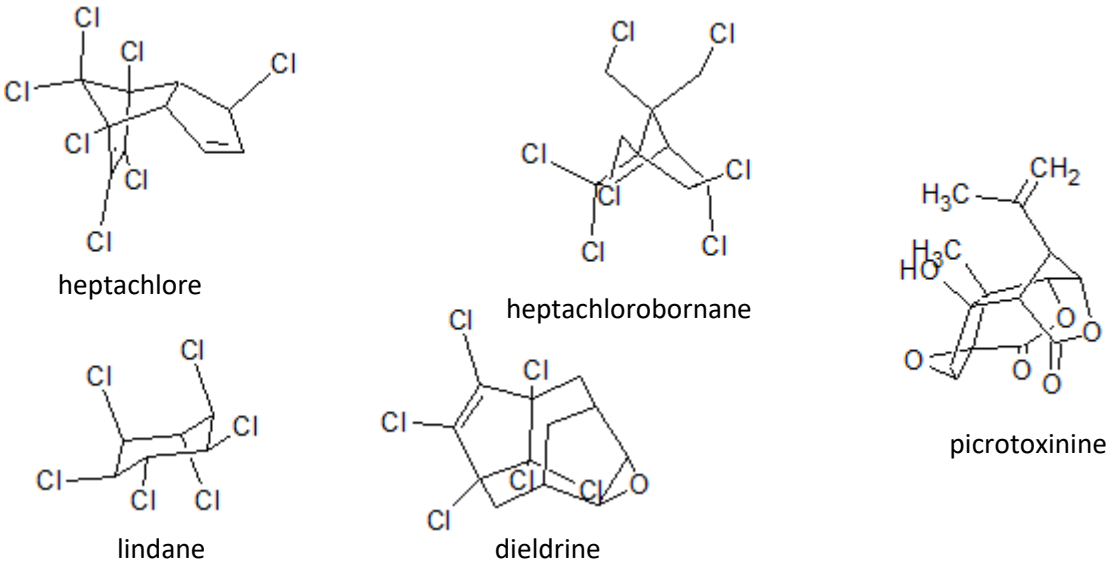


Figure 30 : insecticides alicycliques chlorés et picrotoxinine.

D'autre part, la caféine **23** agit aussi probablement sur ces récepteurs à GABA, au niveau du site à la benzodiazépine. Cependant ce mécanisme est mineur pour la caféine (et les méthylxanthines en général) qui module plusieurs canaux à calcium, soit en les maintenant fermés, soit en les activant, selon les canaux et selon la dose. D'autre part, la caféine est un antagoniste de l'adénosine **24** au niveau de ses différents récepteurs, ce qui a pour effets une dépression de l'activité électrique spontanée des neurones, une inhibition de la transmission synaptique et la libération des neurotransmetteurs (Nehlig *et al.*, 1992). Apparemment, la caféine perturbe aussi la régulation du calcium dans les plantes, mais les mécanismes sont moins bien étudiés (IAC Publishing Labs, n.d.).

Enfin, il a été observé chez la grenouille que l'uréthane **25** avait un effet d'agoniste ou d'antagoniste sur plusieurs canaux à ions activés par neurotransmetteur (récepteurs au $GABA_A$, à l'ACh nicotiniques, à la glycine **26**, au N-méthyl-D-aspartate **27** et à l'acide α-amino-3-hydroxy-5-méthyl-4-isoxazole propionique **28**), ce qui explique ses effets anesthétiques (Hara and Harris, 2002).

g) Autres mécanismes

Il existe encore une grande variété de mécanismes d'interaction spécifique avec une protéine active, qui sont parfois spécifiques à une molécule particulière. Les mécanismes identifiés en sein de ce travail seront renseignés ici, mais pas forcément détaillés. Nous sommes conscients que cette revue n'est pas complète, car beaucoup de mécanismes spécifiques existent, et peuvent être exercés par peu de molécules de manière efficace, notamment pour les substances pharmaceutiques.

i) Mécanisme de l'acide salicylique et ses dérivés

L'acide salicylique **29** et ses dérivés, comme le 2-hydroxybenzamide, inhibent les cyclooxygénases (COX-1 et COX-2), ce qui est à l'origine de leurs propriétés analgésiques (soulagement de la douleur), antipyrétiques (diminution de la fièvre) et anti-inflammatoires. Cependant, à haute dose, ils provoquent un saignement gastrointestinal (probablement dû à l'inhibition de COX-1) (DrugBank, 2016; Vane and Botting, 1998). Chez le poisson, ces substances ont aussi des interactions spécifiques avec des protéines actives ayant pour conséquence de perturber le fonctionnement de la réponse au stress du poisson, et ont aussi des propriétés de perturbateurs endocriniens, avec des effets sur la reproduction (Ji *et al.*, 2013).

ii) Herbicides à mécanisme d'action particulier

Le glyphosate **30**, un herbicide couramment utilisé, inhibe la synthèse d'acide aminés aromatiques dans les plantes en inhibant la 5-énolpyruvylshikimimate-3-phosphate synthase. Pour les mammifères, il a été classé génotoxique et probable cancérigène, mais les effets et les mécanismes sont encore étudiés et controversés (Mensah *et al.*, 2015).

Un autre herbicide, la trifluraline **31**, inhibe chez les plantes la polymérisation de la tubuline pour former des microtubules, structure essentielle pour la division cellulaire, ce qui stoppe ainsi la mitose (mécanisme schématisé en Figure 31). Ainsi, la plante ne se développe pas. Cette substance fait partie de la famille des dinitroanilines qui exercent ce mécanisme avec des potentiels et spécificités différentes. La structure de la tubuline diffère légèrement d'un organisme à l'autre, bien qu'elle soit présente partout. La trifluraline a une particulièrement bonne spécificité pour la tubuline exprimée par les plantes (Fernandes *et al.*, 2013).

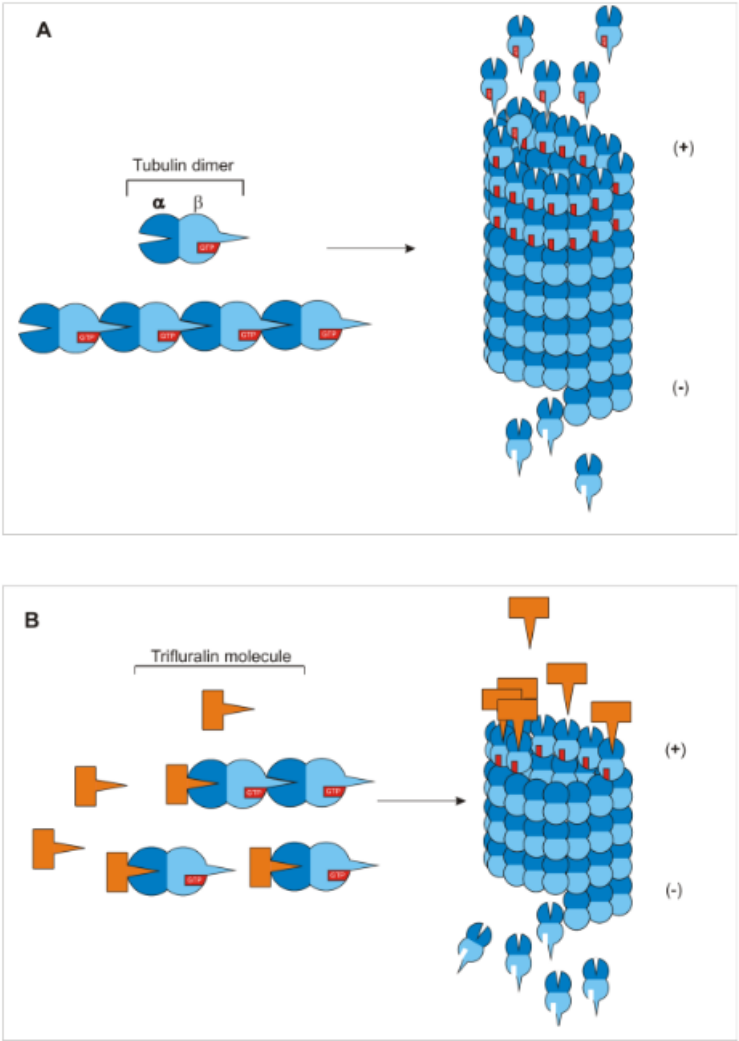


Figure 31 : inhibition de la polymérisation de la tubuline par la trifluraline. Figure tirée de (Fernandes et al., 2013).

Dans les autres organismes que les plantes, cette substance perturbe la régulation du taux de calcium, ce qui a une influence sur beaucoup de procédés, dont la polymérisation de la tubuline justement, ce qui la rend très toxique pour les poissons, mais très peu toxique chez les mammifères (Fernandes *et al.*, 2013), probablement à cause d'un métabolisme et d'une élimination plus efficaces. Il faut d'ailleurs noter que le métabolisme, bien que différent d'une espèce à l'autre, passe généralement par une 2,6-dinitroaniline non N-substituée, une aniline très acide, et le mécanisme de ce type de substances, qui sont assez toxiques, a déjà été discuté dans la section 5.a) Découplage de la phosphorylation oxydative par transport de protons, et le mécanisme des anilines en général dans la section 4.b.ii) Métabolisme des amines et nitros aromatiques.

Le métolachlore **32**, et les chloroacétanilides en général, sont conjugués avec les thiols, probablement spontanément à cause de leur bonne réactivité. Ils réagissent notamment avec la coenzyme A, dont la forme libre vient alors à manquer, bien que beaucoup d'enzymes l'utilisent pour transférer le groupe acétyle (voir Figure 32). Dans les plantes, ce mécanisme leur confère leurs propriétés herbicides, alors que pour les autres organismes, la détoxication par la GSH empêche la toxicité. Cette détoxication est aussi présente chez les plantes, mais apparemment moins efficace dans ce cas (Fuerst, 1987; Rivard, 2003).

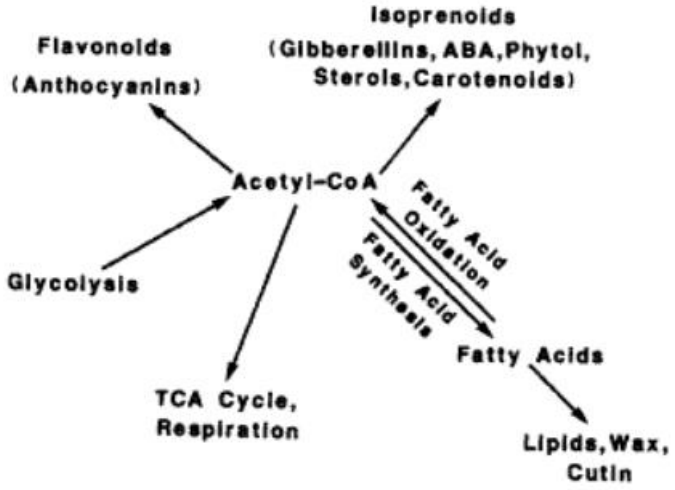


Figure 32 : Rôle central de l'acétyl-coenzyme A dans les plantes. Figure tirée de (Fuerst, 1987).

D'autre part, la clomazone **33** inhibe la biosynthèse des caroténoïdes dans les plantes. Or ces composés sont essentiels à la photosynthèse, et donc à la vie des plantes. Pour les autres organismes, il lui a été trouvé un léger effet d'inhibition de l'AChE (Van Scoy and Tjeerdema, 2014).

Enfin, le 2,4-D **34** (acide 2,4-dichlorophénoxyacétique) est un antagoniste des phytohormones de la classe des auxines, ayant pour effet de stimuler excessivement la croissance de la plante (car la substance n'est éliminée que très lentement par le métabolisme des plantes), et ce plus qu'elle ne peut le supporter, car l'apport en énergie, minéraux, azote, etc. n'est pas assez rapide pour une telle croissance. Il en résulte une mort de la plante. Chez les animaux, cette substance est rapidement éliminée de l'organisme sans effet notable. Des dérivés du 2,4-D sont souvent utilisés comme herbicides, notamment des esters de cet acide (Jervais *et al.*, 2008; Song, 2014).

iii) Mécanisme de la strychnine

La strychnine **35** est un antagoniste de la glycine **26** au niveau de ses récepteurs au niveau des synapses des neurones qui utilisent ce neurotransmetteur (principalement dans le système nerveux central). La glycine a une action inhibitrice, qui est contrecarrée par la strychnine, engendrant une activité excessive des neurones moteurs et des convulsions, mais aussi des réponses exagérées aux stimuli visuels, auditifs et tactiles (Makarovsky *et al.*, 2008). La strychnine a été utilisée comme rodenticide. Ce mécanisme a été identifié chez les mammifères, dont les humains (Makarovsky *et al.*, 2008), et chez les poissons (Enoch *et al.*, 2008).

iv) Mécanisme du dicofol

Le dicofol **36**, molécule semblable au DDT **19**, avec un groupe hydroxyle en plus, a aussi un mécanisme d'action spécifique, mais différent du DDT : plusieurs effets ont été identifiés chez différentes espèces. C'est un antagoniste de l'hormone thyroïdienne, un faible agoniste des estrogènes, un puissant inducteur des cytochromes CYP2B, un faible inhibiteur de l'aromatase, et il peut aussi se lier aux récepteurs androgènes (Sparling *et al.*, 2010). C'est donc un perturbateur endocrinien de multiples manières. Russom *et al.* ont classé ce composé neurotoxique pour les poissons (Enoch *et al.*, 2008).

v) Mécanisme des dioxines, PCBs et polychlorodibenzofuranes

Les dioxines **37**, les PCBs (PolyChloroBiphényles) et les polychlorodibenzofuranes partagent un même mécanisme d'action qui est la liaison aux récepteurs AhR (Aryl hydrocarbon Receptor). L'activation de ces récepteurs induit des enzymes métaboliques, dont le CYP1A1, mais influe aussi sur d'autres processus. Le mécanisme de toxicité de ces substances est complexe et n'est pas encore clairement et totalement élucidé, d'autant plus que la toxicité est très variable d'une espèce à l'autre. Chez les

humains, le symptôme principal est l'apparition de chloracné, des éruptions cutanées qui durent des années après l'exposition, ceci étant probablement dû au fait que ces composés s'accumulent fortement dans l'organisme, faute de métabolisme et élimination efficaces. Des effets cancérigène et reprotoxique sont possibles aussi (Sorg, 2014).

vi) Inhibition du transport et de la réduction de l'oxygène pour la respiration cellulaire

Le monoxyde de carbone se lie à l'hémoglobine à la place de l'oxygène. Ce dernier n'est donc plus transporté dans l'organisme qui meurt par manque d'oxygène, car l'oxygène sert à capter les électrons de la chaîne de transport d'électrons couplée à l'ATP-synthase, pour être réduite en eau. La liaison du monoxyde de carbone à l'hémoglobine est 250 fois plus forte que celle de l'oxygène. Ce mécanisme affecte a priori tous les animaux, où l'oxygène est transporté par ce moyen.

L'anion cyanure quant à lui inhibe la cytochrome c oxydase, qui réduit l'oxygène en molécule d'eau, afin d'alimenter le gradient de protons de la mitochondrie, tels qu'expliqué en 4.e) Métabolisation en perturbateurs directs par docking. Il a été montré que l'anion formate inhibait aussi la cytochrome c oxydase, mais moins efficacement que le cyanure (Liesivuori and Savolainen, 1991).

vii) Mécanisme de l'amanitine

Contenue dans le champignon toxique amanite tue-mouche, l'α-amanitine (molécule macrocyclique non tracée dans ce document) inhibe l'ARN-polymérase II (et dans une moindre mesure la III aussi). Ainsi, l'ADN ne peut presque plus être transcrit en ARN pour synthétiser les protéines encodées. Comme certaines protéines dans l'organisme ont un cycle de vie assez court et ont besoin d'être renouvelées régulièrement, mais que leur nombre est suffisamment important, cette action entraîne la mort au bout de quelques jours, environ 10 chez les humains (Gong *et al.*, 2004).

viii) Mécanisme des stéroïdes et dérivés du diéthyl stilbène, perturbateurs endocriniens

Les perturbateurs endocriniens n'ont pas été étudiés en détail dans ce travail, mais quelques notions seront abordées ici. Les hormones estrogènes et androgènes partagent le fait d'avoir une structure basée sur le squelette du gonane **38**, avec un groupe polaire en position 3 et 17 (un hydroxyle ou un carbonyle). Pour exemple, la structure de l'estradiol, l'hormone estrogène naturelle est présentée en Figure 33.

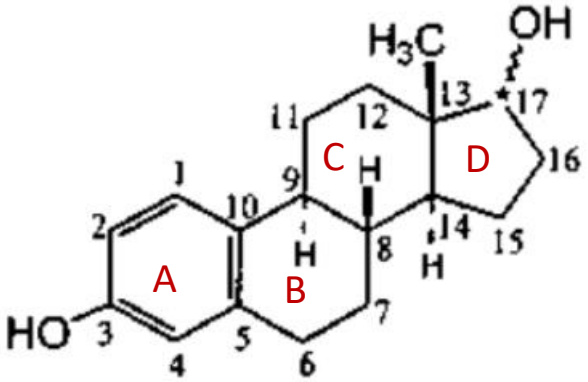


Figure 33 : Structure de l'estradiol. Figure adaptée de (Fang et al., 2001).

Ainsi, des molécules ont été développées pour imiter ces hormones à des fins médicales, et se basent également sur cette structure commune en général. En effet, l'activité auprès des récepteurs des hormones est fortement liée à la structure tridimensionnelle, qui doit correspondre à la poche de liaison. Cependant, des dérivés du stilbène **39** se sont trouvés également très efficaces vis-à-vis des

récepteurs aux hormones, bien que leur structure diffère un peu. Lorsqu'on regarde les structures tridimensionnelles, on constate que ces molécules remplissent également très bien la poche de liaison aux hormones, ce qui explique leur activité (voir par exemple Figure 34).

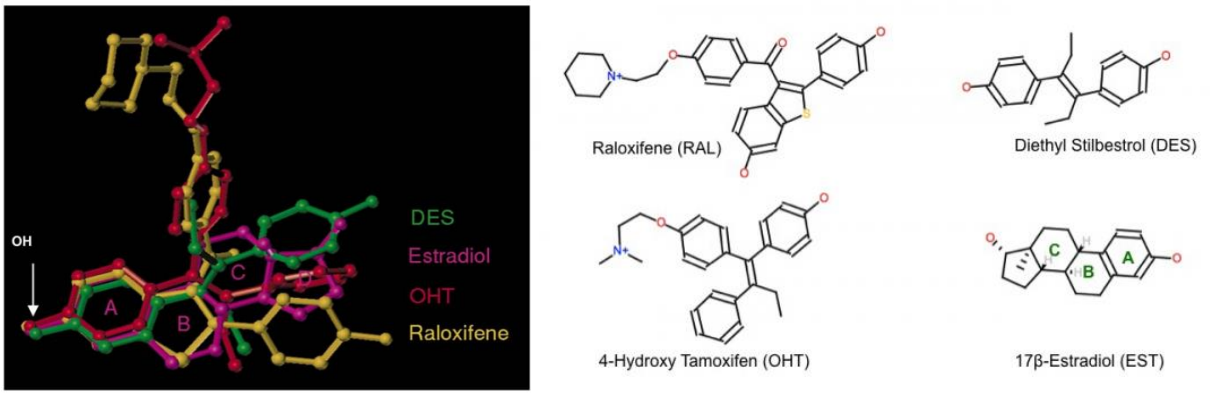


Figure 34 : remplissage de la poche de liaison du récepteur aux estrogènes par différentes molécules actives. Figure adaptée de (Shi et al., 2002).

Pour les récepteurs aux estrogènes, il est important d'avoir un phénol (cycle A, hydroxyle en position 3) pour une bonne liaison au récepteur. Le deuxième groupement polaire en position 17 est moins déterminant (Fang *et al.*, 2001).

Pour les récepteurs aux androgènes, c'est le contraire : un hydroxyle en position 17α est nécessaire pour avoir une bonne activité, alors que le groupement polaire en position 3 est moins important (Fang *et al.*, 2003).

Les molécules qui ont des effets estrogènes ou androgènes peuvent être dangereuses pour les populations si leur exposition n'est pas contrôlée, et ce d'autant plus que des concentrations très faibles suffisent à induire des effets. En effet, ces molécules contrôlent la différentiation sexuelle et la reproduction chez la plupart des vertébrés, y compris les poissons (Pinto *et al.*, 2014).

On peut évoquer également d'autres hormones, comme l'hormone thyroïdienne, qui possède une structure totalement différente des estrogènes et androgènes et n'est pas impliquée dans les mêmes processus physiologiques (pas liée à la reproduction). Nous en avons déjà vu un antagoniste : le dicofol **36**.

ix) Mécanismes peu détaillés

Il a été observé chez les mammifères que le 1,1,1-trichloroéthane sensibilisait le cœur à l'épinéphrine, sans avoir plus de détails sur le mécanisme moléculaire impliqué (ATSDR, 2006).

Le 3-méthoxyphénol, qui a déjà été évoqué plus haut pour sa métabolisation, semble inhiber lui-même la production de mélanine sans passer par ses métabolites, chez les mammifères (Moridani *et al.*, 2003).

Le 2,4,6-triiodophénol est un inhibiteur de la synthèse des leucotriènes B4 chez les humains, mais le mécanisme moléculaire n'a pas été identifié (Matji and Lopez, 2012).

Chapitre 3 : Organisation des mécanismes d'action et discussion de certains mécanismes

Pour évaluer les informations données dans la littérature sur les MechoAs et MoAs, et pour pouvoir aller plus loin que ce qui est publié, les modèles QSAR développées par KREATiS ont été utilisés. Ces modèles sont développés à partir de données expérimentales consciencieusement validées, et les substances sont regroupées par MechoA pour construire des modèles QSAR MechoA-spécifiques. Dans nos modèles d'écotoxicité, la toxicité est reliée à l'activité chimique, au lieu d'être reliée au logP comme c'est fait dans la littérature la plupart du temps. Le concept d'activité chimique, introduit par Mackay *et al.* (Mackay *et al.*, 2009) pour la narcose, est aussi utilisée dans nos modèles pour les autres MechoAs. Plus d'informations sur les méthodes utilisées par KREATiS sont disponibles sur le site internet de KREATiS (KREATiS, n.d.).

1) Une nouvelle classification en mécanismes d'action toxique

Les mécanismes définis dans ce travail sont présentés dans le Tableau 4 et le Tableau 5. C'est suivant la même organisation et dans le même ordre que les mécanismes ont été expliqués dans le chapitre précédent.

Numérotation du MechoA	Nom du MechoA	Description
1	Déstabilisation de membranes	Effet de narcose.
2	Pro-déstabilisation par hydrolyse	Hydrolyse enzymatique produisant en général des MechoA1.
3	Réactivité	Réaction spontanée avec les protéines et l'ADN pour créer des adduits.
4	Pro-activité	Métabolisation générant des substances actives.
5	Perturbation indirecte d'enzymes	Modification de l'environnement d'enzymes spécifiques perturbant leur fonctionnement.
6	Interaction directe dans un site de liaison	Inhibition ou activation d'une protéine active en se liant dans un site de liaison de cette protéine.

Tableau 4 : MechoAs généraux utilisés dans ce travail.

MechoA	Sous-groupe	Détails du mécanisme
1: Déstabilisation de membranes	1.1 Narcose non-polaire 1.2 Narcose polaire 1.3 Narcose cationique	Perturbation physique de l'intégrité des membranes avec différentes affinités pour les non-polaire, polaires et cationiques.
2: Pro-déstabilisation par hydrolyse	2.1 Hydrolyse vers des déstabilisateurs 2.2 Hydrolyse vers des composés actifs	Déstabilisation directe des membranes + hydrolyse enzymatique, générant : 2.1 que des produits narcotiques 2.2 au moins un produit non-narcotique
3: Réactivité	3.1 Réactivité d'électrophile dur 3.2 Réactivité d'électrophile mou 3.3 Rupture homolytique d'une liaison générant des radicaux libres	3.1 formation d'adduits avec les résidus amino et thiol des protéines 3.2 formation d'adduits par addition de Michaël avec les résidus thiol des protéines 3.3 Rupture homolytique d'une liaison faible générant des radicaux libres, ce qui produit du stress oxydatif
4: Pro-activité	4.1 composés facilement détoxifiés 4.3 pro-réactivité 4.4 utilisation d'un cycle RedOx 4.5 Métabolisation vers un perturbateur indirect d'enzymes 4.6 Métabolisation vers un perturbateur direct par docking	4.1 le métabolite agit avec un MechoA 1 4.3 le métabolite agit avec un MechoA 3 4.4 composés subissant un cycle RedOx, générant du stress oxydatif 4.5 le métabolite agit avec un MechoA 5 4.6 le métabolite agit avec un MechoA 6
5: Perturbation indirecte d'enzymes	5.1 Découplage de la phosphorylation oxydative 5.2 Acidification ou basification des cellules 5.3 Autres mécanismes indirects	5.1 transport cyclique d'H+ à travers la membrane mitochondriale interne, empêchant la génération d'ATP 5.2 diffusion à travers les membranes cellulaires sous la forme neutre, puis relargage/captage d'H+ dans le cytosol
6: Interaction directe dans un site de liaison	6.1 inhibition de l'AChE 6.2 liaison aux AChR 6.3 perturbation du transport de la dopamine 6.4 chélation des métaux 6.5 inhibition du transport d'électrons du photosystème II 6.6 modulation des canaux à ions 6.7 autres mécanismes directs	6.1 inhibition de l'AcétylCholinEstérase 6.2 liaison aux récepteurs à l'acétylcholine nicotiniques ou muscariniques, comme agoniste ou antagoniste 6.3 induction du transport de la dopamine dans le synapse et inhibition de son recyclage vers le neurone 6.4 chélation des ions métalliques, comme le Fe^{2+}, bloquant ou dégradant les enzymes contenant cet ion métallique dans leur site actif 6.5 inhibition d'une protéine de la chaîne de transport d'électrons dans le photosystème II des cellules végétales 6.6 liaison dans un site de liaison d'un canal à ion causant son activation ou désactivation

Tableau 5 : MechoAs détaillés. (Sous-catégories des MechoAs généraux)

Les catégories ci-dessus sont différentes de ce qui existe déjà dans la littérature, car nous souhaitions à la fois garder un nombre réduit de mécanismes généraux, et que ceux-ci soient nommés et groupés de manière cohérente, suivant la similarité des mécanismes et des propriétés des molécules qui se retrouvent regroupées. Et enfin, nous souhaitions créer des catégories de sorte qu'à chaque catégorie puisse correspondre un modèle QSAR par espèce et par endpoint pour prédire la toxicité. C'est pour répondre à ce critère que nous avons créé ces MechoAs, dont une partie sont validés par les modèles QSARs d'écotoxicité développés à KREATiS : pour les MechoAs 1.1, 1.2, 2.1, 3.1, 3.2, 5.2. Et pour d'autres MechoAs, nous avons simplement identifié un profil de toxicité clairement différent des QSARs précédents, mais non encore modélisé : pour les MechoAs 2.2, 4.4, 5.1).

Il faut préciser que l'attribution de molécules aux catégories 5.3 et 6.7 (autres mécanismes respectivement indirects et directs), n'est pas une fin, et sert juste à identifier les MechoAs concernés par trop peu de molécules, et catégoriser les molécules pour lesquelles on n'a pas encore assez de données, et pour lesquelles le développement de modèles QSAR MechoA-spécifiques ne peut encore être envisagé. De nouvelles catégories sont amenées à être créées dans le futur pour séparer les MechoAs pour lesquels nous pourrons construire des modèles QSARs.

2) MechoA des amines aliphatiques et ammoniums

Récemment, après l'écriture de la première publication (voir chapitre 4) où ce n'est pas intégré, nous avons acquis une meilleure compréhension du MechoA pour les ammoniums quaternaires, sur la base d'une publication récente (Roberts *et al.*, 2013) et de travaux internes. Nous avons créé pour ce genre de composés une autre catégorie de déstabilisation de la membrane : MechoA 1.3, comme indiqué dans le Tableau 5. En effet, les composés cationiques semblent s'insérer dans les membranes selon une configuration particulière interagissant avec les têtes polaires des lipides membranaires. Cette interaction, plus forte et plus déstabilisante pour la membrane, est la cause de l'excès de toxicité important observé pour ces composés, en comparaison avec les modèles de narcose non-polaire et polaire.

Nous avons également recueilli des données et effectué des analyses QSAR pour les amines aliphatiques, ce qui nous a permis de conclure qu'aucune d'elles ne sont des narcotiques non polaires en écotoxicité, mais sont des narcotiques polaires en général. Ce travail est encore en cours et notre compréhension des MechoAs et MoAs des composés azotés doit encore être améliorée.

3) Discussion sur les modèles QSAR MechoA-spécifiques

Sur la Figure 1, à faible hydrophobie (logP <1), la droite de la narcose non-polaire est nettement en dessous des droites des autres MechoAs. En effet, la CL50 prédite par la narcose non-polaire est environ 10 fois plus grande que pour la narcose polaire, et plus de 100 fois plus grande que pour la toxicité des aldéhydes, donc ce MechoA est bien moins toxique. A forte hydrophobie (logP >4) les droites se rapprochent jusqu'à se rejoindre, car les substances ont une telle préférence pour les membranes plutôt que pour les phases aqueuses, qu'elles s'y accumulent en beaucoup plus grande proportion, et déstabilisent ainsi la membrane à partir d'une concentration aqueuse beaucoup plus faible (CL50 plus faible). Ainsi, le mécanisme de la narcose génère une toxicité aussi forte, voire plus forte que les autres mécanismes, et c'est pourquoi les droites se rejoignent à haut logP.

Comme vu dans le paragraphe sur la narcose polaire (Chapitre 2.1.b) Narcose polaire), la séparation entre narcose polaire et non-polaire est controversée dans la littérature, à cause du fait qu'on obtient deux droites bien distinctes en traçant la toxicité (log(CE50)) en fonction de l'hydrophobie (logP) pour les substances toxiques par ces deux MechoAs.

En revanche, lorsqu'on trace la toxicité en fonction de la solubilité (Figure 35), qui est donc indépendante du choix de l'octanol ou de la DMPC **1**, on obtient également deux droites ce qui maintient l'idée d'une différence de mécanisme d'action. Plusieurs hypothèses peuvent alors être formulées pour expliquer la différence de toxicité :

- Cela peut être dû à une préférence plus forte pour les narcotiques non-polaires que pour les narcotiques polaires pour s'accumuler dans les tissus adipeux plutôt que dans les membranes. Or on peut supposer que l'accumulation de substances dans les tissus adipeux ne contribuent pas à la toxicité, alors que l'accumulation dans les membranes cellulaires contribue directement à la cytotoxicité, ce qui expliquerait qu'une concentration externe plus faible est nécessaire pour obtenir la toxicité dans le cas des narcotiques polaires.
- Une autre possibilité est de considérer que la différence de partition entre les différents tissus n'est pas la cause de la différence de toxicité, mais que les polaires se localisent à des endroits spécifiques des membranes, par exemple autour de certaines protéines membranaires, perturbant leur fonctionnement plus facilement que les narcotiques non-polaires qui s'accumulent de manière indifférente dans la membrane, ou à des endroits différents (Droge *et al.*, 2016).

A notre connaissance, ces hypothèses n'ont pas encore été définitivement validées ou rejetées par des expériences adaptées. On peut remarquer que les deux droites se rejoignent à faible solubilité (hydrophobie élevée), comme expliqué dans le Chapitre 2.1.a) Narcose non-polaire.

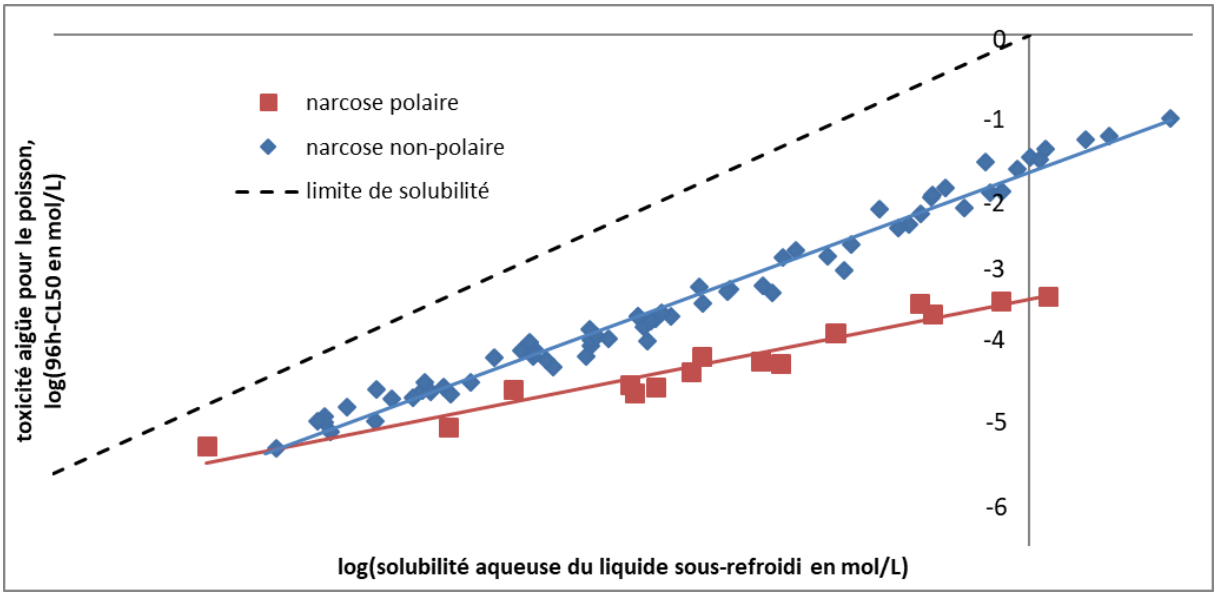


Figure 35 : comparaison de la narcose non-polaire et polaire pour le poisson en fonction de la solubilité. Ces modèles sont la propriété de KREATiS.

Pour ce graphique, le descripteur utilisé est la solubilité aqueuse du liquide sous-refroidi, plutôt que simplement la solubilité aqueuse. Le logarithme de la solubilité aqueuse du liquide sous-refroidi est en effet proportionnel au logarithme de l'activité chimique de la substance dans l'eau, qui, si la durée du test est suffisamment longue pour permettre l'équilibre thermodynamique, correspond à l'activité chimique directement responsable de la toxicité. Solubilité aqueuse et solubilité aqueuse du liquide sous-refroidi sont identiques pour les substances liquides à température ambiante. En revanche, pour les solides, un facteur est appliqué pour tenir compte de l'activité de la substance, qui est réduite dans le cas d'un solide à dissoudre dans l'eau. C'est pour ces solides qu'on parle alors de liquide sous-refroidi, car on ramène ce solide à un état virtuel où il est liquide en dessous de sa température de

solidification. Plus de détails sur cette approche sont donnés dans la littérature (Mackay *et al.*, 2009; Thomas *et al.*, 2015). Les équations qui relient l'activité à la solubilité sont données dans la publication de Mackay *et al.* (Mackay *et al.*, 2009) et sont les suivantes :

La solubilité des liquides vaut $S_L = 1 / (\gamma_{eau}.V_{eau})$ où γ_{eau} est l'activité de la substance dans l'eau et V_{eau} est le volume molaire de l'eau. Pour les solides, la solubilité du liquide sous-refroidi vaut $S_L = S_S / F$, où F est appelé ratio de fugacité et peut être approximé comme $\mathrm{F} = 10^{(0,01(298-T_M))}$ où T_M est la température de fusion du solide.

4) Discussion sur l'excès de toxicité observé pour les esters

La toxicité plus importante des esters par rapport à la narcose de base (voir chapitre 2.2) Narcose des esters ou pronarcose) peut provenir de plusieurs éléments :

- La quantité d'ester qui exerce la narcose est renouvelée en continu, au fur et à mesure de sa disparition par hydrolyse, dans les limites de la vitesse d'absorption et de distribution de l'ester dans l'organisme. En plus de l'ester, l'alcool généré après hydrolyse peut lui aussi s'accumuler dans les membranes et exercer la narcose, et l'acide dans une moindre mesure, car il est ionisé au pH physiologique (pH = -log[H_3O^+], une mesure de l'acidité/basicité du milieu), donc moins hydrophobe. Les membranes cellulaires sont donc soumises à une concentration plus importante de narcotique que lors d'une exposition à des narcotiques simples à même concentration externe.
- Le processus d'hydrolyse génère de l'acidité, tel que montré par la Figure 4. Si la quantité d'ester hydrolysée est suffisamment importante, la quantité d'acide rejetée va baisser le pH du milieu et ainsi perturber le fonctionnement normal de la cellule. (Voir la section 6) Mécanisme responsable de la toxicité et corrosivité des acides et bases pour plus de détails sur les problèmes d'acidification ou basification des organismes).

L'excès de toxicité des esters par rapport à la toxicité de base (narcose non-polaire) n'est en revanche pas constant. En effet, à forte hydrophobie, la droite des esters croise la droite des narcotiques non-polaires (Figure 36). Le rapprochement de la droite des esters de celle de la toxicité de base peut être expliqué par le fait qu'à forte hydrophobie, le mécanisme de narcose devient plus puissant et peut surpasser les autres mécanismes, tel qu'expliqué en a) Narcose non-polaire. Mais cela peut aussi être la conséquence d'une plus faible hydrolyse des esters très hydrophobes, car ils diffusent moins facilement vers les estérases et/ou sont moins facilement liés dans le site actif de ces enzymes (Jaworska *et al.*, 1995). Une corrélation a effectivement pu être établie entre l'hydrolyse abiotique alcaline et l'excès de toxicité des esters. Cependant ces explications vont à l'encontre de l'observation qu'à forte hydrophobie, les deux droites se croisent même, si bien que les esters deviennent moins toxiques que des narcotiques de même hydrophobie. Les explications ci-dessus supposeraient que la toxicité des esters très hydrophobes soient la même que celle prédite par la narcose non-polaire, et non inférieure à cette dernière.

Nous avons introduit dans ce travail le terme de pronarcose – pronarcotique pour signifier que ce mécanisme produit (en général) des narcotiques à partir de la molécule initiale, au même titre que l'appelation « prodrug » en chimie médicinale qui désigne des composés inactifs qui vont être métabolisés en « drug », c'est-à-dire en médicament actif, et au même titre que le terme proélectrophile, utilisé par plusieurs des classifications en MoA présentées plus haut (Bearden and Schultz, 1997; Russom *et al.*, 1997), désignant des composés inactifs qui seront activés par le métabolisme en électrophiles réactifs qui formeront des adduits aux protéines et à l'ADN. Le

mécanisme de la pronarcose hydrolytique est à différencier des mécanismes de toxicité pro-active, car dans le premier, les molécules sont coupées en deux par les enzymes par hydrolyse, alors que les mécanismes de toxicité pro-active gardent la molécule entière en lui faisant subir une oxydation, une réduction ou une conjugaison pour la rendre plus active et/ou plus hydrophile.

D'autre part, les diesters sont un peu plus de 2 fois plus toxiques que les monoesters pour le poisson, mais moins que 2 fois plus toxiques pour la daphnie, et il n'y a pas de différence significative de toxicité pour l'algue, tel que montré en Figure 36 avec les modèles de KREATiS. Cela est probablement aussi lié à l'hydrolyse. Celle-ci n'est vraisemblablement pas présente chez les algues car la droite des esters est à peu près superposée avec la droite de la narcose non-polaire. L'excès de toxicité pour les autres espèces est probablement dû au fait que 2 protons sont générés suite à l'hydrolyse de 2 fonctions ester, donc l'acidification de l'organisme se fait 2 fois plus rapidement. C'est pourquoi on s'attend à une toxicité 2 fois plus importante si celle-ci est totalement dirigée par l'hydrolyse, ou moins que 2 fois, si la toxicité est partiellement due à l'hydrolyse, et partiellement à la narcose. L'excès de toxicité des esters pour les daphnies est plus faible que pour les poissons, ce qui suggère que les daphnies hydrolysent moins facilement les esters que les poissons. Parallèlement, l'excès de toxicité des diesters par rapport aux monoesters est inférieur à 2, ce qui vient corroborer cette hypothèse. En revanche, on trouve même un peu plus de 2 fois plus toxique pour les poissons, ce qui est difficile à rationaliser. Les diesters utilisés dans ces modèles incluent des diesters aliphatiques, avec des positions relatives des deux groupes esters variables, et des diesters aromatiques (phtalates).

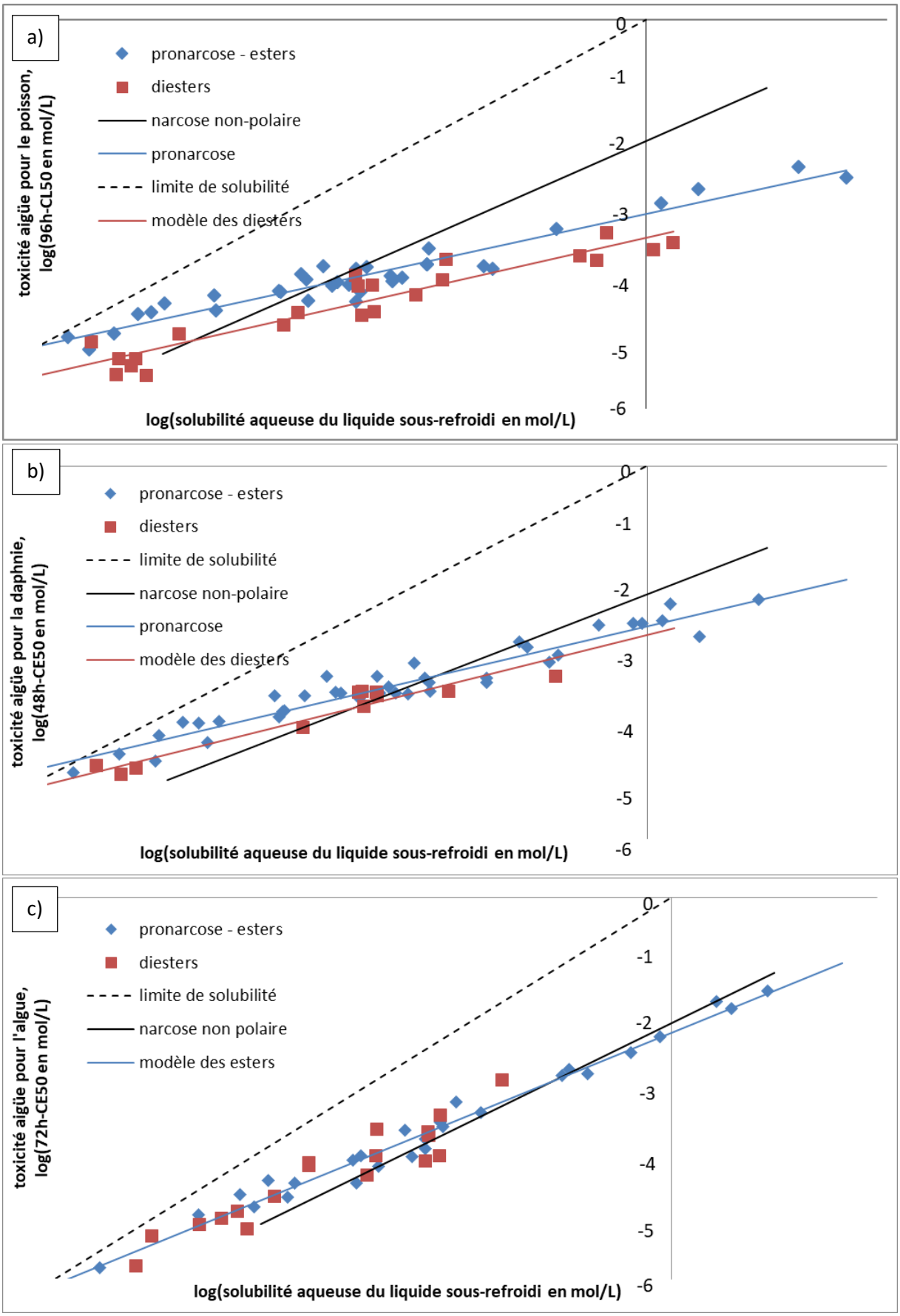


Figure 36 : toxicité des diesters par rapport aux monoesters pour a) le poisson, b) la daphnie, c) l'algue. Ces modèles sont la propriété de KREATiS

5) Réactivité des γ-dicétones en tant qu'électrophiles durs

Nous avons vu que les γ-dicétones (avec certaines restrictions sur la structure détaillées en Chapitre 2.3.a) Electrophiles durs) possèdent une réactivité l'électrophiles durs, bien que les autres cétones ne soient pas toxiques par réactivité. Même si les cétones ne sont pas des groupes électrophiles très réactifs, la formation d'un cycle aromatique, qui est donc plus bas en énergie, favorise la réaction de manière thermodynamique (la fin justifie les moyens : la réaction demande d'investir beaucoup d'énergie, mais celle-ci est restituée par la formation d'un produit très stable, qui ne reviendra pas en arrière). Au contraire, les électrophiles avec un groupe partant font des réactions qui ne demandent pas d'investir beaucoup d'énergie, car l'état de transition de la réaction est stabilisé. Cette différence entre une réaction facilitée par la stabilité du produit final et une réaction facilitée par un état de transition stabilisé est schématisée en Figure 37 de manière qualitative. La faible énergie d'activation (double flèche orange) est la force qui dirige certaines réactions. La grande différence d'énergie entre le produit et le réactif de départ (double flèche bleue) est la force qui dirige d'autres réactions.

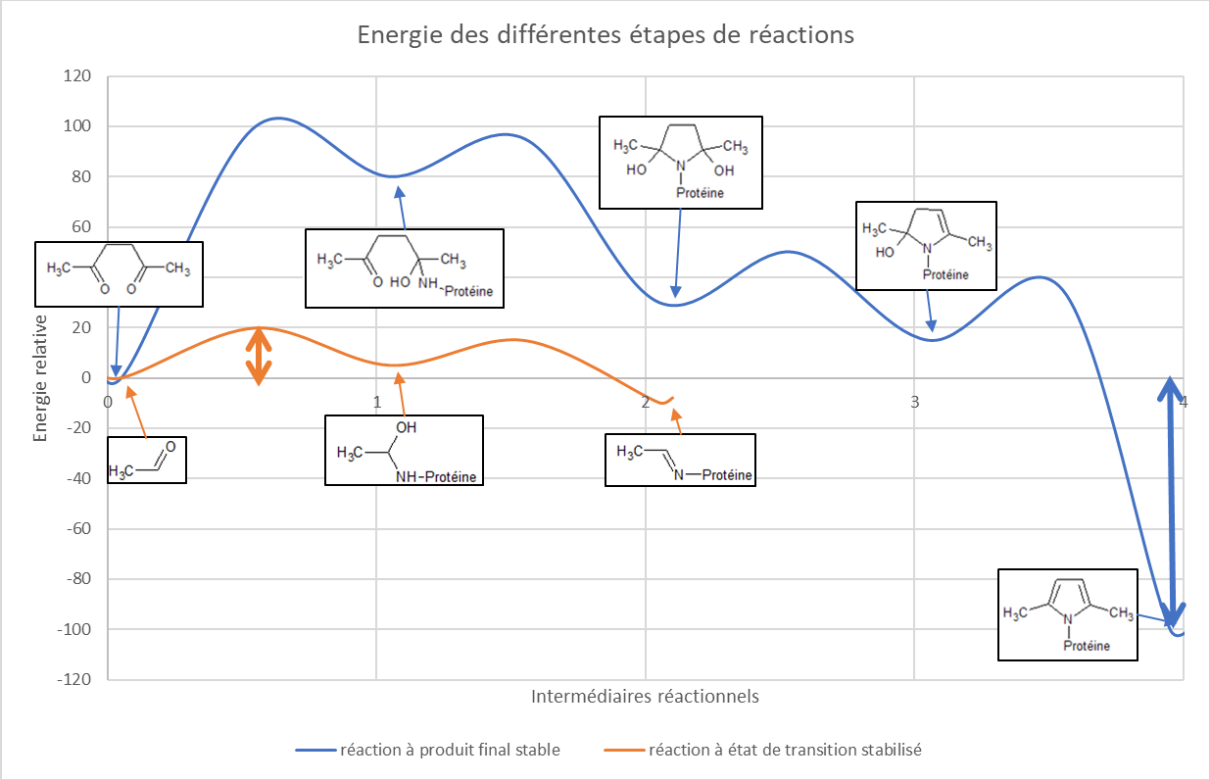


Figure 37 : Réactions favorisées par la stabilité de l'état de transition ou par la stabilité du produit final.

6) Modélisation QSAR pour les découpleurs de la phosphorylation oxydative

Compte-tenu de toutes les informations collectées dans la littérature, l'écotoxicité des phénols acides a été tracée en fonction de leur solubilité, en les regroupant par gamme de pKa (le graphique obtenu pour les poissons est présenté en Figure 38). Une tendance claire de la toxicité en fonction de ces paramètres seulement n'a cependant pas pu être obtenue. L'utilisation de descripteurs rendant

compte de l'encombrement stérique et de son influence sur la formation de dimère pourrait résoudre le problème.

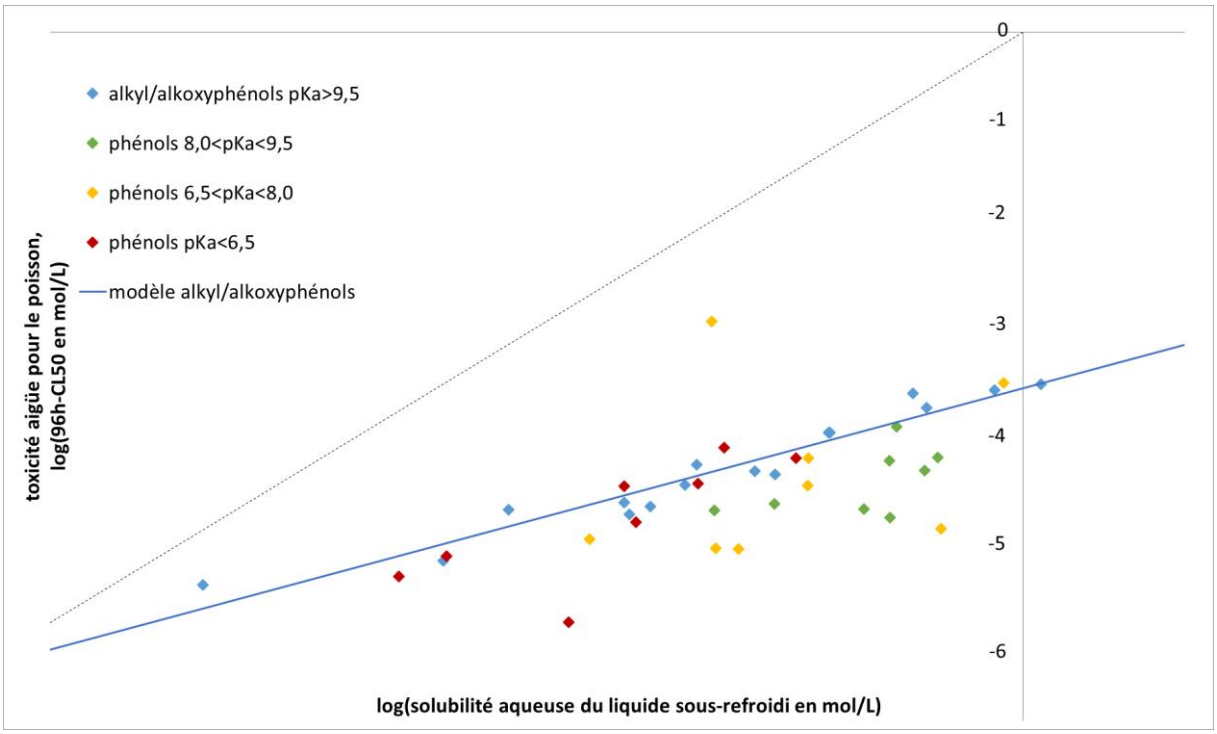


Figure 38 : toxicité des phénols acides pour le poisson en fonction de la solubilité et du pKa. Ces modèles sont la propriété de KREATiS.

D'autres composés ont été identifiés comme ayant l'effet de découpleurs de la phosphorylation oxydative, bien que leur structure, leur pKa, n'indique pas un tel mécanisme. Il s'agit, parmi les molécules étudiées dans ce travail, du 2,4,5-tribromoimidazole **40**, de la 2,4-dinitroaniline et de la pentachloroaniline. Le 2,4,5-tribromoimidazole a un pKa d'environ 1,6 (prédiction avec MarvinSketch 16.1.11, de ChemAxon (ChemAxon, 2016)), ce qui semble trop bas pour que les espèces protonée et non-protonée soient présentes aux pH de la mitochondrie. D'autre part, cette substance possède d'autres effets spécifiques (destruction des neurones, inhibition du quorum sensing) dont les mécanismes n'ont pas été élucidés (Verschoyle *et al.*, 1984). Pour expliquer l'effet de l'aniline et la pyridine qui sont très acides (pKa < 0), on peut peut-être émettre l'hypothèse que ces espèces se coordonnent au fer contenu dans certaines enzymes pour expliquer leur excès de toxicité, mais ceci ne semble pas spécifiquement lié à la phosphorylation oxydative.

7) Mécanisme responsable de la toxicité et corrosivité des acides et bases

La plupart des espèces traversent les membranes cellulaires par diffusion passive. Cette diffusion passive à travers une bicouche lipidique ne se fait efficacement qu'avec les molécules sous leur forme neutre. Cependant, au pH physiologique, la forme neutre n'est pas la forme majoritaire pour les acides et bases expliquées ci-dessus. En effet, le pH extracellulaire est en général de 7,4, et le pH du cytosol est d'environ 7,2 (Casey *et al.*, 2010). Pour ces acides et ces bases, une petite proportion de molécules sous leur forme neutre existe toutefois à l'équilibre, dans les deux compartiments. Ainsi, après que la forme neutre ait traversé la membrane, la molécule arrivant dans le cytosol va capter ou relâcher un proton pour passer sous la forme ionisée, rétablissant l'équilibre acido-basique, qui dépend de la différence entre le pH du milieu et le pKa des espèces chimiques introduites (Figure 39). Sur cette

figure, les acides protonés, électriquement neutres (rouge) peuvent traverser la membrane. Ils relâchent ensuite leur proton, capté par des molécules d'eau (en bleu), acidifiant le cytosol. Remarque : après le transfert de molécules d'acide, un autre acide dans le milieu extracellulaire se reprotone pour rétablir l'équilibre (en bleu). La conséquence directe de ce relargage ou captage de protons dans le cytosol est une acidification ou basification de la cellule, respectivement. Ce mécanisme est en partie limité par les pompes à protons et autres mécanismes qui régulent le pH de la cellule, mais ceux-ci seront dépassés lors d'une forte exposition à un acide ou une base. On peut aussi supposer que ces acides et bases ne vont pas s'arrêter au cytosol pour y modifier le pH, mais vont aussi se diriger vers les différents compartiments cellulaires, dont certains ont des pH très différents du cytosol, ce qui est nécessaire pour leur fonctionnement (mitochondrie à pH 8, endosomes à pH de plus en plus acide jusqu'à obtenir un lysosome de pH 4,7 environ).

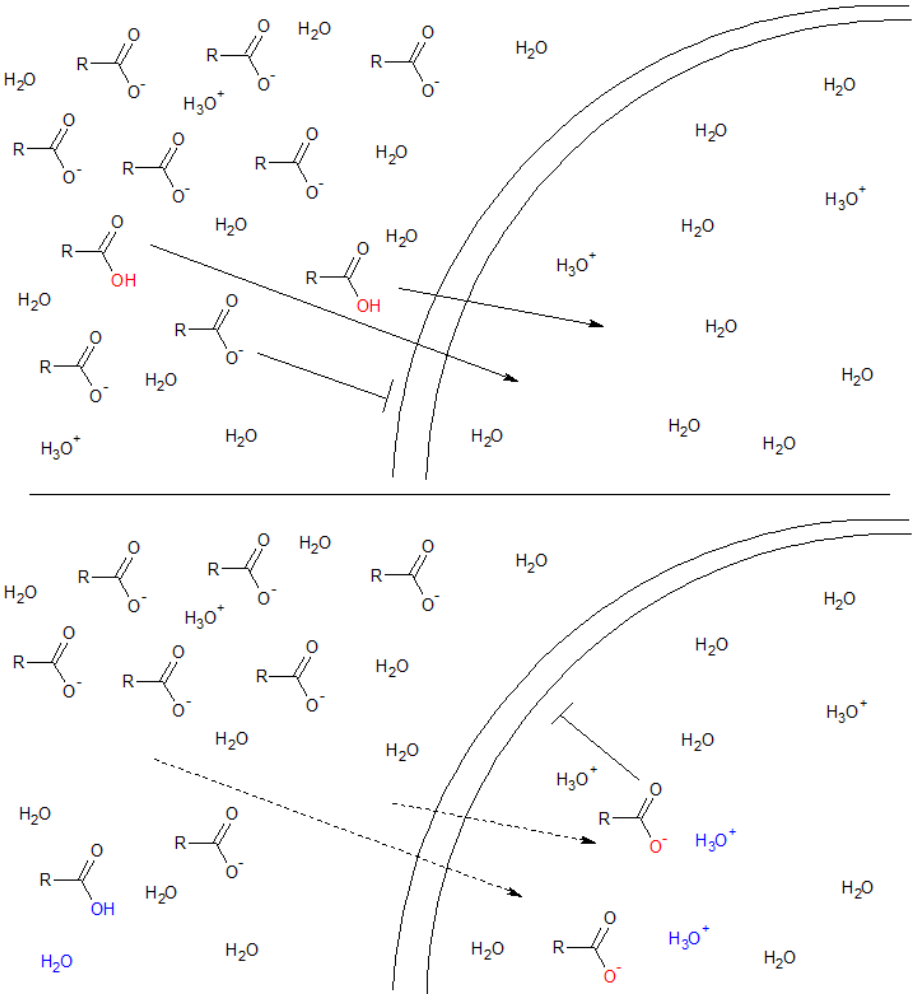


Figure 39 : acidification des cellules par les acides.

L'effet antipaludique de la quinine **41** et de ses analogues a été expliqué récemment par ce mécanisme de basification par Vallabh Minikel (Vallabh Minikel, 2016) : le parasite *Plasmodium falciparum*, responsable de la maladie, stocke les hèmes, qu'il n'arrive pas à dégrader, dans un compartiment acide, où ceux-ci cristallisent spontanément. La basification du compartiment par la quinine entraîne la libération des hèmes, qui, sous cette forme libre, vont être très toxiques pour le parasite et la cellule, en catalysant de manière non contrôlée et non spécifique des réactions d'oxydation, de formation de radicaux libres. Ce même auteur explique qu'une molécule basique comme la quinine va diffuser, sous forme neutre, jusqu'au compartiment le plus acide, afin d'y capter tous les protons. En s'écartant du contexte du paludisme, cela laisse penser que les bases assez fortes en général vont diffuser jusque

dans les lysosomes, qu'elles vont basifier. Cela aura pour conséquence de stopper l'action des protéases qui travaillent dans ces compartiments.

Ainsi, on peut émettre l'hypothèse que c'est en basifiant la cellule et ses différents compartiments que les bases en général sont cytotoxiques, et c'est vraisemblablement le mécanisme responsable de la corrosivité cutanée des bases (qui intervient pour les bases pures ou des solutions de pH>11). En extrapolant encore plus loin, on peut aussi penser que le mécanisme de toxicité des acides est en tous points similaire à celui des bases, si ce n'est qu'ils vont acidifier la cellule et ses compartiments (comme la matrice interne de la mitochondrie, qui est basique).

Chapitre 4 : Développement d'un modèle de prédiction des MechoAs par approche quantique

1) Introduction

Après avoir rassemblé un maximum de mécanismes d'action pour une variété de molécules, nous avons entrepris de développer une méthode de prédiction de ces mécanismes d'action. Nous avons choisi d'utiliser une méthode se basant sur des descripteurs quantiques, obtenus après optimisation de la géométrie des molécules en 3D, dans le but d'étendre le domaine d'applicabilité de la méthode plus loin qu'avec une méthode par alertes structurales sur la molécule en deux dimensions. Le but était également d'éviter les erreurs liées au fait de considérer un même MechoA pour des molécules qui ont la même alerte structurale, alors que leurs propriétés stéréo-électroniques sont différentes à cause d'autres groupements sur la molécule, et des MechoAs différents peuvent en résulter.

Nous avons développé dans un premier temps une méthode de prédiction des MechoAs généraux (voir Tableau 4), et au vu des bons résultats, nous l'avons ensuite étendu au niveau des sous-catégories, qui correspondent aux groupes nécessaires pour dériver des modèles QSARs. En étendant aux MechoAs détaillés, la méthode de classification a dû subir plusieurs modifications. Cette méthode a fait l'objet d'une publication soumise au journal Computational Toxicology, qui est présentée ci-dessous.

Les informations supplémentaires, référées comme « Supporting Information File » dans le manuscrit, sont placées en Annexes (A.1) Informations supplémentaires pour la publication « A New classification algorithm based on Mechanisms of Action. »)

2) Publication

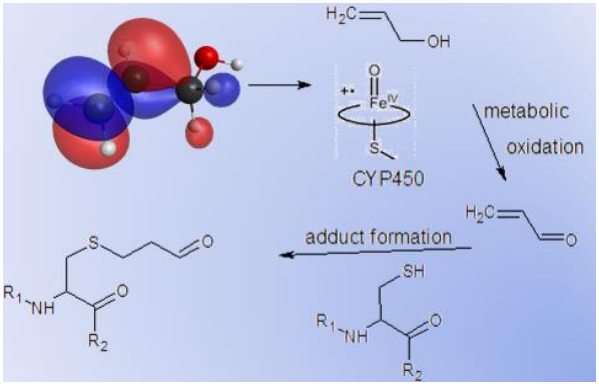
A New classification algorithm based on Mechanisms of Action.

Franklin J. BAUER*[a,b], Paul C. THOMAS[a], Samuel Y. FOUCHARD[b], Serge J.M. NEUNLIST[b]

[a] KREATiS SAS, 23 rue du Creuzat 38080 L'ISLE D'ABEAU, FRANCE
[b] Université de Haute-Alsace, Laboratoire de Chimie Organique et Bioorganique, EA4566, 2 rue des Frères Lumière F-68093 MULHOUSE CEDEX, FRANCE
* corresponding author: franklin.bauer@kreatis.eu ; +33428190106

Abstract

A good understanding of Mechanisms of Action (MechoAs) and appropriate methods to determine them is crucial for the accurate prediction of toxicity using *in silico* techniques. Different MechoAs can be related to different QSAR models to calculate toxicity values. Therefore, we defined a set of MechoAs, based on Molecular Initiating Events, the first step of Adverse Outcome Pathways (AOPs). In the most common classification algorithms used to predict Modes of toxic Action, the prediction domain is often limited by the need to identify known structural alerts. To circumvent this limitation, we developed a new algorithm to predict MechoAs principally based on mammals and fish, using molecular modelling to obtain calculated molecular parameters. Comparing the Verhaar scheme (as modified by Enoch *et al.* (2008)) with the MechoA method for the same validation set, MechoA achieved 69% correct classifications as opposed to 45% for the Verhaar scheme, 17% misclassifications for both, 13% classifications slightly different from the literature for our algorithm. No substances fell into zones where two possible MechoAs couldn't be differentiated from each other, while 1% of the molecules were out of the prediction domain of the algorithm as opposed to 38% using the Verhaar scheme. Thus, this model enhances precision of correct AOP identification for *in silico* toxicity predictions.



Keywords: mechanism of action; classification; QSAR; ecotoxicity; toxicity; quantum calculations

1. Introduction

Under current chemical regulations, the use of *in silico* techniques is encouraged and used to save time, cut costs and reduce animal testing (NC3Rs, n.d.). A good understanding of Mechanisms of Action (MechoAs) and appropriate methods to determine them is crucial for the efficient prediction of toxicity using *in silico* techniques, notably QSARs (Quantitative Structure-Activity Relationships) which are increasingly used to predict ecotoxicity, even occasionally as a regulatory endpoint to replace studies, as can be seen on the ECHA disseminated database (European Chemicals Agency, 2017). In this paper,

the terms Mode of Action (MoA) and MechoA have different meanings. MechoA refers to the molecular interaction that a molecule will undergo and that will be the key starting point of the Adverse Outcome Pathway (AOP) for this substance, i.e. the Molecular Initiating Event (MIE). In this publication, we prefer to use the abbreviation MechoA instead of MIE as we want to underline that these events can serve to classify the toxicity class of molecules (in the same way as MoAs), because they are the basis of how a substance impacts an organism's integrity, at a molecular level. On the other hand, MoA refers to the pathological effects that can be seen at the whole organism level in terms of behaviour or death i.e. at the other end of the AOP (Russom *et al.*, 1997). MechoAs and MoAs are linked by AOPs. But one MoA can be the result of very different MechoAs. Conversely, the same MechoA but different MoAs can happen mainly when comparing the AOPs across different species.

Most of the time, separate models are needed for each MechoA or chemical class to obtain acceptable levels of accuracy. However, in the Verhaar classification (Verhaar *et al.*, 1992), except for MoAs 1 & 2, MoA categories are more broadly defined so a specific QSAR cannot be allocated to a single MoA. Moreover, the most widely used method to determine MoAs, the Verhaar scheme (Verhaar *et al.*, 1992), does not always correctly allocate chemicals to the appropriate categories despite several updates and many common chemical classes still remain "unclassifiable" (MoA5) (Enoch *et al.*, 2008). Another classification scheme was published by Russom *et al.* (Russom *et al.*, 1997). This system is more detailed and could more easily be used to split chemicals into categories that can be predicted with the same QSAR model if it was made publicly available, but the number of categories in this scheme is also far from comprehensive. Finally, several methods to predict MoAs or MechoAs are available but lack reliability or have narrow applicability domains (Asadollahi-Baboli, 2012; Bearden and Schultz, 1997; Karabunarliev *et al.*, 1996; Verhaar *et al.*, 1992). For instance, a classification method using MechoAs, proposed by Bearden and Schultz (Bearden and Schultz, 1997), is quite reliable, with simple and justifiable descriptors, but with an applicability domain limited to benzenes, phenols, anilines and olefinic alcohols and their halo- and nitro- derivatives (together with a few other miscellaneous compounds). Consequently, there is a need for a classification scheme or algorithm with a wider applicability domain based on precise and relevant category definitions corresponding to well-understood mechanisms of action. A model matching these requirements appears useful, if not essential, to the appropriate use of QSARs (Russom *et al.*, 1997) leading to an accurate toxicity prediction. Nevertheless, certain classes of compounds are especially challenging to classify. For instance, non-polar and polar narcosis, two MoAs as defined by Verhaar and co-workers, are still not yet completely understood, even if several hypotheses and likely explanations have been proposed (Vaes *et al.*, 1998). As another example, the different toxicity mechanisms of phenols are not completely described (Rezg *et al.*, 2014) and the criteria to separate out the categories remain somewhat unclear (Netzeva *et al.*, 2008). Moreover, the classification methods already available apply to one species (or group of species), e.g. to fish (Karabunarliev *et al.*, 1996; Martin *et al.*, 2013; Russom *et al.*, 1997), or to aquatic organisms (Bearden and Schultz, 1997; Verhaar *et al.*, 1992), or to rodents (Kazius *et al.*, 2005), but in our work, we compiled data from various species, mostly rodents and fish, and thus, we derive the prediction of MechoAs for rodents and for fish (and additional information can be given for other species, but with lesser weight of evidence).

The aim of this work is to predict the MechoA only from the structure, just as previous methods have attempted to do with MoAs (Verhaar *et al.*, 1992; Russom *et al.*, 1997; Karabunarliev *et al.*, 1996). Consequently, it is reasonable to define chemical categories according to the MIE, because it is the first event and simplest thing we can predict in the whole AOP if we base our prediction only on the structure. Moreover, it may be easier to construct single QSAR models covering complete categories using MechoAs rather than MoAs. Our new MechoAs definitions were chiefly inspired by the work of Russom and co-workers (Russom *et al.*, 1997). In parallel, we developed a new *in silico* method to

predict the MechoA of a molecule from its structure using a tiered selection approach. This new method is applicable to a wide range of organic chemicals (hydrocarbons, halides, ethers, esters, ketones, aldehydes, alcohols, phenols, anilines, nitrobenzenes, carbamates, sulphur-containing compounds, organophosphorus compounds, etc.). The performance of the algorithm developed in this work was compared with that of Verhaar's scheme using a dataset taken from the literature.

1.1. MechoA definitions

For the purpose of this work we defined in a first step five general MechoAs, each one including chemicals with similar MIEs.

The general MechoAs names developed in this work are listed below together with a generic description:

1. Membrane destabilization: accumulation of molecules in cell membranes without specific reaction
2. Hydrolytic prodestabilization: a mixture of both direct accumulation and enzymatic hydrolysis generating membrane destabilizers
3. Reactive toxicity: non-enzymatic reactions with endogenous compounds (proteins, DNA)
4. Pro-active toxicity: metabolic transformation of the molecule into more toxic compounds
5. Indirect enzyme disruption: modification of the environment of an enzyme, preventing its normal activity
6. Direct docking interaction: binding to a docking site of a key protein (enzyme, receptor, ion channel).

Membrane destabilization corresponds to a direct narcotic effect, where cell membrane integrity is impaired by the accumulation of a xenobiotic in the lipid bilayer.

Hydrolytic prodestabilization relates to compounds that are not only membrane destabilizers, but further to enzymatic hydrolysis, generate also new destabilizer products.

Reactive toxicity occurs when a reactive compound creates adducts with proteins and/or DNA by covalent binding, either directly or through generation of reactive oxygen species.

Similarly to hydrolytic prodestabilization, pro-active toxicity is the MechoA of compounds that are first metabolised to become highly toxic compounds, i.e. reactants (MechoA 3), indirect enzyme disruptors (MechoA 5) and direct dockers (MechoA 6).

Enzymes function under specific conditions. Indirect enzyme disruption occurs when these conditions, such as pH, or the relative abundancy of the oxidized and reduced forms of a cofactor, are changed.

Direct docking interaction includes all mechanisms that involve a specific interaction with enzymes, receptors or ion channels by binding to a docking site (which can be the active site or an allosteric site). This can result in the inhibition or blocking of the target protein or on the contrary the induction of it.

MechoA 2, which includes an enzymatic hydrolysis, is actually an example of "Metabolic transformation of the molecule" for MechoA 4, but this was separated out from MechoA 4 for two reasons:

First, hydrolysis splits the molecule into two compounds which are more hydrophilic than the parent, decreasing the narcotic effect, which is mostly a detoxification process. However, hydrolysis can generate highly toxic compounds in some cases. This is the case for dinoseb acetate which is rapidly and quantitatively hydrolysed when administered to rats, releasing dinoseb, which is an oxidative

phosphorylation uncoupler (*Pesticide Residues in Food, 1985 Report Sponsored Jointly by FAO and WHO Evaluation*, 1986).

Second, MechoA 4 concerns the molecules which are activated by metabolism to give more active compounds. However, some exceptions may occur: metabolic oxidation, not hydrolysis, can also generate a less reactive or less toxic compound as is the case for benzaldehyde which is rapidly oxidized to benzoic acid (Nair, 2001).

These two exceptions leading to different outcomes explains the need for subcategorization of the generic MechoA framework described above.

However, even if they share a similar or have the same MIE, the different categories defined in the list above needed to be split further into subgroups as presented in table 1 such that one single MechoA subgroup relates to one QSAR model per endpoint (e.g. daphnid 48h EC50 value). The numbering of the subgroups in MechoA 4 are consistent with the MechoA of the product generated by the metabolisation of the parent compound. Thus, MechoA 4.1 refers to compounds that will be transformed into MechoA 1 products. For MechoA 4.3, the end products are reactants (MechoA 3) and so on. The MechoA of RedOx cycling is assigned the number 4.4 as it is metabolised repeatedly in a cyclic way. This mechanism is harmful, because the xenobiotics will metabolise between a reduced and an oxidized form, generating reactive oxygen species (ROS) each time, and depleting NAD(P)H stocks (Cohen and d'Arcy Doherty, 1987).

MechoA	Subgroup	Mechanism detail
1: membrane destabilization	1.4 Nonpolar narcotics 1.5 polar narcotics	Physical perturbation of membrane integrity with different affinities for non-polar and polar.
2: hydrolytic prodestabilization	2.1 hydrolysis to destabilizers 2.2 hydrolysis to actives	Direct membrane destabilization + Enzymatic hydrolysis, giving: 2.1 only narcotic products 2.2 at least one non-narcotic product
3: reactive toxicity	3.1 hard electrophiles 3.2 soft electrophiles 3.3 spontaneous radical-generating compounds	3.1 adduct formation with amino and thiol protein residues 3.2 adduct formation through Michaël addition with thiol protein residues 3.3 homolytic cleavage of a weak bond generating 2 radicals, generating oxidative stress
4: pro-active toxicity	4.1 readily detoxified compounds 4.3 pro-reactants 4.4 RedOx cyclers 4.5 indirect pro-disruptors 4.6 direct pro-dockers	4.1 the metabolised compound has a MechoA 1 4.3 the metabolised compound has a MechoA 3 4.4 compounds undergoing RedOx cycling, generating oxidative stress 4.5 the metabolised compound has a MechoA 5 4.6 the metabolised compound has a MechoA 6
5: indirect enzyme disruption	5.1 Oxidative Phosphorylation Uncouplers 5.2 acids and bases 5.3 other mechanisms	5.1 cyclic transport of H+ across the mitochondria inner membrane, preventing the generation of ATP 5.2 diffusion across cell membranes in neutral form, and then release/retrieval of H+ in cytosol
6: Direct docking site interaction	6.1 AChE inhibitors 6.2 AChR binders	6.1 inhibition of AcetylCholineEsterase

	6.3 dopamine transport disruptors 6.4 metal chelators 6.5 Photosystem II ET inhibitors 6.6 ion channel modulators 6.7 other mechanisms	6.2 binding to nicotinic or muscarinic AcetylCholine Receptors, as an agonist or antagonist 6.3 induction of the transport of dopamine into the synapse and inhibition of its recycling back into the nerve cell 6.4 chelation of metal ions, such as Fe^{2+}, thus blocking or degrading the enzymes containing this metal ion in their active site 6.5 inhibition of a protein involved in the electron transport chain in the photosystem II of plant cells 6.6 docking to an ion channel's binding site causing its activation or deactivation

Table 1: MechoA subgroups.

The designation of these detailed MechoAs is important, because they separate out compounds that need specific QSAR models to predict their toxicity. However, we will also need to develop the subcategories in the MechoA 6 group which contains various enzyme-specific mechanisms.

Some molecules present several MechoAs for the same or for different species and timeframes, therefore several notation rules (which are included in the supplementary information file) have been used to describe the full MechoA information for each substance: A 'C' before the number of the MechoA means that it is only observable in chronic studies, but not in acute. By default, if no 'C' is present, the MechoA is applicable to the acute timeframe. Some abbreviations are also put before the number of the MechoA to designate species-specific MechoAs. 'm' stands for mammals, 'h' for humans, 'f' for fish, 'p' for plants, 'u' for unicellular organisms, 'd' for daphnia. '#' before these letters means "for all species except these". For example, #up5.2 means MechoA 5.2 for all species but unicellular and plants. '&' means both MechoAs are playing a role. '/' means that it is uncertain whether one or several of the MechoAs are present and are critical for toxicity to occur. They are written in the order of decreasing certainty, where the first one is the most probable. For instance, 1.2 & m4.3 / 3.1 means that this compound is thought to exert polar narcosis (1.2), with metabolism to reactive compounds for mammals (m4.3), and may be able to act through a direct reactive mechanism (3.1). Finally, we defined an order for different species: the MechoA for fish and mammals are put first, the other species information comes afterwards. If equivalent certainty exists for Fish and mammals, fish MechoA is put first.

2. Materials and Methods

2.1. Literature data

A set of chemicals were assigned MechoAs based on outcomes of experimental results from the literature. The data comes from various sources, disseminated publications, often focussing on one compound or chemical group. All the references are given in the supplementary information file (Table S1 and S2). Some experimental studies have focussed on events occurring in the organism at the molecular level (e.g. formation of adducts by Michaël addition with amino residues of proteins), and these have given highly reliable MechoA assignments. Other studies focussed on probable reactivity of molecules coupled with observed organism or tissue response (e.g. animal behaviour such as dizziness, convulsions or liver tissue damage) and these gave less reliable MechoA assignments. Finally,

for some molecules, we could only base our MechoA assignment on EC50s (concentration needed to produce the measured effect on 50% of the animals exposed to the substance), comparing the experimental EC50 to the EC50 that would be predicted using a High-Accuracy QSAR model for non-polar narcotics for instance. In addition, chemical knowledge (in terms of reactivity or vulnerability to a certain enzymatic reaction) about each kind of structure was used to determine more precisely the MechoA even if this level of MechoA assignment is currently less reliable.

To develop the classification algorithm, we used a training set of 201 molecules, containing only organic substances, with compounds from different chemical industries, notably petrochemistry, cosmetics, pesticides and a few pharmaceuticals. In addition, a validation set of 100 molecules was constituted mainly from the comparison set used in Enoch *et al*., 2008 (Enoch *et al.*, 2008). The whole dataset presented in this publication was not used (more than 400 molecules) since it was not considered necessary to construct the model and it would have required significant resources and time to perform the calculations. In fact, as we calculate the optimal geometry for each molecule with the method RHF 6-31G(d,p), the calculation times can be very long (days, with our hardware) with increasing number of atoms, and for this reason we limited the size of the dataset. We used the following rules to select molecules: i) we tried to keep the highest molecular diversity by excluding preferentially homologues of a same family (e.g. butan-1-ol, pentan-1-ol, hexan-1-ol, etc.). ii) when we had the choice, we preferred to keep molecules with lower molecular weight to save time of calculation (e.g. vanillin kept in our dataset and 4-(hexyloxy)-m-anisaldehyde excluded), iii) while trying to keep representatives of the majority of chemical families and MechoAs, for example, resmethrin (MechoA 5.6: keeps sodium channels open in the nerves) was kept in our dataset although it is quite a big molecule and needed a long calculation time. Nevertheless, we added some molecules to this validation set that were not present in the original comparison set of Enoch *et al.* in order to obtain a similar variety in our validation set to that in our training set. The datasets we used for this work can be retrieved from the supplementary information file (Table S1 and S2).

2.2. Computational methods

The software used for the molecular modelling was Firefly (formerly PC-GAMESS) (Granovsky, 2014; Schmidt *et al.*, 1993), and Avogadro (Hanwell *et al.*, 2013, 2012) was used to generate INPUT files for Firefly. The visualisation of molecular orbitals was obtained using wxMacMolPlt (Bode and Gordon, 1998). The equilibrium geometry of molecules was determined using several calculation steps. First, a molecular mechanics method was used in Avogadro to rapidly obtain a pre-optimised structure (steepest descent with UFF force field). Then, this was input into Firefly to calculate the optimal geometry with the semi-empirical method PM3. And finally, to achieve a high level of accuracy, the most time-consuming step was taken, in which the equilibrium geometry (in void) was determined with a Restricted Hartree-Fock (RHF) method, using the basis set 6-31G(d,p). As a result of this calculation, the 3D coordinates of atoms are given and different molecular parameters are calculated by the software, among which are E(LUMO), E(HOMO), qMaxH, qMax nonH, qMax-, which can be retrieved directly in the OUTPUT file of Firefly. These parameters were then used to discriminate between compounds in the different splitting steps. The final calculation method, RHF 6-31G(d,p), was compared with other methods in terms of calculation time and accuracy of results. We used B3LYP and MP2 post-Hartree-Fock method as references for accurate results. These 2 methods were of course more time-consuming than the simple RHF calculation. We tried RHF 6-31G(d) to reduce calculation time, but the loss of accuracy was judged significant, thus we preferred to retain the use of 6-31G(d,p) as a good compromise between calculation time and accuracy.

qMax nonH is the most positive charge not held by a Hydrogen atom. E(LUMO) is the energy of the Lowest Unoccupied Molecular Orbital. The prefix 'a' means that the parameters are calculated for the

activated molecule, whenever applicable. A high qMax nonH indicates that there is a strong electron deficiency on one atom, which makes it a good site for nucleophilic attack. A low aE(LUMO) is an indication for a high electrophilic potential. Both properties, i.e. high qMax nonH and low aE(LUMO), are needed to obtain a very strong electrophile, readily attacked by nucleophiles (e.g. water or amino residues or thiol residues). This combination of parameters is used to make the first split of molecules into different categories. A short explanation of the other molecular parameters used in this classification algorithm is given below:

- qMaxH is the maximal partial charge of a hydrogen atom in the molecule. A high qMaxH characterizes good hydrogen-bond donors. Useful to discriminate between different categories of ethers, alcohols and amines in one group, and also to separate out nitrobenzenes, anilines and phenols in another group.
- qMax- is the maximal negative partial charge in the molecule. A high qMax- characterizes good hydrogen-bond acceptors. High qMax- with high qMax nonH indicates a high substituent polarity. This combination helps to separate out hydrolytic prodestabilizers and AChE inhibitors.
- E(HOMO) is the energy of the highest occupied molecular orbital. E(HOMO), together with E(LUMO) discriminates between hard and soft electrophiles, as already demonstrated in the literature (LoPachin and Gavin, 2014).

We also calculated the acidity constant of the conjugated acid of molecules (pKa(BH^+)), i.e. the basicity of the substance itself. This is the first step which was implemented to separate readily and poorly metabolizable chemicals rather than to define them as ionising and non-ionising. We calculated pKa(BH^+) rather than the basicity constant (pKb) of molecules so as to compare these values more easily with those given in the literature. The pKas were calculated with the software MarvinSketch v. 16.1.11 from ChemAxon (ChemAxon, 2016). However, for some compounds this model was unable to predict a pKa. Consequently, for these molecules, a Read-Across approach was used with one or several similar molecules for which we could obtain a value, and then a final value was extrapolated using knowledge of electronic properties. Examples of this approach are given in the table 2. The figure S21 in the supplementary information file shows the predicted against the experimental pKas and indicates the cut-off value of pKa = -7.5 with the orange dotted line.

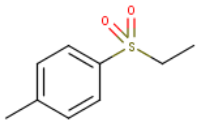
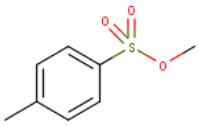
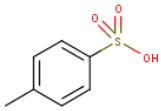
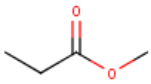
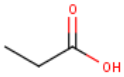
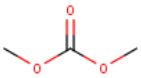
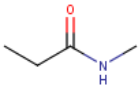
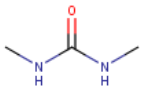
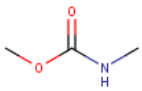
Molecule	Pred. pKa (ChemAxon)	pKa from Read-Across and extrapolation
	-12.88	
	-9.64	
		-9
	-7.02	
		-7
	-9.66	
	-1.61	
	-2.01	
		-3

Table 2: estimation of specific pKas by Read-Across.

2.3. Classification algorithm

We designed the first step of this algorithm to be an activation ability calculation because most of the reactions occurring in living organisms with xenobiotics are thought to be processed by an enzyme, which first activates its substrate with a metallic cation included in the enzyme active site. A molecule bound to a metallic cation is called “activated” because it makes it a stronger electrophile, which can be attacked more easily by a nucleophile than the unbound molecule. In order to model this activation, it was decided to calculate the basicity, i.e. the ability of a molecule to create a bond with a proton. We chose the proton instead of a metallic cation as it is easier to model and because the pKa calculated can then be compared with experimental values, to ensure the method reliability. A cut-off value was empirically determined to be of pKa(BH+) = -7.5. Molecules having a pKa above this cut-off were known to be activated by a metallic cation, such as Mg2+ or Fe3+ in order to be transformed by the enzyme which holds this cation in its active site. Molecules with a pKa below -7.5 and non-protonatable substances do not appear to be readily attacked by enzymes. Consequently, in the next steps of the

algorithm, the parameters of the molecules which had a pKa greater than -7.5 were calculated for the protonated molecule (i.e. the activated molecule), while for molecules with a pKa lower than -7.5, the parameters were calculated for the non-protonated molecule only. Given the poor quality of experimental pKa values below zero (because of the need to carry out measurements in another solvent and then extrapolating into water), we chose to trust the prediction rather than the experimental value in all cases, even if the experimental results are more reliable than predictions for high pKa values (greater than -3).

The chemicals of the training set were then used as markers to designate the categories graphically determined using quantic descriptors. The first splitting step uses 'qMax nonH' and aE(LUMO). Combining these two parameters enables us to split molecules into 5 groups. Then the following steps are different for each group and are presented in Figure 2. An example of a sequence of splitting steps can be found in Figure 3. We will take an example compound, that can be followed using the same figure. The compound is methyl benzoate, which is included in the training set. The structure of methyl benzoate is also shown in Figure 3. The pKa of methyl benzoate, as calculated by the pKa module of MarvinSketch, is -6.88, and the site of protonation is the oxygen of the carbonyl. Consequently, it has to be protonated for the calculation of aE(LUMO). It has a value of aE(LUMO) of -0.1024, so it falls into group E. It has qMaxH = 0.195048 and qMax nonH = 0.846978, therefore it falls into category E.1. It has E(HOMO) = -0.3410, which places it in the top-left side of the final graph, in a position where molecules have MechoA 2.1. Conclusion: MechoA 2.1 (hydrolysis to narcotic products).

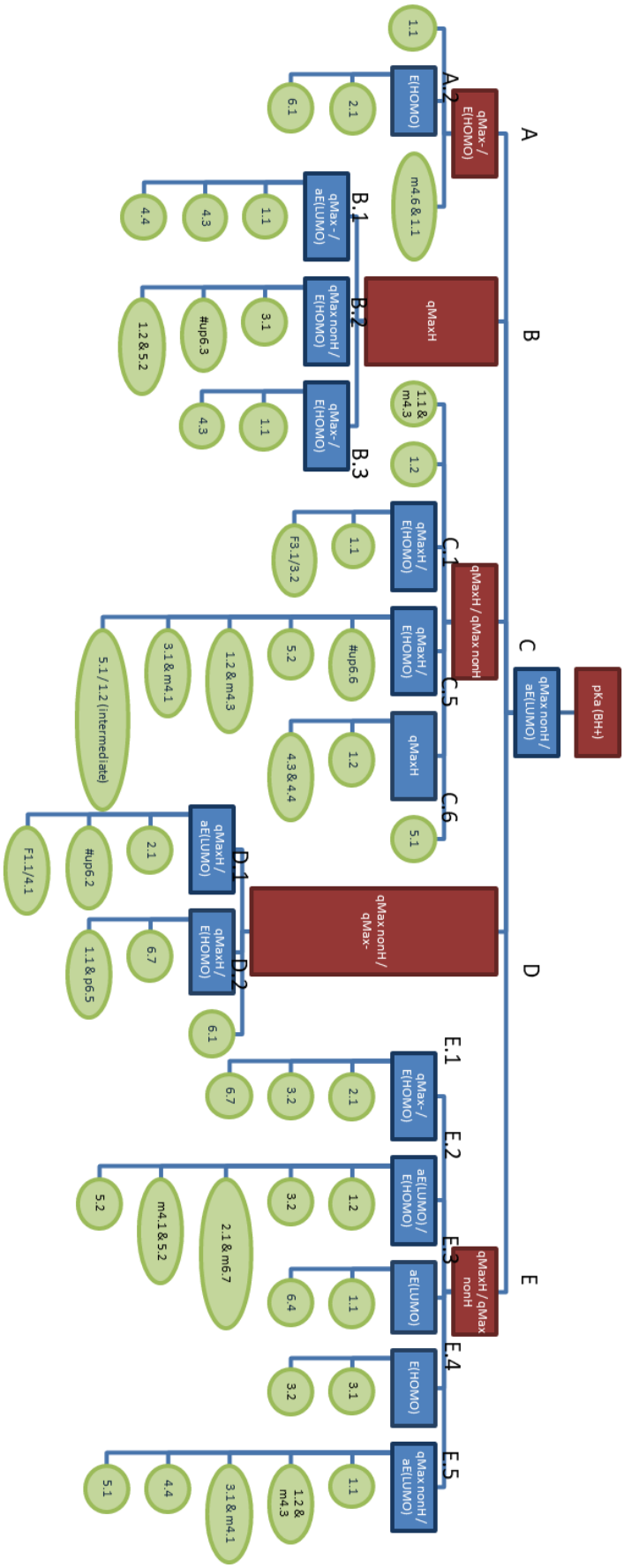


Figure 2: Splitting steps for the classification algorithm. Rectangles = parameter(s) used for splitting; Ellipses: MechoA of this group. Fx/y means Fuzzy zone with MechoA x and y not separated.

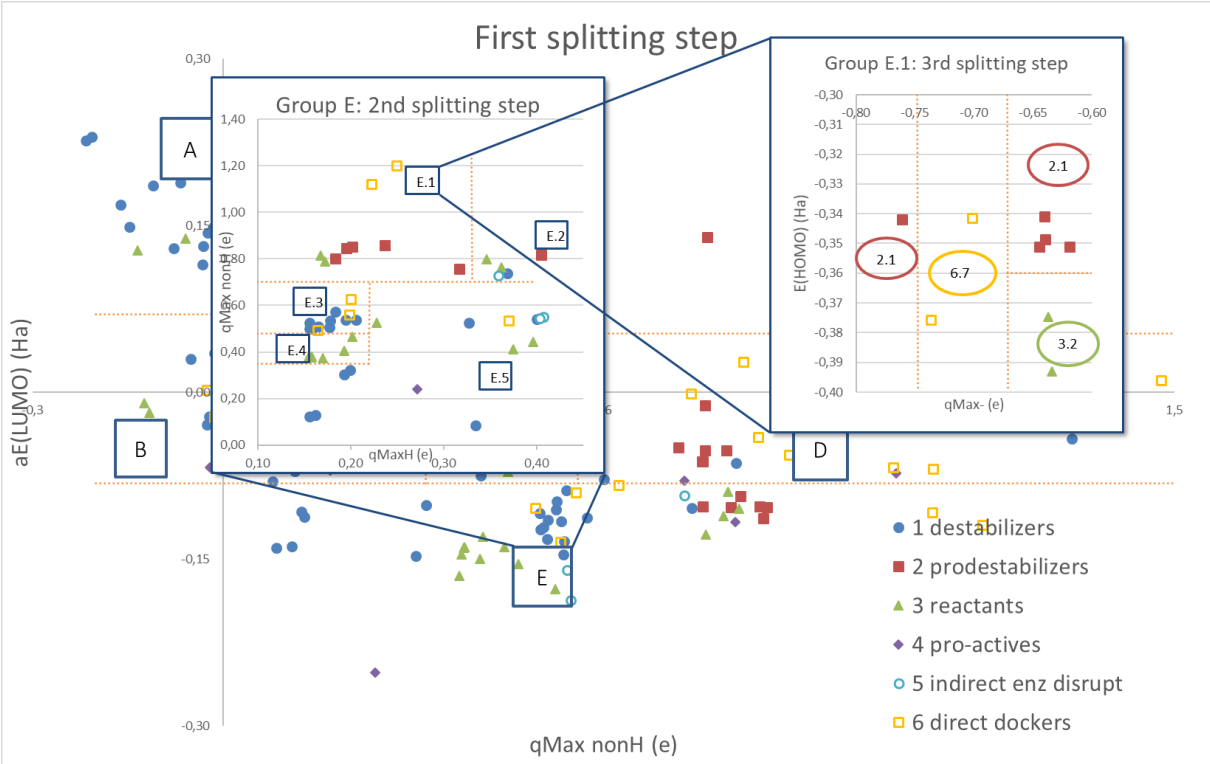


Figure 3: Example of one path in the MechoA mapping. The groups obtained with each step are delimited with orange dotted lines. Energies are expressed in Hartrees (Ha), and partial charges in elementary charge (e).

3. Results and discussion

A maximum of three consecutive splitting steps was needed to classify each molecule. A total of 20 different splitting steps have been used to achieve this degree of molecular coverage. The workflow used for the MechoA mapping is presented in Figure 2. An example of the splitting steps used in one path is given in Figure 3. The individual splitting steps are displayed graphically in the supplementary information file (Figure S1-20), with the points justifying the separation criteria and the numerical values of the thresholds to discriminate between two zones.

Only one molecule from the comparison set (same as validation set of the model) fell outside of the prediction domain (Unclassified). Nevertheless, even with 3 consecutive steps, some groups of molecules could not be separated according to their observed MechoA. These remaining, uncertain zones have been named 'Fuzzy zones' and will be the subject of subsequent research. There are two zones of this kind containing only two to three points each. One of these fuzzy zones contain two similar MechoAs (hard and soft electrophiles). Due to the small number of points and/or the similar values of all the parameters, a method to separate them out that would be applicable to new molecules is still being sought. A restrictive approach would be to classify the compounds in these fuzzy zones as being out of the applicability domain, i.e. Unclassified. However, we thought that it is still better to know that a given molecule acts either by one of two possible MechoAs (e.g. MechoA 3.1 or MechoA 3.2), rather than concluding that this molecule cannot be classified by the algorithm at all.

Five categories were defined to derive the statistics of the algorithm. The rules to decide if one substance should be attributed to one of these categories are given below:

- **Correctly Classified (CC):** if the first MechoA determined by the prediction is the same as the first MechoA specified in the literature referenced MechoA. Also, if all the specified MechoAs correspond, even if they are not in the same order.

- **Slightly Different:** If the general MechoA is correctly predicted (i.e. MechoA 1, 2, 3, 4 or 5), but not the MechoA subgroup (e.g. 1.1 instead of 1.2). This is also the case when the substance falls in a fuzzy zone between two MechoA subgroups included in the same general MechoA, provided the literature MechoA of this substance is one of the MechoAs of the fuzzy zone. A MechoA prediction is also said to be Slightly Different if the predicted MechoA corresponds to an uncertain MechoA from the literature but not to the more probable one (e.g. lit. MechoA = 1.1 / 5.2 ; pred. MechoA = 5.2).
- **Fuzzy zone:** A domain in the classification algorithm where two (or more) MechoAs could not be separated. The substances falling into such zones are assigned the term "fuzzy zone" and are abbreviated as follows: F2.1/5.2, in the case of the MechoAs 2.1 and 5.2 unseparated. Note that the "F" is in upper case, contrarily to the "f" indicating that the MechoA is restricted to fish.
  If a molecule falls into a fuzzy zone while its literature referenced MechoA is different from the MechoAs of the fuzzy zone (e.g. lit. MechoA 5.2 ; pred. MechoA F1.1/4.1), it is assigned "Misclassified".
- **MisClassified:** when the predicted MechoA does not match the literature referenced MechoA.
- **Unclassified:** when one descriptor value of the substance is too far from the values of this descriptor used in the training set. For this part, an outlier detection could be performed, using the k Nearest Neighbours approach. Moreover, if no value could be assigned for the descriptor used in a splitting step (qMaxH can't be assigned if there is no Hydrogen atom in the molecule), the substance is also unclassified.

The statistics on the model are presented in Table 3:

		Percentage of the set	
		This algorithm	**Verhaar's scheme**
Training set (N=201)	Correctly Classified	86.0	
	MisClassified	5.0	
	Slightly Different	7.5	
	Fuzzy Zone	1.5	
	UnClassified	0.0	
Validation set (N=100)	Correctly Classified	69.3	44.6
	MisClassified	16.8	17.8
	Slightly Different	12.9	NA
	Fuzzy Zone	0.0	NA
	UnClassified	1.0	37.6

Table 3: Statistics of the model and comparison of the algorithm developed in this work and Verhaar's method. The expected MechoAs were determined from our literature research. The expected MoAs for Verhaar's scheme were taken from Enoch et al, 2008.

Based on the MoAs given *a priori* by Enoch and co-workers, the statistical success rate of Verhaar's scheme applied to this comparison set were calculated and compared to those obtained with the algorithm developed in this work (Table 3). The algorithm presented here has more statistical power and appears to be a good replacement of Verhaar's scheme. The new method developed in this work is of course more demanding in terms of time and resources. It uses quantum calculations with the RHF method in order to obtain values for molecular parameters, while Verhaar's method has the advantage of being very easy to use because it applies simple rules related to the 2D structure. In fact, we already implemented one structural alert rule in our scheme which is similar to the Verhaar rules.

This takes place before the splitting steps using molecular parameters. The rule directly classifies compounds having an X-X simple bond, X being O (oxygen), N (Nitrogen) or S (Sulfur) as having a MechoA 3.3 for O and S, and MechoA 3.1 & 4.3 for N.

However, some notable errors and limitations were detected in our method for well-defined classes of compounds. Firstly, we couldn't separate out epoxides from other ethers in terms of their quantum parameters, which led to these compounds being classified as non-polar narcotics, while in fact they are hard electrophile reactants. Secondly, aliphatic amines possessing an aromatic cycle may be either neurotoxic or simply neurotransmitters (e.g. serotonin, amphetamine, morphine, nicotine) while others are just polar narcotics (e.g. benzylamine, (3-phenylpropyl)methylamine) (see Tables S1 and S2 for details and references). The difference of activity between these amines comes from their docking ability to specific receptors while their electronic parameters remain very similar and these biological factors are not accounted for in the quantic model. Thirdly, 1,2- and 1,4-dihydroxybenzenes (e.g. catechol, hydroquinone) are predicted, as are simple phenols, to act through polar narcosis while they actually are RedOx cycling compounds which are much more toxic. And finally, benzyl and allyl halides (e.g. a,a'-dichloro-p-xylene, benzyl chloride, 3,4-dichlorobut-1-ene) are predicted as non-polar narcotics while they are hard electrophiles. All these errors could be circumvented with structural alert rules, suggesting that probably neither a model with only quantum calculations nor a model with only structural alert rules would be the ideal way to classify compounds, but an intelligent combination of both may perform best. It should also be noted that the model currently doesn't predict endocrine disruption, which would be included in MechoA 6.

Data used to assign literature MechoAs were sourced from human, mammalian (mainly rat and mouse), fish, invertebrate (mainly daphnid), algae, plant (mainly grass species) and protozoan toxicity data, with rats (and mice) and fish predominate. Therefore, the results from our algorithm would correlate better with these species, while the other species just provide supplementary information, with lower weight of evidence. We observed that some AOPs are well conserved across species, particularly for narcotics, while for certain compounds with non-narcotic MechoAs, the outcome is clearly different between species. This often occurs when the compound metabolises. This is not surprising because the metabolome can vary dramatically between species, for instance, between algae and rats. The essential enzymes between the two species may be slightly different or some enzymatic processes of one organism are not present in another (e.g. algae photosynthesise but rats don't). However, further work is necessary in order to obtain more precise insight into the interspecies character of certain MechoAs, and to identify which chemical classes undergo very different toxicities to different species.

The algorithm is built with, and applicable to, acute fish, *Daphnia* and algae toxicity tests, acute and chronic oral and inhalation toxicity tests on mammals. The scheme can also be applied to other types of study, but should be restricted to tests that allow the whole organism to be involved in the process. We cannot, at this stage, suppose that these mechanisms will play a crucial role in local toxicity tests such as skin irritation for instance, where for example cytotoxicity is the main reason for effects (OECD, 2015).

4. Conclusions

We developed an algorithm to predict Mechanisms of toxic Action (MechoAs), based on experimental results, mainly on mammals and fish. The algorithm is based on quantum calculations which provide high quality structural and electronic parameters, used as input values to filter molecules through a set of consecutive splitting steps. The dataset was divided into groups of common MechoA by using total of 20 splitting processes where a maximum of just three splitting steps were required to obtain a

final classification for each individual substance. Further to statistical analysis of correct and incorrect classification, we achieved 69% correct classifications using our method for our validation set, while the Verhaar scheme achieved only 45% correct classifications for the same dataset. The results of the method presented here are encouraging, and indicate that the widely used Verhaar scheme could be improved upon and extended. We are also developing a structural alerts method, similar to that of the Verhaar scheme, except that it is based on molecular initiating events, which can be adapted to make it fast and usable on any computer. This other method is showing promising results and will be the subject of our next publication.

Acknowledgments

This work is part of a CIFRE doctoral contract funded by the French government.

References

Asadollahi-Baboli, M., 2012. Exploring QSTR analysis of the toxicity of phenols and thiophenols using machine learning methods. Environ. Toxicol. Pharmacol. 34, 826–831. doi:10.1016/j.etap.2012.09.003

Bearden, A.P., Schultz, T.W., 1997. Structure-activity relationships for Pimephales and Tetrahymena: A mechanism of action approach. Environ. Toxicol. Chem. 16, 1311–1317. doi:10.1002/etc.5620160629

Bode, B.M., Gordon, M.S., 1998. Macmolplt: a graphical user interface for GAMESS. J. Mol. Graph. Model. 16, 133–138. doi:10.1016/S1093-3263(99)00002-9

Bode, B.M., Gordon, M.S., n.d. wxMacMolPlt.

ChemAxon, 2016. MarvinSketch. ChemAxon (http://www.chemaxon.com).

Cohen, G.M., d'Arcy Doherty, M., 1987. Free radical mediated cell toxicity by redox cycling chemicals. Br. J. Cancer. Suppl. 8, 46–52.

Enoch, S.J., Hewitt, M., Cronin, M.T.D., Azam, S., Madden, J.C., 2008. Classification of chemicals according to mechanism of aquatic toxicity: An evaluation of the implementation of the Verhaar scheme in Toxtree. Chemosphere 73, 243–248. doi:10.1016/j.chemosphere.2008.06.052

European Chemicals Agency, 2017. Registered substances [WWW Document]. ECHA. URL https://echa.europa.eu/information-on-chemicals/registered-substances (accessed 1.16.17).

Granovsky, A.A., 2014. Firefly.

Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E., Hutchison, G.R., 2013. Avogadro.

Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E., Hutchison, G.R., 2012. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J. Cheminformatics 4, 17. doi:10.1186/1758-2946-4-17

Karabunarliev, S., Mekenyan, O.G., Karcher, W., Russom, C.L., Bradbury, S.P., 1996. Quantum-chemical Descriptors for Estimating the Acute Toxicity of Substituted Benzenes to the Guppy (Poecilia reticulata) and Fathead Minnow (Pimephales promelas). Quant. Struct.-Act. Relatsh. 15, 311–320. doi:10.1002/qsar.19960150406

Kazius, J., McGuire, R., Bursi, R., 2005. Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem. 48, 312–320. doi:10.1021/jm040835a

LoPachin, R.M., Gavin, T., 2014. Molecular Mechanisms of Aldehyde Toxicity: A Chemical Perspective. Chem. Res. Toxicol. 27, 1081–1091. doi:10.1021/tx5001046

Martin, T.M., Grulke, C.M., Young, D.M., Russom, C.L., Wang, N.Y., Jackson, C.R., Barron, M.G., 2013. Prediction of Aquatic Toxicity Mode of Action Using Linear Discriminant and Random Forest Models. J. Chem. Inf. Model. 53, 2229–2239. doi:10.1021/ci400267h

Nair, B., 2001. Final report on the safety assessment of Benzyl Alcohol, Benzoic Acid, and Sodium Benzoate. Int. J. Toxicol. 20 Suppl 3, 23–50.

NC3Rs, n.d. The 3Rs [WWW Document]. URL https://www.nc3rs.org.uk/the-3rs (accessed 6.13.16).

Netzeva, T.I., Pavan, M., Worth, A.P., 2008. Review of (Quantitative) Structure–Activity Relationships for Acute Aquatic Toxicity. QSAR Comb. Sci. 27, 77–90. doi:10.1002/qsar.200710099

OECD, 2015. Test No. 439: In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method. Organisation for Economic Co-operation and Development, Paris.

Pesticide Residues in Food, 1985 Report Sponsored Jointly by FAO and WHO Evaluation: Toxicology, 1986. . Food & Agriculture Org.

Rezg, R., El-Fazaa, S., Gharbi, N., Mornagui, B., 2014. Bisphenol A and human chronic diseases: Current evidences, possible mechanisms, and future perspectives. Environ. Int. 64, 83–90. doi:10.1016/j.envint.2013.12.007

Russom, C.L., Bradbury, S.P., Broderius, S.J., Hammermeister, D.E., Drummond, R.A., 1997. Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (Pimephales promelas). Environ. Toxicol. Chem. 16, 948–967. doi:10.1002/etc.5620160514

Schmidt, M.W., Baldridge, K.K., Boatz, J.A., Elbert, S.T., Gordon, M.S., Jensen, J.H., Koseki, S., Matsunaga, N., Nguyen, K.A., Su, S., Windus, T.L., Dupuis, M., Montgomery, J.A., 1993. General atomic and molecular electronic structure system. J. Comput. Chem. 14, 1347–1363. doi:10.1002/jcc.540141112

Vaes, W.H., Ramos, E.U., Verhaar, H.J., Hermens, J.L., 1998. Acute toxicity of nonpolar versus polar narcosis: Is there a difference? Environ. Toxicol. Chem. 17, 1380–1384.

Verhaar, H.J.M., van Leeuwen, C.J., Hermens, J.L.M., 1992. Classifying environmental pollutants. Chemosphere 25, 471–491. doi:10.1016/0045-6535(92)90280-5

3) Conclusion

Dans cette publication, nous avons ainsi exposé une méthode de prédiction des MechoAs basée uniquement sur la structure. Ce travail était basé sur 22 mécanismes d'action détaillés, présentés dans le tableau 1 de la publication. Les MechoAs de chaque substance ont été obtenu à partir de diverses références de la littérature, telles que données dans le chapitre 2. De plus, les MechoAs ont pu être complétés en utilisant des résultats d'études écotoxicologiques, couplés avec les modèles QSAR de KREATiS. Chaque modèle étant associé à un mécanisme d'action, si les données expérimentales correspondent à la prédiction selon un modèle donné, cela suggère que la substance agit selon le MechoA des substances utilisées pour construire ce modèle. Ensuite, pour prédire les MechoAs ainsi obtenus de la littérature, nous avons fait des calculs quantiques pour optimiser la géométrie de la structure tridimensionnelle de chaque molécule, afin de générer des paramètres électroniques tels que les énergies des orbitales moléculaires la plus haute occupée (E(HOMO)) et la plus basse vacante (E(LUMO)), et des charges partielles. La méthode de calcul utilisée était RHF (Restricted Hartree-Fock) avec le set de base 6-31G(d,p). Ces propriétés électroniques ont servi à séparer les molécules en les plaçant dans des zones délimitées par des valeurs de ces propriétés. Des étapes de séparation successives ont été implémentées afin d'obtenir un maximum de zones contenant un seul MechoA. Nous nous sommes limités à un maximum de 3 étapes successives.

Le set d'entraînement qui a servi à définir ces zones était constitué de 201 molécules. La méthode a obtenu pour ce set 86% de classifications correctes. Nous avons ensuite validé cette approche avec un set de validation constitué de 100 molécules, qui a aussi servi à comparer nos résultats avec le schéma de Verhaar. Pour ce set de validation, notre méthode a obtenu 69% de classifications correctes, contre 45% pour le schéma de Verhaar, avec 17% de mauvaises classifications pour notre méthode contre 18% pour Verhaar, et 13% de classifications légèrement différentes de la littérature pour notre méthode. Enfin, 1% des molécules étaient en dehors du domaine d'applicabilité de notre méthode, alors que 37% étaient hors du domaine d'applicabilité pour le schéma de Verhaar.

Ces résultats montrent une meilleure performance de notre méthode, avec particulièrement un domaine d'applicabilité fortement augmenté par rapport au schéma de Verhaar. Cependant, les calculs engagés pour obtenir ces résultats sont lourds, et les résultats étant quand-même moyens avec 69% de classifications correctes, cela a appelé à tenter d'autres méthodes de prédictions plus légères en calculs, telle celle de Verhaar justement, mais en exploitant la connaissance accumulée sur les mécanismes d'action et les relations structure-activité pour obtenir de bons résultats.

Chapitre 5 : Développement d'un modèle de prédiction des MechoAs par alertes structurales.

1) Introduction

Nous avons défini 23 MechoAs détaillés, qui sont les sous-catégories de 6 MechoAs généraux. Ces catégories ont été définies dans le chapitre 3.1) Pour évaluer les informations données dans la littérature sur les MechoAs et MoAs, et pour pouvoir aller plus loin que ce qui est publié, les modèles QSAR développées par KREATiS ont été utilisés. Ces modèles sont développés à partir de données expérimentales consciencieusement validées, et les substances sont regroupées par MechoA pour construire des modèles QSAR MechoA-spécifiques. Dans nos modèles d'écotoxicité, la toxicité est reliée à l'activité chimique, au lieu d'être reliée au logP comme c'est fait dans la littérature la plupart du temps. Le concept d'activité chimique, introduit par Mackay *et al.* (Mackay *et al.*, 2009) pour la narcose, est aussi utilisée dans nos modèles pour les autres MechoAs. Plus d'informations sur les méthodes utilisées par KREATiS sont disponibles sur le site internet de KREATiS (KREATiS, n.d.).

Une nouvelle classification en mécanismes d'action toxique. En se basant sur la connaissance des mécanismes d'action (MechoAs) qui ont été compilés et rassemblés dans le fichier d'informations supplémentaires de la publication présentée dans le chapitre précédent, nous avons construit des règles structurales liées à chaque MechoA. Nous avons en effet pu trouver des relations fortes entre la structure moléculaire et le MechoA des substances, qui est le point de départ de leur voie à conséquence nocive (AOP), qui décrit comment la toxicité se produit dans un organisme vivant. Dans un premier temps, nous avons essayé de prédire les MechoAs de la structure à l'aide de paramètres quantiques pour séparer les composés en groupes avec des propriétés électroniques distinctes (énergies des orbitales moléculaires, charges partielles). Nous avons obtenu avec cette première méthode des résultats mitigés, avec 69% de classifications correctes pour le set de validation. En outre, la méthode était assez lourde à exécuter, en utilisant l'optimisation de la géométrie avec la méthode Hartree-Fock restreinte et l'ensemble de base 6-31G (d, p). Afin d'obtenir des résultats satisfaisants avec un niveau beaucoup plus bas d'utilisation de l'UC, nous avons développé cette nouvelle méthode avec des alertes structurales simples, ne nécessitant pas de calculer des paramètres quantiques. Notre intention est de mettre en place une méthode de prédiction et une classification permettant un accès facile à l'ensemble de la communauté des écotoxicologues et toxicologues, et donc créer une méthode qui, à l'avenir, puisse être exécutée dans un logiciel sur n'importe quel ordinateur. C'est pourquoi elle a rapidement fait l'objet d'un manuscrit qui sera prochainement soumis au journal *Computational Toxicology*.

Plusieurs méthodes de prédiction basées sur des alertes structurales pour prédire les modes d'action (MoA) existent déjà, et certaines sont listées ci-dessous. La plus célèbre dans le domaine de l'écotoxicologie est l'arbre de décision de Verhaar (Enoch *et al.*, 2008; Verhaar *et al.*, 2000, 1992), qui prédit des MoA pour un large ensemble de molécules organiques. De même, la classification en MoA de Russom et ses collègues (Russom *et al.*, 1997) a été associée à une méthode de prédiction basée sur des règles topologiques, suivie de l'utilisation de QSAR spécifiques à chaque MoA. Le MoA final sélectionné est celui pour lequel le modèle QSAR donne la CL50 la plus basse. Un ensemble d'alertes structurales a également été développé pour prédire la cancérogénicité (Benigni *et al.*, 2013) et un autre pour la mutagénicité (Kazius *et al.*, 2005). Certaines de ces règles structurales sont limitées à une activité biologique spécifique, ou focalisées sur un MechoA spécifique, d'autres visent à prédire différents MoAs censés couvrir les principales voies toxiques existantes, mais ces méthodes manquent de connaissance de MoA ou MechoA pour un certain nombre de classes chimiques (par exemple, les esters dans le cas du schéma de Verhaar) et sont donc incomplètes. Il y a donc un besoin d'une classification générale qui unifie tous ces domaines et qui prend en compte l'ensemble des familles de substances et des voies toxiques connues. De plus, il est plus logique de prédire les MechoAs à partir de la structure, plutôt que les MoAs, car les MechoAs sont liés à la première étape clé d'une AOP, alors

que la définition de MoA peut être source de confusion, et plutôt reliée à la fin de l'AOP, basée sur les effets au niveau de l'organisme entier.

2) Matériels et Méthodes

Le set d'entraînement utilisé pour développer des alertes structurales et l'arbre de décision en découlant est composé de 301 molécules, qui proviennent du set de données (à la fois les sets d'entraînement et de validation) de notre première méthode. Notre set de validation pour la méthode présentée ici contient 491 molécules. Il se compose de nouvelles molécules non utilisées précédemment, issues de la base de données interne de KREATiS, et pour lesquelles nous avions des données de LC50 aiguë sur poisson. Le set a ensuite été complété par certaines substances avec des MechoAs spécifiques qui ne sont pas couverts par notre base de données. Les données du set d'entraînement peuvent être trouvées dans le fichier d'information de supplémentaire de notre publication, aussi placé en annexe A.1) Informations supplémentaires pour la publication « A New classification algorithm based on Mechanisms of Action. », et les données pour le set de validation sont présentées dans l'annexe A.2) Informations supplémentaires pour la méthode de prédiction des MechoAs par alertes structurales. Dans les sets d'entraînement et de validation, les informations sur les MechoAs ont été obtenues à partir de diverses sources dans la littérature publiée, dont (Enoch *et al.*, 2008; Fukuto, 1990; IARC, 2010; Kazius *et al.*, 2005). Nous avons utilisé des données issues de tests avec des mammifères (principalement des rats et des souris), des poissons, des invertébrés (principalement des daphnies), des algues, des plantes (principalement des herbes) et des protozoaires, mais la majorité des données sont sur les rats et souris et sur les poissons. Les références pour chaque attribution de MechoA sont données dans les tableaux des annexes A.1) et A.2).

Pour bien caractériser les relations entre la structure et le MechoA, nous avons la plupart du temps utilisé des alertes structurales uniquement basées sur la structure 2D et la connexion entre les atomes (règles topologiques), mais nous avons aussi eu recours à des alertes structurales basées sur la structure 3D pour évaluer la distance entre les groupes chimiques d'intérêt, pouvant conférer aux substances de bonnes capacités de liaison dans un site de liaison de protéine. À cette fin, les calculs de mécanique moléculaire avec le champ de force UFF sont suffisants, donc la méthode de calcul reste encore très rapide pour un calcul 3D. Plusieurs conformères stables existent souvent pour une molécule, mais il se peut que seulement un ou plusieurs d'entre eux satisfassent aux critères de la règle de décision et les autres non. Pour cette raison, l'étape d'optimisation 3D doit être effectuée avec soin pour s'assurer que la présence ou l'absence de ces conformères est bien identifiée. L'optimisation de la mécanique moléculaire et les mesures de distance ont été réalisées avec le logiciel Avogadro (Hanwell *et al.*, 2013, 2012). Cette étape est importante pour les règles 19 et 20 qui détecteront respectivement les opioïdes et les ligands aux récepteurs de l'acétylcholine. Par exemple, prenons l'exemple de la règle 20 concernant les ligands aux récepteurs de l'acétylcholine pour montrer comment les alertes structurales ont été construites. En comparant les caractéristiques communes des ligands forts aux récepteurs de l'acétylcholine, on peut voir qu'un atome d'azote est toujours présent, et sera positivement chargé au pH physiologique. (Soit c'est un ammonium quaternaire, soit une amine protonée). De plus, il y a un fragment riche en électrons, capable d'accepter des liaisons H (par exemple un groupe ester ou alcool) et/ou participant à des interactions pi-pi (c'est-à-dire un cycle aromatique) pour améliorer la force de liaison. Selon les molécules, la distance entre l'azote et le fragment riche en électrons est comprise entre 4,3 et 5,3 Å. L'azote chargé positivement, le fragment riche en électrons et la distance entre les deux, ont alors été considérés comme les facteurs clés nécessaires à la liaison aux récepteurs de l'acétylcholine, et la détection de ces caractéristiques structurales sont utilisées dans la règle 20. Pour mesurer la distance avec des cycles aromatiques sans hétéroatome qui attire les

électrons, nous considérons le centre du cycle, car c'est là où une liaison H sera dirigée. Ces critères sont présentés avec la structure de certains ligands aux récepteurs de l'acétylcholine en Figure 40.

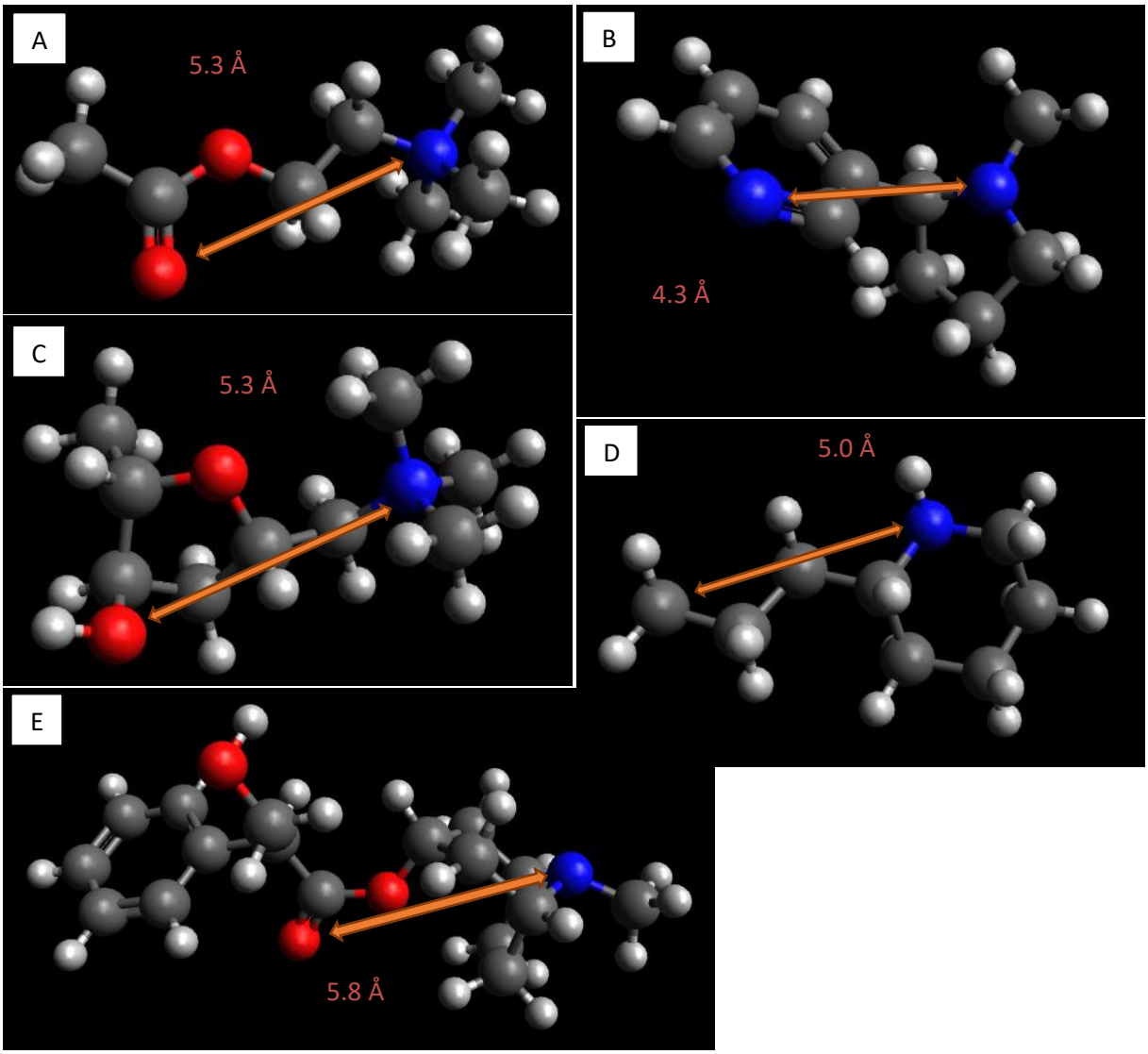


Figure 40 : ligands aux récepteurs de l'acétylcholine. A = acétylcholine. B = nicotine. C = muscarine. D = coniine. E = atropine.

La Figure 40 montre que la coniine et l'atropine ont tous deux des écarts par rapport aux critères mentionnés ci-dessus (l'accepteur de liaison H manque dans le cas de la coniine ou est trop loin de l'azote dans le cas de l'atropine). Néanmoins, ces ligands ont encore une forte affinité de liaison pour les récepteurs d'acétylcholine, l'atropine pour les récepteurs muscariniques et la coniine pour les récepteurs nicotiniques, bien que contrairement aux autres exemples de la Figure 40, les effets sont antagonistes et non agonistes. Cette différence d'effet peut être due à une différence de capacité à former des liaisons H par rapport aux autres ligands. Une règle spécifique pour détecter des molécules comme l'atropine et la coniine a donc été formulée (règle n ° 17 dans l'arbre de décision).

Un autre cas où l'optimisation de la géométrie tridimensionnelle peut être cruciale est l'exemple théorique suivant :

Dans le cas de l'ester ci-dessous (Figure 41), une analyse par alerte structurelle qui ne tient pas compte de la géométrie de la molécule conclura que c'est un ester α,β-insaturé avec une délocalisation des électrons lui conférant une réactivité d'électrophile mou. Cependant, en calculant la géométrie

optimale pour la molécule, on se rend compte qu'à cause des groupes encombrants du cycle, l'ester ne se tient pas dans le même plan que la double liaison du cycle, afin de minimiser les interactions de Van der Waals répulsives. En conséquence, les électrons ne sont pas délocalisés, et la molécule se comportera simplement comme un ester, qui exerce donc la pronarcose.

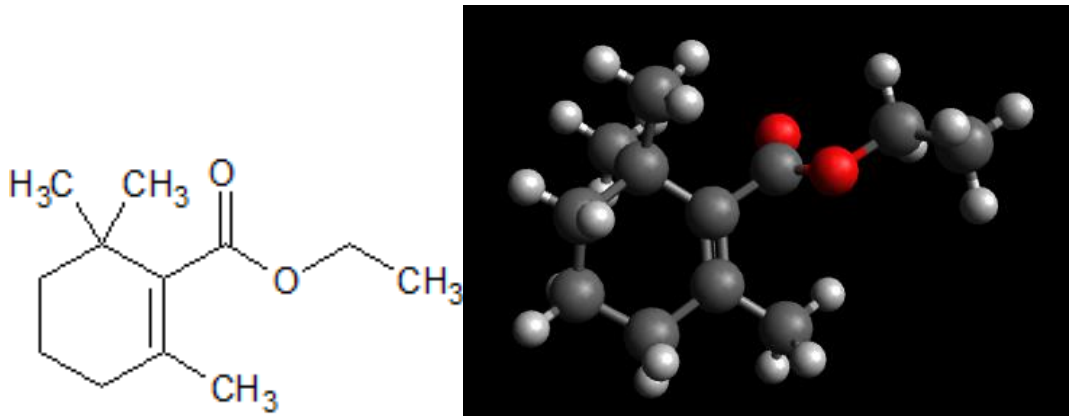


Figure 41 : Un ester α,β-insaturé dont les électrons ne sont pas délocalisés à cause de l'encombrement stérique.

Pour appliquer certaines règles, il faut savoir si un cycle aromatique est pauvre ou riche en électrons. Cette propriété des cycles aromatiques joue un rôle crucial pour leur réactivité et apparemment également pour la facilité du métabolisme. Un cycle aromatique non substitué, comme le benzène, est considéré comme légèrement riche en électrons, car il comporte 6 électrons pi et aucun groupe attracteur d'électrons (GAE) pour les tirer hors du cycle. Cependant, si des GAEs sont présents sur le cycle, tels qu'un nitro, un chloro, un acide, un aldéhyde, un ester ou un amide, branchés du côté du carbonyle, etc., le cycle aromatique devient pauvre en électrons car les électrons sont principalement localisés vers les GAEs et une densité plus faible est présente sur le cycle. À l'inverse, les groupes donneurs d'électrons (GDE) donneront des électrons au cycle. Quelques exemples sont des groupes hydroxyle, alkoxyle, amino et alkyle (dans une moindre mesure). Certaines autres règles doivent prendre en compte le pKa des substances, car le pKa d'un éventuel groupe partant sera un facteur clé pour savoir s'il s'agit d'un bon ou mauvais groupe partant au pH physiologique. D'autre part, le découplage de la phosphorylation oxydative par les phénols dépend de leur capacité à transporter des protons à travers la membrane mitochondriale interne, qui est en partie contrôlée par leur pKa.

Pour construire l'arbre de décision, nous avons dû trier les alertes structurales et les MechoAs correspondants, de sorte que le MechoA le plus toxique soit identifié d'abord et correctement attribué, avant que l'arbre de décision n'atteigne une autre règle qui utilise une autre sous-structure responsable d'un mécanisme moins toxique. Cet arbre de décision a été construit sous une forme linéaire, allant directement de la première question à la dernière question sans aucune ramification. Si une substance n'est caractérisée par aucune règle, elle sera non classée (Résultat de la réponse Non à la dernière question). En fait, deux facteurs importants sont en jeu pour discriminer quel MechoA entraîne la toxicité, lorsque plusieurs MechoA existent pour une même substance. Le premier paramètre est la puissance d'un MechoA donné. En effet, certains mécanismes perturbent des processus critiques au sein de l'organisme (comme la transmission du signal dans les synapses nerveuses), tandis que d'autres perturbent des processus moins essentiels ou nécessitent une plus grande quantité de molécules pour obtenir des effets observables (tels que la narcose). Le deuxième paramètre est la cinétique. Certains mécanismes sont très rapides, tandis que d'autres prennent du temps. Par exemple, certaines substances doivent d'abord être métabolisées pour exercer leur véritable effet toxique. C'est donc plus lent que lorsqu'une substance agit directement. C'est pourquoi les règles qui détectent des mécanismes impliquant le métabolisme sont placées après les règles pour

les mécanismes directs, mais avant les règles de détection de la narcose. La meilleure façon de juger si un mécanisme est plus puissant que l'autre est de prendre une substance qui présente des sous-structures permettant à plusieurs mécanismes de se produire. Mais de telles substances, on peut souvent observer que la toxicité est entraînée uniquement ou principalement par un seul de ces mécanismes. Par exemple, le pentachlorophénol, est un phénol acide et peut agir en tant que découpleur de la phosphorylation oxydative (Schultz, 1987). D'autre part, comme il s'agit d'un phénol, on peut s'attendre à ce qu'il exerce une narcose polaire. Les expériences et les modèles QSAR ont montré que l'ensemble des effets toxiques du pentachlorophénol est davantage dû à son mécanisme de découplage plutôt qu'à la narcose, car il a été démontré qu'il diminue fortement le taux de transfert d'électrons dans les membranes (Escher *et al.*, 1999), tout en ayant un excès de toxicité par rapport à la toxicité qui serait prédite par les modèles de narcose polaire (Enoch *et al.*, 2008; Schultz, 1987). De tels cas nous ont permis de trier les MechoAs et, par exemple, la règle de détection des découpleurs de la phosphorylation oxydative (règle n ° 31) a donc été placée avant la règle de détection des narcotiques polaires (règle n ° 57). Une autre méthode pour comprendre le MechoA critique d'une substance avec plusieurs MechoAs potentiels consiste à utiliser des expériences de toxicité de mélange, pour voir si la valeur de toxicité correspond à un modèle d'addition de concentration (fonctionnant uniquement pour des mélanges de substances avec le même MechoA) ou plutôt un modèle d'indépendance (Altenburger *et al.*, 2000; Backhaus *et al.*, 2000).

Cependant, lorsque plusieurs MechoAs sont identifiés comme jouant un rôle réel pour une même substance, tous les MechoAs sont donnés. Comme ces MechoAs peuvent jouer leur rôle dans différents délais (toxicité aigüe ou chronique) ou espèces, plusieurs règles de notation ont été mises en place pour décrire l'information complète de MechoA pour chaque substance : un 'C' avant le numéro d'un MechoA signifie qu'il est seulement observable dans les études chroniques, mais pas aigües. Par défaut, si aucun 'C' n'est présent, le MechoA s'applique à l'échelle d'un test de toxicité aigüe. Certaines abréviations sont également placées avant le numéro des MechoAs pour désigner des MechoAs spécifiques aux espèces. 'm' signifie pour les mammifères, 'h' pour les humains, 'f' pour les poissons, 'p' pour les plantes, 'u' pour les organismes unicellulaires, 'd' pour les daphnies. '#' Avant ces lettres signifie "pour toutes les espèces sauf celles-ci". Par exemple, #up5.2 signifie MechoA 5.2 pour toutes les espèces, sauf les unicellulaires et les plantes. '&' signifie que les deux MechoAs jouent un rôle. '/' signifie qu'il est incertain si un ou plusieurs des MechoAs sont présents et sont critiques pour la toxicité. Ils sont écrits dans l'ordre de la certitude décroissante, où le premier est le plus probable. Par exemple, 1.2 & m4.3 / 3.1 signifie que ce composé est censé exercer une narcose polaire (1.2), avec une métabolisation en composés réactifs pour les mammifères (m4.3), et pourrait être capable d'agir par un mécanisme réactif direct (3.1). Enfin, nous avons défini un ordre pour les différentes espèces : le MechoA pour les poissons et les mammifères est mis en premier, les informations sur les autres espèces arrivent ensuite. S'il existe une certitude équivalente pour les poissons et les mammifères, le MechoA pour les poissons est placé en premier. Ces règles ont déjà été utilisées dans notre publication précédente (voir chapitre 4).

3) Résultats et discussion

a) Arbre de décision

L'arbre de décision que nous avons construit est présenté dans le Tableau 6. Il est composé de 62 questions. Toutes les alertes structurales, ainsi que la description de chaque MechoA et la référence de la littérature correspondante figurent en annexe A.2) Informations supplémentaires pour la méthode de prédiction des MechoAs par alertes structurales.

Order	Rule	Result if Yes		Result if No
1	Does the molecule contain other elements than H, C, N, O, F, P, S, Cl, Br, I? Or does it not contain any C atom?	UC		Go to 2
2	Is it an epoxide or an aziridine?	3.1	ALL: Nucleophilic addition of proteins and DNA => adducts, impaired function	Go to 3
3	Does the compound have an α,β-unsaturated (C=C) (not aromatic) carbonyl (not ketone) or nitrile?	3.2	ALL: Michaël addition of proteins => adducts, impaired function	Go to 4
4	is the compound a quinone?	3.2 & 4.4	ALL: Michaël addition of proteins and DNA => adducts, impaired function / RedOx cycling	Go to 5
5	Is there a single N-N bond (not included in a heteroaromatic cycle)?	3.1 & 4.3	ALL: binds to ketoacids and vitamin B6, inhibiting its wide actions / metabolised into diazonium, which gives oxidative stress (RNS) and generates Me' radicals. F: reactive	Go to 6
6	Is it an electron-rich benzaldehyde (like benzaldehyde or vanillin)?	3.1 & m4.1	ALL: Schiff-base formation / M: rapidly eliminated by oxidation into acid (and potentially further conjugation)	Go to 7
7	Does the compound have a saturated, not gem-disubstituted gamma-diketone moiety?	3.1	ALL: pyrrole formation with a lysine residue, oxidation and crosslinking with another protein	Go to 8
8	Is the compound a (aliphatic non α,β-unsaturated or electron-poor aromatic) aldehyde?	3.1	ALL: Schiff-base formation with amino residues of proteins (Nucleophilic Addition) => adducts, impaired function	Go to 9
9	Is it an acyl halide?	3.1	ALL: violently hydrolysed in water generating HCl (corrosive), not enzymatic!	Go to 10
10	Does the compound have a leaving group attached to a benzyl or allyl moiety? (LG = Halide except F, O-ester, O-phenyl (phenoxy) (if acidic enough), N if protonated at neutral pH)	3.1 & 4.1	ALL: SN2 adducts / N-acetylcysteine and glutathione conjugates (Nucleophilic Substitution of the LG) => diminution of antioxydant levels	Go to 11
11	Is there a single O-O or S-S bond?	3.3	ALL: Homolytic cleavage giving 2X' radicals, leading to lipid peroxidation (O), RedOx cycling (S)	Go to 12
12	Does the compound have a DDT-like structure? (2 phenyls joined by a mono-substituted methylene, phenyls substituted only in para by not too polar groups)	#up6.6	Animals: prevents the closure of Na+ channel from the nerve, maintaining the signal => causes death by respiratory failure	Go to 13
13	Is it dicofol?	6.7	ALL: thyroid hormone antagonist, weak estrogen agonist, potent inducer of CYP2B, weak aromatase inhibitor, some ability to bind with andogen receptors	Go to 14
14	Is the compound a highly chlorinated (>=4 Cl on a substructure composed of <5 C atoms) alicyclic (non aromatic cyclic hydrocarbon)?	#up6.6	Animals: inhibition of GABAergic Cl- channel into the nerves (binding at the picrotoxinin site), maintaining the signal => causes death by respiratory failure	Go to 15
15	Is the compound a carbamate or (di)thiocarbamate (with alkoxy group as a good enough leaving group, like phenoxy or oxime)?	6.1	Animals: AChE inhibition (reversible bonding to carbonyl) (A: more toxic than esters. Probably inhibits other esterases)	Go to 16
16	Is the compound an organophosphorous compound with the substruture R1P(=O)(R2)LG or R1P(=S)(R2)LG (R1, R2= alkyl, N-alkyl or O-alkyl ; LG less restrictive than in question 10. Can be F, can be not acidic phenoxy, can be thiol)?	6.1	Animals: AChE inhibition (covalent bonding to P) (U&P: Probably inhibits other esterases)	Go to 17
17	Is it atropine? Or is it a 2-alkylpiperidine (best with alkyl=propyl or longer)	#up6.2	Animals: blocks Ach receptors (atropine: muscarinic receptors; piperidines: nicotinic receptors)	Go to 18
18	Is it rotenone of cyanide?	#up6.7	**rotenone:** Animals: inhibition of mitochondrial e- transport at complex I **cyanide:** Animals: cytochrome c oxidase (complex IV) inhibition	Go to 19
19	Is it caffeine?	6.7 & 6.6	Animals: CNS: antagonism of adenosine at its receptors (regulation of adenylate cyclase synthesising cAMP), cyclic nucleotide (cAMP) phosphodiesterase inhibition (=> release of Ca2+), activation of Calcium channels. (Plants: Ca2+ regulation also disrupted)	Go to 20
20	Is it strychnine, alpha-amanitin, carbon monoxide, tartaric acid?	6.7	ALL: various molecule-specific enzyme-specific interactions	Go to 21
21	is the molecule an opioid (e.g. morphine) (common feature: 3° amine - OH and aromatic at 5-7 A distance each)?	#up6.6	Animals: binds to opioid receptors (GPCRs) at presynaptic and postsynaptic neurons, inhibiting the neuronal signal by upregulating the K+ channels and downregulating the Ca2+ channels thereby inhibiting the release of neurotransmitter (presynaptic effect).	Go to 22
22	Does the molecule have a nicotine/acetylcholine-like structure? (N(+) - 7-8 A distance - aromatic / electron-rich site, with H-bond acceptor property (but I find for acetylcholine and nicotine a distance of 4,2-5,3 rather))	#up6.2	Animals: agonism of nicotinic Ach receptors, stimulation of neurotransmitters production, leading to a desensitization and upregulation of nAchRs in chronic	Go to 23
23	Does the compound have a salicylic acid-like structure? (XC(=O)c1c(Y)cccc1 X,Y = N, O, OR)	m+f6.7 & [1.2 if acid, 2.1 if ester or amide]	M: analgesic, antipyretic, anti-inflammatory, certainly mostly due to the inhibition of COX-2 (and COX-1). Main side effect at high doses is gastrointestinal bleeding (thought to be due to COX-1 inhibition) F: other enzyme-specific interactions disturbing the stress response. Other : ?	Go to 24

24	Is the molecule a pyrethroid? ([C]=CC(C1(C)C)C1C(=O)OR see publi for R nature)	#up6.6	Animals: prevents the closure of Na+ channel from the nerve, maintaining the signal / inhibits GABAergic Cl- channel into the nerve / inhibits Ca-ATPase or Ca-Mg-ATPase, disturbing Ca2+ regulation	Go to 25
25	Is it trifluralin or clomazone?	p6.7 & [6.6 for trifluralin, #up6.1 for clomazone]	**Trifluralin:** Plants: inhibit tubulin polymerisation into microtubules (playing a central role in cellular division), stops mitosis. All: disrupts Ca2+ concentration equilibria, thus disrupting a lot of processes in the cell. Very toxic to fish, and very harmless for mammals **Clomazone:** Plants: inhibit carotenoid biosynthesis (essential for photosynthesis). Other: slight AChE inhibition effect	Go to 26
26	Is it a haloacetanilide (like metolachlor)?	4.1 & p6.7	Conjugation with thiol groups (including CoA, inhibiting all the processes using it), not precisely known... Plants: makes it a good herbicide. All: detoxification with GSH.	Go to 27
27	is it a diaminotriazine or glyphosate?	#uC6.7 & [p6.5 for triazine, p6.7 for glyphosate]	**Atrazine:** Plants: inhibits Protein D-1 at photosystem II => blocks production of O2 + ROS generated. Animals: carcinogenic, reproductive and developmental toxicity (several MechoAs are postulated to explain these effects) **Glyphosate:** Plants: inhibit aromatic amino acids synthesis by 5-enolpyruvylshikimimate-3-phosphate synthase inhibition. M: genotoxic and probable carcinogen.	Go to 28
28	Does the compound have a beta-dicarbonyl (that can be parallel) moeity?	6.4	ALL: chelating Fe, degrading cytP450, also reacts slowly with arginine residues to form adducts, causes inactivaton of B vitamins	Go to 29
29	Is the compound a dioxin, polychlorodibenzofuran or PCB?	#u6.7 & 1.1	ALL: induction of cyP450 (binding to AhR, but then not / very slowly metabolised by cytP450 consequently producted) => oxidative stress	Go to 30
30	Does the compound have a steroid-like structure? (gonane with hydroxyl or carboxyl groups in position 3 and 17)	#upC6.7?	Animals: endocrine disruption (sexual differenciation/contraceptive/...)	Go to 31
31	Is the compound a phenol with pKa<8,0? (e.g. with >=2 Cl or >=1 nitro, or mix, or with other EWG)	5.1	ALL: Oxidative Phosphorylation Uncoupling by inner mitochondrial membrane pH gradient suppression (or other vesicles pH gradient also)	Go to 32
32	Is it a carboxylic acid (not abu)?	5.2	ALL: Acids cross membranes in their neutral form, and then release their proton in the cytosol	Go to 33
33	Is the compound a very acidic pyridine or aniline? (e.g. with >=4 Cl or >=2 nitro)	6	ALL: Oxidative Phosphorylation Uncoupling / Electron Transport Chain inhibition (Maybe by coordination to heme, inhibiting them)	Go to 34
34	Is it 2,2,2-trichloroethanol?	m6.6 & 1.1	M: Activate GABA receptor (Cl- channel) (=metabolite of chloral=2,2,2-trichloroacetaldehyde/chloral hydrate)	Go to 35
35	is it 2,2,2-trichloroacetaldehyde or chloral hydrate?	m4.6 & 3.1	M: metabolised into 2,2,2-trichloroethanol, which activates GABA receptor (Cl- channel)	Go to 36
36	Is it dichloromethane?	1.1 & m4.6	M: metabolisation into CO with effects apparent in a few minutes to hours (depending on the dose) => fixing to haemoglobin, diminished O2 supply (coma then death) / directly CNS (not understood)	Go to 37
37	Is it carbon tetrachloride?	1.1 & m4.3	M: cyP450 1e-reduction gives 'CCl3 radical, then quenched by O2 to give 'OOCCl3 peroxy radical, which gives then lipids & proteins peroxidation => hepatotoxicity	Go to 38
38	is the compound an o-phthalate?	2.1 & mC2.2	M: digestive hydolysis of one ester =>chronic=> induction of PPARalpha (peroxisome proliferator-activated receptor, precursor of liver cancer development) and endocrine disruption (if R>C4, can undergo further oxidative metabolism)	Go to 39
39	Is it an ester of 2,4-D acid?	#p2.1 & p2.2	Plants: hydrolysed. Then the acid is an antagonist of phytohormone auxin, overstimulating plant growth, more than it can afford (because too slowly metabolised then). M: after hydrolysis, acid rapidly eliminated.	Go to 40
40	Does the compound have a nitrile group with a H in alpha?	m4.6 & 1.1	M: cytP450 alpha-hydroxylation (cyanohydrin), then non-enzymatic degradation to give CN- and corresponding ketone/aldehyde. F: narcosis	Go to 41
41	Does the compound contain a bipyridinium or phenylpyridinium moiety ? (paraquat, MPP+)	4.4	ALL: RedOx cycling (1e- reduction then give this e- to O2 to give the superoxide anion (ROS) and the parent compound, and so on)	Go to 42
42	is it a 1° or 2° thiol or an aromatic thiol (rich in electrons)?	4.4	ALL: 1° & 2° thiols, in their anionic form, are oxidized into thiyl radical and then enter a redox cycle between the thiyl radical and the disulfides. 3° are not reactive enough due to steric hindrance. Aromatics are more reactive (conjugation stabilization), EWG are stabilizing the anion, therefore decreasing the reactivity, while EDG will stabilize the thiyl radical and favor the oxidation.	Go to 43

43	Is it but-3-yn-1-ol, 2,2,2-trifluoroethanol?	4.3	ALL: metabolisation leads to reactive aldehyde => adducts	Go to 44
44	Is the compound hydroquinone, catechol (potentially substituted)?	4.3 & 4.4	ALL: Oxidation by tyrosinase / cytP450 to quinone => protein and DNA adducts / RedOx cycling	Go to 45
45	Is it 3-methoxyphenol or any methyl ether (potentially substituted, with one OH still present) of hydroquinone or catechol?	1.2 & m4.3	M: Dealkoxylation / 4-hydroxylation, then RedOx cycle between reduced and oxidized quinone	Go to 46
46	Does the compound contain a diazo bond (not in a heteroaromatic ring)? (-N=N-)	m4.3	M: diazo bond cleaved (metabolism mainly by anaerobic intestinal bacteria), generating anilines. Then classical metabolism to nitroso.	Go to 47
47	is the compound a furan or thiophene, not substituted in both positions 2 and 5 and not fused with aromatic?	4.3	ALL: metabolisaton into (Z)-but-2-en-1,4-dial => proteins and DNA adducts after glutathione is totally consumed => carcinogenic and mutagen	Go to 48
48	is the compound a 1° allyl (not substituted) or propargyl alcohol?	4.3	ALL: ox to corresponding abu aldehyde => adducts	Go to 49
49	is it hexane, hexan-2-one?	1.1 & mC4.3	M: CNS seizure: after metabolisation into hexan-2,5-dione, pyrrole formation with a lysine residue, oxidation and crosslinking with another protein	Go to 50
50	Does the compound contain a vinyl alkoxy group?	[1.1, 1.2 or 2.1 depending on the answer to next questions] & C4.3	ALL: metabolism transforming vinyl to epoxide, then adduct to DNA	Go to 51
51	Is it an ester of tartaric acid?	2.2	ALL: Hydrolysed into tartaric acid. Tartaric acid inhibits fumarase thus stopping the Krebs cycle.	Go to 52
52	Is it an ester, amide, phosphate, carbonate (not α,β-unsaturated)?	2.1	ALL: hydrolysis to acid and alcohol/amine + "ester narcosis"	Go to 53
53	Is it a quaternary ammonium?	1.3 & #up6.1	ALL but U: for certain ammonium (better with an ester-like moiety) AChE inhibition. Without specific group (like phosphoester), it's a reversible inhibition Unicellular: disturbs membrane	Go to 54
54	Is it a substituted urea?	p6.5 & [1.1 or 1.2 depending on the answer to next questions]	Plants: blocks electron transfer at the photosystem II => inhibits O2 production	Go to 55
55	Is it an aniline not N-substituted (aniline=NH2-Aryl in a wide sense) (not too acidic)?	1.2 & m4.3	M: metabolisation into nitroso => protein adducts, glatter, spleen and liver damage / chronic => oxidative stress, DNA adducts, cancer development Aq: polar narcosis	Go to 56
56	Is it an aliphatic amine?	1.2 & 5.2	ALL: Directly applied, it is corrosive, and it alkalinize the cells, entering into them in their neutral form and then become protonated, all simply because it is a strong base.	Go to 57
57	Is it a phenol with pKa>8,0? e.g. alkyl-, monohalophenols (=aromatic alcohol in a wide sense)	1.2	ALL: membrane destabilizing: polar narcosis	Go to 58
58	Is it a PAH (one PI system, planar)?	1.1 & m4.3	M: metabolism with cytP1A1 gives epoxide and then quinones in a few hours, adducts to proteins and DNA Aqua: non-polar narcosis	Go to 59
59	Is it a N-substituted aniline (aniline=N(R1)(R2)-Aryl, R1, R2 = H, Alkyl in a wide sense) or a nitrobenzene?	1.1 & m4.3	M: metabolisation into nitroso => protein adducts, glatter, spleen and liver damage / chronic => oxidative stress, DNA adducts, cancer development Aqua: non-polar narcosis	Go to 60
60	Is it trichloroethylene?	1.1 & mC4.3	M: activation by GST gives reactive metabolites (likely dichlorovinyl-cysteine) which cause DNA mutation leading to renal cancer Aqua: non-polar narcosis	Go to 61
61	is it benzyl alcohol?	1.1 & 4.1	ALL: narcosis / rapidly eliminated by oxidation into benzoic acid, conjugated then with glycin to form hippuric acid	Go to 62
62	Does it belong to one of these categories? aliphatic and mono-aromatic hydrocarbons (except hexane); monoketones (except hexan-2-one) and polyketones (not beta or gamma); non vinylic/unsubstituted allylic/propargylic 1° alcohols and ethers (except epoxides); non vinylic 2° and 3° alcohols (non aromatic); non vinylic/allylic/propargylic thiols (3° aliphatic or poor aromatic) and sulfides and oxidized derivatives (sulfoxide, sulfonate, sulfate); heteroaromatic (except furans and thiophenes not substituted in 2 and 5); nitriles without H in alpha; all of that possibly with halides (except benzyl/allyl halides (not F), carbon tetrachloride and dichloromethane, 2,2,2-trifluoroethanol and 2,2,2-trichloroethanol)	1.1	ALL: membrane destabilizing: non-polar narcosis	UC

Tableau 6 : Arbre de décision pour la prédiction des MechoAs. UC = Non classifié.

Comme on peut voir dans ce tableau, plusieurs MechoAs peuvent être prédits par l'arbre de décision, selon que la sous-structure détectée par la règle soit associée à plusieurs MechoAs, qui peuvent s'appliquer à des espèces différentes souvent. Par exemple, la règle n° 39 indique que pour les esters d'acide 2,4-dichlorophenoxyacetique **34**, le MechoA sera 2.2 (hydrolyse vers des produits non-narcotiques) pour les plantes et 2.1 (hydrolyse vers des produits narcotiques) pour les autres organismes.

b) Validation statistique du modèle

Quatre catégories ont été définies pour calculer les statistiques de l'algorithme. Les règles pour déterminer si une substance doit être attribuée à l'une ou l'autre de ces catégories sont données ci-dessous :

- Correctement classée : si le premier MechoA spécifié dans la prédiction est le même que le premier MechoA spécifié dans le MechoA total de la littérature. Aussi, si tous les MechoAs spécifiés correspondent, même s'ils ne sont pas dans le même ordre.
- Légèrement différente : si le MechoA général est correctement prédit (c'est-à-dire MechoA 1, 2, 3, 4 ou 5), mais pas le sous-groupe de MechoA (par exemple, 1.1 au lieu de 1.2). Une prédiction de MechoA est également dite légèrement différente si le MechoA prédit correspond à un MechoA incertain de la littérature, mais pas au plus probable (par exemple, MechoA littérature = 1.1 / 5.2 ; MechoA prédit = 5.2).
- Mal Classées : lorsque le MechoA prédit ne correspond pas au MechoA de la littérature.
- Non classifié (UC) : si la molécule contient d'autres atomes que H, C, N, O, F, P, S, Cl, Br, I ou si elle ne contient aucun atome de carbone (réponse Oui à la règle n ° 1) ou si la substance n'est caractérisée par aucune règle (réponse Non à la règle n ° 62).

Les statistiques du modèle sont présentées dans le Tableau 7 :

Set de données	Classifications	Pourcentage du set
Set d'entraînement (N=301)	Correctes	92.0
	Légèrement différentes	6.0
	Fausses	1.0
	Non classées	1.0
Set de validation (N=491)	Correctes	92.3
	Légèrement différentes	3.4
	Fausses	4.3
	Non classées	0.0

Tableau 7 : Statistiques de notre arbre de décision. Les MechoAs de référence ont été déterminés à partir de notre recherche bibliographique.

Le taux de réussite du schéma de Verhaar appliqué à ce set de comparaison a été calculé et comparé à ceux obtenus avec l'algorithme développé dans ce travail (Tableau 8). L'algorithme présenté ici a plus de pouvoir statistique et semble être un bon remplacement du schéma de Verhaar.

D'autre part, nous avons comparé nos résultats avec la méthode de Russom (Russom *et al.*, 1997), en utilisant le module « MOA classification by OASIS » du logiciel OECD QSAR Toolbox (OECD and ECHA, 2017), qui prédit une classification en modes d'action basé sur le travail de Russom. Cependant, les catégories utilisées dans ce module sont légèrement différentes et moins complètes que celles décrites dans la publication de Russom (Russom *et al.*, 1997), et sont les suivantes. Les MoAs sont organisés en MoAs généraux et en MoAs détaillés. Les MoAs généraux au nombre de 7 sont donnés dans la liste ci-

dessous, avec les sous-catégories associées entre parenthèses, et une courte explication des catégories.

- Narcotiques de base (pas de sous-catégories)
- Amines narcotiques (pas de sous-catégories). Leur toxicité est très variable et se situe entre la narcose non-polaire et la polaire.
- Esters (pas de sous-catégories). Défini comme un troisième type de narcose, spécifique aux esters.
- Phénols et anilines (acides faibles découpleurs de la respiration, électrophiles mous, narcotiques polaires)
- Alcools α,β-insaturés. Ce sont des proélectrophiles conduisant aux aldéhydes α,β-insaturés correspondants, qui réagiront avec les protéines.
- Aldéhydes. Réactivité non spécifique, formation d'adduits bases de Schiff avec les protéines.
- Réactifs non spécifiés (narcose polaire, proélectrophiles/électrophiles, MOA inconnu). Cette catégorie regroupe toutes les familles de substances connues pour avoir un excès de toxicité.

De plus, la méthode de prédiction des MoAs attribue par défaut un MoA de narcotique de base pour les molécules pour lesquelles aucun MoA n'a été identifié (c'est-à-dire en dehors du domaine d'applicabilité). Les statistiques du modèle qui en découlent sont présentées dans le Tableau 8 sur le set de validation de la méthode par alertes structurales, constitué de 491 molécules, et comparées aux statistiques de Verhaar et de notre méthode par alertes structurales :

Classifications	Schéma de Verhaar	Méthode de Russom (non classé = narcose de base)	Méthode de Russom (non classés séparés)	Notre méthode par alertes structurales
Correctes	54,0	82,5	70,3	92,2
Légèrement différentes	NA	6,3	6,3	3,5
Fausses	13,4	11,2	10,0	4,3
Non classées	32,6	0,0	13,4	0,0

Tableau 8 : comparaison des méthodes de classification. Résultats exprimés en pourcentage du set de validation (491 molécules). NA = Non Applicable.

Par conséquent, la méthode de Russom est bien plus performante que celle de Verhaar, tel qu'attendu, mais moins performante que notre seconde méthode. Pourtant, plus on détaille de sous-catégories, plus le risque d'obtenir une classification fausse, ou au moins légèrement différente, augmente. Or, ici, notre méthode possède plus de sous-catégories que celle de Russom, qui en possède plus que celle de Verhaar, et les résultats sont également décroissants. On peut constater que l'approximation de Russom de classer en toxicité de base les substances hors du domaine d'applicabilité est profitable pour les statistiques de sa méthode, puisque la majorité des substances non classées sont effectivement des narcotiques non-polaires. Cependant, cette approche est dangereuse dans le cas d'une évaluation de risque, puisqu'on risque de grandement sous-estimer la toxicité de certains composés, et mettre ainsi en danger la population.

Comme les données utilisées pour attribuer les MechoAs de la littérature provenaient principalement de données de toxicité pour les mammifères (principalement les rats et les souris) et les poissons, les résultats de notre algorithme seront plus pertinents pour ces espèces, tandis que les autres espèces devraient plutôt être considérées comme des informations supplémentaires, qui pourront juste servir pour appuyer un argumentaire basé sur d'autres données. Nous avons observé que certaines AOP sont très similaires dans toutes les espèces, en particulier la narcose, alors que pour certains composés avec

des MechoAs non narcotiques, le résultat est clairement différent entre les espèces. Bien sûr, il faut relativiser, étant donné que deux organismes peuvent être très différents dans de nombreux aspects. Par exemple, les algues, qui sont des plantes unicellulaires vivant dans l'eau, ont peu à voir avec les rats, qui sont des animaux, respirant de l'air, digérant toutes sortes d'aliments en partie en utilisant des bactéries intestinales, ayant également un foie équipé de diverses enzymes métaboliques, ayant des cellules hautement spécialisées telles que les neurones, *etc*. Les AOP qui restent très similaires dans la plupart des espèces sont si bien conservées car elles attaquent des fonctions communes à tous les organismes vivants de la terre, comme une membrane à fluidité et hydrophobie contrôlées, maintenant le cytosol et limitant les flux de molécules, ou encore l'ADN conservant toute l'information pour construire l'organisme et un système de transcription pour le convertir en protéines. Les différences de MechoA se produisent souvent lorsque le composé est métabolisé. Cela n'est pas surprenant car le métabolome peut varier considérablement entre les espèces, par exemple entre une herbe et un poisson. Les enzymes essentielles entre les deux espèces peuvent être légèrement différentes ou certains processus enzymatiques d'un organisme ne sont pas présents dans un autre (par exemple, la photosynthèse opérée par l'herbe mais pas par le poisson).

Il faut également noter que le modèle prédit actuellement un potentiel de perturbation du système endocrinien si la substance a une structure de stéroïde mais ne va pour l'instant pas plus loin dans ce domaine. L'algorithme est construit avec des tests de toxicité aiguë pour les poissons, les daphnies et les algues et des tests de toxicité aigüe et chronique, orale et par inhalation, sur les mammifères. Le schéma peut également être appliqué à d'autres types d'étude, mais devrait être limité aux tests qui permettent à l'organisme entier d'être impliqué dans le processus. Nous ne pouvons pas à ce stade supposer que ces mécanismes joueront un rôle crucial dans les tests de toxicité locaux tels que l'irritation cutanée par exemple, où la cytotoxicité est la principale cause des effets (OECD, 2015).

4) Conclusion

La deuxième méthode présentée ci-dessus fonctionne avec l'identification d'alertes structurales associées à des MechoAs, et ne nécessite une optimisation de géométrie 3D que pour quelques alertes structurales qui tiennent compte des distances interatomiques. Nous avons développé un arbre de décision linéaire comportant 62 questions permettant de prédire les MechoA à partir de ces alertes structurales. Les molécules sont classées en 23 MechoAs. Ce modèle a été construit à partir d'un set de données de 301 molécules. Ces MechoAs et la méthode pour les prédire ont été développés sur la base de résultats expérimentaux sur les mammifères (rats et souris principalement) et poissons, avec quelques informations supplémentaires pour d'autres espèces.

Pour le set d'entraînement, la méthode a obtenu 92% de classifications correctes. Cette méthode a ensuite été validée avec un set de 491 molécules et comparé au schéma de Verhaar et à la méthode de prédiction de Russom. Pour ce set de validation, nous avons obtenu 92% de classifications correctes avec notre nouvelle méthode, alors que le schéma de Verhaar en a obtenu 54%, et la méthode de Russom 83%. Pour 4% des molécules, la classification prédite était légèrement différente des MechoAs de la littérature avec notre méthode, et 6% des molécules avec celle de Russom. Nous avons obtenu 4% de classifications fausses et le schéma de Verhaar en a obtenu 13% et la méthode de Russom 11%. Enfin, aucune molécule du set de validation ne s'est trouvée en dehors du domaine d'applicabilité de notre seconde méthode et de la méthode de Russom, alors que 33% d'entre elles étaient en dehors du domaine d'applicabilité pour le schéma de Verhaar.

Par conséquent, nous avons obtenu une méthode à la fois simple et rapide à exécuter, mais également avec d'excellents résultats, qui surpassent largement les résultats du schéma de Verhaar et de la première méthode que nous avions développée, et qui surpassent légèrement les résultats de la

méthode de Russom, mais sans faire l'approximation de Russom que les substances inconnues sont des narcotiques non-polaires. Les résultats de la méthode présentée ici montrent un grand progrès dans le domaine des modes et des mécanismes d'action en écotoxicologie et en toxicologie chez les mammifères et nous encourageons son utilisation pour les prédictions de toxicité dans ces deux domaines. Cette approche permet ainsi de faire le lien entre ces deux domaines dont les savoirs respectifs peuvent être mis en commun afin de mieux comprendre et caractériser les mécanismes de toxicité dans différents organismes. Cette méthode est en cours d'implémentation dans un logiciel, afin que l'arbre de décision soit parcouru automatiquement et que le logiciel fournisse la prédiction de MechoA à partir de la structure.

Chapitre 6 : Expériences en laboratoire de validation et affinement des modèles.

1) Introduction

Plusieurs expériences en laboratoire ont été effectuées au laboratoire de chimie organique et bioorganique (COB, EA 4566) de l'UHA, Mulhouse afin de valider certaines approches *in silico* ou de fournir des données expérimentales afin de générer ensuite des prédictions pour une nouvelle famille de substances chimiques. Ces expériences ont été intégrées dans le travail de cette thèse car ils permettaient de renforcer les hypothèses de relation étroite entre les paramètres physico-chimiques et écotoxicologiques, y compris pour des substances apparaissant à première vue comme des exceptions pour ces relations. Les différents tests ont été effectués en suivant au mieux les lignes directrices de l'OCDE pour chacun de ces tests, afin d'en assurer la qualité. Trois types de tests ont été mis en œuvre :

- Mesures de coefficient de partage octanol-eau (logP). La ligne directrice associée est la ligne directrice OCDE n° 107 avec des modifications dérivées de la ligne directrice OCDE n° 123.
- Mesures de solubilité aqueuse. La ligne directrice associée est la ligne directrice OCDE n° 105 avec des modifications dérivées de la ligne directrice OCDE n° 123.
- Mesures de toxicité aigüe pour les daphnies. La ligne directrice associée est la ligne directrice OCDE n° 202. Pour nourrir l'élevage de daphnies, nous avons également cultivé des algues vertes, selon les méthodes décrites dans la ligne directrice OCDE n° 201 (test de toxicité sur les algues).

Ces différentes lignes directrices sont placées en annexes (A.3) lignes directrices OCDE pour les tests en laboratoire). Nous avons choisi d'effectuer ces tests, car le logP et la solubilité sont les paramètres fondamentaux permettant de prédire la toxicité, et ce sont aussi les paramètres physico-chimiques qui sont souvent les plus problématiques à mesurer (en comparaison avec le point de fusion, d'ébullition ou la pression de vapeur par exemple). Nous avons ensuite choisi de mesurer la toxicité sur les daphnies, car ce sont des invertébrés, qui peuvent donc être testés sans accréditation spécifique, et les tests sont assez rapides à effectuer (durée de 48h). Comme la partie expérimentale ne disposait que d'une plage de temps réduite dans le cadre de ce travail, cela nous a paru le meilleur choix.

Les substances testées ont été choisies soit parce qu'il existait peu ou pas de données pour celles-ci, soit parce que les données n'étaient pas cohérentes, ou contradictoires selon les sources. La plupart de ces substances se sont donc avérées être des substances difficiles à tester, à cause de leurs propriétés physico-chimiques particulières ou d'un comportement inhabituel en solution, ce qui explique les valeurs controversées de la littérature.

2) Mesures de coefficients de partage octanol-eau

a) Introduction

Le logP est un paramètre qui rend compte de l'équilibre de partition d'une substance entre deux solvants : l'octanol et l'eau. L'octanol est un substitut pour modéliser les lipides des membranes cellulaires des organismes vivants, tout en ayant la propriété de ne pas former trop d'émulsions, contrairement à des mimes plus proches des membranes (comme la DiMyristoylPhosphatidylCholine, un composant majoritaire de la plupart des membranes, qui forme très facilement des émulsions), afin de faciliter la manipulation. En effet, les propriétés de l'octanol le rendent plus proche d'une moyenne entre les propriétés de partition avec l'eau des lipides membranaires et des lipides de stockage qui sont plus hydrophobes.

$logP = \log\left(\frac{C_{octanol}(X)}{C_{eau}(X)}\right)$ (sans dimension)

Ce paramètre a pour but de modéliser la capacité d'une substance à passer d'un milieu aqueux (milieu extérieur, circulation sanguine) aux membranes lipidiques des cellules, pour soit s'y accumuler, soit simplement les traverser et se retrouver dans le cytosol. Le logP est donc un paramètre très utilisé en chimie pharmaceutique notamment pour développer des molécules qui atteindront leur site d'action efficacement. Plus il est élevé, plus la substance aura d'affinité pour les lipides, plus il est faible, plus la substance va préférentiellement rester dans la phase aqueuse.

Le logP est également un des « endpoints » demandés pour compléter les dossiers REACH pour les substances qui sont produites ou importées en Europe à raison de plus d'une tonne par an (cf. Annexe VII de la réglementation REACH), car il modélise l'accumulation des substances dans les organismes vivants. Dans la stratégie de modélisation des substances à KREATiS, c'est aussi le premier paramètre qui est prédit à partir de la structure moléculaire, puis les prédictions des autres paramètres en découlent, d'où l'importance d'une prédiction fiable de ce premier paramètre.

C'est dans ce cadre que l'OCDE a mis en place des lignes directrices pour la mesure des différents « endpoints », dont le logP, afin d'avoir des données homogènes et de pouvoir avoir confiance et comparer les données entre elles.

Plusieurs méthodes sont recommandées :

- Méthode Shake-Flask (OECD Technical Guideline n° 107) : une phase octanol et une phase aqueuse sont mises en présence et agitées vigoureusement, avant de mesurer la concentration de la substance test dans les deux phases. Valable pour des substances ayant un logP compris entre -2 et 5.
- Méthode HPLC (OECD Technical Guideline n° 117) : la substance test estpassée dans une colonne HPLC hydrophobe avec un éluant aqueux. Le temps de rétention est corrélé avec le logP. Pour des substances ayant un logP compris entre 0 et 6. Nous avons constaté à KREATiS que cette méthode donnait des résultats très variables, elle n'est pas fiable. En effet, des étalons sont recommandés dans la ligne directrice, mais ne sont pas toujours adaptés pour comparer avec le comportement dans la colonne des substances testées. D'autre part, selon les laboratoires, des colonnes avec un revêtement intérieur de composition différente seront utilisées, ce qui peut expliquer la variabilité des résultats.
- Méthode Slow-Stir (OECD Technical Guideline n° 123) : Comme la méthode Shake-Flask, mais avec une agitation lente et des méthodes d'échantillonage différentes pour limiter de polluer une phase avec une autre. Utile et valable pour des substances ayant un logP supérieur à 5, car ces substances ont tendance à former plus facilement des émulsions que les substances moins hydrophobes. De plus, du fait de leur coefficient de partage très élevé, la concentration à l'équilibre dans l'eau sera très faible, et la présence de micro-goutelettes d'octanol, chargées en substance test, au sein de la phase aqueuse entraînera une mesure de la concentration en phase aqueuse virtuellement beaucoup plus grande qu'en réalité.

b) Protocole utilisé pour les tests

La méthode que nous avons utilisée au laboratoire était une méthode Shake-Flask, à laquelle nous avons apporté quelques modifications, en se basant sur la méthode Slow-Stir, afin de garantir des résultats fiables sur une plus grande gamme de produits. Globalement, ces modifications consistent à adapter les conditions expérimentales pour éviter ou limiter la formation possible d'émulsions métastables.

Le protocole présenté est le protocole final, après ajustements suite aux erreurs qui ont pu être effectuées sur les premières manipulations. Plusieurs de ces mesures ont été effectuées avec des stagiaires de niveau master 1, qui ont participé à évaluer et améliorer le protocole.

Toutes les manipulations ont été effectuées à température constante, à savoir 20°C (+/- 1°C).

i) Préparation des solvants présaturés

Dans une bouteille de 1L, introduire 900 mL d'octanol (analytical grade) et 50mL d'eau distillée. Agiter vigoureusement pendant 24h, bouteille fermée. Puis laisser reposer pour permettre la séparation des phases (24h).

Dans un bidon de grand volume (20 L), introduire de l'eau distillée et une quantité suffisante d'octanol pour saturer l'eau (solubilité de l'octanol dans l'eau : 0,5 g/L, µ=0.824 g/mL). Agiter vigoureusement pendant 24h, puis laisser reposer (24h), bidon fermé. Cette saturation est effectuée au préalable afin d'accélérer le temps d'équilibrage dans le test qui suivra.

Les quantités ont été choisies de sorte qu'il reste toujours un excès du solvant minoritaire, pour s'assurer que le solvant majoritaire est bien arrivé à saturation.

ii) Préparation de la solution mère S_0

Introduire la substance à tester dans V_0 = 60 mL d'octanol présaturé en eau. Le contenant doit être fermé et contenir le minimum d'espace vide au-dessus de la solution pour éviter la volatilisation de la substance. Recouvrir également le flacon d'aluminium pour éviter une éventuelle dégradation du produit par la lumière. La quantité de substance à introduire sera calculée pour obtenir une concentration un peu inférieure à 0,01M. Éventuellement moins si la solubilité dans l'octanol est trop faible, sachant que la concentration ne doit pas excéder 70% de la solubilité, éventuellement plus, jusqu'à 0,1M, si la substance est très hydrophobe (c'est-à-dire logP > 4). Agiter pour solubiliser la substance et laisser reposer.

Si le logP est inférieur à 1 (la substance est plus soluble dans l'eau que dans l'octanol), alors préparer cette solution mère dans l'eau et non dans l'octanol, suivant le même protocole.

iii) Partition de la substance entre l'eau et l'octanol

Préparer 2 erlenmeyers de 500 mL avec 500 mL d'eau (distillée, présaturée en octanol) et 10 mL de S_0, 2 erlenmeyers de 500 mL avec 500 mL d'eau et 5 mL de S_0, et 2 erlenmeyers de 250 mL avec 250 mL d'eau et 10 mL de S_0. Ces volumes doivent être ajustés au besoin pour des substances plus hydrophiles notamment (diminuer alors le ratio volumique octanol/eau). Inversement, pour une substance très hydrophile, S_0 aura été préparée dans l'eau, et les erlenmeyers sont remplis par exemple avec 50 mL d'octanol et 10 mL de S_0, puis 50 et 5 mL, puis 25 et 10 mL respectivement.

Les erlenmeyers sont fermés et mis sous agitation modérée au moyen d'un agitateur magnétique (de manière à obtenir un vortex profond de 2 cm maximum dans les erlenmeyers de 500 mL) pendant 24h. L'épaisseur de la couche d'octanol devrait être d'au moins 0,5 cm. Les erlenmeyers doivent être recouverts d'aluminium.

Laisser reposer les erlenmeyers pour permettre une séparation grossière des phases (8h). Prélever 3 fois environ 0,3 mL de la phase octanol avec une seringue. Prélever 3 fois environ 13 mL de la phase aqueuse. Pour cela, enfoncer la seringue à travers la phase octanol en éjectant doucement de l'air, prélever dans la phase aqueuse, puis retirer rapidement la seringue. Enlever l'aiguille (contaminée avec l'octanol) avant d'éjecter son contenu.

iv) Analyse des phases

L'analyse d'échantillons standards de concentration connue doit être effectuée préalablement pour pouvoir déduire du chromatogramme la quantité de substance en fonction des aires de pic. L'extraction doit aussi préalablement être étalonnée avec la substance test (ou un étalon externe similaire) à des concentrations proches de celle de l'expérience pour en déterminer le taux de récupération.

Pour les phases octanol : introduire chaque prélèvement dans un tube Eppendorf et centrifuger à 2000 RPM pendant 5 minutes. Prélever ensuite dans chaque tube exactement 0,1 mL de la solution avec une seringue et la verser dans un flacon intermédiaire. Y ajouter 0,9 mL du solvant d'extraction utilisé pour les phases aqueuses. Ajouter un peu de sulfate de magnésium pour sécher l'échantillon. Enfin, prélever le liquide séché avec une seringue sèche, puis équiper la seringue avec un filtre (pour ne pas verser de grains de $MgSO_4$) et vider son contenu dans un pilulier pour GC (chromatographie en phase gazeuse).

Pour les phases aqueuses : les prélèvements sont introduits dans des tubes de 30 mL pour centrifugation. Centrifuger ces échantillons à 8000 RPM pendant 10 minutes pour permettre une bonne séparation des phases. Exactement 10 mL sont prélevés de chaque tube pour être introduits dans des fioles jaugées du même volume. Extraire la substance de la phase aqueuse avec 0,5 mL, 0,4 mL et 0,3 mL de solvant organique volatil dans lequel la substance est plus soluble que dans l'eau, et qui n'est pas lui-même miscible dans l'eau (Acétate d'éthyle, cyclohexane, éther de pétrole). Pour cela, les fioles contenant la phase aqueuse et la phase organique sont agitées, puis mises à reposer. Les volumes de solvant ajoutés sont dégressifs, pour un total de 1,2 mL à cause de la solubilisation du solvant dans l'eau. Ces volumes peuvent être ajustés, le but étant qu'environ 0,9 mL de phase organique puissent être récupérés après regroupement. Les 3 extraits organiques ainsi obtenus pour chaque fiole sont introduits dans le même pilulier contenant du $MgSO_4$ pour sécher la phase. Ensuite, le surnageant est prélevé avec une seringue et introduit dans un pilulier pour GC en ayant adapté un filtre à la seringue.

v) Traitement des données

Effectuer le bilan de masse des expériences pour vérifier qu'il n'y a pas eu de perte non quantifiée dans l'expérience.

Calculer le logP pour chaque expérience. L'erreur ne doit pas excéder 0,3 unités log. Faire la moyenne des 6 valeurs pour obtenir la valeur finale de logP.

c) Substances testées

Une famille de substances qui a été étudiée en particulier fut celle des cétones α,β-insaturées. En effet, pour ces substances, les différentes valeurs de logP ainsi que les valeurs d'écotoxicité correspondantes semblaient incohérentes les unes par rapport aux autres et ne permettaient pas de conclure sur le mécanisme d'action des cétones α,β-insaturées. En effet, selon que celles-ci aient des valeurs de toxicité correspondant à la narcose non-polaire ou bien qu'elles soient plus toxiques, on peut conclure que le mécanisme d'action est respectivement la narcose non-polaire ou la réactivité d'électrophile mou. En effet, le fait que ces substances possèdent un carbonyle avec une double liaison C=C en position alpha-beta justifie l'hypothèse d'une réactivité, bien que les cétones α,β-insaturées soient reconnues pour être peu réactives (Portoghese *et al.*, 1989). Nous avons fait des mesures de coefficient de partage pour les cétones suivantes :

Nom de la substance	Numéro CAS	LogP les plus fiables de la littérature	Référence	Structure
Oxyde de mésityle	141-79-7	1,19	Valeur prédite (par le modèle de KREATiS qui lie le logP et la solubilité) à partir de la valeur de solubilité aqueuse : 26980 mg/L ; ECHA, 2009, ligne directrice OCDE 105, méthode agitation de flacon, 20°C, K1, GLP	
R-(-)-carvone	6485-40-1	2,54*	Valeur prédite à partir de la valeur de solubilité aqueuse : 1310 mg/L ; EpiSuite, Susuki, 1991, 25°C / Yalkowsky & Dannenfelser, 1992, 18°C, K2	
isophorone	78-59-1	1,66 / 1,67	ECHA, 1989, OCDE 107, méthode agitation de flacon, 23°C, K2 / ECHA, Veith *et al.*, 1980, méthode agitation de flacon, 20°C, K2	
β-méthylionone	127-43-5	4,56	Donnée confidentielle, K1	

*Tableau 9 : cétones α,β-insaturées pour les tests de logP. * : valeur de logP prédite à partir d'une solubilité expérimentale fiable.*

Une autre famille de composés que nous avons étudiée est la famille des lactates, car très peu de données existaient pour ces substances. Nous avons testé les lactates de méthyle, éthyle et propyle. Finalement, quelques composés variés ont aussi été mesurés et sont listés ci-dessous, avec les lactates :

Nom de la substance	Numéro CAS	LogP les plus fiables de la littérature	Référence	Structure
Lactate de méthyle	547-64-8	ND	/	
Lactate d'éthyle	97-64-3	ND	/	

Lactate de propyle	616-09-1	ND	/	
1,8-cinéole	470-82-6	2,74 / 3,4 / 2,16	EpiSuite, Griffin *et al.*, méthode par HPLC, 1999 / ECHA, 1998, OCDE 117, méthode par HPLC, K1, GLP / Valeur prédite à partir de la valeur de solubilité aqueuse : 3500 mg/L ; EpiSuite, Yalkowsky & Dannenfelser, 1992, 21°C	
Dicyclo-pentadiène	77-73-6	2,78	ECHA, IPCS, 2005, 25°C, K2 / ECHA, MITI, 1998, OCDE 107, méthode par agitation de flacon, 25°C, K4	
3-carène	13466-78-9	4,38 / 4,61	ECHA, Griffin *et al.*, 1999, OCDE 117, méthode par HPLC, 37°C, K2 / ECHA, Howard & Meylan, 1997, 25°C, K4	
Myrcène	123-35-3	4,17	ECHA, Howard & Meylan, 1996, K4	
L-(-)-menthol	2216-51-5	3,14	ECHA, 2010, EU method A.8, méthode par agitation de flacon, 25°C, K2	
géraniol	106-24-1	2,6 / 3,56	ECHA, 1995, OCDE 117, méthode par HPLC, 25°C, K1, GLP / EpiSuite, Griffin *et al.*, 1999	

Tableau 10 : substances variées pour les tests de logP. HPLC = chromatographie liquide à haute performance.

d) Résultats et discussion

Une mesure d'une substance de référence a été effectuée afin de valider la méthode. Nous avons choisi comme référence la heptan-2-one :

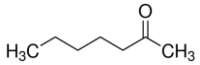
Nom de la substance	Numéro CAS	LogP les plus fiables de la littérature	Référence	Structure
Heptan-2-one	110-43-0	1,98	EpiSuite, Hansch *et al.*, 1995 / ECHA, Tewari *et al.*, 1982, methode generator column, 25°C, K2	
LogP prédit par KREATiS	**logP mesuré dans ce travail**		**Commentaires**	
1,98	2,02 ± 0,03		Mesuré fin 2015	

Tableau 11 : données de la substance de référence.

Les résultats obtenus sur la substance de référence montrent que notre méthode fonctionne correctement, et permet de reproduire les résultats de la littérature sur une substance dont les données sont bien validées.

Avec les moyens utilisés au laboratoire, nous n'avons pas pu obtenir de valeur fiable de logP pour certaines substances. En effet, l'isophorone et le menthol ont des températures d'ébullition légèrement supérieures à celle de l'octanol. Or, comme tous les échantillons contiennent une forte teneur en octanol (même les extractions de phases aqueuses où l'octanol s'est en partie solubilisé), un pic important correspondant à l'octanol apparaît systématiquement sur le chromatogramme. Et comme la méthode d'analyse utilisée était la chromatographie en phase gazeuse, les positions relatives des pics des substances sur le chromatogramme dépendent majoritairement de leur volatilité relative, que l'on peut évaluer grâce à leurs températures d'ébullition. La température d'ébullition de l'octanol est de 194°C. Celle du menthol est de 212°C et celle de l'isophorone de 215°C. Nous avons effectué plusieurs tests avec le menthol en faisant varier les méthodes de montée en température du four. Malgré nos efforts, nous n'avons pas pu obtenir une relation concentration de menthol – aire de pic cohérente, car au moins une partie du menthol est entraînée par l'octanol au moment de sa vaporisation, et ainsi le menthol est caché (en partie au moins) par le pic de l'octanol. La même chose est attendue avec l'isophorone qui n'a donc pas été testée. Nous n'avons pas poussé plus loin les méthodes chromatographiques qui auraient pu permettre de résoudre ce problème.

D'autre part, nous n'avons pas pu obtenir de valeur pour le logP du géraniol, car nous avons constaté que celui-ci se dégradait, aussi bien dans les phases aqueuses que dans les phases organiques, car nous voyions le pic de géraniol fortement diminué, en faveur d'un nouveau pic, absent du chromatogramme du solvant et de l'octanol, qui correspond au produit de dégradation, que nous n'avons pas pu identifier.

Enfin, nous n'avons pas pu obtenir de valeur correcte pour la β-méthylionone et le myrcène, car ces substances sont trop hydrophobes pour être mesurées par cette technique. Les concentrations aqueuses après équilibrage sont trop faibles, et très facilement faussées par la présence de microgouttelettes d'octanol chargées en substance test présentes dans les phases aqueuses, et ceci malgré l'étape de centrifugation. En effet, nous n'avons pu implémenter la méthode Slow-Stir qui nécessite une verrerie spéciale et des temps d'expérience significativement plus longs. D'autre part, la limite de quantification de la chromatographie à notre disposition permettait difficilement de mesurer

les concentrations dans la phase aqueuse, à moins de concentrer préalablement l'échantillon par évaporation du solvant, ce que nous n'avons pas entrepris car ces substances sont quand-même assez volatiles.

Voici les résultats pour les autres substances. Le protocole ayant été affiné au cours du temps, les premières expériences sont moins fiables que les suivantes. Les substances sont triées par famille chimique (cétones, divers, puis lactates).

Nom de la substance	LogP les plus fiables de la littérature	LogP prédit par KREATiS	logP mesuré dans ce travail	Commentaires
Oxyde de mésityle	1,19*	1,15	1,16 ± 0,05	Mesuré fin 2015
R-(-)-carvone	2,54*	2,58	2,52	Mesuré au printemps 2016
1,8-cinéole	2,74 / 3,4 / 2,16*	2,83	2,04 ± 0,25	Mesuré fin 2015
Dicyclo-pentadiène	2,78	3,72	3,34	Mesuré au printemps 2015
3-carène	4,38 / 4,61	4,79	(3,12)	Une des premières mesures, au printemps 2015, non centrifugé = non fiable. Probablement trop hydrophobe pour notre méthode.
Lactate de méthyle	ND	-0,34	-0,61 ± 0,3	Mesuré au printemps 2016
Lactate d'éthyle	ND	0,21	0,12 ± 0,22**	Mesuré au printemps 2015
Lactate de propyle	ND	0,76	(1,29)	Mesuré au printemps 2015. Bilan de masse: 1/3 de la substance a été perdu = non fiable

*Tableau 12 : résultats des mesures de logP. * : valeur de logP prédite à partir d'une solubilité expérimentale fiable. ** : valeur utilisée pour construire le modèle de KREATiS pour les lactates.*

Ainsi, tel que montré avec la substance de référence, cette méthode nous permet d'obtenir des résultats justes et précis, à condition que les propriétés physico-chimiques des substances testées ne soient pas problématiques (logP trop haut, température d'ébullition juste au-dessus de celle de l'octanol, dégradation, *etc.*). En effet, nous avons obtenu des résultats en accord avec les prédictions et/ou les valeurs de la littérature pour une majorité de substances (Tableau 12). Ces mesures ont permis de valider certaines valeurs de la littérature et d'en discréditer d'autres, en se basant sur les prédictions de KREATiS et sur les valeurs mesurées pour d'autres endpoints (solubilité, écotoxicité). Par exemple, notre mesure du logP du 1,8-cinéole a pu permettre de valider la valeur de solubilité de la littérature, en invalidant les valeurs de logP de la littérature. Les deux valeurs de logP de la littérature étaient de plus effectuées avec la méthode par HPLC, que nous avons identifiée comme peu fiable. Les

mesures sur les lactates ont permis d'élargir le domaine d'applicabilité du modèle de prédiction du logP de KREATiS à ce type de substances. L'une des valeurs générées a été intégrée dans le set d'entraînement, alors que les deux autres servent pour la validation.

3) Mesures de solubilités aqueuses

a) Introduction

La solubilité aqueuse d'une substance est un paramètre primordial pour en évaluer le potentiel de toxicité aquatique. En effet, la toxicité des narcotiques particulièrement est directement corrélée avec la solubilité, car cette dernière est responsable de l'activité chimique de la substance en milieu aqueux (l'eau dans laquelle les organismes aquatiques évoluent, mais également la circulation sanguine de n'importe quel organisme ayant une circulation aqueuse destinée à transporter les substances depuis et vers les différentes cellules de l'organisme, comme le sang chez les animaux ou la sève dans les plantes).

La solubilité aqueuse est définie comme la concentration maximale d'une substance qui peut être solubilisée dans de l'eau pure à une température donnée (c'est-à-dire dispersée de manière homogène et stable dans le temps). Pour les dossiers réglementaires, elle est habituellement exprimée en mg/L et mesurée à une température entre 20 et 25°C.

L'OCDE a donc édité une ligne directrice pour mesurer de manière standardisée cet « endpoint » dans la ligne directrice n° 105, où plusieurs méthodes sont recommandées :

- Méthode Shake-Flask : recommandée pour des solubilités supérieures à 10 mg/L
- Méthode élution sur colonne : recommandée pour des solubilités inférieures à 10 mg/L.

Une troisième méthode importante a été développée pour mieux mesurer les substances très peu solubles, en se basant sur une modification de la ligne directrice 123 pour le logP. Cette méthode n'a pas encore été recommandée par l'OCDE, mais est quasiment incontournable si l'on veut mesurer la solubilité de substances liquides très hydrophobes :

- Méthode Slow-Stir : méthode préférentielle pour des solubilités inférieures à 10 mg/L.

b) Protocole utilisé pour les tests

i) Expérience

Peser 5 fois la quantité saturante et l'introduire dans 3 récipients fermés de 10 mL d'eau distillée. Mettre sous agitation vigoureuse à 30 °C. Au bout de 24h, l'un des flacons est mis à l'équilibre à température ambiante (température du laboratoire constante à 20±1°C) pendant 24h avec une agitation occasionnelle.

Puis, centrifuger le contenu à température ambiante toujours et mesurer la concentration de la substance dans la phase aqueuse (par exemple analyse en HPLC-UV/Vis, ou extraction liquide-liquide comme pour le test de logP et analyse en GC) (5 analyses).

Même traitement pour les flacons 2 et 3 après 48h et 72h d'agitation à 30°C. Mesurer le pH de chaque solution.

Remarque : on peut aussi effectuer la période d'agitation à température ambiante, à condition de prouver que selon le temps d'agitation utilisé, la concentration mesurée reste identique (à partir d'un certain temps).

ii) Analyse

Analyse en GC après extraction liquide-liquide : Dans le flacon de solution aqueuse, prélever 3 échantillons de 1 mL. Les centrifuger pour séparer éventuellement l'excès de substance si elle est liquide, ou bien filtrer si elle est solide. Y ajouter 0,4 puis 0,3 puis 0,3 mL de solvant d'extraction. Agiter puis laisser reposer. Prélever les phases organiques et les réunir dans un même flacon. Sécher ce volume sur sulfate de magnésium. Récupérer le liquide en filtrant et analyser en GC.

Analyse par HPLC-UV/Vis : pour les substances qui ont une bande d'absorption spécifique et suffisamment intense dans l'UV ou le visible, la concentration dans l'eau peut directement être suivie dans les échantillons aqueux, en les passant dans une HPLC avec un éluant adapté munie d'un détecteur UV/Vis qui aura été réglé préalablement pour mesurer l'absorption aux longueurs d'ondes d'intérêt.

Chaque méthode d'analyse sera étalonnée au préalable avec des solutions de concentration connue, se rapprochant le plus possible de la limite de solubilité.

Remarque : si la concentration est trop importante, les prélèvements de 1 mL de la solution aqueuse peuvent être dilués pour atteindre 10 mL de phase aqueuse, qui pourra ensuite être extraite avec les mêmes quantités de solvant que ci-dessus.

iii) Traitement des données

La différence entre au moins les deux derniers flacons ne doit pas excéder 15%. Si on voit une augmentation des valeurs en cours de manipulation, le temps d'équilibrage doit être augmenté.

Faire la moyenne de concentration à partir des 3 flacons.

c) Substances testées

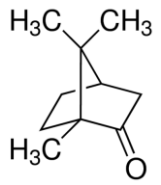
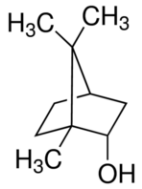
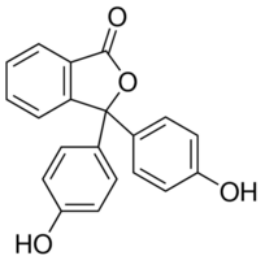
Nom de la substance	Numéro CAS	Solubilités les plus fiables de la littérature (mg/L)	Référence	Structure
Camphre	76-22-2	1537,3 / 1600 / 100	ECHA, 2012, ligne directrice OCDE 107, méthode agitation de flacon, 25°C, K2 / EpiSuite, Yalkowsky & He, 2003, 25°C, K2 / EpiSuite, Riddick *et al.*, 1986, 25°C, K2	
(+)-bornéol	464-43-7	740	EpiSuite, Yalkowsky *et al.*, 1992, 25°C, K2	
Phénolphtaléine	77-09-8	400	ECHA, 2010, OCDE 105, méthode par élution sur colonne, 20°C, pH = 5,8-6,1, K1 / EpiSuite, Yalkowsky & He, 2003, K2	

Tableau 13 : substances pour les tests de solubilité.

Ces substances posent problème car d'après les prédictions de KREATiS, on attend des solubilités respectives de 21, 11 et 29 mg/L, alors que les valeurs mesurées de la littérature sont beaucoup plus élevéess. Les valeurs prédites par KREATiS semblent plus cohérentes avec les valeurs d'écotoxicité (si disponibles) que les valeurs de solubilité de la littérature.

d) Résultats et discussion

Le camphre et le bornéol ont été testés avec la méthode classique d'extraction liquide-liquide puis passage en GC. La phénolphtaléine a été testée par HPLC-UV/Vis, car elle possède une bande d'absorption spécifique assez intense à 276 nm, et une bande moins spécifique mais plus intense à 232 nm. L'éluant choisi pour la colonne était un mélange de 50% d'acétonitrile et 50% d'eau, l'acétonitrile contenant 1% d'acide formique. En effet, la phénolphtaléine se déprotone et change de forme vers pH 8,2, ce qui change son spectre d'absorption et qui est à l'origine de ses propriétés d'indicateur de pH. Ainsi, la présence d'acide formique garantit de rester dans la forme acide de la phénolphtaléine.

Nous n'avons pas pu obtenir de valeur de solubilité pour le bornéol, car celui-ci a un comportement atipyque en solution. En effet, les solutions aqueuses les plus concentrées en bornéol moussaient significativement lorsqu'elles étaient agitées. De plus, pour les étalons en solvant et ceux dans l'eau, préparés par dilutions successives, les aires de pic obtenues en GC n'étaient pas bien corrélées avec la concentration théorique en bornéol et décroissaient rapidement avec les dilutions. Tout cela semble indiquer que le bornéol possède des propriétés interfaciales, c'est-à-dire que les molécules de bornéol

ne se répartissent pas de manière homogène dans la solution, mais se concentrent au niveau des interfaces. Soit au niveau de l'interface liquide-air, soit au niveau de l'interface liquide-paroi en verre.

Les résultats pour le camphre et la phénolphtaléine sont donnés dans le tableau ci-dessous :

Nom de la substance	Solubilités les plus fiables de la littérature (mg/L)	Solubilité prédite par KREATiS (mg/L)	Solubilité mesurée dans ce travail (mg/L)	Commentaires
Camphre	1537,3 / 1600 / 100	21	1495 ± 62	Dissolution du camphre solide difficile
Phénolphtaléine	400	29	3,57 ± 0,26	Même résultat obtenu avec les 2 longueurs d'onde

Tableau 14 : résultats de solubilité.

Ainsi, nos mesures semblent valider les données de la littérature pour le camphre, et invalider les prédictions de KREATiS, alors que pour la phénolphtaléine, nos mesures semblent plutôt aller dans le sens de la prédiction de KREATiS, bien qu'étant encore nettement inférieures. Nous n'avons pour l'instant pas de piste pour expliquer ce résultat.

4) Tests de toxicité aigüe pour les daphnies

a) Introduction

Les daphnies sont utilisées comme espèce modèle pour mesurer la toxicité aquatique des substances chimiques en laboratoire. En effet, ces crustacés d'eau douce constituent un niveau trophique intermédiaire, les consommateurs primaires : elles se nourrissent des producteurs comme les algues et sont mangées par d'autres espèces, les consommateurs secondaires, par exemple les poissons. Les daphnies possèdent deux modes de reproduction en fonction des conditions de vie. Lorsque les conditions sont optimales, elles utilisent la parthénogénèse, c'est-à-dire une reproduction asexuée, où chaque daphnie se clone elle-même, produisant 10 à 30 bébés à la fois tous les 3 jours, à partir du huitième jour de vie (maturité sexuelle). Les conditions idéales sont un approvisionnement suffisant en nourriture (algues), une température tempérée (environ 20-25°C), un bon cycle de lumière (16h de jour, 8h de nuit), une qualité d'eau satisfaisante (teneur en minéraux adaptées pour les macronutriments et les éléments trace, mais aussi des vitamines pour des conditions optimales). Lorsque ces conditions ne sont plus respectées, ou lors d'un stress particulier, ce qui correspond typiquement à l'automne dans la nature, où la température et la durée du jour diminuent et la nourriture se fait plus rare, les daphnies passent en mode de reproduction sexuée. Pour cela, il suffit qu'une daphnie mâle apparaisse, qui viendra ensuite fertiliser toutes les autres daphnies (femelles par défaut), qui produiront alors un éphipium, une coque noire résistante contenant deux cellule-œufs, au lieu de produire 10-30 œufs. Cet ephipium peut résister à des conditions extrêmes (gel, obscurité, sécheresse, etc.) et s'ouvrira uniquement lorsque les conditions redeviendront bonnes pour la vie des daphnies, libérant ainsi deux bébés daphnies.

Les tests en laboratoire sont effectués avec des clones, en maintenant des conditions idéales pour rester en parthénogénèse et conserver ainsi le même patrimoine génétique permettant une bonne reproductibilité des tests, et permettant l'approvisionnement suffisant en jeunes daphnies pour lancer régulièrement des tests.

Le paramètre physiologique mesuré pour quantifier la toxicité des substances sur les daphnies est l'immobilisation et non la mortalité, qui est plus difficile à évaluer. Une daphnie est considérée immobile lorsqu'une agitation légère du flacon ne provoque pas un déplacement actif d'au moins 15 secondes. On renseigne généralement une CE50, c'est-à-dire la concentration de substance qui produit l'effet (immobilisation) pour 50% des daphnies dans le test.

b) Matériel et méthodes

Pour ces tests, un protocole spécifique n'a pas été écrit, car la ligne directrice de l'OCDE est bien détaillée et laisse peu de liberté quant à la manière de procéder, contrairement aux tests de logP et de solubilité. La ligne directrice OCDE n°202 laisse par contre le choix sur le milieu dans lequel les daphnies sont élevées et dans lequel les tests de toxicité sont effectués. On a le choix entre utiliser une eau du réseau contrôlée et ajustée si besoin en fonction des teneurs en différents minéraux et particules organiques, ou bien une eau reconstituée à partir d'eau déminéralisée et de sels de minéraux. Plusieurs compositions sont alors proposées, un milieu dit M4 et un milieu dit M7 sont recommandés et leur préparation est détaillée dans les annexes de la ligne directrice. Nous avons choisi de préparer du milieu M4, qui donne en général plus facilement des daphnies en bonne santé, probablement grâce à un apport en minéraux mieux équilibré.

Nous avons suivi le pH et la teneur en oxygène dissous avec un multimètre VWR® pHenomenal® 6100H S2. Le pH doit rester proche de la neutralité, et l'oxygène dissous > 3mg/L. La concentration de la substance à t=0 (début du test) et t=48h (fin du test) a été mesurée pour certaines substances par analyse en GC après extraction liquide-liquide, suivant le même protocole que dans les tests de solubilité.

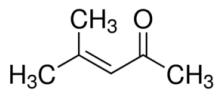
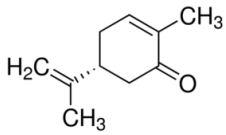
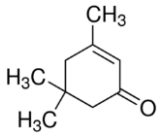
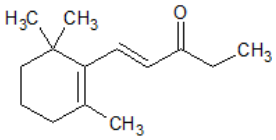
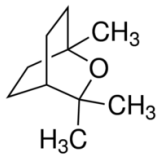
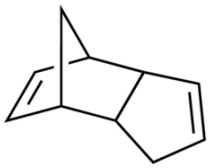
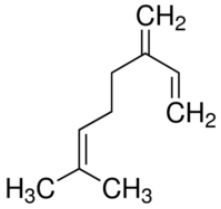
Les daphnies utilisées étaient des *Daphnia magna*, clone 5. Les algues cultivées pour les nourrir sont des algues vertes *Pseudokirchneriella subcapitata*. Ces organismes ont été récupérés chez le laboratoire LPL (Laboratoire des Pyrénées et des Landes. N° SIRET 418 814 059 00014). La croissance des algues était suivie en comptant les individus au microscope dans une cellule de Thoma.

L'élevage des daphnies était maintenu dans des béchers de 2-3 L contenant 2 L de liquide, couverts pour éviter l'entrée de poussières, mais non hermétiquement fermés. Lest tests de toxicité étaient effectués dans des flacons de 20 mL fermés avec un volume d'air faible (<3 mL), pour éviter la vaporisation de la substance test. Les algues étaient cultivées en conditions stériles dans des erlenmeyers fermés avec un bullage d'air filtré à travers des filtres de diamètre de pore de 0,22 µm.

L'élevage des daphnies et les tests de toxicité sur les daphnies ont été maintenus à une température de 20°C ± 1°C. Les algues étaient cultivées dans une chambre thermostatée à 30°C, avec une lampe de culture, pour une croissance plus rapide.

c) Substances testées

Les cétones α,β-insaturées ont aussi fait l'objet de mesures de toxicité pour les daphnies car les valeurs de toxicité et les valeurs de logP étaient incohérentes. D'autres substances, communes aux tests de logP ou non ont été testées car la relation hydrophobie-écotoxicité semblait incohérente, ou des données manquaient.

Nom de la substance	Numéro CAS	Toxicité fiable de la littérature : 48h-CE50 (mg/L)	Référence	Structure
Oxyde de mésityle	141-79-7	89,1	ECHA, 2008, OCDE 202, méthode statique, avec solvant, concentration nominale, 21-22°C, K1, GLP	
R-(-)-carvone	6485-40-1	38	ECHA, 2001, OCDE 202, méthode statique, avec solvant, K2, GLP	
isophorone	78-59-1	120	ECHA/US EPA, LeBlanc, 1980, mortalité mesurée, méthode statique, concentration nominale, 22°C, intervalle de confiance : [72-170], K2	
β-méthylionone	127-43-5	/	Pas de données	
1,8-cinéole	470-82-6	307,04 (50 à 130 poissons et algues)	LDP, 2012, OCDE 202, méthode semi-statique, concentration mesurée, 20°C, K1	
Dicyclo-pentadiène	77-73-6	0,62 / 0,823 / 8	ECHA, 1995, OCDE 202, statique, nominale, avec solvant Tween, 21°C, confiance [0,53-0,72], K2, GLP / ECHA, 2016, OCDE 202, semi-statique, fermé sans espace d'air, mesurée, confiance [0,740-0,909], 20°C, K1, GLP / ECHA, 1995, OCDE 202, semi-statique, nominale, confiance [6,8-9,5], K4	
myrcène	123-35-3	1,47	LDP, 2014, OECD 202, semi-statique, mesurée, fermé, K1	

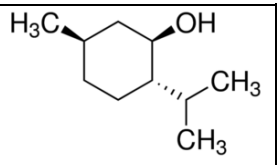
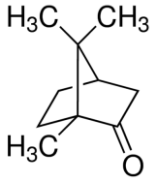
L-(-)-menthol	2216-51-5	26,6	ECHA, 2002, OCDE 202, statique, mesurée, 19,7°C, K2, GLP	
camphre	76-22-2	/	Pas de données	

Tableau 15 : substances pour les tests de toxicité aigüe pour les daphnies. Toxicité exprimée en 48h-CE50 où l'effet mesuré est l'immobilisation (sauf mention contraire).

Pour le 1,8-cinéole, les valeurs de toxicité pour les poissons et les algues ont été données à titre indicatif pour souligner le fait que la valeur donnée pour la daphnie est anormalement haute, sachant que les toxicités pour les trois espèces sont normalement du même ordre de grandeur pour un mécanisme d'action narcotique, qui est attendu ici.

d) Résultats et discussion

D'après la ligne directrice, un test ne peut être considéré valide si le taux en oxygène dissous descend en dessous de 3 mg/L. Lors du test avec notre substance de référence, le menthol, le taux d'oxygène est resté à des taux normaux, et la valeur de toxicité correspondait à la valeur attendue d'après la littérature (voir Tableau 16). Cette situation est arrivée plusieurs fois dans nos expériences, mais les résultats sont quand-même présentés dans le tableau ci-dessous. Les valeurs de toxicité prédites par KREATiS le sont avec le modèle de la narcose non-polaire. En effet, toutes ces substances sont supposées agir selon ce mécanisme / mode d'action. Seule la concentration du menthol a pu être analysée pour connaître la concentration réelle dans les flacons au début et à la fin du test.

Nom de la substance	Toxicité fiable de la littérature (mg/L)	Toxicité prédite par KREATiS (mg/L)	Toxicité mesurée dans ce travail (mg/L)	Commentaires
Oxyde de mésityle	89,1	344	95,5	OD = 1,8 mg/L à la fin. Concentration nominale.
R-(-)-carvone	38	34,3	10,2	Concentration nominale.
isophorone	120	69,4	102	OD = 2,1 mg/L à la fin. Concentration nominale.
β-méthylionone	/	1,67	CE50 > 6,6	OD = 3 mg/L à la fin. Nominale. L'immobilisation la plus forte était 20%. La substance était en solution dans de l'acétate d'éthyle.
1,8-cinéole	307,04 (50 à 130 poissons et algues)	78,4*	111	OD = 1,7 mg/L à la fin. Nominale.
Dicyclo-pentadiène	0,62 / 0,823 / 8	2,94	4,18	Nominale.
myrcène	1,47	0,446	0,874	Piégeage en surface. Nominale.
L-(-)-menthol	26,6	16,1	26,0 (n) / 17,0 (m)	n = concentration nominale, m = mesurée (moyenne géométrique entre la concentration initiale et la finale)
camphre	/	372	CE50 < 93	Nominale. L'immobilisation la plus faible était 100%. Très difficile à dissoudre.

*Tableau 16 : résultats des mesures de toxicité aïgue pour les daphnies. Les toxicités sont exprimées en 48h-CE50, où l'effet mesuré est l'immobilisation. * = valeur prédite à partir d'une solubilité expérimentale fiable de la littérature.*

Ainsi, pour l'oxyde de mésityle, la valeur obtenue dans ce travail confirme plutôt la valeur de la littérature, qui montre une toxicité plus forte que la toxicité de base prédite par KREATiS. Mais cette conclusion doit être balancée par le fait que la concentration en oxygène dissous (OD) a fortement chutée.

La CE50 à 48h mesurée pour la carvone est plus faible que les CE50 à 48h de la littérature et prédite par KREATiS. D'autant plus que c'est une valeur nominale, et que la CE50 réelle qui serait obtenue avec un suivi de la concentration serait plus faible encore. Nous n'avons pas de piste pour expliquer cela.

Pour l'isophorone, la valeur que nous avons obtenue est entre la valeur de la littérature et la valeur de KREATiS, et les deux valeurs expérimentales sont plus hautes que la prédiction, mais sont également des valeurs nominales, donc la valeur réelle doit être plus faible et se rapprocherait ainsi de la valeur

prédite. Dans ce cas, il semble que la prédiction suivant le mécanisme/mode d'action de la narcose non-polaire soit correcte.

La valeur mesurée pour le 1,8-cinéole semble confirmer l'hypothèse de KREATiS que la toxicité pour les daphnies n'est pas due à un mécanisme différent de la narcose non-polaire observée chez les poissons. Ce résultat reste à prendre avec précaution puisque l'OD était sous le seuil d'acceptabilité à la fin du test.

La valeur mesurée pour le dicyclopentadiène, le myrcène et le menthol se rapproche également de la prédiction de KREATiS.

Le camphre est probablement plus toxique à cause d'effets physiques dus à un mauvais équilibre dans l'eau (comportement un peu similaire à celui du bornéol observé dans les tests de solubilité, mais en moins prononcé).

5) Conclusion

Pour certaines substances, nous avons pu valider les hypothèses de KREATiS, à savoir que les substances agissaient bien selon un MoA de narcose non-polaire, et que la valeur de toxicité attendue devait être proche de celle prédite par les modèles QSAR de KREATiS. C'est le cas pour le 1,8-cinéole, le dicyclopentadiène et le myrcène, pour lesquels nous avons pu mesurer des toxicités pour les daphnies similaires à celles prédites par KREATiS, qui allaient à l'encontre des valeurs expérimentales de la littérature. Nous avons pu valider les prédictions de paramètres physico-chimiques de KREATiS (dont découlent les prédictions de toxicité) pour l'oxyde de mésityle et la carvone. Aussi, pour le 1,8-cinéole, nous avons confirmé avec notre mesure de logP la valeur de solubilité expérimentale, qui, d'après les modèles de KREATiS, est bien corrélée avec les valeurs de toxicité expérimentales, bien qu'en désaccord avec les valeurs de logP de la littérature.

Pour d'autres substances, nous avons plutôt confirmé les valeurs de la littérature montrant une toxicité plus importante que celle prédite par les QSARs de toxicité de base, suggérant un MechoA différent d'une simple perturbation des membranes. C'est le cas pour l'oxyde de mésityle. En revanche, pour le camphre, nous avons confirmé la valeur de solubilité expérimentale nettement plus élevée que la valeur prédite par KREATiS, valeur qui devrait déboucher sur une toxicité plus faible par le mécanisme de la narcose. Or notre test de toxicité sur cette substance a montré au contraire une toxicité plus forte qu'attendue, ce qui s'oppose à la valeur de solubilité mesurée. Cette substance reste donc problématique.

D'autre part, pour certaines substances, nous n'avons pas pu améliorer l'évaluation du devenir expérimental, mais nous avons rencontré des problèmes expérimentaux empêchant une mesure fiable dans notre laboratoire. Cela a néanmoins permis de comprendre pourquoi les valeurs de la littérature pouvaient être incohérentes. C'est le cas pour le menthol, l'isophorone et le géraniol (tests de logP), le bornéol (solubilité) et le camphre (toxicité).

Chapitre 7 : Discussion générale et conclusion.

L'objectif de cette thèse était de mieux caractériser les modes/mécanismes d'action que ne le faisait le schéma de Verhaar (Enoch *et al.*, 2008; Verhaar *et al.*, 1992) majoritairement utilisé par les écotoxicologues, et que les autres méthodes de prédiction générales. Nous voulions aussi unifier les domaines de la toxicologie (santé humaine) et de l'écotoxicologie en proposant une classification et une méthode de prédiction utilisable pour des approches de prédiction de toxicité dans ces deux domaines. Le but était de couvrir la plus grande partie possible des substances organiques présentes sur le marché. Ce travail devait se baser sur la structure des molécules et servir de première étape importante pour l'utilisation de modèles QSAR adaptés afin de prédire l'écotoxicité et la toxicité.

La première étape était de constituer une base de données avec des informations sur les mécanismes et modes d'action. Dès le départ, nous avons choisi de nous baser sur les mécanismes d'action, plutôt que les modes d'action, car ils sont plus directement et fortement liés à la structure des molécules et à leurs propriétés, et de ces mécanismes d'action découlent tout le reste des AOPs. Ce travail s'est fait progressivement. Nous avons commencé à développer des méthodes de prédiction avec un set de données de 173 molécules pour lequel nous avions identifié le MechoA, et nous avions classifié les MechoAs uniquement en catégories générales, au nombre de 5. Nous avons finalement abouti avec un set de données de 794 molécules avec un MechoA connu, et des MechoA classés en 6 catégories générales et 23 sous-catégories. Ce travail de classification sera bien sûr poursuivi au-delà de cette thèse pour caractériser un nombre toujours plus grand de molécules et de mécanismes.

Nous avons d'abord tenté de répondre à cette problématique d'une meilleure caractérisation en développant une méthode de prédiction faisant appel à des calculs quantiques. En effet, l'idée de départ était que le taux d'erreurs (14% en moyenne) et de non-classification (35% en moyenne) obtenu par le schéma de Verhaar était grandement dû au fait d'utiliser des alertes structurales connues pour classer les composés. Cette limitation pouvait en théorie être contournée grâce aux descripteurs électroniques générés par les calculs quantiques. En effet, une sous-structure différente d'une alerte structurale connue mais qui possède une répartition électronique semblable serait attendue d'avoir le même mécanisme d'action. Et inversement, si la sous-structure est la même, mais que l'environnement moléculaire de cette sous-structure entraîne une modification importante des propriétés stéréo-électroniques, les calculs quantiques devraient le repérer, et détecter un changement de mécanisme d'action.

La méthode quantique s'est avérée assez complexe et coûteuse en temps de calcul pour obtenir des paramètres électroniques précis, et a obtenu des résultats certes meilleurs que ceux du schéma de Verhaar, mais avec un peu moins de 70% de classifications correctes des mécanismes détaillés et 17% de mauvaises classifications, le bilan était mitigé. Les résultats étaient plus encourageants lorsque la méthode avait été développée pour prédire les 5 MechoAs généraux initiaux, mais lorsque les sous-catégories se sont multipliées, ainsi que le nombre de molécules dans les sets de données, les performances de la méthode ont aussi chuté significativement. En effet, une variabilité plus grande dans les sets de données en termes de structures chimiques et en termes de MechoAs implique qu'il est plus difficile de trouver des valeurs de paramètres quantiques qui caractérisent bien toute cette variabilité.

C'est pourquoi nous avons décidé de développer une méthode par alertes structurales, en se rapprochant finalement de la méthode utilisée par le schéma de Verhaar. Cependant, la différence résidait dans une connaissance plus profonde des mécanismes d'action détaillés, et ce pour une plus grande diversité de structures chimiques. Des analyses structure-activité couplées à un jugement d'expert ont finalement permis de développer un arbre de décision, composé de 62 règles de décision, classifiant les molécules en MechoAs détaillés au nombre de 23, avec 92% de classification correcte, pour un temps de calcul très réduit, puisque l'optimisation de géométrie des molécules n'est

nécessaire que pour quelques questions, pour lesquelles un niveau de calcul basique est suffisant. Cette optimisation de géométrie sert juste à mesurer la distance entre certains groupements clés de la molécule.

Ainsi, cette dernière méthode est satisfaisante et répond bien au cahier des charges. Sa simplicité permettra également de l'intégrer au moins partiellement dans un logiciel entièrement automatisé. « Intégrer partiellement » signifie sans les parties qui requièrent un calcul de la géométrie, car celles-ci sont plus complexes à intégrer dans un nouveau logiciel. Elles pourraient être remplacées par des messages d'alertes lorsqu'elles seraient nécessaires, indiquant à l'utilisateur que pour la molécule présente, une étude de la géométrie de la molécule serait nécessaire afin d'être sûr du MechoA.

Finalement, nous avons comparé nos résultats avec la méthode de Russom (Russom *et al.*, 1997), qui a été intégrée dans le module « MOA classification by OASIS » du logiciel OECD QSAR Toolbox (OECD and ECHA, 2017). Sur le set de validation, composé de 491 molécules, la méthode de Russom a obtenu 83% de classifications correctes, 6% de classifications légèrement différentes de celles de la littérature, 11% de fausses classifications et 0% non classées, puisque la méthode attribue par défaut un mode d'action de narcotique non-polaire aux substances pour lesquelles une classification réelle n'a pu être trouvée.

Par conséquent, la méthode de Russom est bien plus performante que celle de Verhaar, tel qu'attendu, mais moins performante que notre seconde méthode. En revanche, elle semble plus performante que notre première méthode par calculs quantiques. En conclusion, on peut retenir que la première méthode, utilisant des paramètres quantiques, n'était pas assez performante compte-tenu des moyens de calcul engagés. En revanche, la deuxième méthode développée dans ce travail, qui utilise des alertes structurales, a obtenu de très bons résultats, qui surpassent ceux du schéma de Verhaar et de la méthode de prédiction des MoAs de Russom. Cette méthode est rapide et simple à utiliser et classe les substances en fonction de leur mécanisme d'action dans les mammifères et dans les organismes aquatiques.

Enfin, la partie expérimentale effectuée au laboratoire a permis de lever l'incertitude pesant sur certaines substances pour lesquelles la relation entre les propriétés physico-chimiques et la toxicité aquatique ne semblait pas cohérente. Plus d'expériences seraient évidemment nécessaires pour éclaircir davantage la situation pour certaines substances, mais cela restait un objectif mineur de la présente thèse et ce travail pourra être poursuivi dans le cadre d'autres projets.

Une comparaison des résultats avec d'autres méthodes que celle de Verhaar et de Russom est aussi envisagée. De tels travaux de comparaison ont déjà été effectués par d'autres groupes et ont été publiés récemment (Kienzler *et al.*, 2017). Les résultats de ces travaux de comparaison ont montré un taux de non-classification assez important et variable selon les méthodes, et une concordance des prédictions de MoA pour 42% des molécules classifiées par les 3 méthodes testées. Cela montre l'importance d'une bibliographie précise et qui recoupe plusieurs références, si possible, pour l'attribution des mécanismes ou modes d'action, et confirme le besoin d'une méthode de classification qui fonctionne pour une grande variété de substances chimiques afin de permettre une évaluation de risque correcte pour un maximum de substances.

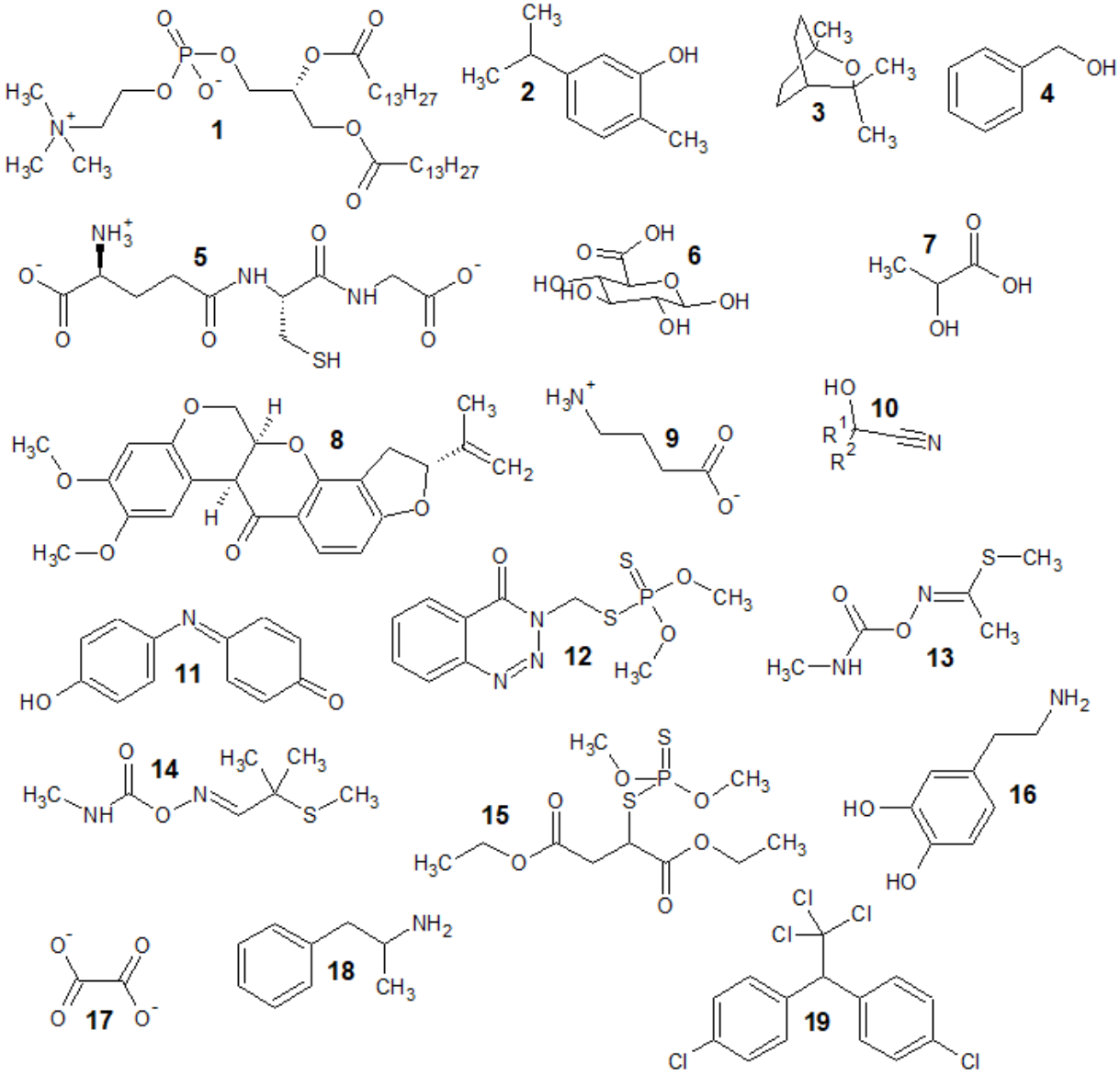
Ce travail pourra être étendu dans le futur aux molécules inorganiques et organométalliques, car un bon nombre d'entre elles possèdent également des mécanismes d'action toxique. Ce type de substances a été écarté du cadre de ce projet de thèse car leurs mécanismes sont *a priori* très différents de ceux que peuvent exercer les substances organiques, et cela aurait complexifié le travail pour une première approche. De même, un travail poussé sur les perturbateurs endocriniens devra être effectué afin d'en caractériser les différentes structures actives et les mécanismes précis associés. Ce domaine

est encore un domaine chargé d'incertitude et en pleine croissance, et des réponses continueront d'être apportées dans la littérature durant les prochaines années. D'autres mécanismes doivent aussi être caractérisés plus en détails, notamment les mécanismes variés déjà identifiés présents dans les sous-catégories 5.3 et 6.7 (autres mécanismes respectivement indirects et directs). Ainsi, le nombre de sous-catégories devra être augmenté pour rendre compte de ces mécanismes.

Pour conclure, cette thèse a permis de montrer qu'il était possible de classer la grande majorité des substances organiques en fonction de leur mécanisme d'action toxique, en se limitant à un nombre limité de MechoAs généraux, pour multiplier plutôt des sous-catégories. Nous avons pu créer une classification qui regroupe les mécanismes qui interviennent chez les mammifères, chez les poissons, et aussi d'autres organismes dont des invertébrés et des plantes. Nous avons développé une méthode de prédiction de ces mécanismes qui obtient de meilleurs résultats que les autres méthodes existantes testées, aussi bien en termes de domaine d'applicabilité qu'en termes de taux d'erreurs. Par conséquent, les objectifs de cette thèse ont été atteints, et ses résultats pourront permettre une meilleure évaluation de risque toxicologique et environnemental.

Planche de structures moléculaires

La figure ci-dessous renseigne les structures moléculaires des substances nommées dans le texte, et qui n'ont pas été explicitées dans des figures précédentes. Un numéro est associé à chacune de ces structures, dans cette figure et dans le texte.



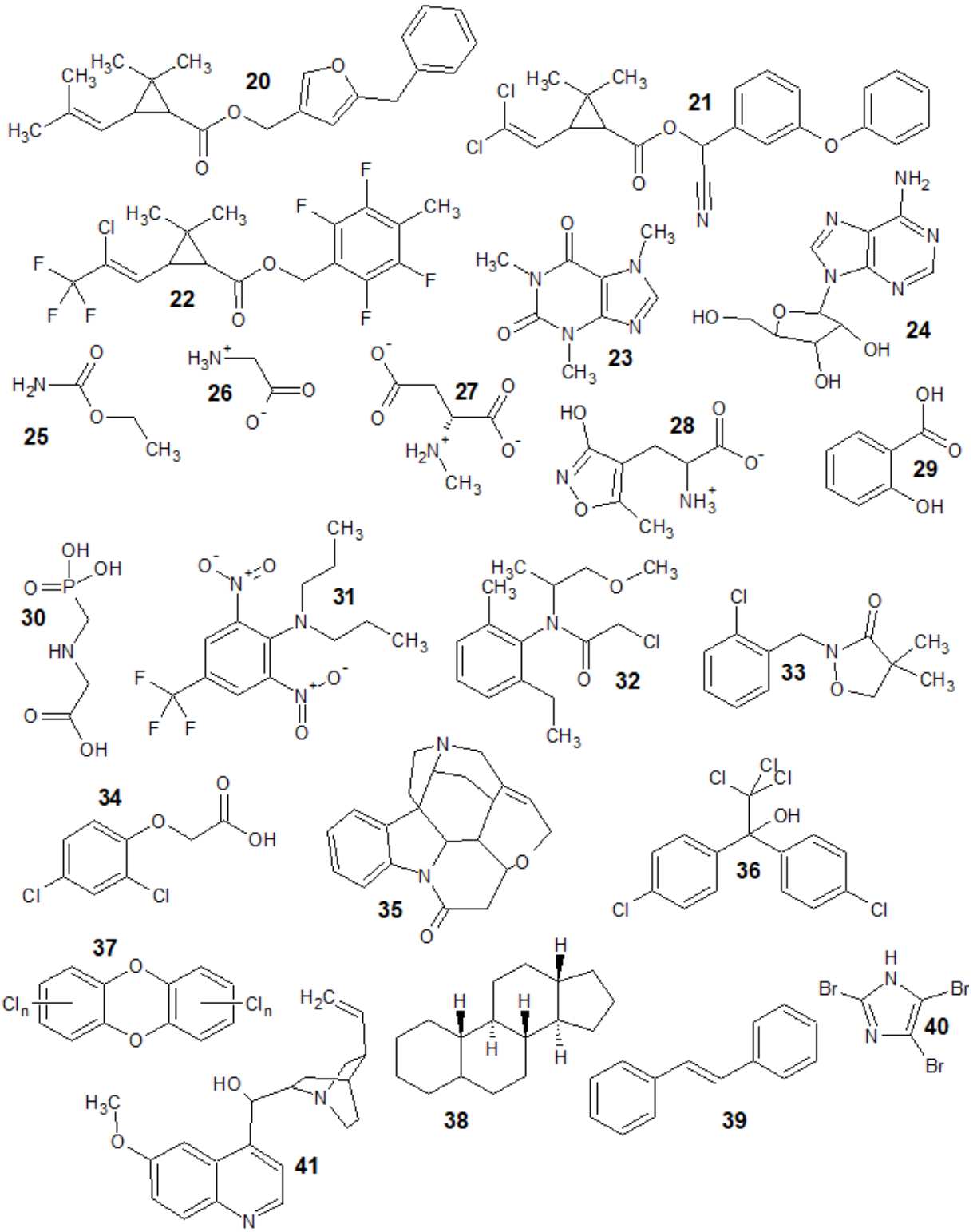


Figure 42 : planche de structures moléculaires

Références bibliographiques

Adams, W.J., Biddinger, G.R., Robillard, K.A., Gorsuch, J.W., 1995. A summary of the acute toxicity of 14 phthalate esters to representative aquatic organisms. Environ. Toxicol. Chem. 14, 1569–1574. doi:10.1002/etc.5620140916

Adeva-Andany, M., López-Ojén, M., Funcasta-Calderón, R., Ameneiros-Rodríguez, E., Donapetry-García, C., Vila-Altesor, M., Rodríguez-Seijas, J., 2014. Comprehensive review on lactate metabolism in human health. Mitochondrion 17, 76–100. doi:10.1016/j.mito.2014.05.007

Airaksinen, M.M., Rosenberg, P.H., Tammisto, T., 1970. A Possible Mechanism of Toxicity of Trifluoroethanol and Other Halothane Metabolites. Acta Pharmacol. Toxicol. (Copenh.) 28, 299–304. doi:10.1111/j.1600-0773.1970.tb00556.x

Altenburger, R., Backhaus, T., Boedeker, W., Faust, M., Scholze, M., Grimme, L.H., 2000. Predictability of the toxicity of multiple chemical mixtures to Vibrio fischeri: Mixtures composed of similarly acting chemicals. Environ. Toxicol. Chem. 19, 2341–2347. doi:10.1002/etc.5620190926

Aptula, A.O., Roberts, D.W., 2006. Mechanistic Applicability Domains for Nonanimal-Based Prediction of Toxicological End Points: General Principles and Application to Reactive Toxicity. Chem. Res. Toxicol. 19, 1097–1105. doi:10.1021/tx0601004

APVMA, 2010. Atrazine Toxicity: Analysis of Potential Modes of Action. APVMA.

ATSDR, 2014. Medical Management Guidelines (MMGs): Methylene Chloride.

ATSDR, 2006. Toxicological Profile: 1,1,1-Trichloroethane.

ATSDR, 1999. Toxicological Profile: n-Hexane.

ATSDR, 1997a. Toxicological Profile: Chloroform [WWW Document]. URL https://www.atsdr.cdc.gov/toxprofiles/tp.asp?id=53&tid=16 (accessed 10.26.16).

ATSDR, 1997b. Toxicological Profile: Hydrazines [WWW Document]. URL http://www.atsdr.cdc.gov/ToxProfiles/tp.asp?id=502&tid=89 (accessed 7.24.15).

Backhaus, T., Altenburger, R., Boedeker, W., Faust, M., Scholze, M., Grimme, L.H., 2000. Predictability of the toxicity of a multiple mixture of dissimilarly acting chemicals to Vibrio fischeri. Environ. Toxicol. Chem. 19, 2348–2356. doi:10.1002/etc.5620190927

Barron, A.B., Søvik, E., Cornish, J.L., 2010. The Roles of Dopamine and Related Compounds in Reward-Seeking Behavior Across Animal Phyla. Front. Behav. Neurosci. 4. doi:10.3389/fnbeh.2010.00163

Barron, M.G., Lilavois, C.R., Martin, T.M., 2015. MOAtox: A comprehensive mode of action and acute aquatic toxicity database for predictive model development. Aquat. Toxicol. 161, 102–107. doi:10.1016/j.aquatox.2015.02.001

Bearden, A.P., Schultz, T.W., 1997. Structure-activity relationships for Pimephales and Tetrahymena: A mechanism of action approach. Environ. Toxicol. Chem. 16, 1311–1317. doi:10.1002/etc.5620160629

Benigni, R., Bossa, C., Netzeva, T., Rodomonte, A., Tsakovska, I., 2007. Mechanistic QSAR of aromatic amines: New models for discriminating between homocyclic mutagens and nonmutagens, and validation of models for carcinogens. Environ. Mol. Mutagen. 48, 754–771. doi:10.1002/em.20355

Benigni, R., Bossa, C., Tcheremenskaia, O., 2013. Nongenotoxic Carcinogenicity of Chemicals: Mechanisms of Action and Early Recognition through a New Set of Structural Alerts. Chem. Rev. 113, 2940–2957. doi:10.1021/cr300206t

Berg, J.M., Tymoczko, J.L., Stryer, L., 2002. A Proton Gradient Powers the Synthesis of ATP, in: Biochemistry. W.H. Freeman, New-York.

Blanco-Ayala, T., Andérica-Romero, A.C., Pedraza-Chaverri, J., 2014. New insights into antioxidant strategies against paraquat toxicity. Free Radic. Res. 48, 623–640. doi:10.3109/10715762.2014.899694

Bolton, J.L., Trush, M.A., Penning, T.M., Dryhurst, G., Monks, T.J., 2000. Role of Quinones in Toxicology. Chem. Res. Toxicol. 13, 135–160. doi:10.1021/tx9902082

Bonnard, N., Brondeau, M.-T., Jargot, D., Miraval, S., Pillière, F., Robert, S., Schneider, O., Serre, P., 2013a. Méthacrylate de méthyle - Fiche toxicologique n°62.

Bonnard, N., Falcy, M., Miraval, S., Robert, S., Schneider, O., 2013b. Acrylate de méthyle - Fiche toxicologique n° 181.

Brown, M.A., Vito, S.C.D., 1993. Predicting azo dye toxicity. Crit. Rev. Environ. Sci. Technol. 23, 249–324. doi:10.1080/10643389309388453

Brüning, T., Bolt, H.M., 2000. Renal toxicity and carcinogenicity of trichloroethylene: key results, mechanisms, and controversies. Crit. Rev. Toxicol. 30, 253–285. doi:10.1080/10408440091159202

Buckpitt, A., Boland, B., Isbell, M., Morin, D., Shultz, M., Baldwin, R., Chan, K., Karlsson, A., Lin, C., Taff, A., West, J., Fanucchi, M., Van Winkle, L., Plopper, C., 2002. Naphthalene-induced respiratory tract toxicity: metabolic mechanisms of toxicity. Drug Metab. Rev. 34, 791–820. doi:10.1081/DMR-120015694

Calipari, E.S., Ferris, M.J., 2013. Amphetamine Mechanisms and Actions at the Dopamine Terminal Revisited. J. Neurosci. Off. J. Soc. Neurosci. 33, 8923–8925. doi:10.1523/JNEUROSCI.1033-13.2013

Casey, J.R., Grinstein, S., Orlowski, J., 2010. Sensors and regulators of intracellular pH. Nat. Rev. Mol. Cell Biol. 11, 50–61. doi:10.1038/nrm2820

Chahl, L.A., 1996. Experimental and Clinical Pharmacology: Opioids - mechanisms of action. Aust. Prescr. 19, 63–65. doi:10.18773/austprescr.1996.063

ChemAxon, 2016. MarvinSketch. ChemAxon (http://www.chemaxon.com).

Coats, J.R., 1990. Mechanisms of toxic action and structure-activity relationships for organochlorine and synthetic pyrethroid insecticides. Environ. Health Perspect. 87, 255–262.

Committee on Acute Exposure Guideline Levels, Committee on Toxicology, Board on Environmental Studies and Toxicology, Division on Earth and Life Studies, National Research Council, 2012. Piperidine, in: Acute Exposure Guideline Levels for Selected Airbone Chemicals. National Academies Press (US).

de Sousa, J.P., Torres, R. de A., de Azerêdo, G.A., Figueiredo, R.C.B.Q., Vasconcelos, M.A. da S., de Souza, E.L., 2012. Carvacrol and 1,8-cineole alone or in combination at sublethal concentrations induce changes in the cell morphology and membrane permeability of Pseudomonas fluorescens in a vegetable-based broth. Int. J. Food Microbiol. 158, 9–13. doi:10.1016/j.ijfoodmicro.2012.06.008

Dearden, J.C., 2002. Prediction of environmental toxicity and fate using quantitative structure-activity relationships (QSARs). J. Braz. Chem. Soc. 13, 754–762. doi:10.1590/S0103-50532002000600005

Di Francesco, A.M., Ward, T.H., Butler, J., 2004. Diaziridinylbenzoquinones, in: Quinones and Quinone Enzymes. Academic Press, pp. 181–182.

Droge, S.T.J., Hermens, J.L.M., Rabone, J., Gutsell, S., Hodges, G., 2016. Phospholipophilicity of CxHyN+ amines: chromatographic descriptors and molecular simulations for understanding partitioning into membranes. Environ. Sci. Process. Impacts 18, 1011–1023. doi:10.1039/C6EM00118A

Droual, B., 2011. Chapter 8 - Synaptic Transmission and Neural Integration [WWW Document]. Courses Taught Dr Droual Modesto Jr. Coll. URL http://droualb.faculty.mjc.edu/Course%20Materials/Physiology%20101/Chapter%20Notes/Fall%202011/chapter_8%20Fall%202011.htm (accessed 7.18.17).

DrugBank (Ed.), 2016. Salicylamide. DrugBank.

Dunn, M.F., Ramírez-Trujillo, J.A., Hernández-Lucas, I., 2009. Major roles of isocitrate lyase and malate synthase in bacterial and fungal pathogenesis. Microbiology 155, 3166–3175. doi:10.1099/mic.0.030858-0

Elersek, T., Filipic, M., 2011. Organophosphorus Pesticides - Mechanisms of Their Toxicity, in: Stoytcheva, M. (Ed.), Pesticides - The Impacts of Pesticides Exposure. InTech.

Ellison, C.M., Madden, J.C., Cronin, M.T.D., Enoch, S.J., 2015. Investigation of the Verhaar scheme for predicting acute aquatic toxicity: Improving predictions obtained from Toxtree ver. 2.6. Chemosphere 139, 146–154. doi:10.1016/j.chemosphere.2015.06.009

Endo, S., Brown, T.N., Goss, K.-U., 2013. General Model for Estimating Partition Coefficients to Organisms and Their Tissues Using the Biological Compositions and Polyparameter Linear Free Energy Relationships. Environ. Sci. Technol. 47, 6630–6639. doi:10.1021/es401772m

Enoch, S.J., Hewitt, M., Cronin, M.T.D., Azam, S., Madden, J.C., 2008. Classification of chemicals according to mechanism of aquatic toxicity: An evaluation of the implementation of the Verhaar scheme in Toxtree. Chemosphere 73, 243–248. doi:10.1016/j.chemosphere.2008.06.052

Escher, B.I., Hermens, J.L.M., 2002. Modes of Action in Ecotoxicology: Their Role in Body Burdens, Species Sensitivity, QSARs, and Mixture Effects. Environ. Sci. Technol. 36, 4201–4217. doi:10.1021/es015848h

Escher, B.I., Hunziker, R., Schwarzenbach, R.P., Westall, J.C., 1999. Kinetic Model To Describe the Intrinsic Uncoupling Activity of Substituted Phenols in Energy Transducing Membranes. Environ. Sci. Technol. 33, 560–570. doi:10.1021/es980545h

European Commission - DG Joint Reseach Centre, U.S. Environmental Protection Agency, n.d. AOPs [WWW Document]. AOPWiki. URL https://aopwiki.org/aops (accessed 4.28.17).

Fang, H., Tong, W., Branham, W.S., Moland, C.L., Dial, S.L., Hong, H., Xie, Q., Perkins, R., Owens, W., Sheehan, D.M., 2003. Study of 202 Natural, Synthetic, and Environmental Chemicals for Binding to the Androgen Receptor. Chem. Res. Toxicol. 16, 1338–1358. doi:10.1021/tx030011g

Fang, H., Tong, W., Shi, L.M., Blair, R., Perkins, R., Branham, W., Hass, B.S., Xie, Q., Dial, S.L., Moland, C.L., Sheehan, D.M., 2001. Structure−Activity Relationships for a Large Diverse Set of Natural, Synthetic, and Environmental Estrogens. Chem. Res. Toxicol. 14, 280–294. doi:10.1021/tx000208y

Ferguson, J., 1939. The Use of Chemical Potentials as Indices of Toxicity. Proc. R. Soc. Lond. B Biol. Sci. 127, 387–404.

Fernandes, T.C., Pizano, M.A., Marin-Morales, M.A., 2013. Characterization, Modes of Action and Effects of Trifluralin: A Review, in: Price, A.J., Kelton, J.A. (Eds.), Herbicides - Current Research and Case Studies in Use. InTech.

Food and Drug Administration, 2004. Docket No. 2004N-0205 Furan in Food, Thermal Treatment; Request for Data and Information.

Franks, N.P., Lieb, W.R., 1982. Molecular mechanisms of general anaesthesia. Nature 300, 487–493. doi:10.1038/300487a0

Fuerst, E.P., 1987. Understanding the Mode of Action of the Chloroacetamide and Thiocarbamate Herbicides. Weed Technol. 1, 270–277.

Fukuto, T.R., 1990. Mechanism of action of organophosphorus and carbamate insecticides. Environ. Health Perspect. 87, 245–254.

Gobas, F.A.P.C., Mackay, D., Shiu, W.Y., Lahittete, J.M., Garofalo, G., 1988. A novel method for measuring membrane-water partition coefficients of hydrophobic organic chemicals: Comparison with 1-octanol-water partitioning. J. Pharm. Sci. 77, 265–272. doi:10.1002/jps.2600770317

Gong, X.Q., Nedialkov, Y.A., Burton, Z.F., 2004. α-Amanitin Blocks Translocation by Human RNA Polymerase II. J. Biol. Chem. 279, 27422–27427. doi:10.1074/jbc.M402163200

Grogan, J., DeVito, S.C., Pearlman, R.S., Korzekwa, K.R., 1992. Modeling cyanide release from nitriles: prediction of cytochrome P 450-mediated acute nitrile toxicity. Chem. Res. Toxicol. 5, 548–552. doi:10.1021/tx00028a014

Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E., Hutchison, G.R., 2013. Avogadro.

Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E., Hutchison, G.R., 2012. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J. Cheminformatics 4, 17. doi:10.1186/1758-2946-4-17

Hara, K., Harris, R.A., 2002. The Anesthetic Mechanism of Urethane: The Effects on Neurotransmitter-Gated Ion Channels: Anesth. Analg. 94, 313–318. doi:10.1213/00000539-200202000-00015

IAC Publishing Labs, n.d. How does caffeine affect plant growth? [WWW Document]. Reference.com. URL https://www.reference.com/science/caffeine-affect-plant-growth-7e511b9d85fbdfdb (accessed 1.20.17).

IARC, 2010. General Discussion of common mechanisms for aromatic amines. IARC Monogr. Eval. Carcinog. Risks Hum. 99, 41–54.

Ioannides, C., 1996. The CYP2E Subfamily - Evolution, in: Cytochromes P450: Metabolic and Toxicological Aspects. CRC Press, p. 213.

IPCS Inchem, 2015. Vanillin SIDS.

IPCS Inchem, 2002. Monograph on Atropine - Antidotes for poisoning by organophosphorus pesticides [WWW Document]. URL http://www.inchem.org/documents/antidote/antidote/atropine.htm#7.1 (accessed 4.28.17).

IPCS Inchem, 1998. Diphenylamine (addendum), in: JMPR Evaluations. Presented at the Toxicological and environmental evaluations, Rome.

IPCS Inchem, 1980. Ethyl lactate, in: WHO Food Additives Series 15. Rome.

Jacobs, A., 1997. Hard and soft nucleophiles, in: Understanding Organic Reaction Mechanisms. Cambridge University Press, pp. 45–46.

Jaworska, J.S., Hunter, R.S., Schultz, T.W., 1995. Quantitative structure-toxicity relationships and volume fraction analyses for selected esters. Arch. Environ. Contam. Toxicol. 29, 86–93. doi:10.1007/BF00213091

Jervais, G., Luukinen, B., Buhl, K., Stone, D., 2008. 2,4-D Technical Fact Sheet [WWW Document]. Natl. Pestic. Infomation Cent. URL http://npic.orst.edu/factsheets/archive/2,4-DTech.html (accessed 12.16.16).

Ji, K., Liu, X., Lee, S., Kang, S., Kho, Y., Giesy, J.P., Choi, K., 2013. Effects of non-steroidal anti-inflammatory drugs on hormones and genes of the hypothalamic-pituitary-gonad axis, and reproduction of zebrafish. J. Hazard. Mater. 254, 242–251. doi:10.1016/j.jhazmat.2013.03.036

Karabunarliev, S., Mekenyan, O.G., Karcher, W., Russom, C.L., Bradbury, S.P., 1996. Quantum-chemical Descriptors for Estimating the Acute Toxicity of Substituted Benzenes to the Guppy (Poecilia reticulata) and Fathead Minnow (Pimephales promelas). Quant. Struct.-Act. Relatsh. 15, 311–320. doi:10.1002/qsar.19960150406

Kazius, J., McGuire, R., Bursi, R., 2005. Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem. 48, 312–320. doi:10.1021/jm040835a

Kienzler, A., Barron, M.G., Belanger, S.E., Beasley, A., Embry, M.R., 2017. Mode of Action (MOA) Assignment Classifications for Ecotoxicology: An Evaluation of Approaches. Environ. Sci. Technol. doi:10.1021/acs.est.7b02337

King, M.W., 2015a. Ethanol (Alcohol) Metabolism: Acute and Chronic Toxicities [WWW Document]. Med. Biochem. Page. URL http://themedicalbiochemistrypage.org/ethanol-metabolism.php (accessed 9.15.15).

King, M.W., 2015b. Biochemistry of Neurotransmitters and Nerve Transmission [WWW Document]. URL http://themedicalbiochemistrypage.org/nerves.php (accessed 9.25.15).

Koleva, Y., Barzilov, I., 2010. Comparative study of mechanism of action of allyl alcohols for different endpoints 49, 55–59.

Koleva, Y.K., Cronin, M.T.D., Madden, J.C., Schwöbel, J.A.H., 2011. Modelling acute oral mammalian toxicity. 1. Definition of a quantifiable baseline effect. Toxicol. In Vitro 25, 1281–1293. doi:10.1016/j.tiv.2011.04.015

KREATiS, n.d. Downloads and useful links [WWW Document]. KREATiS.eu. URL http://kreatis.eu/en/qsars-downloads.php (accessed 10.12.17).

Kulma, A., Szopa, J., 2007. Catecholamines are active compounds in plants. Plant Sci. 172, 433–440. doi:10.1016/j.plantsci.2006.10.013
Li, X., 2009. Glutathione and Glutathione-S-Transferase in Detoxification Mechanisms, in: General, Applied and Systems Toxicology. John Wiley & Sons, Ltd. doi:10.1002/9780470744307.gat166
Liesivuori, J., Savolainen, A.H., 1991. Methanol and Formic Acid Toxicity: Biochemical Mechanisms. Pharmacol. Toxicol. 69, 157–163. doi:10.1111/j.1600-0773.1991.tb01290.x
Liska, D.J., 1998. The detoxification enzyme systems. Altern. Med. Rev. J. Clin. Ther. 3, 187–198.
LoPachin, R.M., Gavin, T., 2014. Molecular Mechanisms of Aldehyde Toxicity: A Chemical Perspective. Chem. Res. Toxicol. 27, 1081–1091. doi:10.1021/tx5001046
Lugli, A.K., Yost, C.S., Kindler, C.H., 2009. Anaesthetic mechanisms: update on the challenge of unravelling the mystery of anaesthesia. Eur. J. Anaesthesiol. 26, 807–820. doi:10.1097/EJA.0b013e32832d6b0f
Mache, R., 2015. Photosynthèse et biomasse. Encycl. Energ.
Mackay, D., Arnot, J.A., Petkova, E.P., Wallace, K.B., Call, D.J., Brooke, L.T., Veith, G.D., 2009. The physicochemical basis of QSARs for baseline toxicity. SAR QSAR Environ. Res. 20, 393–414. doi:10.1080/10629360902949153
Madden, J.C., 2010. Chapter 2:Introduction to QSAR and Other In Silico Methods to Predict Toxicity, in: In Silico Toxicology. pp. 11–30. doi:10.1039/9781849732093-00011
Makarovsky, I., Markel, G., Hoffman, A., Schein, O., Brosh-Nissimov, T., Tashma, Z., Dushnitsky, T., Eisenkraft, A., 2008. Strychnine - A Killer from the Past. IMAJ Isr. Med. Assoc. J. 10, 142–145.
Martin, T.M., Grulke, C.M., Young, D.M., Russom, C.L., Wang, N.Y., Jackson, C.R., Barron, M.G., 2013. Prediction of Aquatic Toxicity Mode of Action Using Linear Discriminant and Random Forest Models. J. Chem. Inf. Model. 53, 2229–2239. doi:10.1021/ci400267h
Matji, T.J.A., Lopez, B.P.J., 2012. Micronized composition of 2,4,6-triiodophenol. EP2174654 B1.
McHale, C.M., Zhang, L., Smith, M.T., 2012. Current understanding of the mechanism of benzene-induced leukemia in humans: implications for risk assessment. Carcinogenesis 33, 240–252. doi:10.1093/carcin/bgr297
Mensah, P.K., Palmer, C.G., Odume, O.N., 2015. Ecotoxicology of Glyphosate and Glyphosate-Based Herbicides — Toxicity to Wildlife and Humans, in: Larramendy, M.L., Soloneski, S. (Eds.), Toxicity and Hazard of Agrochemicals. InTech.
Moridani, M.Y., Cheon, S.S., Khan, S., O'Brien, P.J., 2003. Metabolic activation of 3-hydroxyanisole by isolated rat hepatocytes. Chem. Biol. Interact. 142, 317–333. doi:10.1016/S0009-2797(02)00125-4
Morotomi, M., Sakai, K., Yazawa, K., Suegara, N., Kawai, Y., Mutai, M., 1981. Effect and fate of orally administered lactic acid in rats. J. Nutr. Sci. Vitaminol. (Tokyo) 27, 117–128.
Munday, R., 1989. Toxicity of thiols and disulphides: Involvement of free-radical species. Free Radic. Biol. Med. 7, 659–673. doi:10.1016/0891-5849(89)90147-0
Nair, B., 2001. Final report on the safety assessment of Benzyl Alcohol, Benzoic Acid, and Sodium Benzoate. Int. J. Toxicol. 20 Suppl 3, 23–50.
National Center for Biotechnology Information, 2017. CONIINE | C8H17N [WWW Document]. PubChem Compd. Database. URL https://pubchem.ncbi.nlm.nih.gov/compound/441072 (accessed 12.21.16).
NC3Rs, n.d. The 3Rs [WWW Document]. URL https://www.nc3rs.org.uk/the-3rs (accessed 6.13.16).
Nehlig, A., Daval, J.-L., Debry, G., 1992. Caffeine and the central nervous system: mechanisms of action, biochemical, metabolic and psychostimulant effects. Brain Res. Rev. 17, 139–170. doi:10.1016/0165-0173(92)90012-B
O'Donoghue, J.L., 2001. Ketones of Four or Five Carbons, in: Patty's Toxicology. John Wiley & Sons, Inc.
OECD, 2015. Test No. 439: In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method. Organisation for Economic Co-operation and Development, Paris.
OECD, ECHA, 2017. QSAR Toolbox.
Olsen, R.W., DeLorey, T.M., 1999. GABA Receptor Physiology and Pharmacology.
OpenStax, 2014. Carbohydrate Metabolism, in: Anatomy & Physiology.

Parlement Européen, Conseil de l'Union Européenne, 2006. Règlement (CE) N° 1907/2006 du parlement européen et du conseil.

Patlewicz, G., Jeliazkova, N., Safford, R., Worth, A., Aleksiev, B., 2014. Toxtree. Ideaconsult Ltd.

Patlewicz, G., Jeliazkova, N., Safford, R.J., Worth, A.P., Aleksiev, B., 2008. An evaluation of the implementation of the Cramer classification scheme in the Toxtree software. SAR QSAR Environ. Res. 19, 495–524. doi:10.1080/10629360802083871

Peoples, R.W., Weight, F.F., 1994. Trichloroethanol potentiation of gamma-aminobutyric acid-activated chloride current in mouse hippocampal neurones. Br. J. Pharmacol. 113, 555–563.

Petering, D.H., Fowler, B.A., 1986. Roles of metallothionein and related proteins in metal metabolism and toxicity: problems and perspectives. Environ. Health Perspect. 65, 217–224.

Pinto, P.I.S., Estêvão, M.D., Power, D.M., 2014. Effects of Estrogens and Estrogenic Disrupting Compounds on Fish Mineralized Tissues. Mar. Drugs 12, 4474–4494. doi:10.3390/md12084474

Portoghese, P.S., Kedziora, G.S., Larson, D.L., Bernard, B.K., Hall, R.L., 1989. Reactivity of glutathione with α,β-unsaturated ketone flavouring substances. Food Chem. Toxicol. 27, 773–776. doi:10.1016/0278-6915(89)90106-3

Rafii, F., Cerniglia, C.E., 1995. Reduction of azo dyes and nitroaromatic compounds by bacterial enzymes from the human intestinal tract. Environ. Health Perspect. 103, 17–19.

Recknagel, R.O., Glende Jr., E.A., Dolak, J.A., Waller, R.L., 1989. Mechanisms of carbon tetrachloride toxicity. Pharmacol. Ther. 43, 139–154. doi:10.1016/0163-7258(89)90050-8

Ren, S., 2003. Phenol mechanism of toxic action classification and prediction: a decision tree approach. Toxicol. Lett. 144, 313–323. doi:10.1016/S0378-4274(03)00236-4

Rivard, L., 2003. Environmental fate of metolachlor. Chemistry 51218, 45–2.

Roberts, D.W., Roberts, J.F., Hodges, G., Gutsell, S., Ward, R.S., Llewellyn, C., 2013. Aquatic toxicity of cationic surfactants to Daphnia magna. SAR QSAR Environ. Res. 24, 417–427. doi:10.1080/1062936X.2013.781538

Russom, C.L., Bradbury, S.P., Broderius, S.J., Hammermeister, D.E., Drummond, R.A., 1997. Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (Pimephales promelas). Environ. Toxicol. Chem. 16, 948–967. doi:10.1002/etc.5620160514

Schultz, T.W., 1987. The use of the ionization constant (pKa) in selecting models of toxicity in phenols. Ecotoxicol. Environ. Saf. 14, 178–183.

Schwöbel, J.A.H., Madden, J.C., Cronin, M.T.D., 2011. Application of a computational model for Michael addition reactivity in the prediction of toxicity to Tetrahymena pyriformis. Chemosphere 85, 1066–1074. doi:10.1016/j.chemosphere.2011.07.037

Shaw, W., 2002. Chapter 3 Organic Acid Testing, Byproducts of Yeast and their Relationship to Autism, in: Biological Treatments for Autism and PDD. Causes and Biomedical Therapies for Autism & PDD.

Sherer, T.B., Richardson, J.R., Testa, C.M., Seo, B.B., Panov, A.V., Yagi, T., Matsuno-Yagi, A., Miller, G.W., Greenamyre, J.T., 2007. Mechanism of toxicity of pesticides acting at complex I: relevance to environmental etiologies of Parkinson's disease. J. Neurochem. 100, 1469–1479. doi:10.1111/j.1471-4159.2006.04333.x

Shi, L., Tong, W., Fang, H., Xie, Q., Hong, H., Perkins, R., Wu, J., Tu§, M., Blair, R.M., Branham, W.S., Waller, C., Walker, J., Sheehan, D.M., 2002. An integrated "4-phase" approach for setting endocrine disruption screening priorities–phase I and II predictions of estrogen receptor binding affinity. SAR QSAR Environ. Res. 13, 69–88. doi:10.1080/10629360290002235

Shima, M., Kotera, A., 1963. Copolymerization characteristics of methyl acrylate and methyl methacrylate. I. Study of copolymerization kinetics by a deuterium tracer method. J. Polym. Sci. A 1, 1115–1121. doi:10.1002/pol.1963.100010404

Sikkema, J., de Bont, J.A., Poolman, B., 1995. Mechanisms of membrane toxicity of hydrocarbons. Microbiol. Rev. 59, 201–222.

Smith, G.F., 2011. Designing Drugs to Avoid Toxicity, in: Witty, G.L. and D.R. (Ed.), Progress in Medicinal Chemistry. Elsevier, pp. 1–47.

Song, Y., 2014. Insight into the mode of action of 2,4-dichlorophenoxyacetic acid (2,4-D) as an herbicide. J. Integr. Plant Biol. 56, 106–113. doi:10.1111/jipb.12131

Sorg, O., 2014. AhR signalling and dioxin toxicity. Toxicol. Lett., Environmental contaminants and target organ toxicities 230, 225–233. doi:10.1016/j.toxlet.2013.10.039

Sparling, D.W., Linder, G., Bishop, C.A., Krest, S., 2010. 10.3.2 Organochloride pesticides, in: Ecotoxicology of Amphibians and Reptiles, Second Edition. CRC Press, pp. 303–308.

Spycher, S., Netzeva, T.I., Worth, A.P., Escher, B.I., 2008. Mode of action-based classification and prediction of activity of uncouplers for the screening of chemical inventories. SAR QSAR Environ. Res. 19, 433–463. doi:10.1080/10629360802348803

Thomas, P., Dawick, J., Lampi, M., Lemaire, P., Presow, S., van Egmond, R., Arnot, J.A., Mackay, D., Mayer, P., Galay Burgos, M., 2015. Application of the Activity Framework for Assessing Aquatic Ecotoxicology Data for Organic Chemicals. Environ. Sci. Technol. 49, 12289–12296. doi:10.1021/acs.est.5b02873

Tomson, T., Tybring, G., Bertilsson, L., 1983. Single-dose kinetics and metabolism of carbamazepine-10,11-epoxide. Clin. Pharmacol. Ther. 33, 58–65. doi:10.1038/clpt.1983.8

Trushin, N., Alam, S., El-Bayoumy, K., Krzeminski, J., Amin, S.G., Gullett, J., Meyers, C., Prokopczyk, B., 2012. Comparative metabolism of benzo[a]pyrene by human keratinocytes infected with high-risk human papillomavirus types 16 and 18 as episomal or integrated genomes. J. Carcinog. 11, 1. doi:10.4103/1477-3163.92309

Tsuruda, L.S., Lamé, M.W., Jones, A.D., 1995. Formation of epoxide and quinone protein adducts in B6C3F1 mice treated with naphthalene, sulfate conjugate of 1,4-dihydroxynaphthalene and 1,4-naphthoquinone. Arch. Toxicol. 69, 362–367. doi:10.1007/s002040050185

Vaes, W.H., Ramos, E.U., Verhaar, H.J., Hermens, J.L., 1998. Acute toxicity of nonpolar versus polar narcosis: Is there a difference? Environ. Toxicol. Chem. 17, 1380–1384.

Vallabh Minikel, E., 2016. A mechanism of action hypothesis [WWW Document]. CureFFI.org. URL http://www.cureffi.org/2016/05/08/a-mechanism-of-action-hypothesis/ (accessed 9.19.16).

van Rensen, J.J.S., 1982. Molecular mechanisms of herbicide action near photosystem II. Physiol. Plant. 54, 515–521. doi:10.1111/j.1399-3054.1982.tb00719.x

Van Scoy, A.R., Tjeerdema, R.S., 2014. Environmental fate and toxicology of clomazone. Rev. Environ. Contam. Toxicol. 229, 35–49. doi:10.1007/978-3-319-03777-6_3

Vane, J.R., Botting, R.M., 1998. Mechanism of Action of Nonsteroidal Anti-inflammatory Drugs. Am. J. Med. 104, 2S–8S. doi:10.1016/S0002-9343(97)00203-9

Veith, G.D., Broderius, S.J., 1987. Structure-Toxicity Relationships for Industrial Chemicals Causing Type (II) Narcosis Syndrome, in: Kaiser, K.L.E. (Ed.), QSAR in Environmental Toxicology - II. Springer Netherlands, pp. 385–391. doi:10.1007/978-94-009-3937-0_29

Verhaar, H.J.M., Solbé, J., Speksnijder, J., van Leeuwen, C.J., Hermens, J.L.M., 2000. Classifying environmental pollutants: Part 3. External validation of the classification system. Chemosphere 40, 875–883. doi:10.1016/S0045-6535(99)00317-3

Verhaar, H.J.M., van Leeuwen, C.J., Hermens, J.L.M., 1992. Classifying environmental pollutants. Chemosphere 25, 471–491. doi:10.1016/0045-6535(92)90280-5

Verschoyle, R.D., Brown, A.W., Thompson, C.A., 1984. The toxicity and neuropathology of 2,4,5-tribromoimidazole and its derivatives in rats. Arch. Toxicol. 56, 109–112. doi:10.1007/BF00349081

Vetter, J., 2004. Poison hemlock (Conium maculatum L.). Food Chem. Toxicol. 42, 1373–1382. doi:10.1016/j.fct.2004.04.009

Way, J.L., Leung, P., Cannon, E., Morgan, R., Tamulinas, C., Leong-Way, J., Baxter, L., Nagi, A., Chui, C., 2007. The Mechanism of Cyanide Intoxication and its Antagonism - Ciba Foundation Symposium 140 - Cyanide Compounds in Biology - Way - Wiley Online Library. Presented at the Ciba Foudation Symposium 140 - Cyanide Compounds in Biology, D. Evered and S. Hamett, pp. 232–248.

Wiley-VCH Verlag, 2002a. Organic peroxides [MAK Value Documentation, 1992], in: The MAK-Collection for Occupational Health and Safety. Wiley-VCH Verlag GmbH & Co. KGaA. doi:10.1002/3527600418.mb0opexgrpe0003

Wiley-VCH Verlag, 2002b. Benzaldehyde [MAK Value Documentation, 2002], in: The MAK-Collection for Occupational Health and Safety. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 13–36.

Wiley-VCH Verlag, 2002c. Triethylene glycol [MAK Value Documentation, 2007], in: The MAK-Collection for Occupational Health and Safety. Wiley-VCH Verlag GmbH & Co. KGaA.

Annexes

A.1) Informations supplémentaires pour la publication « A New classification algorithm based on Mechanisms of Action. »

Les informations supplémentaires associées à l'article sont des tableaux de données et des graphiques. Les tableaux étant trop larges pour une feuille A4, ils ont été séparés en deux parties. Une première partie pour les valeurs numériques associées à chaque substance, et une deuxième partie pour les mécanismes d'action.

Table S1: training set. Part 1.

N°	Molecule	SMILES	CAS	Exp pKa(BH+)	pKa ref	pred pKa (Chemaxon+RA)	qMaxH	qMax nonH	qMax-	aE(LUMO)	E(HOMO)
1	thioterpineol	SC(C)(C)C1CC=C(C)CC1	83150-78-1	-6	b	-6	0,146610	0,016927	-0,342642	0,1850	-0,3289
2	geraniol	OCC=C(C)CCC=C(C)C	106-24-1	-2	a	-2,22	0,335335	0,108942	-0,655412	-0,1383	-0,3252
3	calone	O=C1COc2ccc(C)cc2OC1	28940-11-6	-7	a	-4,64	0,176471	0,497722	-0,693378	-0,0476	-0,3129
4	benzyl fluoride	c1ccccc1CF	350-50-5	NA	NA	NA	0,174946	0,248783	-0,413636	0,1389	-0,3293
5	1,2-dibromobenzene	Brc1ccccc1Br	583-53-9	NA	NA	NA	0,183089	-0,032079	-0,143058	0,1149	-0,3362
6	2,2,5,5-tetramethyltetrahydrofuran	O1C(C)(C)CCC1(C)C	15045-43-9	-3	a	-4,2	0,124447	0,300795	-0,696760	0,0171	-0,3895
7	2',4'-dichloroacetophenone	CC(=O)c1c(Cl)cc(Cl)cc1	2234-16-4	-7	a	-7,51	0,206041	0,537430	-0,506222	-0,1455	-0,3520
8	2,5-dimethylhexa-2,4-diene	CC(C)=CC=C(C)C	764-13-6	NA	NA	NA	0,131686	-0,030831	-0,361347	0,1318	-0,28
9	2-methylbut-3-yn-2-ol	CC(O)(C)C#C	115-19-5	-2	a	-2,66	0,338827	0,249593	-0,636150	-0,0489	-0,3866
10	3,3-dimethylbutan-2-one	CC(=O)C(C)(C)C	75-97-8	-7	a	-7,4	0,155512	0,525819	-0,531026	-0,1052	-0,3916
11	3,6-dithiaoctane	CCSCCSCC	5395-75-5	-5,4	b	-5,4	0,156005	0,121369	-0,385973	-0,0390	-0,3292
12	4-methyloxazole	O1C=NC(C)=C1	693-93-6	0,8	c	0,5	0,178428	0,404955	-0,537545	-0,0684	-0,3369
13	4-phenylpyridine	n1ccc(c2ccccc2)cc1	939-23-1	5,5	c	5,08	0,155602	0,123645	-0,541237	-0,1071	-0,3149
14	a,a,a-trifluoro-o-tolunitrile	FC(F)(F)c1ccccc1C#N	447-60-9	-10	b	-10	0,200358	1,219591	-0,434534	0,0679	-0,3712
15	acetone	CC(=O)C	67-64-1	-7,2	b	-7,19	0,159259	0,500707	-0,515673	-0,1234	-0,4063
16	acetophenone	c1ccccc1C(=O)C	98-86-2	-6	a	-7,39	0,194895	0,538524	-0,540207	-0,1341	-0,3413
17	butan-2-ol	CC(O)CC	78-92-2	-2	a	-1,62	0,333267	0,203276	-0,648212	-0,0356	-0,4280
18	butanol	OCCCC	71-36-3	-2	a	-1,94	0,336158	0,105681	-0,642139	-0,0457	-0,4335
19	cyclohexane	C1CCCCC1	110-82-7	NA	NA	NA	0,111415	-0,216157	-0,216157	0,2267	-0,4231
20	E-hex-3-en-1-ol	OCC/C=C/CC	928-97-2	-2	a	-2,02	0,335902	0,117675	-0,651220	-0,0420	-0,3385
21	ethylbenzene	CCc1ccccc1	100-41-4	NA	NA	NA	0,146662	-0,000608	-0,322259	0,1493	-0,3188
22	octan-2-one	CC(=O)CCCCCC	111-13-7	-7	a	-7,27	0,15574	0,499629	-0,528117	-0,1089	-0,3977
23	octanol	OCCCCCCCC	68526-83-0	-2	a	-1,99	0,334735	0,124532	-0,649446	-0,0431	-0,4300
24	o-xylene	Cc1c(C)cccc1	95-47-6	NA	NA	NA	0,145142	-0,015121	-0,345618	0,1487	-0,3123
25	pentan-1-ol	OCCCCC	71-41-0	-2	a	-1,99	0,334803	0,124097	-0,649173	-0,0443	-0,4332
26	pyrrole	N1C=CC=C1	109-97-7	-3,8	d	-3,8	0,312041	0,078249	-0,647029	-0,0802	-0,2927
27	t-butanol	OC(C)(C)C	75-65-0	-2	a	-1,45	0,329206	0,257441	-0,650329	-0,0289	-0,4245
28	tert-butyl methyl ether	CC(C)(C)OC	1634-04-4	-3	a	-4,14	0,124942	0,277418	-0,649306	-0,0018	-0,4037
29	tetrahydrofuran	O1CCCC1	109-99-9	-3	a	-4,12	0,130324	0,109884	-0,631769	-0,0218	-0,4101
30	2-chloroethanol	OCCCl	107-07-3	-2	a	-2,86	0,345333	0,113758	-0,643727	-0,0709	-0,4343
31	hex-1-en-3-ol	C=CC(O)CCC	4798-44-1	-2	a	-1,66	0,339176	0,169969	-0,651938	-0,0501	-0,3627
32	N-vinylcarbazole	C=Cn1c2ccccc2c3ccccc1	1484-13-5	-3,8	d	-3,8	0,163776	0,365242	-0,891449	-0,0457	-0,2748
33	diethyl ether	CCOCC	60-29-7	-3,6	b	-4,12	0,126072	0,126797	-0,633816	-0,0039	-0,4105
34	tetrachloroethylene	ClC(Cl)=C(Cl)Cl	127-18-4	NA	NA	NA	NA	0,096748	-0,193497	0,1120	-0,3721
35	p-dimethoxybenzene	COc1ccc(OC)cc1	150-78-7	-6	a	-4,51	0,165019	0,389692	-0,667491	-0,0157	-0,2842
36	dimethylsulfoxide	CS(=O)C	67-68-5	-7	a	-7,98	0,172171	0,944282	-0,785703	0,2023	-0,3581
37	1-bromobutane	BrCCCC	109-65-9	NA	NA	NA	0,16724	-0,161003	-0,333822	0,1686	-0,3894
38	2-amino-4-chloro-6-methylpyrimidine	n1c(N)nc(Cl)cc1C	5600-21-5	1,73	e	4,16	0,324543	0,809079	-0,755394	-0,0636	-0,3301
39	anisole	c1ccccc1OC	100-66-3	-6	a	-4,82	0,16256	0,412630	-0,661758	-0,0279	-0,3047
40	p-toluenesulfonic acid	OS(=O)(=O)c1ccc(C)cc1	104-15-4	NA	a	NA	0,388918	1,640622	-0,678665	0,0965	-0,3577
41	1,4-cineole	CC(C)C12CCC(C)(O1)CC2	470-67-7	-3	a	-4,21	0,136167	0,263752	-0,666261	0,0149	-0,4048
42	1,6-dichlorohexane	ClCCCCCCCl	2163-00-0	NA	NA	NA	0,167474	-0,110219	-0,294432	0,1859	-0,4229
43	4-tert-butylcyclohexanol	OC1CCC(C(C)(C)C)CC1	98-52-2	-2	a	-1,45	0,330922	0,235008	-0,660546	-0,0256	-0,4076
44	a-damascone	CC=CC(=O)C1C(C)=CCC	35044-68-9	-6	a	-4,72	0,183265	0,574432	-0,610237	-0,1124	-0,3302
45	Benzene	c1ccccc1	71-43-2	NA	NA	NA	0,147657	-0,147657	-0,147657	0,1491	-0,3307
46	Bromobenzene	Brc1ccccc1	108-86-1	NA	NA	NA	0,175622	-0,077384	-0,146927	0,1295	-0,3304
47	Chlorobenzene	Clc1ccccc1	108-90-7	NA	NA	NA	0,176292	-0,007342	-0,159153	0,1315	-0,3347
48	cyclohexenone	O=C1C=CCCC1	930-68-7	-6	a	-4,65	0,164903	0,511047	-0,551076	-0,1316	-0,3740
49	damascenone total	CC=CC(=O)C1=C(C)C=CC	23696-85-7	-6	a	-4,73	0,176988	0,505952	-0,563114	-0,1213	-0,3037
50	D-damascone	CC=CC(=O)C1C(C)CC=CC	71048-82-3	-6	a	-4,63	0,177573	0,533447	-0,578245	-0,1156	-0,3406
51	diethylene glycol dibutyl ether	CCCCOCCOCCOCCCC	112-73-2	-3	a	-3,67	0,121987	0,149499	-0,670022	0,0499	-0,4143
52	fluorobenzene	Fc1ccccc1	462-06-6	NA	NA	NA	0,170695	0,447708	-0,390581	0,1335	-0,3336
53	indane	c1cccc2c1CCC2	496-11-7	NA	NA	NA	0,14866	-0,024211	-0,246281	0,1434	-0,3099
54	isopentene	C=CC(C)C	563-45-1	NA	NA	NA	0,124188	-0,067497	-0,319465	0,1886	-0,3562
55	isophorone	O=C1C=C(C)CC(C)(C)C1	78-59-1	-6	a	-4,53	0,156795	0,513021	-0,561377	-0,1144	-0,3600
56	m-xylene	Cc1cc(C)ccc1	108-38-3	NA	NA	NA	0,145849	0,004912	-0,342548	0,1459	-0,3138
57	nerol	OCC=C(C)CCC=C(C)C	106-25-2	-2	a	-2,22	0,334666	0,084533	-0,643112	-0,1400	-0,3247
58	p-cymene	CC(C)c1ccc(C)cc1	99-87-6	NA	NA	NA	0,142714	0,006093	-0,341079	0,1493	-0,3081
59	p-dichlorobenzene	Clc1ccc(Cl)cc1	105-46-7	NA	NA	NA	0,186759	0,008053	-0,164398	0,1152	-0,3382
60	pentachlorobenzene	Clc1c(Cl)c(Cl)c(Cl)cc1Cl	608-93-5	NA	NA	NA	0,218392	0,099929	-0,131845	0,0812	-0,358
61	Styrene	C=Cc1ccccc1	100-42-5	NA	NA	NA	0,150165	-0,006985	-0,300101	0,1114	-0,3016
62	Trifluorotoluene	FC(F)(F)c1ccccc1	98-08-8	NA	NA	NA	0,184558	1,170687	-0,370471	0,1173	-0,3515
63	butyl trimethyl ammonium	CCCC[N+](C)(C)C	14251-72-0	NA	NA	NA	0,192350	-0,050863	-0,606735	0,0300	-0,5996
64	benzyl alcohol	c1ccccc1CO	100-51-6	-2	a	-2,79	0,341251	0,148050	-0,653723	-0,0636	-0,3181

Table S1: training set. Part 1. (continued)

65	ethanol	OCC	64-17-5	-2,4	b	-2,16	0,335191	0,105548	-0,639099	-0,0519	-0,4342
66	1-(2-chloroethyl)pyrrolidine	ClCCN1CCCC1	7250-67-1	10,43	f	8,56	0,178564	0,003794	-0,626240	0,0012	-0,3599
67	N,N-diethylethanolamine	OCCN(CC)CC	100-37-8	9,8	g	9,55	0,336904	0,130165	-0,653874	0,0299	-0,3433
68	triethylamine	CCN(CC)CC	121-44-8	10,65	f	10,21	0,122516	-0,013208	-0,597086	0,0351	-0,3401
69	tripropargylamine	C#CCN(CC#C)CC#C	6921-29-5	8,31	f	6,69	0,252689	0,114176	-0,588170	-0,0083	-0,3730
70	N,N-dimethyl-4-aminopyridine	CN(C)c1ccncc1	1122-58-3	9,2	h	8,78	0,149409	0,456774	-0,732515	-0,0575	-0,3023
71	acenaphthene	c12cccc(CC3)c1c3ccc2	83-32-9	NA	NA	NA	0,149496	0,033339	-0,227964	0,1105	-0,274
72	carbon tetrachloride	ClC(Cl)(Cl)Cl	56-23-5	NA	NA	NA	NA	0,090966	-0,363864	0,0940	-0,4638
73	1-naphthol	Oc1cccc2ccccc12	90-15-3	-6	a	-5,59	0,351081	0,418124	-0,661334	-0,0598	-0,2767
74	2,4,6-trinitrotoluene	O=[N+]([O-])c1cc([N+](=	118-96-7	-11	b	-9,79	0,283431	0,533142	-0,455859	0,0114	-0,4222
75	2,6-dinitrotoluene	O=[N+]([O-])c1cccc([N+	606-20-2	-11	b	-9,24	0,228748	0,493501	-0,453932	0,0452	-0,3864
76	m-nitrotoluene	Cc1cc([N+](=O)[O-])ccc	99-08-1	-11	b	-8,83	0,215845	0,524968	-0,473446	0,0635	-0,3524
77	Nitrobenzene	O=[N+]([O-])c1ccccc1	98-95-3	-11,3	b	-8,63	0,218464	0,525740	-0,471470	0,0607	-0,3655
78	1,3-dichloro-4,6-dinitrobenzene	Clc1cc(Cl)c([N+](=O)[O-	3698-83-7	-11	b	-9,55	0,277172	0,496292	-0,444398	0,0185	-0,4054
79	N,N-diisopropylaniline	CC(C)N(C(C)C)c1ccccc1	4107-98-6	5,6	f	5,84	0,141011	0,416080	-0,772430	-0,0078	-0,2672
80	1,1,1-trichloroethane	ClC(Cl)(Cl)C	71-55-6	NA	NA	NA	0,166084	0,034291	-0,306950	0,1267	-0,4458
81	hexane	CCCCCC	110-54-3	NA	NA	NA	0,111473	-0,206517	-0,332497	0,2299	-0,4369
82	diuron	CN(C)C(=O)Nc1cc(Cl)c(C	330-54-1	0	f	-3,24	0,321188	1,088156	-0,885185	-0,0298	-0,3118
83	pentachloroethane	ClC(Cl)(Cl)C(Cl)Cl	76-01-7	NA	NA	NA	0,2531	0,074272	-0,298775	0,1150	-0,448
84	pyridine	n1ccccc1	110-86-1	5,14	f	5,12	0,161992	0,128150	-0,535821	-0,1122	-0,3441
85	triphenylphosphine oxide	c1ccccc1P(=O)(c2ccccc2	791-28-6	-6	a	-7,45	0,195067	1,338917	-0,756587	-0,0414	-0,3337
86	5-chloro-2-mercaptobenzothiazole	Sc1nc2cc(Cl)ccc2s1	5331-91-9	2,5	c	-2,33	0,199713	0,320818	-0,509142	-0,1013	-0,3212
87	Tris(benzyltriazolylmethyl)amine	c1ccccc1CN2N=NC(=C2	510758-28-8	9,48	f	2,17	0,207861	0,174160	-0,608816	0,0476	-0,3288
88	solketal	CC1(C)OC(CO)CO1	100-79-8	-2	a	-2,98	0,333435	0,601119	-0,672682	-0,0783	-0,4036
89	4-chloro-3-methylphenol	Oc1cc(C)c(Cl)cc1	59-50-7	-6	a	-6,26	0,350592	0,396536	-0,652587	-0,0436	-0,3111

Table S1: training set. Part 1. (continued)

89	4-chloro-3-methylphenol	Oc1cc(C)c(Cl)cc1	59-50-7	-6	a	-6,26	0,350592	0,396536	-0,652587	-0,0436	-0,3111
90	o-cresol	Oc1c(C)cccc1	95-48-7	-6	a	-5,46	0,34098	0,381964	-0,661583	-0,0353	-0,3029
91	p-chlorophenol	Oc1ccc(Cl)cc1	106-48-9	-6	a	-6,26	0,351007	0,389455	-0,651654	-0,0482	-0,3149
92	Phenol	Oc1ccccc1	108-95-2	-6,7	b	-5,48	0,347971	0,391009	-0,655794	-0,0414	-0,3089
93	2-cyanopyridine	n1c(C#N)cccc1	100-70-9	1,45	f	0,98	0,193393	0,303933	-0,529443	-0,1468	-0,3683
94	3-cyano-4,6-dimethyl-2-hydroxypyridine	n1c(C)cc(C)c(C#N)c1O	769-28-8	1,45	f	-4,54	0,368452	0,739251	-0,682914	-0,1038	-0,3368
95	m-cresol	Oc1cc(C)ccc1	108-39-4	-6	a	-5,48	0,347558	0,400178	-0,657004	-0,0378	-0,3051
96	p-cresol	Oc1ccc(C)cc1	106-44-5	-6	a	-5,44	0,346832	0,380427	-0,658769	-0,0376	-0,2987
97	putrescine	NCCCCN	110-60-1	10,84	f	10,51	0,264615	-0,022024	-0,708076	-0,0216	-0,3784
98	2-methylimidazole	N1C(C)=NC=C1	693-98-1	7,75	f	7,52	0,314364	0,513899	-0,669570	-0,0263	-0,3087
99	tert-butylamine	NC(C)(C)C	75-64-9	10,55	f	10,65	0,258421	0,088574	-0,678239	-0,0151	-0,3786
100	diethylamine	CCNCC	109-89-7	10,98	f	10,58	0,258801	-0,007058	-0,640130	0,0001	-0,3589
101	diisopropylamine	CC(C)NC(C)C	108-18-9	11,05	f	10,72	0,259494	0,082309	-0,645651	0,0149	-0,3520
102	pentan-2-amine	CC(N)CCC	96-15-1	10,56	f	10,43	0,263271	0,043213	-0,690715	-0,0117	-0,3774
103	2-chloroaniline	Nc1c(Cl)cccc1	95-51-2	2,62	f	2,79	0,309463	0,326134	-0,750191	-0,0414	-0,2993
104	aniline	Nc1ccccc1	62-53-3	4,6	f	4,64	0,302381	0,336348	-0,799294	-0,0317	-0,2770
105	2-chloro-4-nitroaniline	Nc1ccc([N+](=O)[O-])cc	121-87-9	0,98	f	0,12	0,327396	0,526838	-0,766428	-0,0978	-0,3297
106	4-amino-2-nitrophenol	Oc1ccc(N)cc1[N+](=O)[(	119-34-6	2,45	f	2,92	0,400213	0,541275	-0,735094	-0,0878	-0,3049
107	3-methoxyphenol	Oc1cc(OC)ccc1	150-19-6	-6	a	-4,84	0,3489	0,433072	-0,660361	-0,0300	-0,3053
108	N-allylaniline	C=CCNc1ccccc1	589-09-3	4,17	i	4,54	0,296722	0,390836	-0,810571	-0,0231	-0,2738
109	benzylamine	NCc1ccccc1	100-46-9	9,34	i	9,3	0,268863	-0,024784	-0,708024	-0,0287	-0,3180
110	2,4,6-tribromophenol	Oc1c(Br)cc(Br)cc1Br	118-79-6	-6	a	-7,71	0,369213	0,456884	-0,630309	0,0960	-0,3299
111	2,2'-methylenebis(4-chlorophenol) (dichlorophen)	Oc1ccc(Cl)cc1Cc2cc(Cl)c	97-23-4	-6	a	-6,28	0,352792	0,399181	-0,661450	-0,0259	-0,3088
112	2,4,6-triiodophenol	Oc1c(I)cc(I)cc1I	640-23-4	-6	a	-7,28	0,293725	0,406574	-0,589881	-0,0750	-0,3228
113	2,4-D butoxyethanol ester	CCCCOCCOC(=O)COc1c	1929-73-3	-3	a	-4,08	0,204147	0,760558	-0,661258	-0,0118	-0,3240
114	methyl benzoate	c1ccccc1C(=O)OC	93-58-3	-7	a	-6,88	0,195048	0,846978	-0,640513	-0,1024	-0,3410
115	benzamide	NC(=O)c1ccccc1	55-21-0	0	b	-0,36	0,317113	0,756269	-0,760724	-0,1028	-0,3419
116	tributyl phosphate	CCCCOP(=O)(OCCCC)OC	126-73-8	-7	a	-9,06	0,143578	1,647063	-0,744226	0,2263	-0,4409
117	ethyl lactate	CC(O)C(=O)OCC	97-64-3	-2	a	-3,67	0,364548	0,760610	-0,660047	-0,0521	-0,4396

Table S1: training set. Part 1. (continued)

118	benzyl acetate	CC(=O)OCc1ccccc1	140-11-4	-7	a	-7,01	0,16958	0,755293	-0,643000	-0,0620	-0,3278
119	ethyl acetate	CC(=O)OCC	141-78-6	-7	a	-6,99	0,158307	0,757225	-0,625073	-0,0624	-0,4378
120	N-methylacetamide	CC(=O)NC	79-16-3	0	b	-1,59	0,304140	0,718033	-0,708864	-0,0492	-0,3848
121	2-hydroxybenzamide	NC(=O)c1c(O)cccc1	65-45-2	0	b	-0,97	0,404825	0,815626	-0,760515	-0,0930	-0,3152
122	dimethyl phthalate	COC(=O)c1c(C(=O)OC)cc	131-11-3	-7	a	-6,65	0,183684	0,800957	-0,619137	-0,1033	-0,3510
123	γ-decalactone	O=C1CCC(O1)CCCCCC	706-14-9	-7	a	-7,04	0,172254	0,794477	-0,626586	-0,0523	-0,4250
124	dimethyl carbonate	COC(=O)OC	616-38-6	-7	a	-5,54	0,139144	1,046543	-0,605711	-0,0250	-0,4648
125	dimethyl isophthalate	COC(=O)c1cc(C(=O)OC)c	1459-93-4	-7	a	-6,59	0,236881	0,858278	-0,644779	-0,1032	-0,3510
126	dimethyl terephthalate	COC(=O)c1ccc(C(=O)OC	120-61-6	-7	a	-6,6	0,201803	0,852729	-0,639859	-0,1133	-0,3487
127	2,2-dichloroacetamide	ClC(Cl)C(=O)N	683-72-7	0	b	-10,36	0,33421	0,763718	-0,734769	0,1399	-0,433
128	o-nitrobenzaldehyde	O=Cc1c([N+](=O)[O-])cc	552-89-6	-10	a	-7,44	0,22814	0,524488	-0,495565	-0,1769	-0,3733
129	aziridine	N1CC1	151-56-4	8,04	i	8,04	0,275747	-0,116007	-0,567809	-0,0181	-0,3872
130	oxirane	O1CC1	75-21-8	-3	a	-4,15	0,130694	0,014804	-0,552385	-0,0511	-0,4451
131	propanal	O=CCC	123-38-6	-10	a	-6,94	0,157360	0,376269	-0,486798	-0,1460	-0,4120
132	hexanal	O=CCCCCC	66-25-1	-10	a	-6,94	0,154962	0,380459	-0,491357	-0,1394	-0,4079
133	o-fluorobenzaldehyde	O=Cc1c(F)cccc1	446-52-6	-10	a	-7,33	0,201994	0,466248	-0,506535	-0,1546	-0,3516
134	acetyl chloride	CC(=O)Cl	75-36-5	-7	a	-9,09	0,173207	0,395961	-0,421771	0,1403	-0,4503
135	3,4-dichlorobut-1-ene	C=CC(Cl)CCl	760-23-6	NA	NA	NA	0,199591	-0,058982	-0,258877	0,1383	-0,3856
136	a,a'-dichloro-p-xylene	ClCc1ccc(CCl)cc1	623-25-6	NA	NA	NA	0,179583	-0,005944	-0,307250	0,0999	-0,3408
137	benzyl chloride	ClCc1ccccc1	100-44-7	NA	NA	NA	0,176790	-0,005227	-0,305747	0,1207	-0,3363
138	1,3-dibromopropane	BrCCCBr	109-64-8	NA	NA	NA	0,172336	-0,133968	-0,263204	0,1275	-0,3951
139	1,1-dimethylhydrazine	CN(C)N	57-14-7	7,21	f	5,88	0,271633	-0,123708	-0,516831	-0,0095	-0,3282
140	Benzaldehyde	O=Cc1ccccc1	100-52-7	-10	a	-7,11	0,192829	0,405291	-0,507167	-0,1500	-0,3467
141	vanillin	O=Cc1cc(OC)c(O)cc1	121-33-5	-3	a	-4,9	0,374557	0,409727	-0,706392	-0,1296	-0,3109
142	o-methoxybenzaldehyde	O=Cc1ccccc1OC	135-02-4	-3	a	-4,87	0,196616	0,449365	-0,673454	-0,0711	-0,3223
143	salicylaldehyde	O=Cc1ccccc1O	90-02-8	-6	a	-5,99	0,396056	0,443857	-0,660083	-0,1390	-0,3223

Table S1: training set. Part 1. (continued)

144	o-vanillin	O=Cc1c(O)c(OC)ccc1	148-53-8	-3	a	-4,9	0,395351	0,432160	-0,647189	-0,0675	-0,3044
145	acrylic acid	OC(=O)C=C	79-10-7	-7	a	-7	0,362357	0,761921	-0,613551	-0,1280	-0,3987
146	methyl acrylate	COC(=O)C=C	96-33-3	-7	a	-6,77	0,172472	0,789853	-0,633978	-0,1111	-0,3931
147	nonenal	O=CC=CCCCCCC	2463-53-8	-10	a	-4,2	0,158336	0,380190	-0,514667	-0,1396	-0,3744
148	2-hydroxypropyl acrylate	C=CC(=O)OCC(O)C	999-61-1	-2	a	-2,87	0,345911	0,796725	-0,678605	-0,0891	-0,3997
149	acrolein	O=CC=C	107-02-8	-10	a	-4,31	0,169983	0,372608	-0,497741	-0,1650	-0,3974
150	methyl methacrylate	C=C(C)C(=O)OC	80-62-6	-7	a	-6,82	0,167367	0,814188	-0,636897	-0,1045	-0,3746
151	1,4-benzoquinone	O=C1C=CC(=O)C=C1	106-51-4	-6	a	-7,67	0,196222	0,486642	-0,521574	0,0202	-0,4103
152	diphenyl disulfide	c1ccccc1SSc2ccccc2	882-33-7	-5	b	-5	0,175132	0,069467	-0,244060	-0,1002	-0,3256
153	diisopropyl disulfide	CC(C)SSC(C)C	4253-89-8	-5	b	-5	0,162496	0,014765	-0,343185	0,1401	-0,3381
154	benzoyl peroxide	c1ccccc1C(=O)OOC(=O)	94-36-0	-7	a	-7,17	0,199583	0,880795	-0,544026	-0,1039	-0,348
155	diethyl peroxide	CCOOCC	628-37-5	-3	a	-4,8	0,133493	0,141372	-0,372713	0,2417	-0,4045
156	L-(-)-lactic acid	O=C(O)C(O)C	79-33-4	-2	a	-3,79	0,36732	0,727777	-0,655996	-0,0793	-0,4468
157	furan	O1C=CC=C1	110-00-9	-3	a	-2,87	0,166160	0,169005	-0,543308	-0,0699	-0,3183
158	allyl alcohol	OCC=C	107-18-6	-2	a	-2,29	0,338005	0,098825	-0,649290	-0,0665	-0,3575
159	2,2,2-trifluoroethanol	OCC(F)(F)F	75-89-8	-2	a	-4,33	0,351315	1,061492	-0,628282	-0,0728	-0,4801
160	but-2-yne-1,4-diol	OCC#CCO	110-65-6	-2	a	-3,13	0,341719	0,105701	-0,623778	-0,0615	-0,3652
161	hydroquinone	Oc1ccc(O)cc1	123-31-9	-6	a	-5,89	0,346314	0,365150	-0,662902	-0,0398	-0,2909
162	4,5-dichlorocatechol	Oc1cc(Cl)c(Cl)cc1O	3428-24-8	-6	a	-6,93	0,374937	0,375604	-0,683189	-0,0497	-0,3163
163	paraquat	C[n+]1ccc(cc1)c2cc[n+]	4685-14-7	NA	NA	NA	0,271576	0,240402	-0,624971	-0,2521	-0,6459
164	ethanethiol	SCC	75-08-1	-6,8	b	-6,8	0,153312	-0,020884	-0,372869	-0,0677	-0,3536
165	pentachlorophenol	Oc1c(Cl)c(Cl)c(Cl)c(Cl)c1	87-86-5	-6	a	-8,45	0,376975	0,511426	-0,622062	0,0786	-0,3462

Table S1: training set. Part 1. (continued)

166	2,5-dinitrophenol	Oc1cc([N+](=O)[O-])ccc	329-71-5	-6	a	-7,03	0,407798	0,548776	-0,639311	-0,1872	-0,3688
167	2-sec-butyl-4,6-dinitrophenol (dinoseb)	Oc1c(C(C)CC)cc([N+](=C	88-85-7	-6	a	-7,81	0,413833	0,549347	-0,644651	0,0338	-0,3648
168	4,6-dinitro-o-cresol	Oc1c(C)cc([N+](=O)[O-]	534-52-1	-6	a	-8,04	0,412928	0,549837	-0,638957	0,0335	-0,3678
169	2,4,6-trichlorophenol	Oc1c(Cl)cc(Cl)cc1Cl	88-06-2	-6	a	-7,8	0,373584	0,486309	-0,630070	0,0975	-0,3354
170	o-nitrophenol	Oc1c([N+](=O)[O-])cccc	88-75-5	-6	a	-6,88	0,403976	0,542781	-0,648090	-0,1600	-0,3412
171	2,4-dichlorophenol	Oc1c(Cl)cc(Cl)cc1	120-83-2	-6	a	-7,03	0,355975	0,440806	-0,633332	-0,0516	-0,3245
172	2,4,5-tribromoimidazole	[nH]1c(Br)nc(Br)c1Br	2034-22-2	3,7	f	1,59	0,351083	0,457844	-0,636593	-0,0597	-0,3278
173	acetic acid	O=C(O)C	64-19-7	-7	a	-7	0,359526	0,728033	-0,593298	-0,0929	-0,4475
174	quinine	COc(c1)ccc2c1c(C(O)C3	130-95-0	8,52	i	9,05	0,341146	0,407133	-0,666875	-0,0149	-0,2949
175	atrazine	CCNc(nc(Cl)n1)nc1NC(C	1912-24-9	1,6	j	3,2	0,31957	0,893122	-0,764435	-0,0566	-0,3541
176	glyphosate	OP(=O)(O)CNCC(=O)O	1071-83-6	5,69	k	6,34	0,385842	1,480404	-0,709486	0,0105	-0,3876
177	atropine	c1ccccc1C(CO)C(=O)OC	51-55-8	9,43	l	9,39	0,332097	0,821822	-0,651826	0,0268	-0,3306
178	piperidine	N1CCCCC1	110-89-4	11,22	f	10,4	0,267097	-0,024746	-0,637510	0,0021	-0,3573
179	amphetamine	NC(C)Cc1ccccc1	300-62-9	10,13	m	10,01	0,263676	0,062800	-0,702386	-0,0110	-0,3198
180	morphine	Oc1ccc(CC5C3C=CC4O)	57-27-2	9,85	i	9,14	0,363288	0,414201	-0,720659	0,0204	-0,2915
181	resmethrin	CC(C)=CC1C(C1(C)C)C(=(	10453-86-8	-3	a	-2,85	0,187441	0,844203	-0,640358	-0,0405	-0,3057
182	strychnine	c12ccccc1N(C34)C(=O)C	57-24-9	8,26	i	9,27	0,210784	0,739302	-0,796159	-0,0015	-0,2926
183	2-phenylethylamine	c1ccccc1CCN	64-04-4	9,83	f	9,79	0,26742	-0,012374	-0,711283	-0,0224	-0,3179
184	xanthone	c12ccccc1Oc3ccccc3C2	90-47-1	-3	a	-3,75	0,200472	0,624380	-0,710001	-0,0840	-0,3148
185	2,4-dinitroaniline	O=[N+]([O-])c1c(N)ccc([	97-02-9	-1	f	-2,21	0,370802	0,532404	-0,773808	-0,1346	-0,3463
186	p-(tert-butyl)-phenyl-N-methylcarbamate	CNC(=O)Oc1ccc(C(C)(C)(	2626-83-7	0	b	-3	0,319979	1,018702	-0,730798	0,0009	-0,3152
187	azinphos-methyl	COP(=S)(OC)SCN1N=Nc2	86-50-0	-1	c	-3,81	0,223154	1,119092	-0,701005	-0,1085	-0,3415
188	carbophenothion	CCOP(=S)(OCC)SCSc1ccc	786-19-6	-5,4	b	-5	0,207504	1,120655	-0,720686	-0,0688	-0,3395
189	methomyl (lannate)	CNC(=O)ON=C(C)SC	16752-77-5	3	f	0,69	0,312536	1,057302	-0,745108	-0,0675	-0,3346
190	sarin	FP(=O)(OC(C)C)C	107-44-8	-7	a	-8,5	0,17313	1,568608	-0,715380	0,1981	-0,4533

Table S1: training set. Part 1. (continued)

191	VX	CCOP(=O)(C)SCCN(C(C)C	50782-69-9	10,3	f	10,57	0,176051	1,317526	-0,735483	0,0372	-0,3226
192	aldicarb	CNC(=O)ON=CC(C)(C)SC	116-06-3	3	f	1,63	0,316374	1,053692	-0,740759	-0,0190	-0,3430
193	malathion	S=P(OC)(OC)SC(CC(=O)C	121-75-5	-7	a	-6,82	0,237951	1,123023	-0,694892	-0,0125	-0,3443
194	1-benzoylacetone	c1ccccc1C(=O)OCC(=O)	93-91-4	-7	a	-5,55	0,19905	0,557177	-0,504330	-0,0906	-0,3496
195	pentan-2,4-dione	CC(=O)CC(=O)C	123-54-6	-7	a	-5,03	0,164888	0,493448	-0,512428	-0,1045	-0,4070
196	urethane	NC(=O)OCC	51-79-6	0	b	-3	0,324829	1,001488	-0,775954	-0,0086	-0,4173
197	trifluralin	CCCN(CCC)c1c([N+](=O	1582-09-8	-1	f	-1,35	0,249423	1,198403	-0,736180	-0,1192	-0,3759
198	caffeine	CN1C=2N=CN(C)C2C(=O	58-08-2	0,61	f	-0,92	0,204564	1,089576	-0,876935	-0,0304	-0,3146
199	benzoic acid	c1ccccc1C(=O)O	65-85-0	-7	a	-6,9	0,363778	0,807481	-0,616002	-0,1164	-0,1831
200	Acetonitrile	CC#N	75-05-8	-10,1	b	-10,1	0,180110	0,285240	-0,454267	0,2120	-0,4634

a	Myers, B. pKa table. https://www2.onu.edu/~b-myers/organic/2511_Files/pKa%20table.pdf (accessed Jan 16, 2017).
b	Moore, J. S. Specific pKas Table. 2011. http://www.chemistry2011.org/ResourceFiles/specific_pKa.pdf (accesssed Jan 16, 2017).
c	Baran, P.; Richter. Essentials of Heterocyclic Chemistry - I. 2009. http://www.scripps.edu/baran/heterocycles/Essentials1-2009.pdf (accesssed Jan 16, 2017)
d	Katritzky, A. R.; Ramsden, C. A.; Joule, J. A.; Zhdankin, V. V. 3.3 Reactivity of Five-Membered Rings with One Heteroatom. In *Handbook of Heterocyclic Chemistry* ; Elsevier, 2010; p 398.
e	Zhang, S. A Reliable and Efficient First Principles-Based Method for Predicting pKa Values. 4. Organic Bases. *J. Comput. Chem.* **2012**, *33* (31), 2469–2482 DOI: 10.1002/jcc.23068.
f	Williams, R. pKa Data Compilation. (compiled by W.P. Jencks and added to by F.H. Westheimer). http://www.chem.wisc.edu/areas/reich/pkatable/pKa_compilation-1-Williams.pdf (accessed Jan, 16th, 2017).
g	Rayer, A. V.; Sumon, K. Z.; Jaffari, L.; Henni, A. Dissociation Constants (pKa) of Tertiary and Cyclic Amines: Structural and Temperature Dependences. *J. Chem. Eng. Data* **2014**, *59* (11), 3805–3813 DOI: 10.1021/je500680q.
h	Ripin, D. H.; Evans, D. A. pKa Table. April 11, 2005. http://evans.rc.fas.harvard.edu/pdf/evans_pKa_table.pdf (accessed Jan 17, 2017)
i	Haynes, W. M. Dissociation Constants of Organic Acids and Bases. In *CRC Handbook of Chemistry and Physics* ; CRC Press, 2010.
j	MacBean C, ed; e-Pesticide Manual. 15th ed., ver. 5.1, Alton, UK; British Crop Protection Council. Atrazine (1912-24-9) (2008-2010)
k	Peixoto, M. M.; Bauerfeldt, G. F.; Herbst, M. H.; Pereira, M. S.; da Silva, C. O. Study of the Stepwise Deprotonation Reactions of Glyphosate and the Corresponding pKa Values in Aqueous Solution. *J. Phys. Chem. A* **2015**, *119* (21), 5241–5249 DOI: 10.1021/jp5099552.
l	Sangster J; LOGKOW Databank. Sangster Res. Lab., Montreal Quebec, Canada (1994)
m	Perrin DD; Dissociation constants of organic bases in aqueous solution. IUPAC Chem Data Ser, Buttersworth, London (1965)

Table S1: training set. Part 2.

N°	Molecule	literature MechoA	MechoA detail	literature MechoA reference	pred MechoA
1	thioterpineol	1.1	3° thiols are not reactive enough due to steric hindrance.	Munday, R. Toxicity of Thiols and Disulphides: Involvement of Free-Radical Species. Free Radic. Biol. Med. 1989, 7 (6), 659–673 DOI: 10.1016/0891-5849(89)90147-0.	1.1
2	geraniol	1.1	Aq: non-polar narcosis	Based on experimental ecotoxicity values, compared to the water solubility of the substance with proprietary QSAR models, it could be assigned a probable mode of action.	1.1
3	calone	1.1	Aq: similar substances seem MoA1	Based on experimental ecotoxicity values, compared to the water solubility of the substance with proprietary QSAR models, it could be assigned a probable mode of action.	1.1
4	benzyl fluoride	1.1	C-F bond very stable		1.1
5	1,2-dibromobenzene	1.1	F: narcosis	A	1.1
6	2,2,5,5-tetramethyltetrahydrofuran	1.1	F: narcosis	A	1.1
7	2',4'-dichloroacetophenone	1.1	F: narcosis	A	1.1
8	2,5-dimethylhexa-2,4-diene	1.1	F: narcosis	A	1.1
9	2-methylbut-3-yn-2-ol	1.1	F: narcosis	A	1.1
10	3,3-dimethylbutan-2-one	1.1	F: narcosis	A	1.1
11	3,6-dithiaoctane	1.1	F: narcosis	A	1.1
12	4-methyloxazole	1.1	F: narcosis	A	1.1
13	4-phenylpyridine	1.1	F: narcosis	A	1.1
14	a,a,a-trifluoro-o-tolunitrile	1.1	F: narcosis	A	1.1
15	acetone	1.1	F: narcosis	A	1.1
16	acetophenone	1.1	F: narcosis	A	1.1
17	butan-2-ol	1.1	F: narcosis	A	1.1
18	butanol	1.1	F: narcosis	A	1.1
19	cyclohexane	1.1	F: narcosis	A	1.1
20	E-hex-3-en-1-ol	1.1	F: narcosis	A	1.1
21	ethylbenzene	1.1	F: narcosis	A	1.1
22	octan-2-one	1.1	F: narcosis	A	1.1
23	octanol	1.1	F: narcosis	A	1.1
24	o-xylene	1.1	F: narcosis	A	1.1
25	pentan-1-ol	1.1	F: narcosis	A	1.1
26	pyrrole	1.1	F: narcosis	A	1.2 & 5.2
27	t-butanol	1.1	F: narcosis	A	1.1
28	tert-butyl methyl ether	1.1	F: narcosis	A	1.1
29	tetrahydrofuran	1.1	F: narcosis	A	1.1
30	2-chloroethanol	1.1	F: reactive	A	1.1
31	hex-1-en-3-ol	1.1	F: reactive	A	1.1
32	N-vinylcarbazole	1.1	F: reactive	A	1.1
33	diethyl ether	1.1	H: CNS. F: narcosis	A	1.1
34	tetrachloroethylene	1.1	M/F: narcosis	A / Wheeler, H. Tetrachloroethylene (UK PID). IPCS Inchem March 1996.	1.1
35	p-dimethoxybenzene	1.1	M: apparently not significantly transformed into quinone, because no particular effects. (PubMed) F: narcosis	A / PubChem record citing [Altmann H-J et al; Archives of Toxicol Suppl 8: 114-6 (1985)]	1.1
36	dimethylsulfoxide	1.1	M: lots of effects maybe all due to its degreasing effect, increasing permeability of membranes. F: narcosis	A / Gurtovenko, A. A.; Anwar, J. Modulating the Structure and Properties of Cell Membranes: The Molecular Mechanism of Action of Dimethyl Sulfoxide. J. Phys. Chem. B 2007, 111 (35), 10453–10460 DOI: 10.1021/jp073113e.	1.1
37	1-bromobutane	1.1	M: narcosis (formation of GS-Bu, depletion of GSH) F: narcosis	A / Khan, S.; O'Brien, P. J. 1-Bromoalkanes as New Potent Nontoxic Glutathione Depletors in Isolated Rat Hepatocytes. Biochem. Biophys. Res. Commun. 1991, 179 (1), 436–441.	1.1
38	2-amino-4-chloro-6-methylpyrimidine	1.1	may have interactions with thiamine (vitamin B1) transporters. F: narcosis	A	1.1 & p6.5
39	anisole	1.1	phenyl hydroxylation? SNAr		1.1
40	p-toluenesulfonic acid	1.1	sulfonates are primarily tensioactives, just disturbing membranes as narcotics do		1.1
41	1,4-cineole	1.1			1.1
42	1,6-dichlorohexane	1.1			1.1
43	4-tert-butylcyclohexanol	1.1			1.1
44	a-damascone	1.1		Based on experimental ecotoxicity values, compared to the water solubility of the substance with proprietary QSAR models, it could be assigned a probable mode of action.	1.1
45	Benzene	1.1			1.1
46	Bromobenzene	1.1			1.1
47	Chlorobenzene	1.1			1.1
48	cyclohexenone	1.1		Based on experimental ecotoxicity values, compared to the water solubility of the substance with proprietary QSAR models, it could be assigned a probable mode of action.	1.1
49	damascenone total	1.1		Based on experimental ecotoxicity values, compared to the water solubility of the substance with proprietary QSAR models, it could be assigned a probable mode of action.	1.1
50	D-damascone	1.1		Based on experimental ecotoxicity values, compared to the water solubility of the substance with proprietary QSAR models, it could be assigned a probable mode of action.	1.1
51	diethylene glycol dibutyl ether	1.1			1.1
52	fluorobenzene	1.1			1.1
53	indane	1.1			1.1
54	isopentene	1.1			1.1
55	isophorone	1.1		Based on experimental ecotoxicity values, compared to the water solubility of the substance with proprietary QSAR models, it could be assigned a probable mode of action.	1.1
56	m-xylene	1.1			1.1
57	nerol	1.1			1.1
58	p-cymene	1.1			1.1
59	p-dichlorobenzene	1.1			1.1
60	pentachlorobenzene	1.1			1.1
61	Styrene	1.1			1.1
62	Trifluorotoluene	1.1			1.1
63	butyl trimethyl ammonium	1.1 & #up6.1	B: disturbance of membrane / for certain ammonium (better with an ester-like moiety) AChE inhibition. Without specific group (like phosphoester), it's a reversible inhibition	Čolović, M. B., Krstić, D. Z., Lazarević-Pašti, T. D., Bondžić, A. M. & Vasić, V. M. Acetylcholinesterase Inhibitors: Pharmacology and Toxicology. Curr. Neuropharmacol. 11, 315–335 (2013).	1.1
64	benzyl alcohol	1.1 & 4.1	M: eliminated by oxidation into benzoic acid, conjugated then with glycin to form hippuric acid	Nair, B. Final report on the safety assessment of Benzyl Alcohol, Benzoic Acid, and Sodium Benzoate. Int. J. Toxicol. 20 Suppl 3, 23–50 (2001).	1.1

Table S1: training set. Part 2. (continued)

65	ethanol	1.1 & 4.3 & 4.	H: Oxidation into acetaldehyde then acetate consumes NAD+ to produce NADH, this consumption goes the same direction as Krebs cycle, it is balanced by the oxidative phosphorylation. As a consequence, glycolysis and fatty acid oxidation rates are decreased, less energy is produced and there is an overconsumption of O2 by the Electron Transport Chain. Moreover acetaldehyde can form adducts before being oxidized into acetate. F: non-polar narcosis	A / King, Michael W. 2015. 'Ethanol (Alcohol) Metabolism: Acute and Chronic Toxicities.' The Medical Biochemistry Page. August 18. http://themedicalbiochemistrypage.org/ethanol-metabolism.php.	1.1
66	1-(2-chloroethyl)pyrrolidine	1.1 & 5.2	M: not highly toxic F: narcosis	A / Vallabh Minikel, E. A mechanism of action hypothesis http://www.cureffi.org/2016/05/08/a-mechanism-of-action-hypothesis/ (accessed Sep 19, 2016).	1.1
67	N,N-diethylethanolamine	1.1 & 5.2	Structure very similar to choline. M: Transformed into phosphatidylcholine, except if the enzyme Phtdcholine Methyltransferase is missing (mouse embryos). If not transformed, causes developmental abnormalities. F: narcosis	A / Fisher, M. C.; Zeisel, S. H.; Mar, M.-H.; Sadler, T. W. Perturbations in Choline Metabolism Cause Neural Tube Defects in Mouse Embryos in Vitro. FASEB J. 2002 DOI: 10.1096/fj.01-0564fje.	1.1
68	triethylamine	1.1 & 5.2		Vallabh Minikel, E. A mechanism of action hypothesis http://www.cureffi.org/2016/05/08/a-mechanism-of-action-hypothesis/ (accessed Sep 19, 2016).	1.1
69	tripropargylamine	1.1 & 5.2 / 3.1	F: narcosis	A / Vallabh Minikel, E. A mechanism of action hypothesis http://www.cureffi.org/2016/05/08/a-mechanism-of-action-hypothesis/ (accessed Sep 19, 2016).	1.2 & 5.2
70	N,N-dimethyl-4-aminopyridine	1.1 & m4.1 / r	M: metabolism mainly into N-oxide, less toxic compound, but also after demethylation into nitroso, results in methaemoglobinaemia => lack of O2 supply.	IARC. 2010. 'General Discussion of Common Mechanisms for Aromatic Amines.' IARC Monographs on the Evaluation of Carcinogenic Risks to Humans 99: 41–54.	1.1
71	acenaphthene	1.1 & m4.3	F: narcosis	A / Tsuruda, L. S., Lamé, M. W. & Jones, A. D. Formation of epoxide and quinone protein adducts in B6C3F1 mice treated with naphthalene, sulfate conjugate of 1,4-dihydroxynaphthalene and 1,4-naphthoquinone. Arch. Toxicol. 69, 362–367 (1995). / Buckpitt, A. et al. Naphthalene-induced respiratory tract toxicity: metabolic mechanisms of toxicity. Drug Metab. Rev. 34, 791–820 (2002).	1.1
72	carbon tetrachloride	1.1 & m4.3	M: cyt P (1e-reduction gives 'CCl3 radical, then quenched by O2 to give 'OOCCl3 peroxy radical, which gives then lipids & proteins peroxidation => hepatotoxicity) Aq: narcosis (KREATiS)	Recknagel, Richard O., Eric A. Glende Jr., James A. Dolak, and Robert L. Waller. 1989. 'Mechanisms of Carbon Tetrachloride Toxicity.' Pharmacology & Therapeutics 43 (1): 139–54. doi:10.1016/0163-7258(89)90050-8.	1.1
73	1-naphthol	1.1 & m4.3	M: metabolised into 1,4- and 1,2-naphthoquinone, which form adducts to proteins and undergo RedOx cycling F: narcosis	A / Doherty, M. D.; Cohen, G. M.; Smith, M. T. Mechanisms of Toxic Injury to Isolated Hepatocytes by 1-Naphthol. Biochem. Pharmacol. 1984, 33 (4), 543–549 DOI: 10.1016/0006-2952(84)90305-8.	1.2 & m4.3
74	2,4,6-trinitrotoluene	1.1 & m4.3	M: metabolism into nitroso, results in methaemoglobinaemia => lack of O2 supply.	IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.1 & m4.3
75	2,6-dinitrotoluene	1.1 & m4.3	M: metabolism into nitroso, results in methaemoglobinaemia => lack of O2 supply.	IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.1 & m4.3
76	m-nitrotoluene	1.1 & m4.3	M: metabolism into nitroso, results in methaemoglobinaemia => lack of O2 supply.	IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.1 & m4.3
77	Nitrobenzene	1.1 & m4.3	M: metabolism into nitroso, results in methaemoglobinaemia => lack of O2 supply. F: narcosis	IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.1 & m4.3
78	1,3-dichloro-4,6-dinitrobenzene	1.1 & m4.3	M: metabolism into nitroso, results in methaemoglobinaemia => lack of O2 supply. F: reactive	A / IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.1 & m4.3
79	N,N-diisopropylaniline	1.1 & m4.3	M: metabolism mainly into N-oxide, less toxic compound, but also after demethylation into nitroso, results in methaemoglobinaemia => lack of O2 supply.	IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.1
80	1,1,1-trichloroethane	1.1 & m6.7	M: Rapid elimination and recovery / sensitization of heart to epinephrine (fast elimination >90% without metabolism), CNS depression at very high doses. F: narcosis	A / ATSDR. Toxicological Profile: 1,1,1-Trichloroethane. July 2006.	1.1
81	hexane	1.1 & mC4.3	M: chronic CNS seizure: after metabolisation into hexan-2,5-dione, pyrrole formation with a lysine residue, oxidation and crosslinking with another protein. F: narcosis	A / ATSDR. Toxicological Profile: N-Hexane. July 1999.	1.1
82	diuron	1.1 & p6.5	F: narcosis, rapidly eliminated with almost no metabolism (metabolised into 3,4-dichloroaniline, which should be OPU) Plants: blocks electron transfer at the photosystem II => inhibits O2 production	A / van Rensen, J. J. S. Molecular mechanisms of herbicide action near photosystem II. Physiol. Plant. 54, 515–521 (1982). / Call, D. J.; Brooke, L. T.; Kent, R. J.; Knuth, M. L.; Poirier, S. H.; Huot, J. M.; Lima, A. R. Bromacil and Diuron Herbicides: Toxicity, Uptake, and Elimination in Freshwater Fish. Arch. Environ. Contam. Toxicol. 1987, 16 (5), 607–613 DOI: 10.1007/BF01055817.	1.1 & p6.5
83	pentachloroethane	1.1 / 4.3	F: narcosis	A	1.1
84	pyridine	1.1 / 6	F: narcosis 2. M: CNS depression have been reported, but not significantly.	A / HSDB. PYRIDINE. 2005. http://toxnet.nlm.nih.gov/cgi-bin/sis/search/a?dbs+hsdb:@term+@DOCNO+118 (accessed Jul 23, 2015).	1.1
85	triphenylphosphine oxide	1.1 / 6.1	M: Organophosphorus compounds can disturb so various targets, most of them acting in the signaling pathways (inhibit AChE, carboxylases, DAG-lipase indirectly, PKC (probably indirect by ROS generation), steroid androgen receptor, activate Ca2+/cAMP response element binding protein). F: narcosis	A / Elersek, T.; Filipic, M. Organophosphorous Pesticides - Mechanisms of Their Toxicity. In Pesticides - The Impacts of Pesticides Exposure; Stoytcheva, M., Ed.; InTech, 2011.	6.1
86	5-chloro-2-mercaptobenzothiazole	1.1 / 6.4	may have some process inhibitor effects because it is a good chelatant of metallic ions. F: narcosis	A / De Wever, H.; Verachtert, H. Biodegradation and Toxicity of Benzothiazoles. Water Res. 1997, 31 (11), 2673–2684 DOI: 10.1016/S0043-1354(97)00138-3.	1.1
87	Tris(benzyltriazolylmethyl)amine	1.1 / 6.4	narcosis / Copper ions chelating (good chelatant in chemistry) (no exp tox data) (triazolyl not a good LG because N not protonated)	Chan, T. R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V. Polytriazoles as Copper(I)-Stabilizing Ligands in Catalysis. Org. Lett. 2004, 6 (17), 2853–2855 DOI: 10.1021/ol0493094.	1.1
88	solketal	1.1 / 6.7	F: narcosis (maybe a weak MEK (MAPK) inhibitor)	A / Toronto Research Chemicals. (S)-(+)-2,2-Dimethyl-1,3-dioxolane-4-methanol http://www.trc-canada.com/product-detail/?CatNum=D472635&CAS=22323-82-6&Chemical_Name=(S)-(+)-2,2-Dimethyl-1,3-dioxolane-4-methanol&Mol_Formula=C%E2%82%86H%E2%82%81%E2%82%82O%E2%82%83 (accessed Jan 19, 2017).	F1.1/4.1
89	4-chloro-3-methylphenol	1.2	F,Pz: narcosis 2	A	1.2

Table S1: training set. Part 2. (continued)

90	o-cresol	1.2	F,Pz: narcosis 2	A	1.2
91	p-chlorophenol	1.2	F,Pz: narcosis 2	A	1.2
92	Phenol	1.2	F,Pz: narcosis 2	A	1.2
93	2-cyanopyridine	1.2	F: narcosis 2	A	1.1
94	3-cyano-4,6-dimethyl-2-hydroxypyridine	1.2	F: narcosis 2	A	1.2
95	m-cresol	1.2	Pz: narcosis 2	A	1.2
96	p-cresol	1.2	Pz: narcosis 2	A	1.2
97	putrescine	1.2 & 5.2	Directly applied, it is corrosive, and it alkalinize the cells, entering into them in their neutral form and then become protonated, all simply because it is a strong base.	Vallabh Minikel, E. A mechanism of action hypothesis http://www.cureffi.org/2016/05/08/a-mechanism-of-action-hypothesis/ (accessed Sep 19, 2016). / Committee on Acute Exposure Guideline Levels; Committee on Toxicology; Board on Environmental Studies and Toxicology; Division on Earth and Life Studies; National Research Council. Piperidine. In Acute Exposure Guideline Levels for selected Airbone Chemicals; National Academies Press (US), 2012; Vol. 13. / Cyclohexylamine [MAK Value Documentation, 2006]. In The MAK-Collection for Occupational Health and Safety; Wiley-VCH Verlag GmbH & Co. KGaA, 2002.	1.2 & 5.2
98	2-methylimidazole	1.2 & 5.2	M: apparently no acute effects (vey hydrophilic). F: reactive	A / IARC. 2-Methylimidazole Monograph. IARC Monogr. Eval. Carcinog. Risks Hum. 2012, 101, 435–445.	1.2
99	tert-butylamine	1.2 & 5.2	no special effects, no rapid metabolism. Directly applied, it is corrosive, and it alkalinize the cells, entering into them in their neutral form and then become protonated, all simply because it is a strong base.	Vallabh Minikel, E. A mechanism of action hypothesis http://www.cureffi.org/2016/05/08/a-mechanism-of-action-hypothesis/ (accessed Sep 19, 2016). / Committee on Acute Exposure Guideline Levels; Committee on Toxicology; Board on Environmental Studies and Toxicology; Division on Earth and Life Studies; National Research Council. Piperidine. In Acute Exposure Guideline Levels for selected Airbone Chemicals; National Academies Press (US), 2012; Vol. 13. / Cyclohexylamine [MAK Value Documentation, 2006]. In The MAK-Collection for Occupational Health and Safety; Wiley-VCH Verlag GmbH & Co. KGaA, 2002.	1.2 & 5.2
100	diethylamine	1.2 & 5.2		Vallabh Minikel, E. A mechanism of action hypothesis http://www.cureffi.org/2016/05/08/a-mechanism-of-action-hypothesis/ (accessed Sep 19, 2016).	1.2 & 5.2
101	diisopropylamine	1.2 & 5.2		Vallabh Minikel, E. A mechanism of action hypothesis http://www.cureffi.org/2016/05/08/a-mechanism-of-action-hypothesis/ (accessed Sep 19, 2016).	1.2 & 5.2
102	pentan-2-amine	1.2 & 5.2		Vallabh Minikel, E. A mechanism of action hypothesis http://www.cureffi.org/2016/05/08/a-mechanism-of-action-hypothesis/ (accessed Sep 19, 2016).	1.2 & 5.2
103	2-chloroaniline	1.2 & m4.3	F: narcosis 2	A	1.2
104	aniline	1.2 & m4.3	M: metabolism into nitroso, results in methaemoglobinaemia => lack of O2 supply. F: narcosis 2	A / IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.2
105	2-chloro-4-nitroaniline	1.2 & m4.3	metabolism into nitroso? F: narcosis 2	A / IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.2 & m4.3
106	4-amino-2-nitrophenol	1.2 & m4.3	metabolism into nitroso? F: narcosis 2 Pz: soft elec	A / IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.2 & m4.3
107	3-methoxyphenol	1.2 & m4.3 &	M: 6-hydroxylation then ox to quinone / parent compound apparently inhibit melanin production. F,Pz: narcosis 2	A / Moridani, M. Y.; Cheon, S. S.; Khan, S.; O'Brien, P. J. Metabolic Activation of 3-Hydroxyanisole by Isolated Rat Hepatocytes. Chem. Biol. Interact. 2003, 142 (3), 317–333 DOI: 10.1016/S0009-2797(02)00125-4.	1.2 & m4.3
108	N-allylaniline	1.2 & m4.3 / 3	M: metabolism by cytP450 into nitroso, results in methaemoglobinaemia => lack of O2 supply. F: narcosis / narcosis 2	A / IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.2
109	benzylamine	1.2 / 3.1	benzyl activation / narcosis. Toxicity value for fish corresponds to polar narcosis (Russom model)	Russom, C. L.; Bradbury, S. P.; Broderius, S. J.; Hammermeister, D. E.; Drummond, R. A. Predicting Modes of Toxic Action from Chemical Structure: Acute Toxicity in the Fathead Minnow (Pimephales Promelas). *Environ. Toxicol. Chem.* **1997**, *16* (5), 948–967 DOI: 10.1002/etc.5620160514.	#up6.3
110	2,4,6-tribromophenol	1.2 / 5.1	F: narcosis Pz: narcosis 2	A	5.1 / 1.2 (inte
111	2,2'-methylenebis(4-chlorophenol) (dichlorophen)	1.2 / 6	MoA 2 toxicity for aquatic species F: reactive	A / Based on experimental ecotoxicity values, compared to the water solubility of the substance with proprietary QSAR models, it could be assigned a probable mode of action.	1.2
112	2,4,6-triiodophenol	1.2 / m6.7 / m	H: inhibitor of the Leukotriene B4 synthesis / M: specific non-competitive myosin VI ATPase inhibitor F: narcosis	A / Matji, T. J. A.; Lopez, B. P. J. Micronized Composition of 2,4,6-Triiodophenol. EP2174654 B1, August 22, 2012.	1.2
113	2,4-D butoxyethanol ester	#p2.1 & p2.2	Plants: hydrolysed. Then the acid is an antagonist of phytohormone auxin, overstimulating plant growth, more than it can afford (because too slowly metabolised then). H/M: after hydrolysis, acid rapidly eliminated.	Song, Y. Insight into the Mode of Action of 2,4-Dichlorophenoxyacetic Acid (2,4-D) as an Herbicide. J. Integr. Plant Biol. 2014, 56 (2), 106–113 DOI: 10.1111/jipb.12131. / Jervais, G.; Luukinen, B.; Buhl, K.; Stone, D. 2,4-D Technical Fact Sheet http://npic.orst.edu/factsheets/archive/2,4-DTech.html (accessed Dec 16, 2016).	2.1
114	methyl benzoate	2.1	?SNAr	Jaworska, J. S.; Hunter, R. S.; Schultz, T. W. Quantitative Structure-Toxicity Relationships and Volume Fraction Analyses for Selected Esters. Arch. Environ. Contam. Toxicol. 1995, 29 (1), 86–93 DOI: 10.1007/BF00213091.	2.1
115	benzamide	2.1	F: narcosis	A	2.1
116	tributyl phosphate	2.1	F: narcosis (seems to have no more effects than narcosis of the parent and its hydrolysed forms)	A / Tributyl Phosphate [MAK Value Documentation, 2002]. In The MAK-Collection for Occupational Health and Safety; Wiley-VCH Verlag GmbH & Co. KGaA, 2002.	2.1
117	ethyl lactate	2.1	M: + fast metabolism into Krebs cycle	IPCS Inchem. Ethyl Lactate. WHO Food Additives Series. 1980. / Adeva-Andany, M.; López-Ojén, M.; Funcasta-Calderón, R.; Ameneiros-Rodríguez, E.; Donapetry-García, C.; Vila-Altesor, M.; Rodríguez-Seijas, J. Comprehensive Review on Lactate Metabolism in Human Health. Mitochondrion 2014, 17, 76–100 DOI: 10.1016/j.mito.2014.05.007.	2.1

Table S1: training set. Part 2. (continued)

118	benzyl acetate	2.1		Jaworska, J. S.; Hunter, R. S.; Schultz, T. W. Quantitative Structure-Toxicity Relationships and Volume Fraction Analyses for Selected Esters. Arch. Environ. Contam. Toxicol. 1995, 29 (1), 86–93 DOI: 10.1007/BF00213091.	2.1
119	ethyl acetate	2.1		Jaworska, J. S.; Hunter, R. S.; Schultz, T. W. Quantitative Structure-Toxicity Relationships and Volume Fraction Analyses for Selected Esters. Arch. Environ. Contam. Toxicol. 1995, 29 (1), 86–93 DOI: 10.1007/BF00213091.	2.1
120	N-methylacetamide	2.1			2.1
121	2-hydroxybenzamide	2.1 & m6.7 / #	H: hydrolysis to form salicylic acid / analgesic (pain-relieve) (inhibit COX 1&2) & antipyretic (drops fever)). F: narcosis. Pz: narcosis 2	A / Salicylamide. DrugBank; DrugBank, Ed.; 2016.	2.1 & m6.7
122	dimethyl phthalate	2.1 & mC2.2	M: 1 ester is mainly hydrolysed. The monoester induces carcinogenesis. F,D,A: narcosis? (uncertain data)	Adams, W. J.; Biddinger, G. R.; Robillard, K. A.; Gorsuch, J. W. A Summary of the Acute Toxicity of 14 Phthalate Esters to Representative Aquatic Organisms. Environ. Toxicol. Chem. 1995, 14 (9), 1569–1574 DOI: 10.1002/etc.5620140916.	2.1
123	γ-decalactone	2.1 / 2.2	Unicellular Fungi: lactone much more toxic than the free acid form, disturbance of membranes. => ester narcosis? F: narcosis	A / Feron, G.; Dufosse, L.; Pierard, E.; Bonnarme, P.; Quere, J. L.; Spinnler, H. Production, Identification, and Toxicity of (Gamma)-Decalactone and 4-Hydroxydecanoic Acid from Sporidiobolus Spp. Appl. Environ. Microbiol. 1996, 62 (8), 2826–2831.	2.1
124	dimethyl carbonate	2.1 / 6.1	similar to esters (or inhibits AChE as carbamates do?)		2.1
125	dimethyl isophthalate	2.1 / mC2.2	Aq: narcosis? (uncertain data)		2.1
126	dimethyl terephthalate	2.1 / mC2.2	Aq: narcosis? (uncertain data)		2.1
127	2,2-dichloroacetamide	2.2	dichloroacetate inhibits pyruvate dehydrogenase kinase F: reactive	A / Kato, M.; Li, J.; Chuang, J. L.; Chuang, D. T. Distinct Structural Mechanisms for Inhibition of Pyruvate Dehydrogenase Kinase Isoforms by AZD7545, Dichloroacetate, and Radicicol. Struct. Lond. Engl. 1993 2007, 15 (8), 992–1004 DOI: 10.1016/j.str.2007.07.001.	2.1
128	o-nitrobenzaldehyde	3.1	F: reactive, Schiff-base formation	A / Aptula, A. O.; Roberts, D. W. Mechanistic Applicability Domains for Nonanimal-Based Prediction of Toxicological End Points: General Principles and Application to Reactive Toxicity. Chem. Res. Toxicol. 2006, 19 (8), 1097–1105 DOI: 10.1021/tx0601004.	5.1
129	aziridine	3.1	M: AN	Aptula, A. O.; Roberts, D. W. Mechanistic Applicability Domains for Nonanimal-Based Prediction of Toxicological End Points: General Principles and Application to Reactive Toxicity. Chem. Res. Toxicol. 2006, 19 (8), 1097–1105 DOI: 10.1021/tx0601004.	3.1
130	oxirane	3.1	M: AN	Aptula, A. O.; Roberts, D. W. Mechanistic Applicability Domains for Nonanimal-Based Prediction of Toxicological End Points: General Principles and Application to Reactive Toxicity. Chem. Res. Toxicol. 2006, 19 (8), 1097–1105 DOI: 10.1021/tx0601004.	1.1
131	propanal	3.1	M: AN (protein terminal NH2 - Schiff base = imine formation)	Aptula, A. O.; Roberts, D. W. Mechanistic Applicability Domains for Nonanimal-Based Prediction of Toxicological End Points: General Principles and Application to Reactive Toxicity. Chem. Res. Toxicol. 2006, 19 (8), 1097–1105 DOI: 10.1021/tx0601004.	3.1 (Schiff ba
132	hexanal	3.1	Schiff-base formation F: reactive	A / Aptula, A. O.; Roberts, D. W. Mechanistic Applicability Domains for Nonanimal-Based Prediction of Toxicological End Points: General Principles and Application to Reactive Toxicity. Chem. Res. Toxicol. 2006, 19 (8), 1097–1105 DOI: 10.1021/tx0601004.	3.1 (Schiff ba
133	o-fluorobenzaldehyde	3.1	Schiff-base formation F: reactive	A / Aptula, A. O.; Roberts, D. W. Mechanistic Applicability Domains for Nonanimal-Based Prediction of Toxicological End Points: General Principles and Application to Reactive Toxicity. Chem. Res. Toxicol. 2006, 19 (8), 1097–1105 DOI: 10.1021/tx0601004.	3.1 (Schiff ba
134	acetyl chloride	3.1	violently hydrolysed in water generating HCl (corrosive), not enzymatic!	PubChem, citing [Sax, N.I. Dangerous Properties of Industrial Materials. 6th ed. New York, NY: Van Nostrand Reinhold, 1984., p. 106]	F3.1/3.2
135	3,4-dichlorobut-1-ene	3.1 & 4.1	3-Chloride leaves to obtain a reactive allyl+ F: reactive	A	1.1
136	a,a'-dichloro-p-xylene	3.1 & 4.1	Cl- as a good leaving group F: reactive	A / α-Chlorinated Toluenes [MAK Value Documentation, 1993]. In The MAK-Collection for Occupational Health and Safety; Wiley-VCH Verlag GmbH & Co. KGaA, 2002.	1.1
137	benzyl chloride	3.1 & 4.1	M: N-acetylcysteine and glutathione conjugates (Nucleophilic Substitution of the LG) => diminution of antioxydant levels	α-Chlorinated Toluenes [MAK Value Documentation, 1993]. In The MAK-Collection for Occupational Health and Safety; Wiley-VCH Verlag GmbH & Co. KGaA, 2002.	1.1
138	1,3-dibromopropane	3.1 & 4.1 / 1.1	? F: reactive	A	1.1
139	1,1-dimethylhydrazine	3.1 & 4.3	M: binds to ketoacids and vitamin B6, inhibiting its wide actions / metabolised into diazonium, which gives oxidative stress (RNS) and generates Me' radicals. F: reactive	A / ATSDR. Toxicological Profile: Hydrazines http://www.atsdr.cdc.gov/ToxProfiles/tp.asp?id=502&tid=89 (accessed Jul 24, 2015).	3.1 & 4.3
140	Benzaldehyde	3.1 & m4.1	F: reactive (gill irritation reported by Russom et al, as it happened with the reference electrophile: acrolein) (M: Schiff base formation / very fast metabolism/elimination)	Russom, C. L.; Bradbury, S. P.; Broderius, S. J.; Hammermeister, D. E.; Drummond, R. A. Predicting Modes of Toxic Action from Chemical Structure: Acute Toxicity in the Fathead Minnow (Pimephales Promelas). *Environ. Toxicol. Chem.* **1997**, *16* (5), 948–967 DOI: 10.1002/etc.5620160514. / Benzaldehyde [MAK Value Documentation, 2002]. In The MAK-Collection for Occupational Health and Safety; Wiley-VCH Verlag GmbH & Co. KGaA, 2002; Vol. 17, pp 13–36.	3.1 (Schiff ba
141	vanillin	3.1 & m4.1	M: rapidly and quasi totally metabolised into vanillic acid, rapidly eliminated. F: reactive. Pz: narcosis 2	A / IPCS Inchem. Vanillin SIDS http://www.inchem.org/documents/sids/sids/121335.pdf (accessed Jul 23, 2015).	3.1 & m4.1
142	o-methoxybenzaldehyde	3.1 & m4.1	Schiff-base formation	Benzaldehyde [MAK Value Documentation, 2002]. In The MAK-Collection for Occupational Health and Safety; Wiley-VCH Verlag GmbH & Co. KGaA, 2002; Vol. 17, pp 13–36. / IPCS Inchem. Vanillin SIDS http://www.inchem.org/documents/sids/sids/121335.pdf (accessed Jul 23, 2015).	F3.1/3.2
143	salicylaldehyde	3.1 & m4.1	Schiff-base formation F: reactive	A / Benzaldehyde [MAK Value Documentation, 2002]. In The MAK-Collection for Occupational Health and Safety; Wiley-VCH Verlag GmbH & Co. KGaA, 2002; Vol. 17, pp 13–36. / IPCS Inchem. Vanillin SIDS http://www.inchem.org/documents/sids/sids/121335.pdf (accessed Jul 23, 2015).	3.1 & m4.1

Table S1: training set. Part 2. (continued)

144	o-vanillin	3.1 & m4.1	Schiff-base formation F: reactive Pz: narcosis 2	A / Benzaldehyde [MAK Value Documentation, 2002]. In The MAK-Collection for Occupational Health and Safety; Wiley-VCH Verlag GmbH & Co. KGaA, 2002; Vol. 17, pp 13–36. / IPCS Inchem. Vanillin SIDS http://www.inchem.org/documents/sids/sids/121335.pdf (accessed Jul 23, 2015).	3.1 & m4.1
145	acrylic acid	3.2	M: Michael	Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037.	3.2
146	methyl acrylate	3.2	M: Michael	Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037.	3.2
147	nonenal	3.2	M: Michael	Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037.	3.2
148	2-hydroxypropyl acrylate	3.2	M: Michael. F: reactive	A / Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037.	3.2
149	acrolein	3.2	M: Michael. F: reactive	A / Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037.	3.2
150	methyl methacrylate	3.2 & 4.1	M: Easy Metabolism/Michael	Bonnard, N.; Brondeau, M.-T.; Jargot, D.; Miraval, S.; Pillière, F.; Robert, S.; Schneider, O.; Serre, P. Méthacrylate de méthyle - Fiche toxicologique n°62. INRS 2013. / Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037.	3.2
151	1,4-benzoquinone	3.2 & 4.4		Bolton, J. L.; Trush, M. A.; Penning, T. M.; Dryhurst, G.; Monks, T. J. Role of Quinones in Toxicology. Chem. Res. Toxicol. 2000, 13 (3), 135–160 DOI: 10.1021/tx9902082.	F3.1/3.2
152	diphenyl disulfide	3.3	F: reactive	A / Morgott, D.; Lewis, C.; Bootman, J.; Banton, M. Disulfide Oil Hazard Assessment Using Categorical Analysis and a Mode of Action Determination. Int. J. Toxicol. 2014, 33 (1 Suppl), 181S–198S DOI: 10.1177/1091581813504227. / Kazius, J.; McGuire, R.; Bursi, R. Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem. 2005, 48 (1), 312–320 DOI: 10.1021/jm040835a.	3.3
153	diisopropyl disulfide	3.3	M: homolytic cleavage. F: narcosis	A / Morgott, D.; Lewis, C.; Bootman, J.; Banton, M. Disulfide Oil Hazard Assessment Using Categorical Analysis and a Mode of Action Determination. Int. J. Toxicol. 2014, 33 (1 Suppl), 181S–198S DOI: 10.1177/1091581813504227. / Kazius, J.; McGuire, R.; Bursi, R. Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem. 2005, 48 (1), 312–320 DOI: 10.1021/jm040835a.	3.3
154	benzoyl peroxide	3.3	ROS	Kazius, J.; McGuire, R.; Bursi, R. Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem. 2005, 48 (1), 312–320 DOI: 10.1021/jm040835a.	3.3
155	diethyl peroxide	3.3	ROS	Kazius, J.; McGuire, R.; Bursi, R. Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem. 2005, 48 (1), 312–320 DOI: 10.1021/jm040835a.	3.3
156	L-(-)-lactic acid	4.1	M: highly metabollised, high elimination and high incorporation into proteins with high amounts.	Morotomi, M.; Sakai, K.; Yazawa, K.; Suegara, N.; Kawai, Y.; Mutai, M. Effect and Fate of Orally Administered Lactic Acid in Rats. J. Nutr. Sci. Vitaminol. (Tokyo) 1981, 27 (2), 117–128.	F1.1/4.1
157	furan	4.3	M: metabolism into (Z)-but-2-en-1,4-dial => proteins and DNA adducts after glutathione is totally consumed => caricnogenic and mutagen. F: reactive	A / Food and Drug Administration. Docket No. 2004N-0205 Furan in Food, Thermal Treatment; Request for Data and Information; 2004.	4.3
158	allyl alcohol	4.3	M: ox to acrolein => adducts. F: reactive	A / Koleva, Y.; Barzilov, I. Comparative Study of Mechanism of Action of Allyl Alcohols for Different Endpoints. 2010, 49, 55–59.	4.3
159	2,2,2-trifluoroethanol	4.3	M: ox to trifluoroacetaldehyde => adducts. F: reactive	A / Airaksinen, M. M.; Rosenberg, P. H.; Tammisto, T. A Possible Mechanism of Toxicity of Trifluoroethanol and Other Halothane Metabolites. Acta Pharmacol. Toxicol. (Copenh.) 1970, 28 (4), 299–304 DOI: 10.1111/j.1600-0773.1970.tb00556.x.	F1.1/4.1
160	but-2-yne-1,4-diol	4.3	metabolism into (Z)-but-2-en-1,4-dial? F: reactive	A / Koleva, Y.; Barzilov, I. Comparative Study of Mechanism of Action of Allyl Alcohols for Different Endpoints. 2010, 49, 55–59.	4.3
161	hydroquinone	4.3 & 4.4	M: Oxidation by tyrosinase / cytP450 to quinone => protein and DNA adducts. Pz: pro-elec	A / Bonnard, N.; Pillière, F.; Protois, J.-C.; Schneider, O. Fiche toxicologique de l'hydroquinone. INRS 2006. / Bolton, J. L.; Trush, M. A.; Penning, T. M.; Dryhurst, G.; Monks, T. J. Role of Quinones in Toxicology. Chem. Res. Toxicol. 2000, 13 (3), 135–160 DOI: 10.1021/tx9902082.	1.2
162	4,5-dichlorocatechol	4.3 & 4.4	Ox to o-quinone. F: reactive	A / Bolton, J. L.; Trush, M. A.; Penning, T. M.; Dryhurst, G.; Monks, T. J. Role of Quinones in Toxicology. Chem. Res. Toxicol. 2000, 13 (3), 135–160 DOI: 10.1021/tx9902082.	4.3 & 4.4
163	paraquat	4.4	ALL: RedOx cycling	Cohen, G. M.; d'Arcy Doherty, M. Free Radical Mediated Cell Toxicity by Redox Cycling Chemicals. Br. J. Cancer. Suppl. 1987, 8, 46–52.	4.4
164	ethanethiol	4.4	M: 1° & 2° thiols, in their anionic form, are oxidized into thiyl radical and then enter a redox cycle between the thiyl radical and the disulfides. Aq: more toxic than baseline	Munday, R. Toxicity of Thiols and Disulphides: Involvement of Free-Radical Species. Free Radic. Biol. Med. 1989, 7 (6), 659–673 DOI: 10.1016/0891-5849(89)90147-0.	4.4
165	pentachlorophenol	5.1	F,Pz: OPU(H+)	A / Schultz, T. W. The Use of the Ionization Constant (pKa) in Selecting Models of Toxicity in Phenols. Ecotoxicol. Environ. Saf. 1987, 14 (2), 178–183.	5.1

Table S1: training set. Part 2. (continued)

166	2,5-dinitrophenol	5.1	F: reactive Pz: resp unc	A / Schultz, T. W. The Use of the Ionization Constant (pKa) in Selecting Models of Toxicity in Phenols. Ecotoxicol. Environ. Saf. 1987, 14 (2), 178–183.	5.1
167	2-sec-butyl-4,6-dinitrophenol (dinoseb)	5.1	F: uncoupler	A / Schultz, T. W. The Use of the Ionization Constant (pKa) in Selecting Models of Toxicity in Phenols. Ecotoxicol. Environ. Saf. 1987, 14 (2), 178–183.	5.1
168	4,6-dinitro-o-cresol	5.1	F: uncoupler Pz: resp unc	A / Schultz, T. W. The Use of the Ionization Constant (pKa) in Selecting Models of Toxicity in Phenols. Ecotoxicol. Environ. Saf. 1987, 14 (2), 178–183.	5.1
169	2,4,6-trichlorophenol	5.1 / 1.2	OPU(H+). F,Pz: narcosis 2	A / Schultz, T. W. The Use of the Ionization Constant (pKa) in Selecting Models of Toxicity in Phenols. Ecotoxicol. Environ. Saf. 1987, 14 (2), 178–183.	5.1
170	o-nitrophenol	5.1 / 1.2	OPU(H+). F: narcosis 2. Pz: soft elec	A / Schultz, T. W. The Use of the Ionization Constant (pKa) in Selecting Models of Toxicity in Phenols. Ecotoxicol. Environ. Saf. 1987, 14 (2), 178–183.	5.1
171	2,4-dichlorophenol	5.1 / 1.2	OPU(H+). Pz: narcosis 2	A / Schultz, T. W. The Use of the Ionization Constant (pKa) in Selecting Models of Toxicity in Phenols. Ecotoxicol. Environ. Saf. 1987, 14 (2), 178–183.	5.1 / 1.2 (inte
172	2,4,5-tribromoimidazole	5.1 / 5	M: OPU + destruction of neurones. F: reactive. Quorum sensing inhibitor (mainly bacteria)	A /Verschoyle, R. D.; Brown, A. W.; Thompson, C. A. The Toxicity and Neuropathology of 2,4,5-Tribromoimidazole and Its Derivatives in Rats. Arch. Toxicol. 1984, 56 (2), 109–112 DOI: 10.1007/BF00349081.	5.1 / 1.2 (inte
173	acetic acid	5.2	Acids cross membranes in their neutral form and then release their proton in the cytosol.	Royce, L. A.; Boggess, E.; Fu, Y.; Liu, P.; Shanks, J. V.; Dickerson, J.; Jarboe, L. R. Transcriptomic Analysis of Carboxylic Acid Challenge in Escherichia Coli : Beyond Membrane Damage. PLOS ONE 2014, 9 (2), e89580 DOI: 10.1371/journal.pone.0089580.	5.2
174	quinine	5.2	crosses the membranes of lisosomes unprotonated, and then become protonated in it. Thus it becomes ionised and cannot escape anymore from this organelle, and it increases the pH so that proteases can't work anymore. (similar MechoA against malaria parasite)	Vallabh Minikel, E. A mechanism of action hypothesis http://www.cureffi.org/2016/05/08/a-mechanism-of-action-hypothesis/ (accessed Sep 19, 2016).	5.2
175	atrazine	#uC6 & p6.5	Plants: inhibits Protein D-1 at photosystem II => blocks production of O2 + ROS generated. Other: carcinogenic, reproductive and developmental toxicity (several MechoAs are postulated to explain these effects)	APVMA. Atrazine Toxicity: Analysis of Potential Modes of Action; APVMA, 2010; p 49. / van Rensen, J. J. S. Molecular Mechanisms of Herbicide Action near Photosystem II. Physiol. Plant. 1982, 54 (4), 515–521 DOI: 10.1111/j.1399-3054.1982.tb00719.x.	1.1 & p6.5
176	glyphosate	#uC6 & p6.7 /	Plants: inhibit aromatic amino acids synthesis by 5-enolpyruvylshikimimate-3-phosphate synthase inhibition. M: genotoxic and probable carcinogen.	Mensah, P. K.; Palmer, C. G.; Odume, O. N. Ecotoxicology of Glyphosate and Glyphosate-Based Herbicides — Toxicity to Wildlife and Humans. In Toxicity and Hazard of Agrochemicals; Larramendy, M. L., Soloneski, S., Eds.; InTech, 2015.	6.1
177	atropine	#up6.2	blocks muscarinic Ach receptors	Atropine. DrugBank; DrugBank, Ed.; 2016.	#up6.2
178	piperidine	#up6.2	M: blocks nicotinic Ach receptors, but rapidly eliminated (and partially metabolised). Directly applied, it is corrosive, and it alkalinize the cells, entering into them in their neutral form and then become protonated, all simply because it is a strong base.	Committee on Acute Exposure Guideline Levels; Committee on Toxicology; Board on Environmental Studies and Toxicology; Division on Earth and Life Studies; National Research Council. Piperidine. In Acute Exposure Guideline Levels for selected Airbone Chemicals; National Academies Press (US), 2012; Vol. 13.	1.2 & 5.2
179	amphetamine	#up6.3	M: forces release of dopamine at nerve junctions while competitively inhibiting the transport of dopamine to the cell. F: neurotox	A / Calipari, E. S.; Ferris, M. J. Amphetamine Mechanisms and Actions at the Dopamine Terminal Revisited. J. Neurosci. Off. J. Soc. Neurosci. 2013, 33 (21), 8923–8925 DOI: 10.1523/JNEUROSCI.1033-13.2013.	#up6.3
180	morphine	#up6.6	binds to some GPCRs at presynaptic and postsynaptic neurons, inhibiting the neuronal signal by upregulating the K+ channels and downregulating the Ca2+ channels.	Chahl, L. A. Experimental and Clinical Pharmacology: Opioids - Mechanisms of Action. Aust. Prescr. 1996, 19 (3), 63–65 DOI: 10.18773/austprescr.1996.063.	#up6.6
181	resmethrin	#up6.6	keeps sodium channels open in the nerves F: neurotox	A / Jackson, D.; Luukinen, B.; Buhl, K.; Stone, D. Resmethrin Technical Fact Sheet - National Pesticide Information Center http://npic.orst.edu/factsheets/archive/ResTech.html (accessed Feb 10, 2016).	2.1
182	strychnine	#up6.7	antagonist of glycine in the glycine receptor of nerve jonctions F: neurotox	A / Makarovsky, I.; Markel, G.; Hoffman, A.; Schein, O.; Brosh-Nissimov, T.; Tashma, Z.; Dushnitsky, T.; Eisenkraft, A. Strychnine - A Killer from the Past. IMAJ Isr. Med. Assoc. J. 2008, 10, 142–145.	6.7
183	2-phenylethylamine	#upC6.3 & 1.2	different neurotransmitter effects (serotoninergic and catecholaminergic), but quickly eliminated.	Sloviter, R. S.; Connor, J. D.; Drust, E. G. Serotonergic Properties of Beta-Phenethylamine in Rats. Neuropharmacology 1980, 19 (11), 1071–1074.	#up6.3
184	xanthone	6 / 1.1	? F: reactive	A	6.4
185	2,4-dinitroaniline	6 / 5.1	F: OPU (how? Maybe by coordination to heme, inhibiting them)	A / Bearden, A. P.; Schultz, T. W. Structure-Activity Relationships for Pimephales and Tetrahymena: A Mechanism of Action Approach. Environ. Toxicol. Chem. 1997, 16 (6), 1311–1317 DOI: 10.1002/etc.5620160629. / Martin, T. M.; Grulke, C. M.; Young, D. M.; Russom, C. L.; Wang, N. Y.; Jackson, C. R.; Barron, M. G. Prediction of Aquatic Toxicity Mode of Action Using Linear Discriminant and Random Forest Models. J. Chem. Inf. Model. 2013, 53 (9), 2229–2239 DOI: 10.1021/ci400267h.	5.1
186	p-(tert-butyl)-phenyl-N-methylcarbamate	6.1	AChE inhibition (reversible bonding to carbonyl). F: reactive	A / Fukuto, T. R. Mechanism of Action of Organophosphorus and Carbamate Insecticides. Environ. Health Perspect. 1990, 87, 245–254.	6.1
187	azinphos-methyl	6.1	AChE inhibition F: AChE	A / Elersek, T.; Filipic, M. Organophosphorus Pesticides - Mechanisms of Their Toxicity. In Pesticides - The Impacts of Pesticides Exposure; Stoytcheva, M., Ed.; InTech, 2011.	6.7
188	carbophenothion	6.1	AChE inhibition F: AChE	A / Elersek, T.; Filipic, M. Organophosphorus Pesticides - Mechanisms of Their Toxicity. In Pesticides - The Impacts of Pesticides Exposure; Stoytcheva, M., Ed.; InTech, 2011.	6.1
189	methomyl (lannate)	6.1	AChE inhibition F: AChE	A / Fukuto, T. R. Mechanism of Action of Organophosphorus and Carbamate Insecticides. Environ. Health Perspect. 1990, 87, 245–254.	6.1
190	sarin	6.1	Animals: AChE inhibition (covalent bond with P formed) (U&P: Probably inhibits other esterases)	Elersek, T.; Filipic, M. Organophosphorus Pesticides - Mechanisms of Their Toxicity. In Pesticides - The Impacts of Pesticides Exposure; Stoytcheva, M., Ed.; InTech, 2011. / Gonçalves, A. da S.; França, T. C. C.; Wilter, A.; Figueroa-Villar, J. D. Molecular Dynamics of the Interaction of Pralidoxime and Deazapralidoxime with Acetylcholinesterase Inhibited by the Neurotoxic Agent Tabun. J. Braz. Chem. Soc. 2006, 17 (5), 968–975 DOI: 10.1590/S0103-50532006000500022.	6.1

Table S1: training set. Part 2. (continued)

191	VX	6.1	Animals: AChE inhibition (covalent bond with P formed) (U&P: Probably inhibits other esterases)	Elersek, T.; Filipic, M. Organophosphorus Pesticides - Mechanisms of Their Toxicity. In Pesticides - The Impacts of Pesticides Exposure; Stoytcheva, M., Ed.; InTech, 2011. / Gonçalves, A. da S.; França, T. C. C.; Wilter, A.; Figueroa-Villar, J. D. Molecular Dynamics of the Interaction of Pralidoxime and Deazapralidoxime with Acetylcholinesterase Inhibited by the Neurotoxic Agent Tabun. J. Braz. Chem. Soc. 2006, 17 (5), 968–975 DOI: 10.1590/S0103-50532006000500022.	6.1
192	aldicarb	6.1	F: AChE inhibtion	A / Fukuto, T. R. Mechanism of Action of Organophosphorus and Carbamate Insecticides. Environ. Health Perspect. 1990, 87, 245–254.	6.1
193	malathion	6.1	F: AChE inhibtion	A / Elersek, T.; Filipic, M. Organophosphorus Pesticides - Mechanisms of Their Toxicity. In Pesticides - The Impacts of Pesticides Exposure; Stoytcheva, M., Ed.; InTech, 2011.	6.1
194	1-benzoylacetone	6.4	chelating Fe, degrading cytP450 F: reactive	A / O'Donoghue, J. L. Ketones of Four or Five Carbons. In Patty's Toxicology; John Wiley & Sons, Inc., 2001.	6.4
195	pentan-2,4-dione	6.4	M: chelating Fe, degrading cytP450, also reacts slowly with arginine residues to form adducts, causes inactivaton of B vitamins. F: reactive	A / O'Donoghue, J. L. Ketones of Four or Five Carbons. In Patty's Toxicology; John Wiley & Sons, Inc., 2001.	1.1
196	urethane	6.6	Frog: activates / inhibits various neurotransmitter-gated ion channels F: narcosis	A / Hara, K.; Harris, R. A. The Anesthetic Mechanism of Urethane: The Effects on Neurotransmitter-Gated Ion Channels: Anesth. Analg. 2002, 94 (2), 313–318 DOI: 10.1213/00000539-200202000-00015.	6.1
197	trifluralin	6.7	Plants: inhibit tubulin polymerisation into microtubules (playing a central role in cellular division), stops mitosis. All: disrupts Ca2+ concentration equilibria, thus disrupting a lot of processes in the cell. Very toxic to fish, and very harmless for mammals	Fernandes, T. C.; Pizano, M. A.; Marin-Morales, M. A. Characterization, Modes of Action and Effects of Trifluralin: A Review. In Herbicides - Current Research and Case Studies in Use; Price, A. J., Kelton, J. A., Eds.; InTech, 2013.	6.7
198	caffeine	6.7 & 6.6	M: CNS (antagonism of adenosine at its receptors (regulation of adenylate cyclase synthesising cAMP), cyclic nucleotide (cAMP) phosphodiesterase inhibition (=> release of Ca2+), activation of Calcium channels). (Plants: Ca2+ regulation also disrupted) F: neurotox	A / Nehlig, A.; Daval, J.-L.; Debry, G. Caffeine and the Central Nervous System: Mechanisms of Action, Biochemical, Metabolic and Psychostimulant Effects. Brain Res. Rev. 1992, 17 (2), 139–170 DOI: 10.1016/0165-0173(92)90012-B. / IAC Publishing Labs. How does caffeine affect plant growth? https://www.reference.com/science/caffeine-affect-plant-growth-7e511b9d85fbdfdb (accessed Jan 20, 2017).	6.7
199	benzoic acid	m4.1 & 5.2	M: Condensation with amino residues => very fast metabolism/elimination. Acids cross membranes in their neutral form and then release their proton in the cytosol.	Nair, B. Final Report on the Safety Assessment of Benzyl Alcohol, Benzoic Acid, and Sodium Benzoate. Int. J. Toxicol. 2001, 20 Suppl 3, 23–50.	m4.1 & 5.2
200	Acetonitrile	m4.6 & 1.1	M: cytP450 alpha-hydroxylation (cyanohydrin), then non-enzymatic degradation to give CN- and acetaldehyde / fast elimination due to high polarity, hydrophilicity. F: narcosis	A / Grogan, J.; DeVito, S. C.; Pearlman, R. S.; Korzekwa, K. R. Modeling Cyanide Release from Nitriles: Prediction of Cytochrome P 450-Mediated Acute Nitrile Toxicity. Chem. Res. Toxicol. 1992, 5 (4), 548–552 DOI: 10.1021/tx00028a014.	A.2
				A: Enoch, S. J.; Hewitt, M.; Cronin, M. T. D.; Azam, S.; Madden, J. C. Classification of Chemicals according to Mechanism of Aquatic Toxicity: An Evaluation of the Implementation of the Verhaar Scheme in Toxtree. Chemosphere 2008, 73 (3), 243–248 DOI: 10.1016/j.chemosphere.2008.06.052. (using Russom et al., 1997 and Schlutz et al., 1997 data)	

Table S2: validation set. Part 1.

N°	Molecule	SMILES	CAS	pred pKa (Chemaxon+RA)	qMaxH	qMax nonH	qMax-	aE(LUMO)	E(HOMO)
1	1,4-dioxane	O1CCOCC1	123-91-1	-3,85	0,125034	0,087588	-0,637819	-0,0241	-0,3872
2	1,2,3-trichloropropane	ClCC(Cl)CCl	96-18-4		0,214006	-0,043891	-0,298020	0,1450	-0,4399
3	1,8-cineole	CC12CCC(C(C)(C)O1)CC2	470-82-6	-4,21	0,123404	0,304737	-0,704458	0,0205	-0,3746
4	1-bromopropane	BrCCC	106-94-5		0,166851	-0,161062	-0,347493	0,1666	-0,3906
5	2-picoline	Cc1ncccc1	109-06-8	5,81	0,159632	0,275051	-0,574147	-0,0984	-0,3323
6	B-ionone	CC(=O)C=CC1=C(C)CCC(	79-77-6	-4,69	0,185444	0,509951	-0,551818	-0,1215	-0,3281
7	butan-2-one	CC(=O)CC	78-93-3	-7,26	0,158336	0,564326	-0,522312	-0,1122	-0,4027
8	diisopropyl ether	CC(C)OC(C)C	108-20-3	-4,16	0,122613	0,208845	-0,659564	0,0126	-0,3981
9	ditert-butyl sulfide	CC(C)(C)SC(C)(C)C	107-47-1	-5	0,132487	0,115354	-0,322391	0,2050	-0,3253
10	hexan-1-ol	OCCCCCC	111-27-3	-1,99	0,334745	0,124521	-0,649374	-0,0437	-0,4331
11	isopropanol	CC(O)C	67-63-0	-1,81	0,332112	0,198215	-0,644688	-0,0390	-0,4290
12	L-(-)-menthol	OC1C(C(C)C)CCC(C)C1	2216-51-5	-0,81	0,332638	0,239875	-0,663054	-0,0211	-0,4036
13	p-xylene	Cc1ccc(C)cc1	106-42-3		0,141354	-0,012487	-0,340992	0,1458	-0,3078
14	triethyleneglycol	OCCOCCOCCO	112-27-6	-2,75	0,335799	0,120001	-0,665691	0,0517	-0,4147
15	2-acetyl-1-methylpyrrole	Cn1cccc1C(=O)C	932-16-1	-3,8	0,164225	0,523149	-0,658292	-0,1316	-0,3055
16	deca-1,9-diene	C=CCCCCCCC=C	1647-16-1		0,126968	-0,101802	-0,290795	0,1805	-0,3540
17	p-tert-butylstyrene	C=Cc1ccc(C(C)(C)C)cc1	1746-23-2		0,149956	0,026076	-0,322412	0,1148	-0,2933
18	1,2-dichloropropane	ClCC(Cl)C	78-87-5		0,193636	-0,075023	-0,332986	0,1733	-0,4320
19	o-tolunitrile	N#Cc1c(C)cccc1	529-19-1	-10	0,185813	0,263604	-0,458256	0,0910	-0,3457
20	diphenyl ether	c1ccccc1Oc2ccccc2	101-84-8	-3,69	0,168819	0,370147	-0,717291	-0,0260	-0,3063
21	dodecanol	OCCCCCCCCCCCC	112-53-8	-1,99	0,334685	0,124631	-0,649502	-0,0427	-0,4166
22	dynascone	C=CCCC(=O)C1=CCCC(C	56973-85-4	-4,77	0,180704	0,553096	-0,560157	-0,1099	-0,3500
23	m-dichlorobenzene	Clc1cc(Cl)ccc1	541-73-1		0,201594	-0,080323	-0,156788	0,1155	-0,3444
24	mesityl oxide	CC(=O)C=C(C)C	141-79-7	-4,28	0,159167	0,523536	-0,568793	-0,1255	-0,3550
25	o-dichlorobenzene	Clc1c(Cl)cccc1	95-50-1		0,184019	0,037679	-0,142824	0,1177	-0,3423
26	Toluene	Cc1ccccc1	108-88-3		0,146692	-0,002129	-0,342713	0,1475	-0,3187
27	benzyl butyl dimethyl ammonium	c1ccccc1C[N+](CCCC)(C	139-07-1		0,198117	-0,046670	-0,618194	-0,0020	-0,4574
28	trimethylamine	CN(C)C	75-50-3	9,57	0,116767	-0,128529	-0,563140	-0,0025	-0,3458
29	naphthalene	c12ccccc1cccc2	91-20-3		0,151564	-0,033963	-0,157990	0,1030	-0,2869
30	p-nitrotoluene	O=[N+]([O-])c1ccc(C)cc	99-99-0	-8,82	0,217695	0,524909	-0,475041	0,0647	-0,3598
31	1-chloro-3-nitrobenzene	O=[N+]([O-])c1cc(Cl)ccc	121-73-3	-8,88	0,241025	0,529262	-0,466053	0,0483	-0,3658
32	N,N-dimethylaniline	CN(C)c1ccccc1	121-69-7	5,02	0,144941	0,399401	-0,732615	-0,0189	-0,2693
33	dichloromethane	ClCCl	75-09-2		0,21319	-0,033758	-0,358864	0,1540	-0,4473
34	trichloroethylene	ClC(Cl)=CCl	79-01-6		0,218059	0,083568	-0,202342	0,1256	-0,3718
35	hexan-2-one	CC(=O)CCCC	591-78-6	-7,27	0,155853	0,499010	-0,527698	-0,1102	-0,3989

Table S2: validation set. Part 1. (continued)

36	bromacil	CC(CC)N1C(=O)NC(C)=C	314-40-9	-4,76	0,341152	1,164697	-0,877808	-0,0508	-0,3397
37	flavone	c12ccccc1C(=O)C=C(O2	525-82-6	-5,36	0,201056	0,591198	-0,713173	-0,1092	-0,3197
38	2-chlorophenol	Oc1ccccc1Cl	95-57-8	-6,73	0,35334	0,442156	-0,636405	-0,0490	-0,3192
39	2,4,6-trimethylphenol	Oc1c(C)cc(C)cc1C	527-60-6	-5,41	0,354091	0,338600	-0,673003	-0,0355	-0,2901
40	2,4-dimethylphenol	Oc1ccc(C)cc1C	105-67-9	-5,42	0,347158	0,371204	-0,664343	-0,0354	-0,2942
41	2-bromo-3-pyridinol	n1cccc(O)c1Br	6602-32-0	1,67	0,354918	0,382296	-0,632475	-0,1242	-0,3268
42	3-methylindole	[nH]1cc(C)c2ccccc12	83-34-1	-3,8	0,308849	0,309042	-0,757653	-0,1270	-0,2750
43	m-chlorophenol	Oc1cc(Cl)ccc1	108-43-0	-6,3	0,352432	0,401187	-0,649988	0,1268	-0,3221
44	cylcohexylamine	NC1CCCCC1	108-91-8	10,45	0,260985	0,058196	-0,695017	-0,0105	-0,3719
45	(3-phenylpropyl)methylamine	c1ccccc1CCCNC	23580-89-4	10,27	0,262903	0,003721	-0,643164	-0,0036	-0,3176
46	2-methylallylamine	NCC(C)=C	2878-14-0	9,67	0,267876	0,052725	-0,702842	-0,0272	-0,3484
47	3-amino-5,6-dimethyl-1,2,4-triazine	n1nc(N)nc(C)c1C	17584-12-2	3,57	0,325595	0,808962	-0,766682	-0,0777	-0,3231
48	4-chloroaniline	Nc1ccc(Cl)cc1	106-47-8	3,49	0,288906	0,271015	-0,740331	-0,0419	-0,2954
49	4-fluoroaniline	Nc1ccc(F)cc1	371-40-4	3,8	0,303249	0,404544	-0,802234	-0,0361	-0,2811
50	4,5-dichloroguaiacol	Oc1c(OC)cc(Cl)c(Cl)c1	2460-49-3	-6,93	0,352574	0,404632	-0,646592	0,1256	-0,3094
51	3-benzyloxyaniline	Nc1cccc(OCc2ccccc2)c1	1484-26-0	3,98	0,304973	0,452544	-0,797380	-0,0273	-0,2760
52	diphenylamine	c1ccccc1Nc2ccccc2	122-39-4	0,78	0,302598	0,329654	-0,840192	-0,0275	-0,2695
53	4-nitrophenol	Oc1ccc([N+](=O)[O-])cc	100-02-7	-7,13	0,357573	0,526297	-0,639235	-0,0957	-0,3490
54	m-bromobenzamide	NC(=O)c1cc(Br)ccc1	22726-00-7	-0,43	0,308144	0,831374	-0,761462	-0,1107	-0,3397
55	dibutyl isophthalate	CCCCOC(=O)c1cc(C(=O)	3126-90-7	-6,6	0,235954	0,929254	-0,668781	-0,0963	-0,3486
56	tert-butyl acetate	CC(=O)OC(C)(C)C	540-88-5	-7,01	0,161204	0,766429	-0,659728	-0,0428	-0,4258
57	phenyl acetate	CC(=O)Oc1ccccc1	122-79-2	-7,14	0,172708	0,835784	-0,655734	-0,0630	-0,3429
58	diethyl phthalate	CCOC(=O)c1c(C(=O)OCC	84-66-2	-6,66	0,182207	0,943809	-0,632509	-0,0971	-0,3493
59	N,N-diphenylformamide	O=CN(c1ccccc1)c2ccccc	607-00-1	-4,37	0,176844	0,577438	-0,781916	-0,0491	-0,3075
60	diethyl benzylphosphonate	c1ccccc1CP(=O)(OCC)O	1080-32-6	-7,96	0,197187	1,661949	-0,755203	0,0086	-0,3245
61	vinyl acetate	CC(=O)OC=C	108-05-4	-7,12	0,192774	0,747995	-0,609283	-0,0788	-0,3622
62	diethyl tartrate	CCOC(=O)C(O)C(O)C(=O	87-91-2	-4,33	0,374033	0,884132	-0,647094	-0,0395	-0,4293
63	cyclohexene oxide	C1(O2)C2CCCC1	286-20-4	-4,2	0,138579	0,140273	-0,582674	-0,0220	-0,4092
64	a,a,a-trifluoro-m-tolualdehyde	O=Cc1cc(C(F)(F)F)ccc1	454-89-7	-7,12	0,221679	1,212625	-0,497366	-0,1621	-0,3654

Table S2: validation set. Part 1. (continued)

65	acetaldehyde	CC=O	75-07-0	-6,87	0,16073	0,416847	-0,472344	-0,1540	-0,4180
66	pentafluorobenzaldehyde	O=Cc1c(F)c(F)c(F)c(F)c1F	653-37-2	-7,59	0,151033	0,473558	-0,469574	0,0525	-0,3794
67	a,a,a',a'-tetrabromo-o-xylene	BrC(Br)c1c(C(Br)Br)cccc1	13209-15-9		0,23437	0,014005	-0,247514	0,0767	-0,3507
68	4-dimethylaminocinnamaldehyde	O=CC=Cc1ccc(N(C)C)cc1	6203-18-5	4,74	0,160447	0,411423	-0,734706	-0,0576	-0,2723
69	allyl methacrylate	C=C(C)C(=O)OCC=C	96-05-9	-6,84	0,165103	0,820060	-0,662777	-0,0980	-0,3684
70	isobutyl acrylate	C=CC(=O)OCC(C)C	106-63-8	-6,78	0,171917	0,793890	-0,658564	-0,1043	-0,3920
71	2-methyl-1,4-naphthoquinone	O=C1C(C)=CC(=O)c2ccc	58-27-5	-7,16	0,200886	0,547530	-0,543808	-0,1193	-0,3531
72	dipropyl disulfide	CCCSSCCC	629-19-6	-5	0,157484	0,019847	-0,366441	-0,0787	-0,3444
73	ditert-butyl disulfide	CC(C)(C)SSC(C)(C)C	110-06-5	-5	0,138918	0,006996	-0,326884	0,1460	-0,3235
74	p-dimethylaminobenzaldehyde	O=Cc1ccc(N(C)C)cc1	100-10-7	3,48	0,189635	0,411856	-0,733168	-0,0571	-0,2875
75	4-(diethylamino)salicylaldehyde	O=Cc1ccc(N(CC)CC)cc1	17754-90-4	-6,01	0,396557	0,457979	-0,737511	-0,0546	-0,2891
76	metolachlor	Cc1cccc(CC)c1N(C(COC)	51218-45-2	-5,88	0,192375	0,755618	-0,788192	-0,0297	-0,3371
77	but-3-yn-1-ol	OCCC#C	927-74-2	-2,21	0,335305	0,125102	-0,643783	-0,0561	-0,3837
78	p-phenylazophenol	Oc1ccc(N=Nc2ccccc2)cc	1689-82-3	0,72	0,35114	0,399710	-0,650246	-0,1339	-0,2910
79	3-bromothiophene	S1C=C(Br)C=C1	872-31-1	-4	0,200937	0,312535	-0,358642	0,1200	-0,3335
80	catechol	Oc1c(O)cccc1	120-80-9	-6,29	0,369973	0,415913	-0,690899	-0,0278	-0,2998
81	but-2-yn-1-ol	OCC#CC	764-01-2	-3,38	0,339838	0,104872	-0,622917	-0,0619	-0,3624
82	1-methyl-4-phenylpyridinium	C[n+]1ccc(cc1)c2ccccc2	36913-39-0		0,238687	0,225636	-0,642892	-0,0986	-0,4479
83	butan-2-thiol	CC(S)CC	513-53-1	-5	0,148696	-0,024312	-0,342836	0,1820	-0,3509

Table S2: validation set. Part 1. (continued)

84	2,4-dinitrophenol	Oc1c([N+](=O)[O-])cc([N	51-28-5	-8,06	0,411478	0,533782	-0,631349	0,0300	-0,3781
85	2,3,4,6-tetrachlorophenol	Oc1c(Cl)c(Cl)c(Cl)cc1Cl	58-90-2	-8,12	0,375219	0,498085	-0,625498	0,0878	-0,3413
86	3-trifluoromethyl-4-nitrophenol	Oc1cc(C(F)(F)F)c([N+](=(	88-30-2	-7,17	0,360171	1,249020	-0,636006	-0,1004	-0,3626
87	tetrachlorocatechol	Oc1c(O)c(Cl)c(Cl)c(Cl)c1	1198-55-6	-8,02	0,382445	0,443501	-0,678608	0,0980	-0,3337
88	butanoic acid	O=C(O)CCC	107-92-6	-7	0,359296	0,750622	-0,603742	-0,0811	-0,4440
89	dicofol (kelthane)	ClC(Cl)(Cl)C(O)(c1ccc(Cl)	115-32-2	-4,51	0,362018	0,360755	-0,643247	-0,0563	-0,3337
90	coniine	N1CCCCC1CCC	458-88-8	10,58	0,263988	0,071400	-0,642792	0,0152	-0,3555
91	nicotine	c1ncccc1C2CCCN2C	54-11-5	8,58	0,169127	0,144540	-0,603083	-0,0106	-0,3295
92	pentachloropyridine	n1c(Cl)c(Cl)c(Cl)c(Cl)c1C	2176-62-7	-6,43	NA	0,161586	-0,482939	0,0624	-0,3654
93	carbaryl (sevin)	CNC(=O)Oc1cccc2ccccc	63-25-2	-3	0,319096	1,048355	-0,736695	-0,0294	-0,2842
94	tabun	N#CP(=O)(OCC)N(C)C	77-81-6	-1,8	0,177036	1,579839	-0,756128	-0,0442	-0,4034
95	disulfoton	CCOP(=S)(OCC)SCCSCC	298-04-4	-5	0,204408	1,168454	-0,739663	-0,0257	-0,3338
96	aminocarb	CNC(=O)Oc1cc(C)c(N(C)	2032-59-9	4,57	0,319227	1,025751	-0,729874	-0,0062	-0,3149
97	serotonin	NCCc(cn1)c2cc(O)ccc12	50-67-9	10	0,344611	0,370670	-0,769941	-0,0200	-0,2780
98	clomazone	Clc1ccccc1CN2OCC(C)(	81777-89-1	-4,67	0,197264	0,809032	-0,593665	-0,0093	-0,3315
99	4-methoxyphenol	Oc1ccc(OC)cc1	150-76-5	-4,81	0,345964	0,424477	-0,667309	-0,0218	-0,2875
100	propionitrile	CCC#N	107-12-0	-10	0,175874	0,291867	-0,459729	0,1955	-0,4580
101	2,2,2-trichloroethanol	OCC(Cl)(Cl)Cl	115-20-8	-3,89	0,356102	0,141520	-0,616558	0,1255	-0,4502

Table S2: validation set. Part 2.

N°	Molecule	literature MechoA	MechoA detail	literature MechoA reference	pred MechoA
1	1,4-dioxane	1.1	(M: metabolised into (2-hydroxyethoxy)-acetic acid). F: narcosis	A / ATSDR. Toxicological Profile: 1,4-Dioxane http://www.atsdr.cdc.gov/ToxProfiles/tp.asp?id=955&tid=199 (accessed Jul 22, 2015).	1.1
2	1,2,3-trichloropropane	1.1	F: narcosis	A	1.1
3	1,8-cineole	1.1	F: narcosis	A	1.1
4	1-bromopropane	1.1	F: narcosis	A	1.1
5	2-picoline	1.1	F: narcosis	A	1.1
6	B-ionone	1.1	F: narcosis	A	1.1
7	butan-2-one	1.1	F: narcosis	A	1.1
8	diisopropyl ether	1.1	F: narcosis	A	1.1
9	ditert-butyl sulfide	1.1	F: narcosis	A	1.1
10	hexan-1-ol	1.1	F: narcosis	A	1.1
11	isopropanol	1.1	F: narcosis	A	1.1
12	L-(-)-menthol	1.1	F: narcosis	A	1.1
13	p-xylene	1.1	F: narcosis	A	1.1
14	triethyleneglycol	1.1	F: narcosis. (M: metabolised into more toxic compound such as oxalic acid, ethylene dioxyacetic acid, and possibly ethylene glycol (leading to glycoaldehyde, glycolate, glyoxylate and oxalate. "Theses metabolites are OPU, inhibit glucose and serotonin metabolism, protein synthesis, DNA replication, ribosomal RNA formation." These effecs may all be due to acidosis generated, and oxalate crystallising with calcium, leading to deposits blocking circulation etc. and eventually hypocalcemia))	A / Triethylene Glycol [MAK Value Documentation, 2007]. In The MAK-Collection for Occupational Health and Safety; Wiley-VCH Verlag GmbH & Co. KGaA, 2002.	1.1
15	2-acetyl-1-methylpyrrole	1.1	F: reactive	A	1.1
16	deca-1,9-diene	1.1	F: reactive	A	1.1
17	p-tert-butylstyrene	1.1	F: reactive	A	1.1
18	1,2-dichloropropane	1.1	M: no specific effect. F: narcosis	A / ATSDR. Toxicological Profile: 1,2-Dichloropropane http://www.atsdr.cdc.gov/toxprofiles/tp.asp?id=831&tid=162 (accessed Jul 16, 2015).	1.1
19	o-tolunitrile	1.1	M: There is no H in alpha to the nitrile, so cytP450 can't attack it to generate CN- => just narcosis. F: narcosis	A / Grogan, J.; DeVito, S. C.; Pearlman, R. S.; Korzekwa, K. R. Modeling Cyanide Release from Nitriles: Prediction of Cytochrome P 450-Mediated Acute Nitrile Toxicity. Chem. Res. Toxicol. 1992, 5 (4), 548–552 DOI: 10.1021/tx00028a014.	1.1
20	diphenyl ether	1.1	No information on toxicity mechanism. M: not highly toxic F: narcosis	A / HSDB. DIPHENYL ETHER https://toxnet.nlm.nih.gov/cgi-bin/sis/search/a?dbs+hsdb:@term+@DOCNO+934 (accessed Jan 20, 2017).	1.1
21	dodecanol	1.1			1.1
22	dynascone	1.1		Based on experimental ecotoxicity values, compared to the water solubility of the substance with proprietary QSAR models, it could be assigned a probable mode of action.	1.1
23	m-dichlorobenzene	1.1			1.1
24	mesityl oxide	1.1			1.1
25	o-dichlorobenzene	1.1			1.1
26	Toluene	1.1			1.1
27	benzyl butyl dimethyl ammonium	1.1 & #up6.1	B: disturbance of membrane / for certain ammonium (better with an ester-like residue) AChE inhibition. Without specific group (like phosphoester), it's a reversible inhibition	Čolović, M. B., Krstić, D. Z., Lazarević-Pašti, T. D., Bondžić, A. M. & Vasić, V. M. Acetylcholinesterase Inhibitors: Pharmacology and Toxicology. Curr. Neuropharmacol. 11, 315–335 (2013).	1.1
28	trimethylamine	1.1 & 5.2		Vallabh Minikel, E. A mechanism of action hypothesis http://www.cureffi.org/2016/05/08/a-mechanism-of-action-hypothesis/ (accessed Sep 19, 2016).	1.1
29	naphthalene	1.1 & m4.3	M: cytP, gives epoxide and quinones in a few hours, adducts to proteins and DNA. Aq: narcotic	Tsuruda, L. S.; Lamé, M. W.; Jones, A. D. Formation of Epoxide and Quinone Protein Adducts in B6C3F1 Mice Treated with Naphthalene, Sulfate Conjugate of 1,4-Dihydroxynaphthalene and 1,4-Naphthoquinone. Arch. Toxicol. 1995, 69 (6), 362–367 DOI: 10.1007/s002040050185. / Buckpitt, A.; Boland, B.; Isbell, M.; Morin, D.; Shultz, M.; Baldwin, R.; Chan, K.; Karlsson, A.; Lin, C.; Taff, A.; West, J.; Fanucchi, M.; Van Winkle, L.; Plopper, C. Naphthalene-Induced Respiratory Tract Toxicity: Metabolic Mechanisms of Toxicity. Drug Metab. Rev. 2002, 34 (4), 791–820 DOI: 10.1081/DMR-120015694.	1.1
30	p-nitrotoluene	1.1 & m4.3	M: metabolism into nitroso, results in methaemoglobinaemia => lack of O2 supply.	IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.1 & m4.3
31	1-chloro-3-nitrobenzene	1.1 & m4.3	M: metabolism into nitroso, results in methaemoglobinaemia => lack of O2 supply. F: narcosis	A / IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.1 & m4.3
32	N,N-dimethylaniline	1.1 & m4.3	M: metabolism mainly into N-oxide, less toxic compound, but also after demethylation into nitroso, results in methaemoglobinaemia => lack of O2 supply. F: narcosis	A / IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.1
33	dichloromethane	1.1 & m4.6	M: metabolisation into CO with effects apparent in a few minutes to hours (depending on the dose) => fixing to haemoglobin, diminished O2 supply (coma then death) / directly CNS (not understood). F: narcosis	A / ATSDR. Medical Management Guidelines (MMGs): Methylene Chloride. October 21, 2014.	1.1
34	trichloroethylene	1.1 & mC4.3	M: chronic activation by GST gives reactive metabolites (likely dichlorovinyl-cysteine) which cause DNA mutation leading to renal cancer. F: narcosis	A / Brüning, T.; Bolt, H. M. Renal Toxicity and Carcinogenicity of Trichloroethylene: Key Results, Mechanisms, and Controversies. Crit. Rev. Toxicol. 2000, 30 (3), 253–285 DOI: 10.1080/10408440091159202.	1.1
35	hexan-2-one	1.1 & mC4.3	M: chronic CNS seizure: after metabolisation into hexan-2,5-dione, pyrrole formation with a lysine residue, oxidation and crosslinking with another protein. F: narcosis	A / ATSDR. Toxicological Profile: N-Hexane. July 1999.	1.1

Table S2: validation set. Part 2. (continued)

36	bromacil	1.1 & p6.5	F: narcosis, rapidly eliminated with almost no metabolism. Plants: inhibits Protein D-1 at photosystem II => blocks production of O2 + ROS generated.	A / Call, D. J.; Brooke, L. T.; Kent, R. J.; Knuth, M. L.; Poirier, S. H.; Huot, J. M.; Lima, A. R. Bromacil and Diuron Herbicides: Toxicity, Uptake, and Elimination in Freshwater Fish. Arch. Environ. Contam. Toxicol. 1987, 16 (5), 607–613 DOI: 10.1007/BF01055817. / Dube, S.; Lesoli, M. S.; Fatunbi, A. O. The Efficacy and Safety of Bromacil Based Herbicide for the Control of the Invasive Bush Species in South African Rangelands. Afr. J. Biotechnol. 2009, 8 (9), 1776–1781 DOI: 10.5897/AJB2009.000-9253.	1.1 & p6.5
37	flavone	1.1 / 6 / 3.2	F: narcosis	A / IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.1
38	2-chlorophenol	1.2	F,Pz: narcosis 2	A	5.1 / 1.2 (inte
39	2,4,6-trimethylphenol	1.2	F,Pz: narcosis 2	A	1.2
40	2,4-dimethylphenol	1.2	F,Pz: narcosis 2	A	1.2
41	2-bromo-3-pyridinol	1.2	F: narcosis 2	A	3.1 & m4.1
42	3-methylindole	1.2	F: reactive	A	1.1
43	m-chlorophenol	1.2	Pz: narcosis 2	A	1.2
44	cylcohexylamine	1.2 & 5.2	no special effects, no rapid metabolism. Directly applied, it is corrosive, and it alkalinize the cells, entering into them in their neutral form and then become protonated, all simply because it is a strong base.	Cyclohexylamine [MAK Value Documentation, 2006]. In The MAK-Collection for Occupational Health and Safety; Wiley-VCH Verlag GmbH & Co. KGaA, 2002. / Vallabh Minikel, E. A mechanism of action hypothesis http://www.cureffi.org/2016/05/08/a-mechanism-of-action-hypothesis/ (accessed Sep 19, 2016).	1.2 & 5.2
45	(3-phenylpropyl)methylamine	1.2 & 5.2		Vallabh Minikel, E. A mechanism of action hypothesis http://www.cureffi.org/2016/05/08/a-mechanism-of-action-hypothesis/ (accessed Sep 19, 2016).	#up6.3
46	2-methylallylamine	1.2 & 5.2 / 3.1		Vallabh Minikel, E. A mechanism of action hypothesis http://www.cureffi.org/2016/05/08/a-mechanism-of-action-hypothesis/ (accessed Sep 19, 2016).	1.2 & 5.2
47	3-amino-5,6-dimethyl-1,2,4-triazine	1.2 & m4.3	M: metabolism into nitroso, results in methaemoglobinaemia => lack of O2 supply. F: narcosis 2	A / IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.1 & p6.5
48	4-chloroaniline	1.2 & m4.3	M: metabolism into nitroso, results in methaemoglobinaemia => lack of O2 supply. F: narcosis 2	A / IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.2 & 5.2
49	4-fluoroaniline	1.2 & m4.3	M: metabolism into nitroso, results in methaemoglobinaemia => lack of O2 supply. F: reactive	A / IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.2
50	4,5-dichloroguaiacol	1.2 & m4.3	Ox to corresponding catechol? F: narcosis	A / Riley, P. A. Mechanism of Pigment-Cell Toxicity Produced by Hydroxyanisole. J. Pathol. 1970, 101 (2), 163–169 DOI: 10.1002/path.1711010211.	1.2
51	3-benzyloxyaniline	1.2 & m4.3 / 3	M: metabolism into nitroso, results in methaemoglobinaemia => lack of O2 supply / 3-aminophenolate as a good leaving group F: narcosis 2	A / IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.2
52	diphenylamine	1.2 & m4.6 / r	M: OPU, Metabolism gives hydroxylated products and indophenol (quinoneimine). F: narcosis	A / IPCS Inchem. Diphenylamine (Addendum). In JMPR Evaluations; Rome, 1998. / IARC. General Discussion of Common Mechanisms for Aromatic Amines. IARC Monogr. Eval. Carcinog. Risks Hum. 2010, 99, 41–54.	1.2
53	4-nitrophenol	1.2 / 5.1	Certainly a combination of both effects! F: narcosis 2 Pz: soft elec	A / Schultz, T. W. The Use of the Ionization Constant (pKa) in Selecting Models of Toxicity in Phenols. Ecotoxicol. Environ. Saf. 1987, 14 (2), 178–183.	1.2 & m4.3
54	m-bromobenzamide	2.1	F: narcosis	A	2.1
55	dibutyl isophthalate	2.1	F: narcosis. D,A: narcosis?	A / Jaworska, J. S.; Hunter, R. S.; Schultz, T. W. Quantitative Structure-Toxicity Relationships and Volume Fraction Analyses for Selected Esters. Arch. Environ. Contam. Toxicol. 1995, 29 (1), 86–93 DOI: 10.1007/BF00213091.	2.1
56	tert-butyl acetate	2.1	M: really hydrolized in the organism F: narcosis	A	2.1
57	phenyl acetate	2.1			2.1
58	diethyl phthalate	2.1 & mC2.2	M: 1 ester is mainly hydrolysed. The monoester induces carcinogenesis. F: narcosis. D,A: narcosis?	A	2.1
59	N,N-diphenylformamide	2.1 / 4.3	F: narcosis	A	2.1
60	diethyl benzylphosphonate	2.1 / 6.1	? M: Can disturb so various targets, most of them acting in the signaling pathways (inhibit AChE, carboxylases, DAG-lipase indirectly, PKC (probably indirect by ROS generation), steroid androgen receptor, activate Ca2+/cAMP response element binding protein). F: narcosis	A	6.1
61	vinyl acetate	2.2	after ester hydrolysis, acetaldehyde is generated, giving Schiff-base adducts. (Aqua: more toxic than normal esters)		2.1
62	diethyl tartrate	2.2	tartaric acid inhibits fumarase thus stopping the Krebs cycle.	Shaw, W. Chapter 3 Organic Acid Testing, Byproducts of Yeast and Their Relationship to Autism. In Biological Treatments for Autism and PDD. Causes and Biomedical Therapies for Autism & PDD.; 2002.	2.1
63	cyclohexene oxide	3.1	M: AN	Aptula, A. O.; Roberts, D. W. Mechanistic Applicability Domains for Nonanimal-Based Prediction of Toxicological End Points: General Principles and Application to Reactive Toxicity. Chem. Res. Toxicol. 2006, 19 (8), 1097–1105 DOI: 10.1021/tx0601004.	1.1
64	a,a,a-trifluoro-m-tolualdehyde	3.1	M: Schiff-base formation. F: reactive	A / Aptula, A. O.; Roberts, D. W. Mechanistic Applicability Domains for Nonanimal-Based Prediction of Toxicological End Points: General Principles and Application to Reactive Toxicity. Chem. Res. Toxicol. 2006, 19 (8), 1097–1105 DOI: 10.1021/tx0601004.	3.2

Table S2: validation set. Part 2. (continued)

65	acetaldehyde	3.1	M: Schiff-base formation. F: reactive	A / Aptula, A. O.; Roberts, D. W. Mechanistic Applicability Domains for Nonanimal-Based Prediction of Toxicological End Points: General Principles and Application to Reactive Toxicity. Chem. Res. Toxicol. 2006, 19 (8), 1097–1105 DOI: 10.1021/tx0601004.	3.1 (Schiff ba
66	pentafluorobenzaldehyde	3.1	Schiff-base formation F: reactive	A / Aptula, A. O.; Roberts, D. W. Mechanistic Applicability Domains for Nonanimal-Based Prediction of Toxicological End Points: General Principles and Application to Reactive Toxicity. Chem. Res. Toxicol. 2006, 19 (8), 1097–1105 DOI: 10.1021/tx0601004.	F3.1/3.2
67	a,a,a',a'-tetrabromo-o-xylene	3.1 & 4.1	Br- as a good leaving group F: reactive	A / α-Chlorinated Toluenes [MAK Value Documentation, 1993]. In The MAK-Collection for Occupational Health and Safety; Wiley-VCH Verlag GmbH & Co. KGaA, 2002.	1.1
68	4-dimethylaminocinnamaldehyde	3.2	Michaël F: reactive	A / Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037.	1.1
69	allyl methacrylate	3.2	Michaël F: reactive	A / Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037.	3.2
70	isobutyl acrylate	3.2	Michaël F: reactive	A / Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037.	3.2
71	2-methyl-1,4-naphthoquinone	3.2 & 4.4	Quinone-semiquinone-hydroquinone cycle F: reactive	A / Bolton, J. L.; Trush, M. A.; Penning, T. M.; Dryhurst, G.; Monks, T. J. Role of Quinones in Toxicology. Chem. Res. Toxicol. 2000, 13 (3), 135–160 DOI: 10.1021/tx9902082.	1.1
72	dipropyl disulfide	3.3	M: homolytic cleavage F: reactive	A / Morgott, D.; Lewis, C.; Bootman, J.; Banton, M. Disulfide Oil Hazard Assessment Using Categorical Analysis and a Mode of Action Determination. Int. J. Toxicol. 2014, 33 (1 Suppl), 181S–198S DOI: 10.1177/1091581813504227. / Kazius, J.; McGuire, R.; Bursi, R. Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem. 2005, 48 (1), 312–320 DOI: 10.1021/jm040835a.	3.3
73	ditert-butyl disulfide	3.3 / 1.1	M: homolytic cleavage (maybe not much with tertiary disulfides). F: narcosis	A / Morgott, D.; Lewis, C.; Bootman, J.; Banton, M. Disulfide Oil Hazard Assessment Using Categorical Analysis and a Mode of Action Determination. Int. J. Toxicol. 2014, 33 (1 Suppl), 181S–198S DOI: 10.1177/1091581813504227. / Kazius, J.; McGuire, R.; Bursi, R. Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem. 2005, 48 (1), 312–320 DOI: 10.1021/jm040835a.	3.3
74	p-dimethylaminobenzaldehyde	4.1 & 3.1	AN. F: narcosis	A / Benzaldehyde [MAK Value Documentation, 2002]. In The MAK-Collection for Occupational Health and Safety; Wiley-VCH Verlag GmbH & Co. KGaA, 2002; Vol. 17, pp 13–36. / IPCS Inchem. Vanillin SIDS http://www.inchem.org/documents/sids/sids/121335.pdf (accessed Jul 23, 2015).	F3.1/3.2
75	4-(diethylamino)salicylaldehyde	4.1 & 3.1	oxidation of aldehyde to acid? F,Pz: narcosis	A / Benzaldehyde [MAK Value Documentation, 2002]. In The MAK-Collection for Occupational Health and Safety; Wiley-VCH Verlag GmbH & Co. KGaA, 2002; Vol. 17, pp 13–36. / IPCS Inchem. Vanillin SIDS http://www.inchem.org/documents/sids/sids/121335.pdf (accessed Jul 23, 2015).	3.1 & m4.1
76	metolachlor	4.1 & p6.7 & 3	Conjugation with thiol groups (including CoA, inhibiting all the processes using it), not precisely known... Plants: makes it a good herbicide. Plants & Other: detoxification with GSH.	Fuerst, E. P. Understanding the Mode of Action of the Chloroacetamide and Thiocarbamate Herbicides. Weed Technol. 1987, 1 (4), 270–277. / Rivard, L. Environmental Fate of Metolachlor. Chemistry 2003, 51218, 45–2.	6.7
77	but-3-yn-1-ol	4.3	Frog: Reacts with the O of the heme of cytP to form adducts to the heme, inhibiting the cytP. F: reactive	A / Smith, G. F. Designing Drugs to Avoid Toxicity. In Progress in Medicinal Chemistry; Witty, G. L. and D. R., Ed.; Elsevier, 2011; Vol. 50, pp 1–47.	1.1
78	p-phenylazophenol	4.3	M: diazo bond cleaved (metabolism mainly by anaerobic intestinal bacteria), generating anilines. Then classical metabolism to nitroso. F: uncoupler	A / Brown, M. A.; Vito, S. C. D. Predicting Azo Dye Toxicity. Crit. Rev. Environ. Sci. Technol. 1993, 23 (3), 249–324 DOI: 10.1080/10643389309388453.	3.1 & m4.1
79	3-bromothiophene	4.3	M: same as furan. Epoxidation by cytP450, followed by adducts formation. F: reactive	A / Smith, G. F. Designing Drugs to Avoid Toxicity. In Progress in Medicinal Chemistry; Witty, G. L. and D. R., Ed.; Elsevier, 2011; Vol. 50, pp 1–47.	1.1
80	catechol	4.3 & 4.4 / 1.2	M: Ox to o-quinone by tyrosinase. F: narcosis 2. Pz: pro-elec	A / Bonnard, N.; Pillière, F.; Protois, J.-C.; Schneider, O. Fiche toxicologique de l'hydroquinone. INRS 2006. / Bolton, J. L.; Trush, M. A.; Penning, T. M.; Dryhurst, G.; Monks, T. J. Role of Quinones in Toxicology. Chem. Res. Toxicol. 2000, 13 (3), 135–160 DOI: 10.1021/tx9902082.	#up6.6
81	but-2-yn-1-ol	4.3 / 6.7	F: reactive (ox to acrolein => adducts?) / M: CNS (seems to block the release of ACh in the nerve junction, but how? ...) (sedative, hypnotic) but then easily eliminated	A / Koleva, Y.; Barzilov, I. Comparative Study of Mechanism of Action of Allyl Alcohols for Different Endpoints. 2010, 49, 55–59.	4.3
82	1-methyl-4-phenylpyridinium	4.4	M,P: RedOx cycling	Cohen, G. M.; d'Arcy Doherty, M. Free Radical Mediated Cell Toxicity by Redox Cycling Chemicals. Br. J. Cancer. Suppl. 1987, 8, 46–52.	1.1
83	butan-2-thiol	4.4	M: 1° & 2° thiols, in their anionic form, are oxidized into thiyl radical and then enter a redox cycle between the thiyl radical and the disulfides.	Munday, R. Toxicity of Thiols and Disulphides: Involvement of Free-Radical Species. Free Radic. Biol. Med. 1989, 7 (6), 659–673 DOI: 10.1016/0891-5849(89)90147-0.	1.1

Table S2: validation set. Part 2. (continued)

84	2,4-dinitrophenol	5.1	F,Pz: OPU (H+)	A / Schultz, T. W. The Use of the Ionization Constant (pKa) in Selecting Models of Toxicity in Phenols. Ecotoxicol. Environ. Saf. 1987, 14 (2), 178–183.	5.1
85	2,3,4,6-tetrachlorophenol	5.1	F: narcosis Pz: resp unc for 2,3,5,6- and 2,3,4,5-	A / Schultz, T. W. The Use of the Ionization Constant (pKa) in Selecting Models of Toxicity in Phenols. Ecotoxicol. Environ. Saf. 1987, 14 (2), 178–183.	5.1
86	3-trifluoromethyl-4-nitrophenol	5.1 / 1.2	F: narcosis 2	A / Schultz, T. W. The Use of the Ionization Constant (pKa) in Selecting Models of Toxicity in Phenols. Ecotoxicol. Environ. Saf. 1987, 14 (2), 178–183.	1.2
87	tetrachlorocatechol	5.1 / 4.4	F: uncoupler Pz: proelec	A / Schultz, T. W. The Use of the Ionization Constant (pKa) in Selecting Models of Toxicity in Phenols. Ecotoxicol. Environ. Saf. 1987, 14 (2), 178–183. / Bonnard, N.; Pillière, F.; Protois, J.-C.; Schneider, O. Fiche toxicologique de l'hydroquinone. INRS 2006. / Bolton, J. L.; Trush, M. A.; Penning, T. M.; Dryhurst, G.; Monks, T. J. Role of Quinones in Toxicology. Chem. Res. Toxicol. 2000, 13 (3), 135–160 DOI: 10.1021/tx9902082.	5.1 / 1.2 (inte
88	butanoic acid	5.2	Acids cross membranes in their neutral form and then release their proton in the cytosol.	Royce, L. A.; Boggess, E.; Fu, Y.; Liu, P.; Shanks, J. V.; Dickerson, J.; Jarboe, L. R. Transcriptomic Analysis of Carboxylic Acid Challenge in Escherichia Coli : Beyond Membrane Damage. PLOS ONE 2014, 9 (2), e89580 DOI: 10.1371/journal.pone.0089580.	F1.1/4.1
89	dicofol (kelthane)	#p6.7	M: thyroid hormone antagonist, weak estrogen agonist, potent inducer of CYP2B, weak aromatase inhibitor, some ability to bind with andogen receptors F: neurotox	A / Sparling, D. W.; Linder, G.; Bishop, C. A.; Krest, S. Ecotoxicology of Amphibians and Reptiles, Second Edition; CRC Press, 2010.	1.2
90	coniine	#up6.2	blocks nicotinic Ach receptors	National Center for Biotechnology Information. CONIINE | C8H17N https://pubchem.ncbi.nlm.nih.gov/compound/441072 (accessed Dec 21, 2016). / Vetter, J. Poison Hemlock (Conium Maculatum L.). Food Chem. Toxicol. 2004, 42 (9), 1373–1382 DOI: 10.1016/j.fct.2004.04.009.	1.2 & 5.2
91	nicotine	#up6.2	M: agonism of nicotinic Ach receptors, leading to a desensitization of them. F: neurotox	A / Narahashi, T.; Fenster, C. P.; Quick, M. W.; Lester, R. a. J.; Marszalec, W.; Aistrup, G. L.; Sattelle, D. B.; Martin, B. R.; Levin, E. D. Symposium Overview: Mechanism of Action of Nicotine on Neuronal Acetylcholine Receptors, from Molecule to Behavior. Toxicol. Sci. 2000, 57 (2), 193–202 DOI: 10.1093/toxsci/57.2.193.	1.1
92	pentachloropyridine	6 / 5.1	F: OPU (how? Maybe by coordination to heme, inhibiting them)	A / Bearden, A. P.; Schultz, T. W. Structure-Activity Relationships for Pimephales and Tetrahymena: A Mechanism of Action Approach. Environ. Toxicol. Chem. 1997, 16 (6), 1311–1317 DOI: 10.1002/etc.5620160629. / Martin, T. M.; Grulke, C. M.; Young, D. M.; Russom, C. L.; Wang, N. Y.; Jackson, C. R.; Barron, M. G. Prediction of Aquatic Toxicity Mode of Action Using Linear Discriminant and Random Forest Models. J. Chem. Inf. Model. 2013, 53 (9), 2229–2239 DOI: 10.1021/ci400267h.	UC
93	carbaryl (sevin)	6.1	(A)ChE inhibition F: AChE	A / Fukuto, T. R. Mechanism of Action of Organophosphorus and Carbamate Insecticides. Environ. Health Perspect. 1990, 87, 245–254.	6.1
94	tabun	6.1	AChE inhibition	Elersek, T.; Filipic, M. Organophosphorus Pesticides - Mechanisms of Their Toxicity. In Pesticides - The Impacts of Pesticides Exposure; Stoytcheva, M., Ed.; InTech, 2011. / Gonçalves, A. da S.; França, T. C. C.; Wilter, A.; Figueroa-Villar, J. D. Molecular Dynamics of the Interaction of Pralidoxime and Deazapralidoxime with Acetylcholinesterase Inhibited by the Neurotoxic Agent Tabun. J. Braz. Chem. Soc. 2006, 17 (5), 968–975 DOI: 10.1590/S0103-50532006000500022.	6.1
95	disulfoton	6.1	AChE inhibition F: AChE	A / Elersek, T.; Filipic, M. Organophosphorus Pesticides - Mechanisms of Their Toxicity. In Pesticides - The Impacts of Pesticides Exposure; Stoytcheva, M., Ed.; InTech, 2011.	6.1
96	aminocarb	6.1	M,F: AChE inhibtion	A / Fukuto, T. R. Mechanism of Action of Organophosphorus and Carbamate Insecticides. Environ. Health Perspect. 1990, 87, 245–254.	6.1
97	serotonin	6.7 & 4.1	endogenous neurotransmitter. In reasonable exogenous quantities it is metabolised following the normal pathway.	King, M. W. Biochemistry of Neurotransmitters and Nerve Transmission http://themedicalbiochemistrypage.org/nerves.php (accessed Sep 25, 2015). / Frazer, A.; Hensler, J. G. Serotonin. Basic Neurochemistry: Molecular, Cellular and medical aspects.; Siegel GJ, Agranoff BW, Albers RW, et al., 1999.	1.2
98	clomazone	6.7 / 6.1	Plants: inhibit carotenoid biosynthesis (essential for photosynthesis). Other: slight AChE inhibition effect	Van Scoy, A. R.; Tjeerdema, R. S. Environmental Fate and Toxicology of Clomazone. Rev. Environ. Contam. Toxicol. 2014, 229, 35–49 DOI: 10.1007/978-3-319-03777-6_3.	2.1
99	4-methoxyphenol	m4.3 & 1.2	M: 2-hydroxylation and then ox to o-quinone. F,Pz: narcosis 2	A / Moridani, M. Y.; Cheon, S. S.; Khan, S.; O'Brien, P. J. Metabolic Activation of 3-Hydroxyanisole by Isolated Rat Hepatocytes. Chem. Biol. Interact. 2003, 142 (3), 317–333 DOI: 10.1016/S0009-2797(02)00125-4.	1.2 & m4.3
100	propionitrile	m4.6 & 1.1	M: release of CN- F: narcosis	A / Grogan, J.; DeVito, S. C.; Pearlman, R. S.; Korzekwa, K. R. Modeling Cyanide Release from Nitriles: Prediction of Cytochrome P 450-Mediated Acute Nitrile Toxicity. Chem. Res. Toxicol. 1992, 5 (4), 548–552 DOI: 10.1021/tx00028a014.	m4.6 & 1.1
101	2,2,2-trichloroethanol	m6.6 & 1.1	M: Activate GABA receptor (Cl- channel) (=metabolite of chloral=2,2,2-trichloroacetaldehyde/chloral hydrate). F: narcosis	A / Peoples, R. W.; Weight, F. F. Trichloroethanol Potentiation of Gamma-Aminobutyric Acid-Activated Chloride Current in Mouse Hippocampal Neurones. Br. J. Pharmacol. 1994, 113 (2), 555–563.	6.1

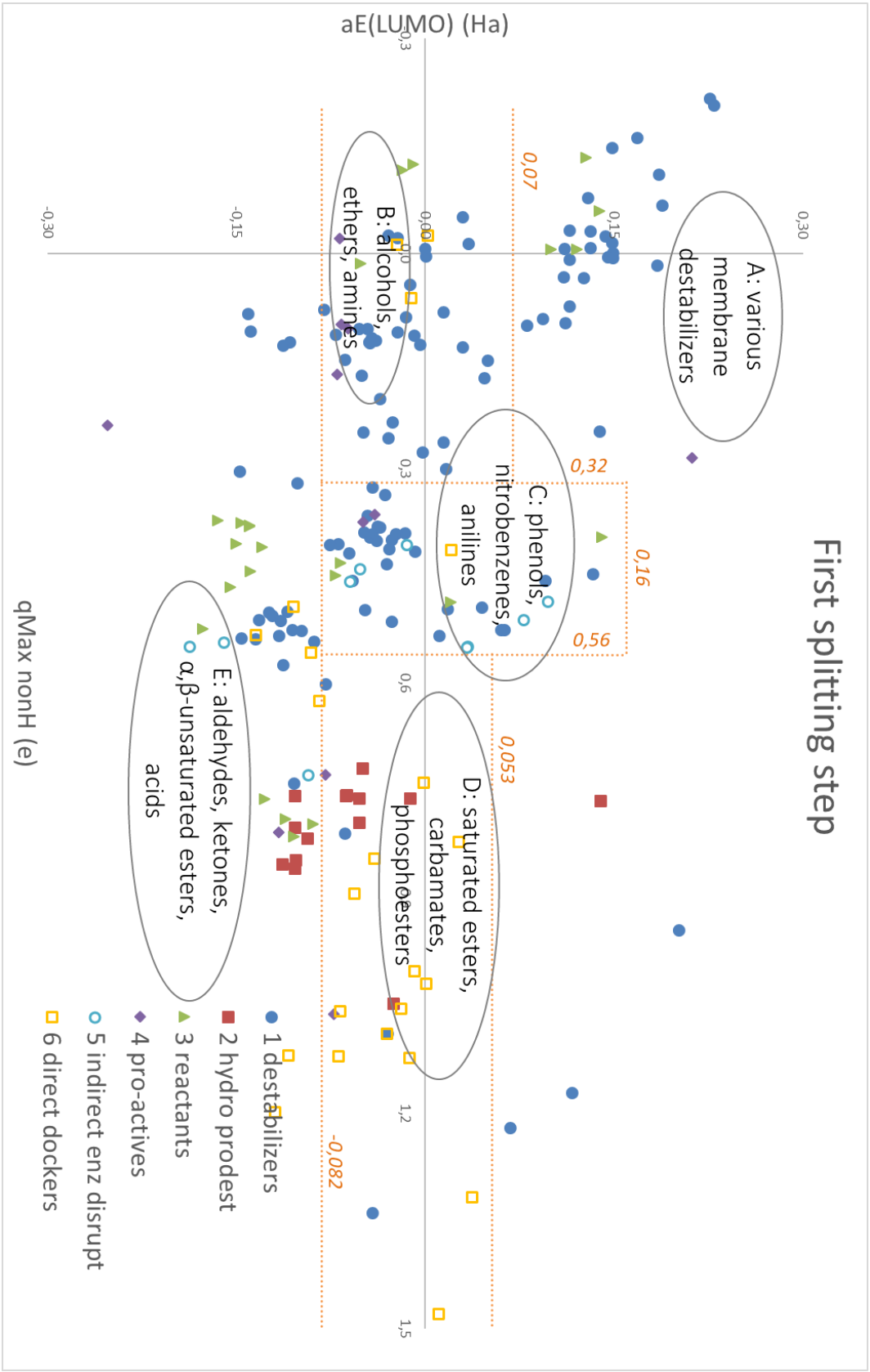


Figure S1: First split.

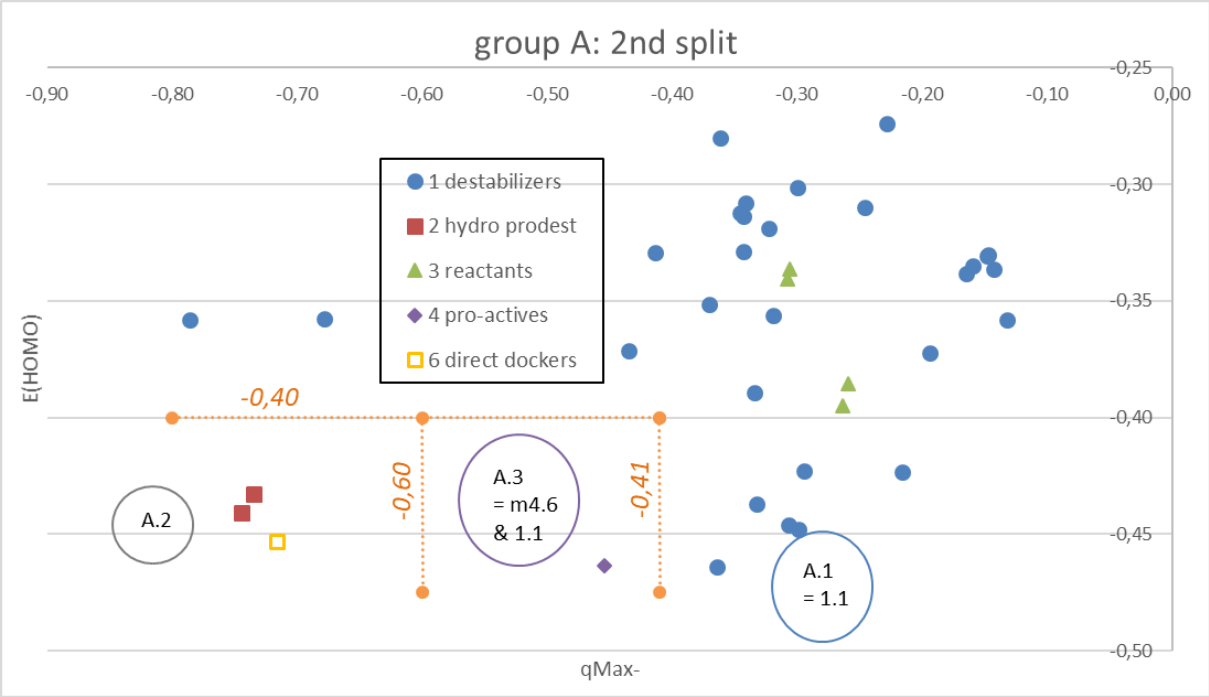


Figure S2: Second splits, group A.

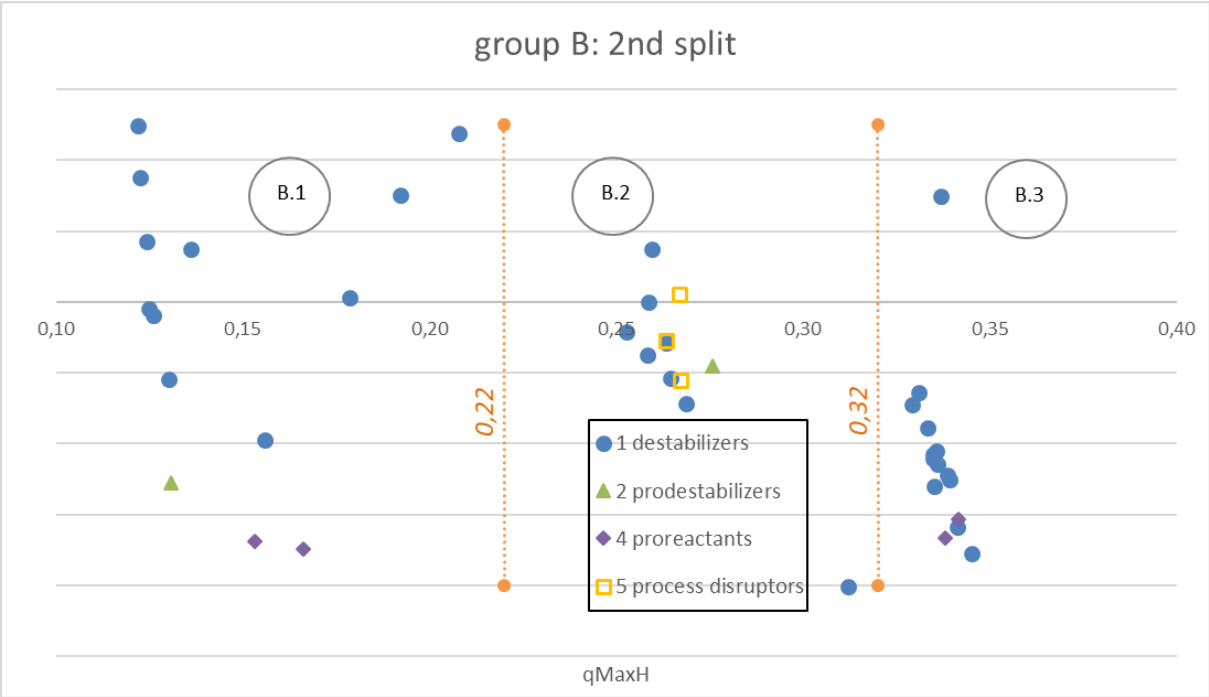


Figure S3: Second splits, group B.

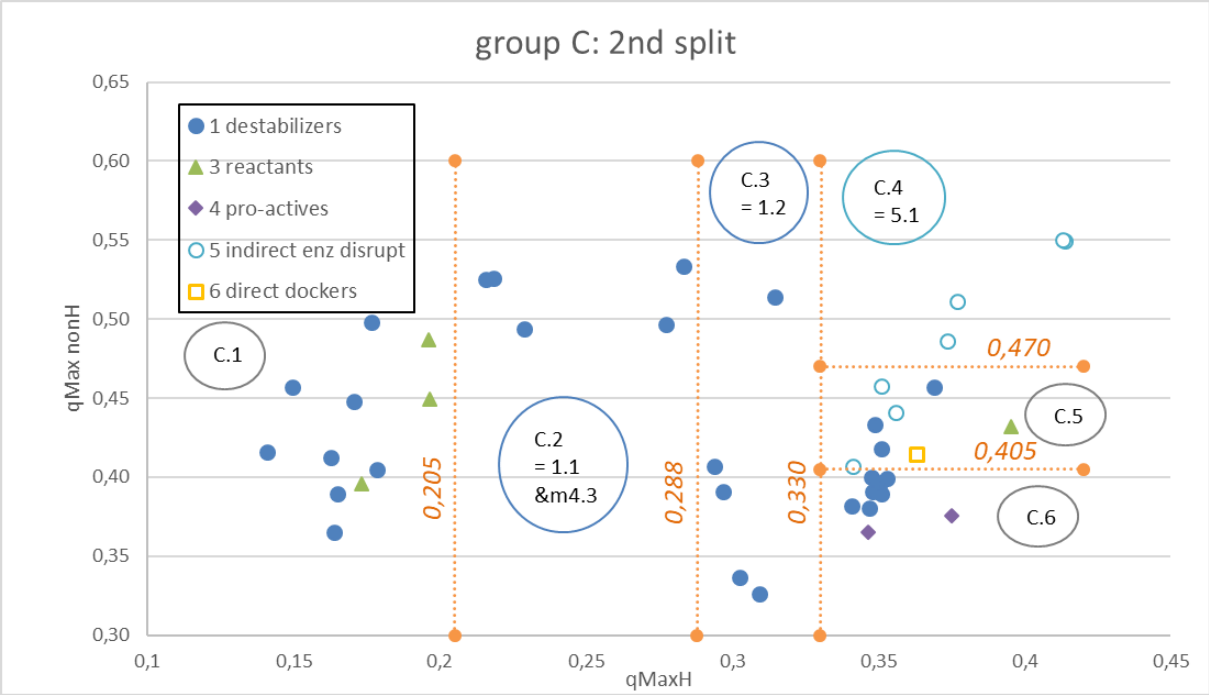


Figure S4: Second splits, group C.

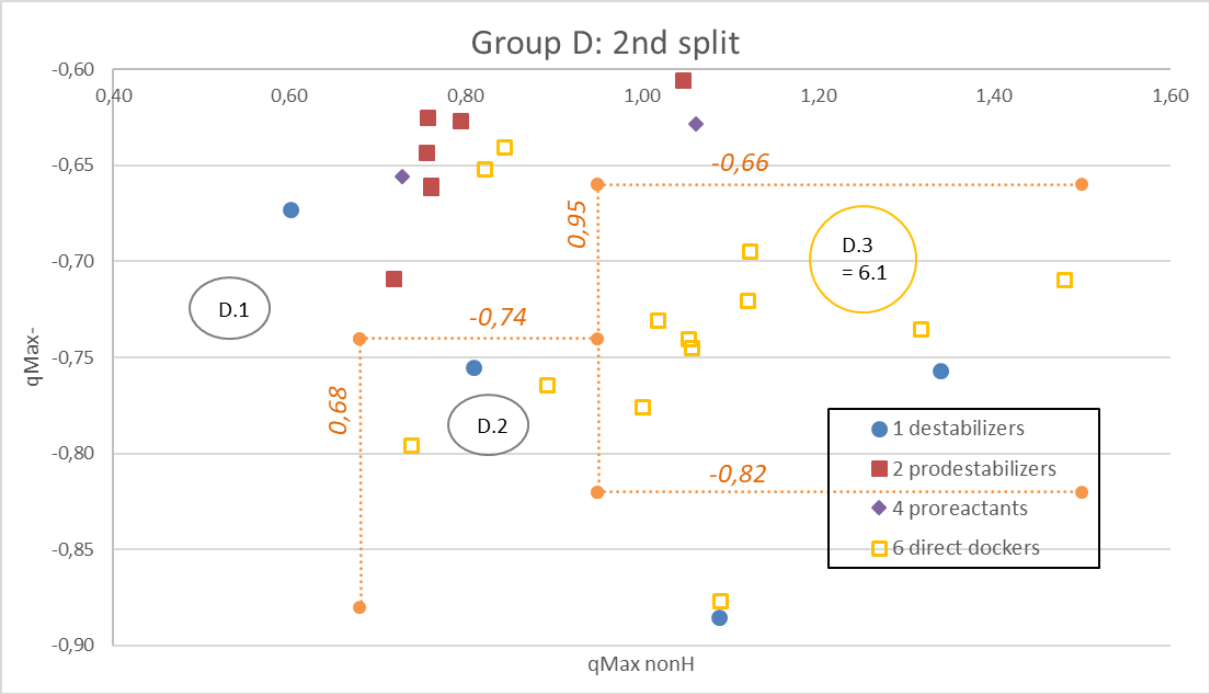


Figure S5: Second splits, group D.

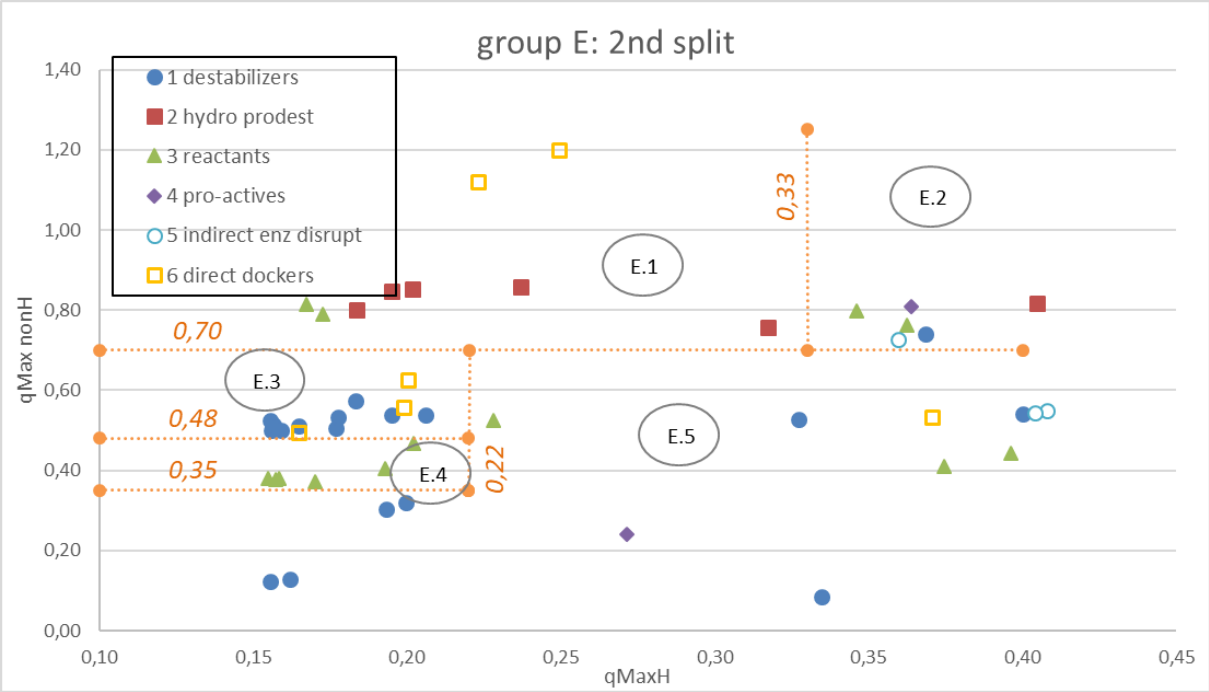


Figure S6: Second splits, group D.

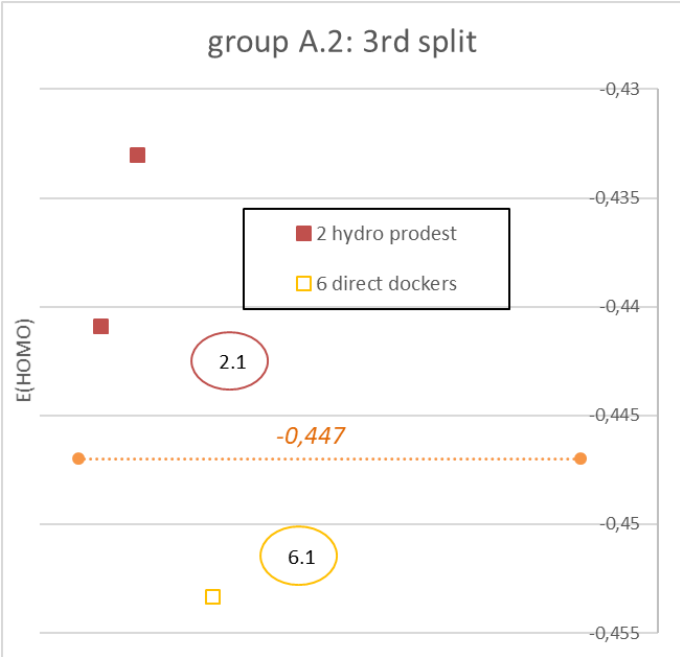


Figure S7: Third splits, group A.2.

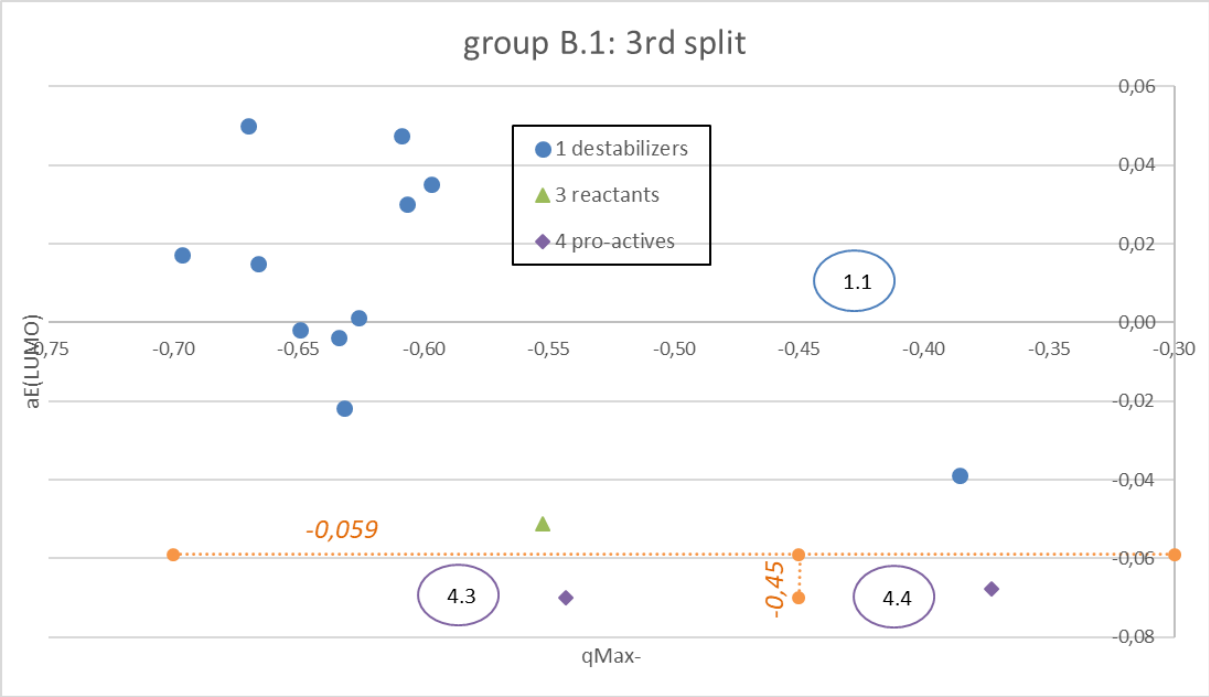


Figure S8: Third splits, group B.1.

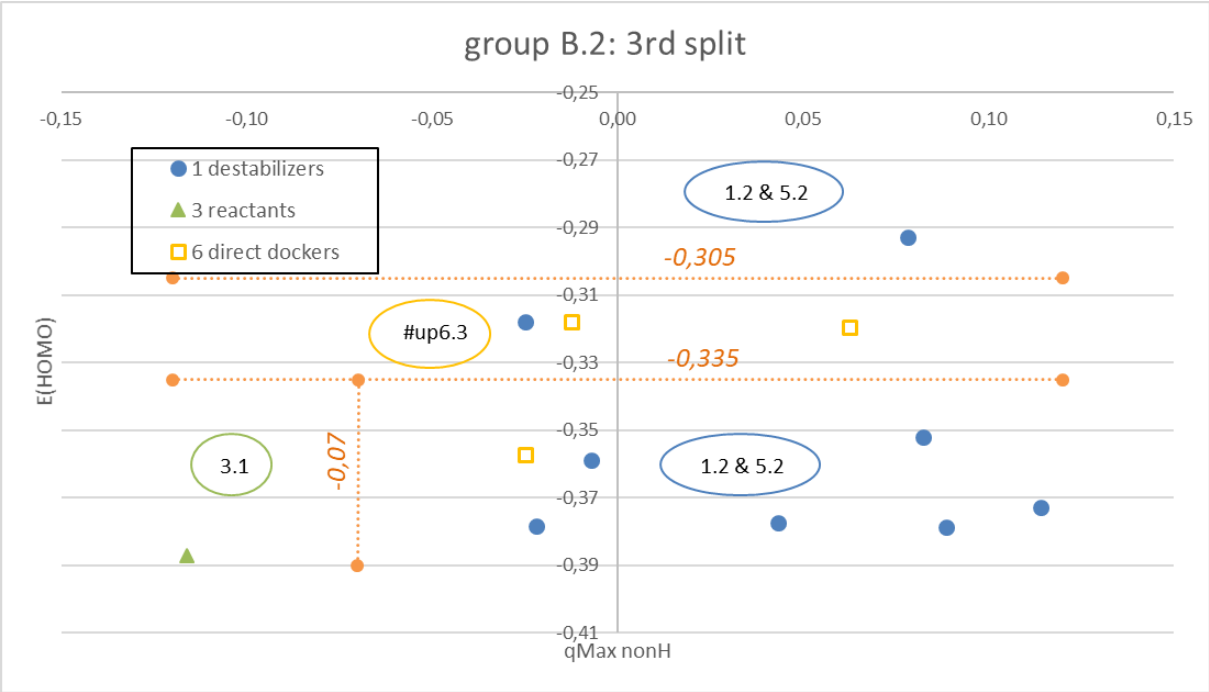


Figure S9: Third splits, group B.2.

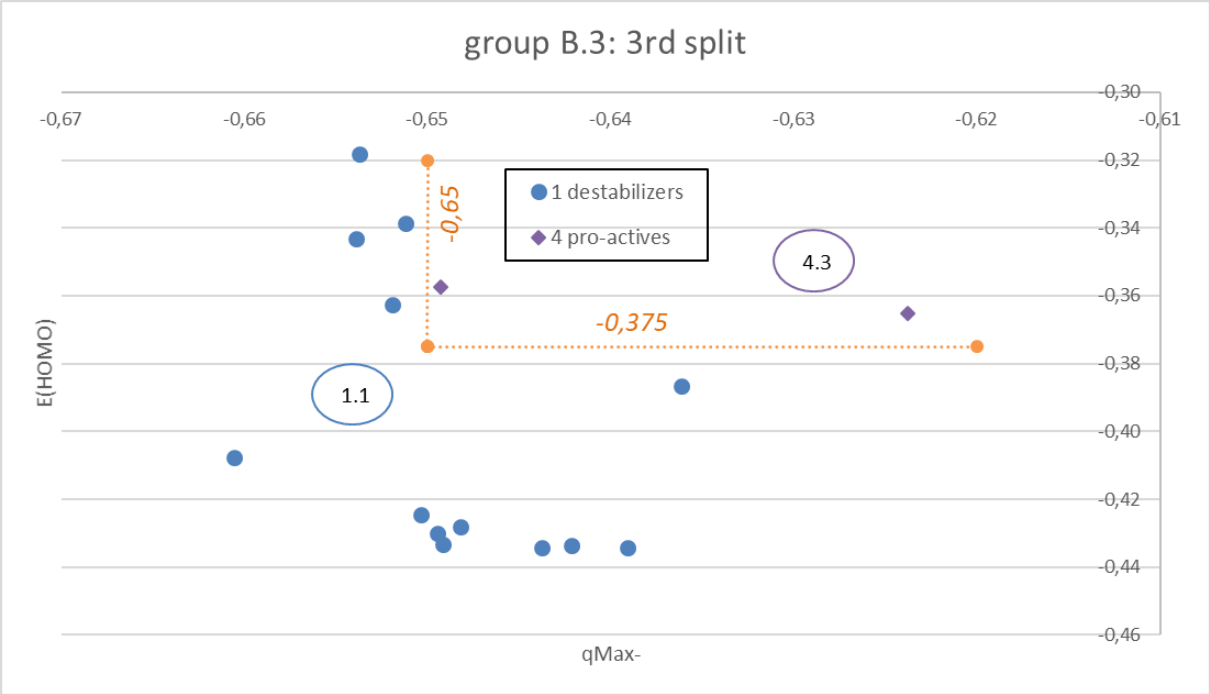


Figure S10: Third splits, group B.3.

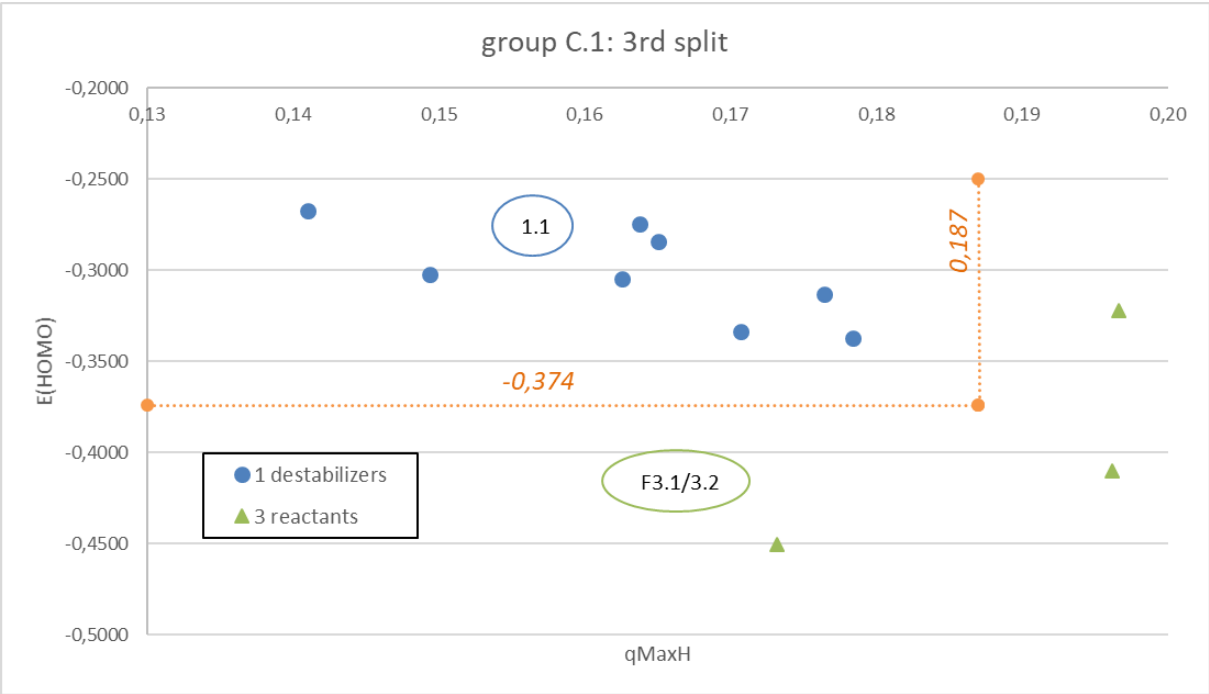


Figure S11: Third splits, group C.1.

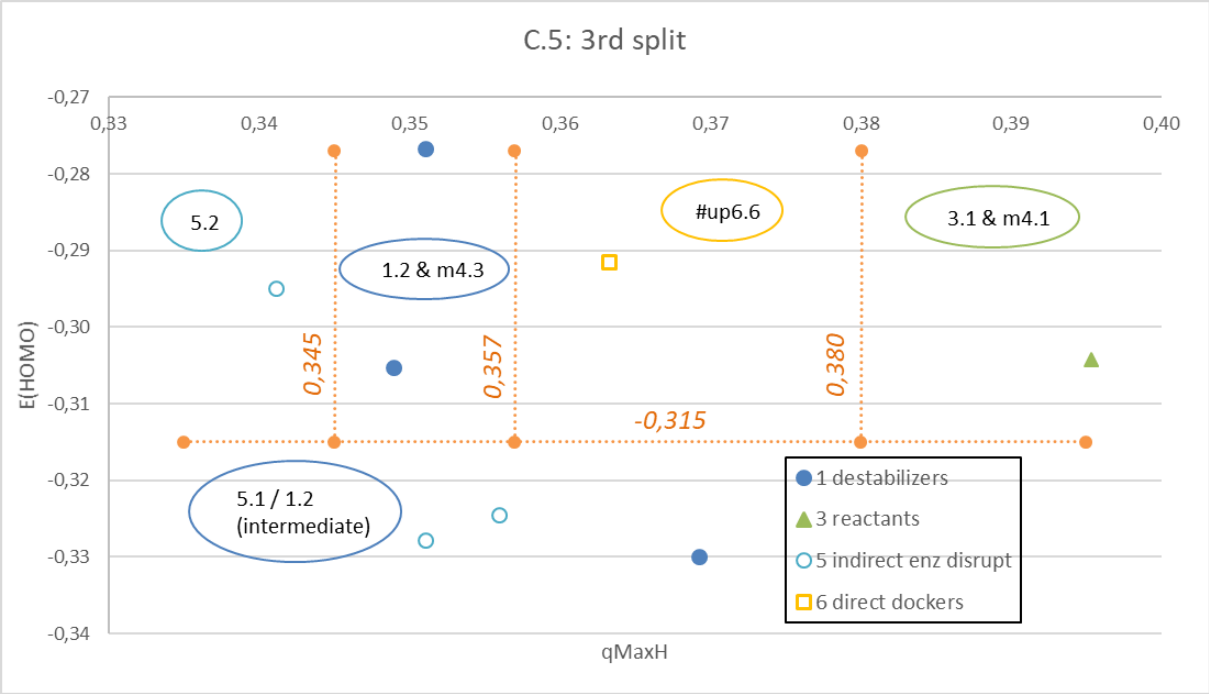


Figure S12: Third splits, group C.5.

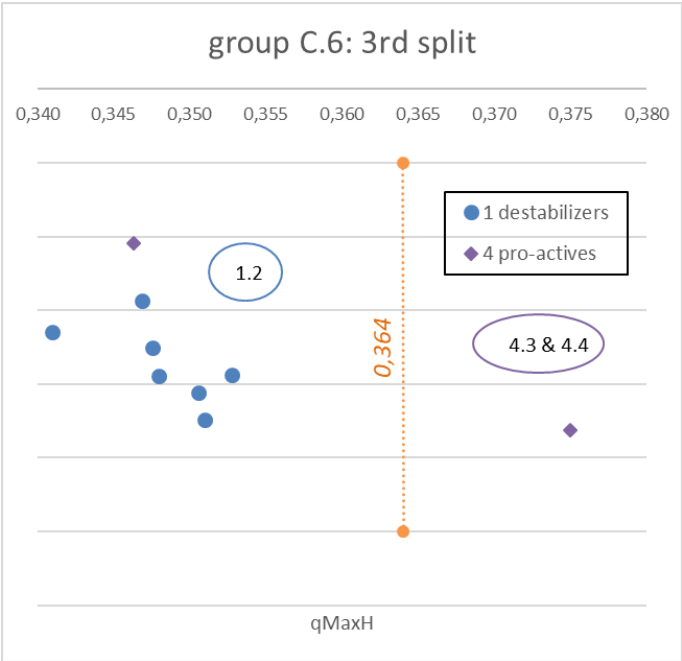


Figure S13: Third splits, group C.6.

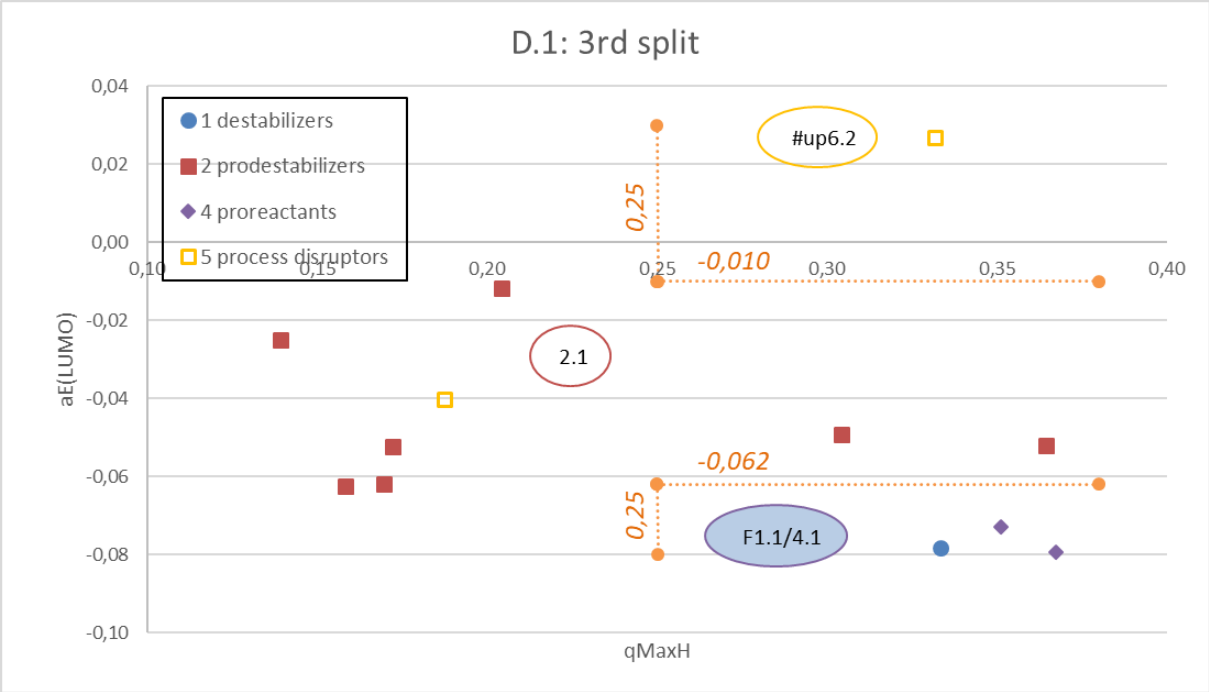


Figure S14: Third splits, group D.1.

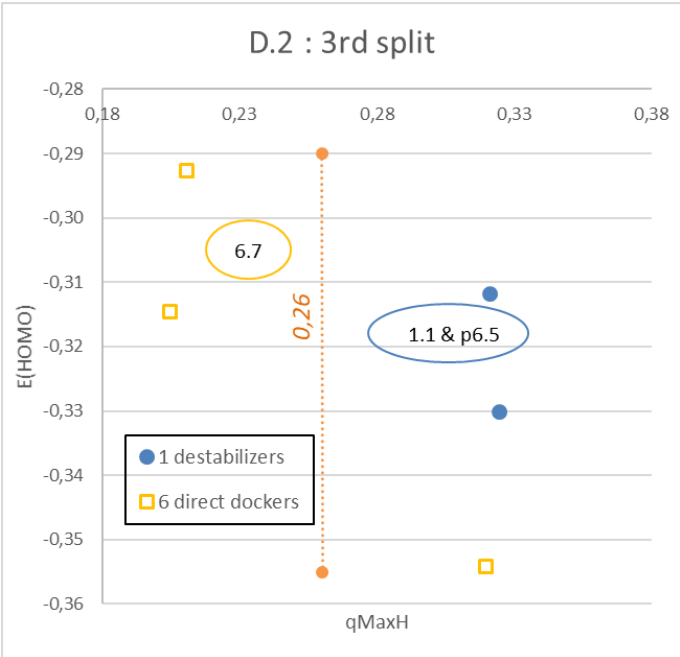


Figure S15: Third splits, group D.2.

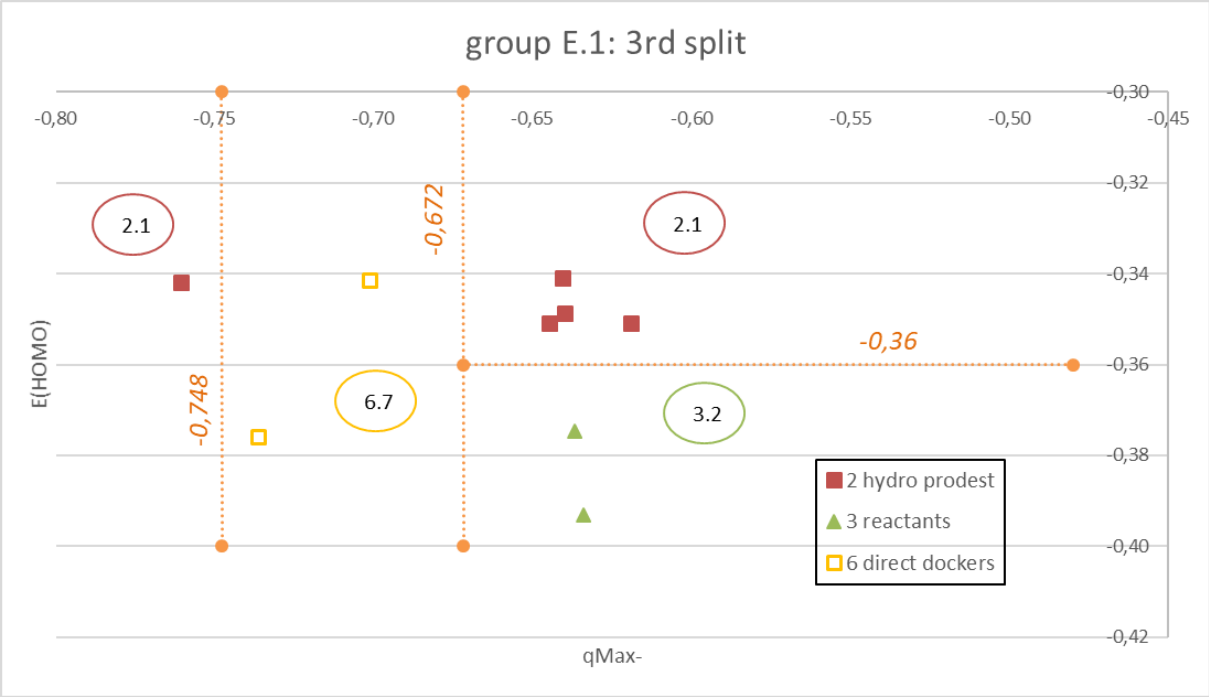


Figure S16: Third splits, group E.1.

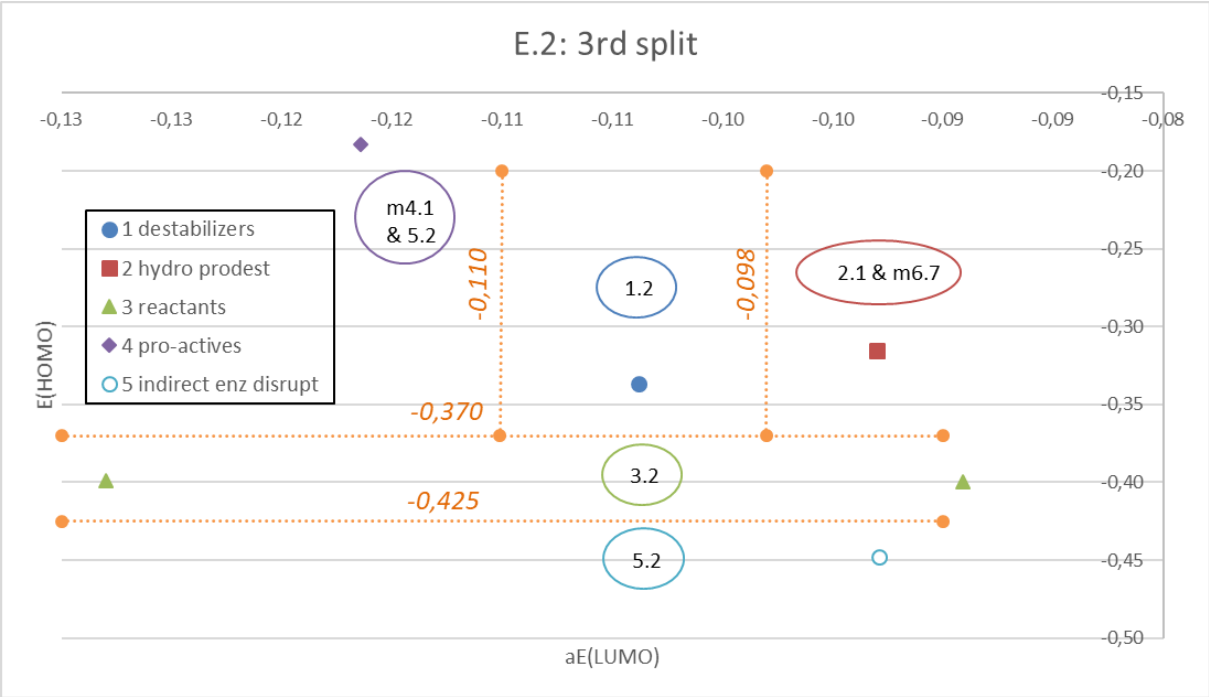


Figure S17: Third splits, group E.2.

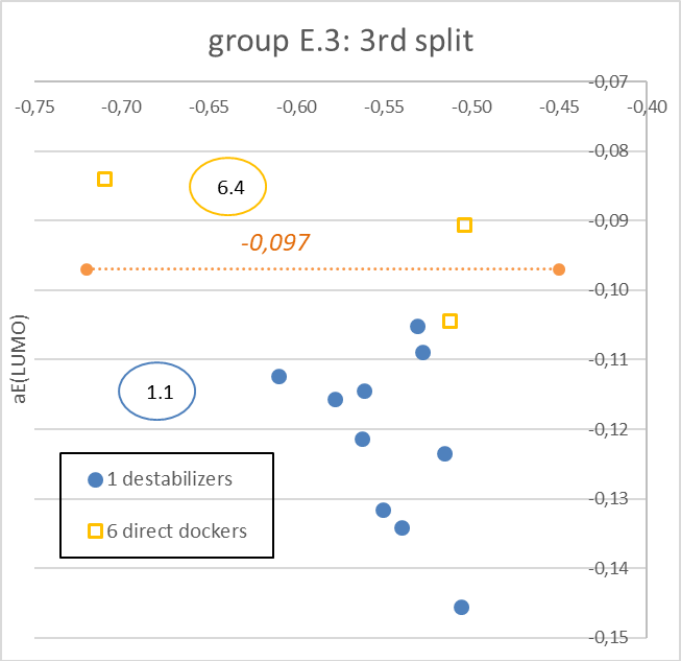


Figure S18: Third splits, group E.3.

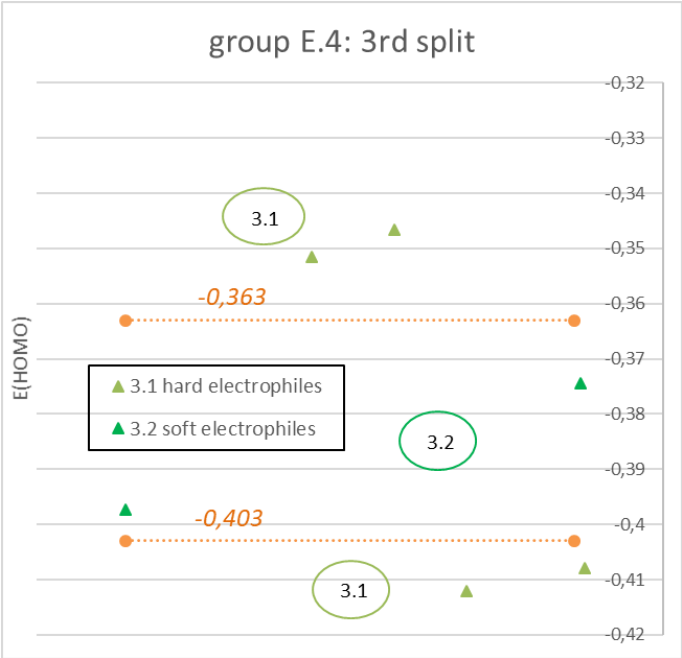


Figure S19: Third splits, group E.4.

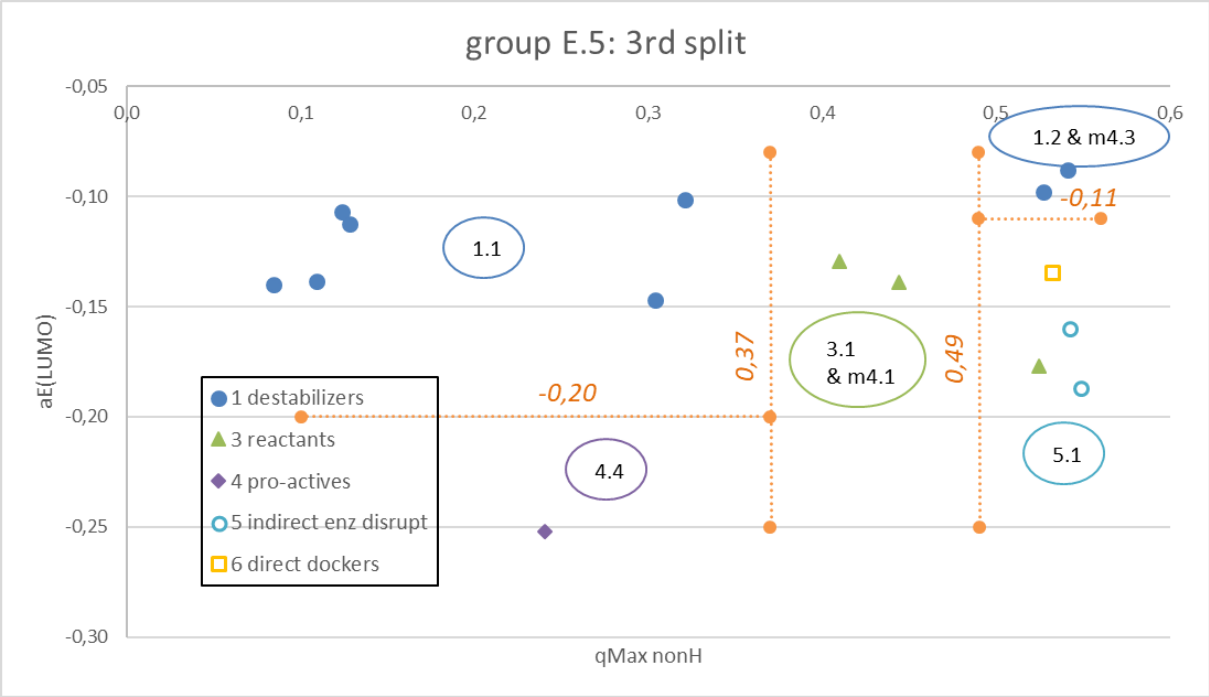


Figure S20: Third splits, group E.5.

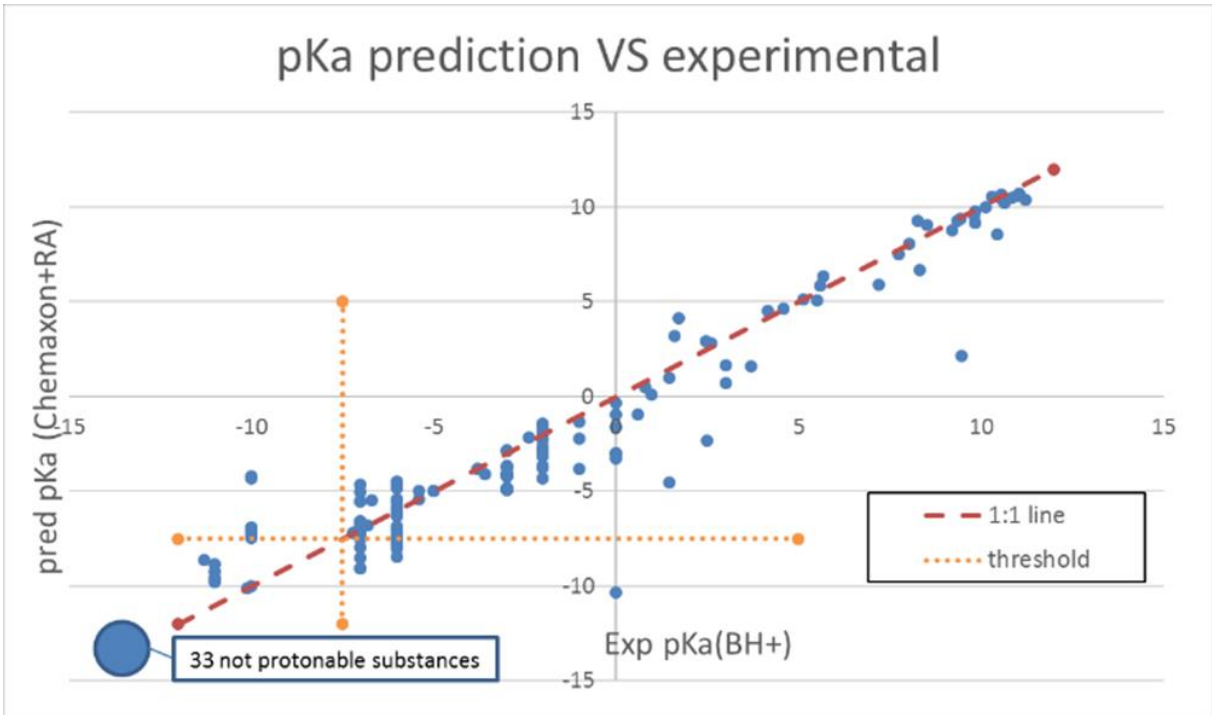


Figure S21: pKa prediction VS experimental.

A.2) Informations supplémentaires pour la méthode de prédiction des MechoAs par alertes structurales

Les informations supplémentaires associées à cette méthode sont 2 tableaux de données. Le deuxième tableau étant trop large pour une feuille A4, il a été séparé en deux parties. Une première partie pour l'identification de chaque substance et les numéros de MoA et MechoAs de la littérature et prédits, et une deuxième partie pour l'explication plus détaillée des mécanismes d'action de la littérature si nécessaire ainsi que les références associées.

Tableau S1 : Règles structurales.

Scaffold	Family	Examples	MechoA	Mechanism of action	Confidence factor	References
e.g. CC=CC, Cc1ccccc1, CC(=O)CCC, CC(O)C, Cn1cccc1, CCCSCCC, N#Cc1ccccc1C	aliphatic and mono-aromatic hydrocarbons (except hexane); monoketones (except hexan-2-one); non vinylic/allylic/propargylic 1° alcohols and ethers (except epoxides); non vinylic 2° and 3° alcohols (non aromatic); non aromatic/vinylic/allylic/propargylic 3° thiols, sulfides and oxidized derivatives (sulfoxide, sulfonate, sulfate); heteroaromatics (except furans and thiophenes); nitriles without H in alpha; all of that possibly with halides (except benzyl/vinyl/allyl halides (not F), carbon tetrachloride and dichloromethane, 2,2,2-trifluoroethanol and 2,2,2-trichloroethanol)	but-2-ene, toluene, pentan-2-one, isopropanol, N-methylpyrrole, dipropyl sulfide, o-tolunitrile	1.1	ALL: membrane destabilizing: non-polar narcosis	1	KREATiS internal experimental database and QSARs. Verhaar, H.J.M., van Leeuwen, C.J., Hermens, J.L.M., 1992. Classifying environmental pollutants. Chemosphere 25, 471–491. doi:10.1016/0045-6535(92)90280-5 Russom, C.L., Bradbury, S.P., Broderius, S.J., Hammermeister, D.E., Drummond, R.A., 1997. Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (Pimephales promelas). Environ. Toxicol. Chem. 16, 948–967. doi:10.1002/etc.5620160514
RC(R')S R, R' = Alkyl or H	1° & 2° thiols	ethanethiol	4.4	ALL: 1° & 2° thiols, in their anionic form, are oxidized into thiyl radical and then enter a redox cycle between the thiyl radical and the disulfides. 3° are not reactive enough due to steric hindrance. Aromatics are more reactive (conjugation stabilization), EWG are stabilizing the anion, therefore decreasing the reactivity, while EDG will stabilize the thiyl radical and favor the oxidation.	2	Munday, R. Toxicity of Thiols and Disulphides: Involvement of Free-Radical Species. Free Radic. Biol. Med. 1989, 7 (6), 659–673 DOI: 10.1016/0891-5849(89)90147-0.
RC(=O)O R=Alkyl or H	carboxylic acids (not abu)	octanoic acid, butanoic acid	5.2	ALL: Acids cross membranes in their neutral form, and then release their proton in the cytosol	2	Royce, L. A. et al. Transcriptomic Analysis of Carboxylic Acid Challenge in Escherichia coli : Beyond Membrane Damage. PLOS ONE 9, e89580 (2014).
c1ccccc1[N+](=O)[O-] or c1ccccc1N(R1)(R2) R1, R2 = H, Alkyl (at least 1 alkyl)	2° & 3° anilines, nitrobenzenes	aniline, m-nitrotoluene	1.1 & m4.3	M: metabolisation into nitroso => protein adducts, glatter, spleen (anilines) or liver damage (nitros) / chronic => oxidative stress, DNA adducts, cancer development	3	IARC. 2010. 'General Discussion of Common Mechanisms for Aromatic Amines.' *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans* 99: 41–54.
PAH and hydroxylated PAH	PAH and hydroxylated PAH	naphthalene, benzo[a]pyrene, 1-naphthol	1.1 & m4.3	M: metabolism with CYP gives epoxide and then quinones in a few hours, adducts to proteins and DNA	1	Tsuruda, L. S., Lamé, M. W. & Jones, A. D. Formation of epoxide and quinone protein adducts in B6C3F1 mice treated with naphthalene, sulfate conjugate of 1,4-dihydroxynaphthalene and 1,4-naphthoquinone. Arch. Toxicol. 69, 362–367 (1995). Buckpitt, A. et al. Naphthalene-induced respiratory tract toxicity: metabolic mechanisms of toxicity. Drug Metab. Rev. 34, 791–820 (2002).
R1[N+](R2)(R3)R4 R1-4= Alkyls	quaternary ammonium	butyl trimethyl ammonium	1.3 & #up6.1	ALL: insertion into membranes with strong electrostatic interaction, accumulating faster than other types of membrane destabilizers. ALL but U: for certain ammonium (better with an ester-like moiety) AChE inhibition. Without specific group (like phosphoester), it's a reversible inhibition.	3	Roberts, D.W., Roberts, J.F., Hodges, G., Gutsell, S., Ward, R.S., Llewellyn, C., 2013. Aquatic toxicity of cationic surfactants to Daphnia magna. SAR QSAR Environ. Res. 24, 417–427. doi:10.1080/1062936X.2013.781538 Čolović, M. B., Krstić, D. Z., Lazarević-Pašti, T. D., Bondžić, A. M. & Vasić, V. M. Acetylcholinesterase Inhibitors: Pharmacology and Toxicology. *Curr. Neuropharmacol.* **11,** 315–335 (2013).
R1NC(=O)NR2	substituted urea	diuron	p6.5 & [1.1 or 1.2 depending on other substructures]	Plants: blocks electron transfer at the photosystem II => inhibits O2 production. Other: This substruture doesn't cause specific toxicity.	1	van Rensen, J. J. S. Molecular mechanisms of herbicide action near photosystem II. *Physiol. Plant.* **54,** 515–521 (1982).
e.g. Oc1ccccc1, Oc1ccc(Cl)cc1	phenols with pKa>8,0. e.g. alkyl-, monohalophenols (=aromatic alcohol in a wide sense)	phenol, p-chlorophenol, o-cresol	1.2	ALL: membrane destabilizing: polar narcosis	1	KREATiS internal experimental database and QSARs. Verhaar, H.J.M., van Leeuwen, C.J., Hermens, J.L.M., 1992. Classifying environmental pollutants. Chemosphere 25, 471–491. doi:10.1016/0045-6535(92)90280-5 Russom, C.L., Bradbury, S.P., Broderius, S.J., Hammermeister, D.E., Drummond, R.A., 1997. Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (Pimephales promelas). Environ. Toxicol. Chem. 16, 948–967. doi:10.1002/etc.5620160514
N(R1)(R2)R3 R1-3 = H, Alkyl	aliphatic amines	putrescine, diisopropylamine	1.2 & 5.2	ALL: Directly applied, it is corrosive, and it alkalinize the cells, entering into them in their neutral form and then become protonated, all simply because it is a strong base.	3	Vallabh Minikel, E. A mechanism of action hypothesis http://www.cureffi.org/2016/05/08/a-mechanism-of-action-hypothesis/ (accessed Sep 19, 2016). Committee on Acute Exposure Guideline Levels; Committee on Toxicology; Board on Environmental Studies and Toxicology; Division on Earth and Life Studies; National Research Council. Piperidine. In Acute Exposure Guideline Levels for selected Airbone Chemicals; National Academies Press (US), 2012; Vol. 13. Cyclohexylamine [MAK Value Documentation, 2006]. In The MAK-Collection for Occupational Health and Safety; Wiley-VCH Verlag GmbH & Co. KGaA, 2002.
c1ccccc1[NH2]	1° anilines, can be substituted with halo- and nitro- groups (not too much chloro...)	4-chloroaniline	1.2 & m4.3	M: metabolisation into nitroso => protein adducts, glatter and spleen damage / chronic => oxidative stress, DNA adducts, cancer development	2	IARC. 2010. 'General Discussion of Common Mechanisms for Aromatic Amines.' *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans* 99: 41–54.
CC(=O)OR, CC(=O)NR, ROC(=O)OR, O=P(OR)3	ester, amide, phosphate, carbonate not abu	ethyl acetate, tributyl phosphate, acetamide	2.1	ALL: hydrolysis to acid and alcohol + "ester narcosis"	3	Jaworska, J. S.; Hunter, R. S.; Schultz, T. W. Quantitative Structure-Toxicity Relationships and Volume Fraction Analyses for Selected Esters. Arch. Environ. Contam. Toxicol. 1995, 29 (1), 86–93 DOI: 10.1007/BF00213091.
COC(=O)C(O)C(O)C(=O)OC	tartrates	diethyl tartrate	2.2	ALL: Hydrolysed into tartaric acid. Tartaric acid inhibits fumarase thus stopping the Krebs cycle.	2	Shaw, W. Chapter 3 Organic Acid Testing, Byproducts of Yeast and Their Relationship to Autism. In *Biological Treatments for Autism and PDD. Causes and Biomedical Therapies for Autism & PDD.* ; 2002.
CC(=O)X X=Hal	acyl halide	acetyl chloride	3.1	ALL: violently hydrolysed in water generating HCl (corrosive), not enzymatic!	1	PubChem, citing [Sax, N.I. Dangerous Properties of Industrial Materials. 6th ed. New York, NY: Van Nostrand Reinhold, 1984., p. 106]
C1OC1 / C1NC1	epoxide, aziridine	oxirane, cylcohexene oxide	3.1	ALL: Nucleophilic addition of proteins and DNA => adducts, impaired function	1	Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320
CC(=O)H	aldehyde (not a,b-u)	acetaldehyde, benzaldehyde	3.1	ALL: Schiff-base formation with amino residues of proteins (Nucleophilic Addition) => adducts, impaired function	1	A.O. Aptula, D.W. Roberts. 'Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity'. Chem. Res. Toxicol., 19 (2006), pp. 1097–1105 LoPachin, R. M.; Gavin, T. Molecular Mechanisms of Aldehyde Toxicity: A Chemical Perspective. Chem. Res. Toxicol. 2014, 27 (7), 1081–1091 DOI: 10.1021/tx5001046.
CC(=O)CCC(=O)C	gamma-diketones	hexane-2,5-dione	3.1	ALL: pyrrole formation with a lysine residue, oxidation and crosslinking with another protein	1	ATSDR. Toxicological Profile: n-Hexane. Toxicological Profile for n-Hexane (1999). Available at: http://www.atsdr.cdc.gov/ToxProfiles/tp.asp?id=393&tid=68. (Accessed: 10th July 2015)
N-N	hydrazines	1,1-dimethylhydrazine	3.1 & 4.3	M,F: binds to ketoacids and vitamin B6, inhibiting its wide actions / metabolised into diazonium, which gives oxidative stress (RNS) and generates Me' radicals	2	ATSDR. 1997. 'Toxicological Profile: Hydrazines'. September. http://www.atsdr.cdc.gov/ToxProfiles/tp.asp?id=502&tid=89.

Tableau S1 : Règles structurales.

C=CC(=O)X X=H, OH, OR, NR	a,b-unsat. Carbonyls (not ketones) and nitriles	acrolein, methyl acrylate	3.2	ALL: Michaël addition of proteins => adducts, impaired function	1	Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037. Schwöbel, Johannes A. H., Judith C. Madden, and Mark T. D. Cronin. 2011. 'Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis.' Chemosphere 85 (6): 1066–74. doi:10.1016/j.chemosphere.2011.07.037.
X-X X=O or S	peroxides, disulfides (maybe not tertiary disulfides)	benzoyl peroxide, dimethyl disulfide	3.3	ALL: Homolytic cleavage giving 2X' radicals, leading to lipid peroxidation (O), RedOx cycling (S)	2	Morgott, D., C. Lewis, J. Bootman, and M. Banton. 2014. 'Disulfide Oil Hazard Assessment Using Categorical Analysis and a Mode of Action Determination'. International Journal of Toxicology 33 (1 Suppl): 181S – 198S. doi:10.1177/1091581813504227. Kazius, Jeroen, Ross McGuire, and Roberta Bursi. 2005. 'Derivation and Validation of Toxicophores for Mutagenicity Prediction'. Journal of Medicinal Chemistry 48 (1): 312–20. doi:10.1021/jm040835a.
c1ccccc1C[LG], C=CC[LG]	benzyl/allyl-leaving group	benzyl chloride, benzyl acetate	3.1 & 4.1	ALL: SN2 adducts / N-acetylcysteine and glutathione conjugates (Nucleophilic Substitution of the LG) => diminution of antioxydant levels	2	'α-Chlorinated Toluenes [MAK Value Documentation, 1993].' 2002. In The MAK-Collection for Occupational Health and Safety. Wiley-VCH Verlag GmbH & Co. KGaA. http://onlinelibrary.wiley.com/doi/10.1002/3527600418.mb0m02mike0006/abstract.
c1ccccc1CO	benzyl alcohol	benzyl alcohol	1.1 & m4.1	M: rapidly eliminated by oxidation into benzoic acid, conjugated then with glycin to form hippuric acid	2	Nair, B. Final report on the safety assessment of Benzyl Alcohol, Benzoic Acid, and Sodium Benzoate. *Int. J. Toxicol.* **20 Suppl 3,** 23–50 (2001).
e.g. c1ccccc1C=O, O=Cc1cc(OC)c(O)c c1	electron-rich benzaldehyde	benzaldehyde, vanillin	3.1 &m 4.1	ALL: Schiff-base formation / M: rapidly eliminated by oxidation into acid (and potentially further conjugation)	2	Benzaldehyde [MAK Value Documentation, 2002]. in The MAK-Collection for Occupational Health and Safety 17, 13–36 (Wiley-VCH Verlag GmbH & Co. KGaA, 2002). IPCS Inchem. Vanillin SIDS. (2015). Available at: http://www.inchem.org/documents/sids/sids/121335.pdf. (Accessed: 23rd July 2015)
[CH2]=[CH]O…	vinyl alkoxy	vinyl carbamates	C4.3	ALL: metabolism transforming vinyl to epoxide, then adduct to DNA	1	Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320
[CH2]=[CH][CH2]O , C#C[CH2]O	1° allyl/propargyl alcohol	allyl alcohol, but-2-yn-1,4-diol, nerol?	4.3	ALL: ox to corresponding abu aldehyde => adducts	1	Koleva, Y. & Barzilov, I. Comparative study of mechanism of action of allyl alcohols for different endpoints. 49, 55–59 (2010).
o1cccc1, s1cccc1	furan, thiophene, not substituted in both positions 2 and 5 and not fused with aromatic	furan, 3-bromothiophene	4.3	ALL: metabolisaton into (Z)-but-2-en-1,4-dial => proteins and DNA adducts after glutathione is totally consumed => carcinogenic and mutagen	3	Food and Drug Administration. 2004. 'Docket No. 2004N-0205 Furan in Food, Thermal Treatment; Request for Data and Information.' May 10. http://www.fda.gov/OHRMS/DOCKETS/98fr/04-10588.htm.
ClC(Cl)(Cl)Cl	carbon tetrachloride	carbon tetrachloride	m4.3 & 1.1	M: cyt P 1e-reduction gives 'CCl3 radical, then quenched by O2 to give 'OOCCl3 peroxy radical, which gives then lipids & proteins peroxidation => hepatotoxicity	2	Recknagel, Richard O., Eric A. Glende Jr., James A. Dolak, and Robert L. Waller. 1989. 'Mechanisms of Carbon Tetrachloride Toxicity.' *Pharmacology & Therapeutics* 43 (1): 139–54. doi:10.1016/0163-7258(89)90050-8.
CCCCCC, CC(=O)CCCC	hexane and hexan-2-one	hexane, hexan-2-one	1.1 & mC4.3	M: CNS seizure: after metabolisation into hexan-2,5-dione, pyrrole formation with a lysine residue, oxidation and crosslinking with another protein	3	ATSDR. 1999. 'Toxicological Profile: N-Hexane.' Toxicological Profile for N-Hexane. July. http://www.atsdr.cdc.gov/ToxProfiles/tp.asp?id=393&tid=68.
c1ccccc1	benzene	benzene	1.1 & mC4.3	M: benzene is metabolised in the liver and lungs (CYP2E1) into hydroquinone, then it is oxidized into benzoquinone in bone marrow (MPO = myeloperoxidase), where it induces leukemia (blood cancer).	1	McHale, C. M.; Zhang, L.; Smith, M. T. Current Understanding of the Mechanism of Benzene-Induced Leukemia in Humans: Implications for Risk Assessment. *Carcinogenesis* **2012**, *33* (2), 240–252 DOI: 10.1093/carcin/bgr297.
FC(F)(F)CO	2,2,2-trifluoroethanol	2,2,2-trifluoroethanol	4.3	ox to trifluoroacetaldehyde => adducts	2	Airaksinen, M. M., Rosenberg, P. H. & Tammisto, T. A Possible Mechanism of Toxicity of Trifluoroethanol and Other Halothane Metabolites. Acta Pharmacol. Toxicol. (Copenh.) 28, 299–304 (1970).
ClC(Cl)=CCl	trichloroethylene	trichloroethylene	1.1 & m4.6 & mC4.3	M: major metabolic pathway is oxidation by CYP450 to generate chloral, 2,2,2-trichloroethanol, which activates GABA receptors. Minor metabolic pathway is the activation by GST to give reactive metabolites (likely dichlorovinyl-cysteine) which cause DNA mutation leading to renal cancer	2	Brüning, T. & Bolt, H. M. Renal toxicity and carcinogenicity of trichloroethylene: key results, mechanisms, and controversies. *Crit. Rev. Toxicol.* **30,** 253–285 (2000).
OCCC#C	but-3-yn-1-ol	but-3-yn-1-ol	4.3	ALL: Reacts with the O of the heme of cytP to form adducts to the heme, inhibiting the cytP	4	Smith, G. F. in Progress in Medicinal Chemistry (ed. Witty, G. L. and D. R.) 50, 1–47 (Elsevier, 2011)
ArN=NR	azo compounds	p-phenylazophenol	m4.3	M: azo bond cleaved (metabolism mainly by anaerobic intestinal bacteria), generating anilines. Then classical metabolism to nitroso.	3	Brown, M. A. & Vito, S. C. D. Predicting azo dye toxicity. *Crit. Rev. Environ. Sci. Technol.* **23,** 249–324 (1993).
Oc1ccc(O)cc1, Oc1ccccc1O	benzene-1,2- and -1,4-diols, potentially substituted	hydroquinone, catechol	4.3 & 4.4	ALL: Oxidation by tyrosinase / CYP450 to quinone => protein and DNA adducts / RedOx cycling	1	Bonnard, N., F. Pillière, J.-C. Protois, and O. Schneider. 2006. 'Fiche toxicologique de l'hydroquinone.' INRS. http://www.inrs.fr/dms/inrs/FicheToxicologique/TI-FT-159/ft159.pdf. Bolton, Judy L., Michael A. Trush, Trevor M. Penning, Glenn Dryhurst, and Terrence J. Monks. 2000. 'Role of Quinones in Toxicology'. Chemical Research in Toxicology 13 (3): 135–60. doi:10.1021/tx9902082.
Oc1ccccc1OC, Oc1ccc(OC)cc1	methyl ethers of hydroquinone and catechol with one -OH still present (potentially substituted)	guaiacol	m4.3 & 1.2	M: Dealkoxylation, then RedOx cycle between reduced and oxidized quinone	2	Riley, P. A. Mechanism of Pigment-Cell Toxicity Produced by Hydroxyanisole. J. Pathol. 1970, 101 (2), 163–169 DOI: 10.1002/path.1711010211.
CCO	ethanol	ethanol	1.1 & 4.3 & 4.5	ALL: Oxidation into acetaldehyde then acetate consumes NAD+ to produce NADH, this consumption goes the same direction as Krebs cycle, it is balanced by the oxidative phosphorylation. As a consequence, glycolysis and fatty acid oxidation rates are decreased, less energy is produced and there is an overconsumption of O2 by the Electron Transport Chain. Moreover acetaldehyde can form adducts before being oxidized into acetate. Chronic: The effects explained in acute generate as a consequence fatty acids / triglycerides deposition in the liver (and their concentration increases in the blood too)	2	King, Michael W. 2015. 'Ethanol (Alcohol) Metabolism: Acute and Chronic Toxicities.' The Medical Biochemistry Page. August 18. http://themedicalbiochemistrypage.org/ethanol-metabolism.php.
Oc1cc(O)ccc1	3-methoxyphenol	3-methoxyphenol	m4.3 & m6.7 & 1.2	M: 6-hydroxylation then ox to quinone / M: parent compound apparently inhibit melanin production	2	Moridani, M. Y., Cheon, S. S., Khan, S. & O'Brien, P. J. Metabolic activation of 3-hydroxyanisole by isolated rat hepatocytes. Chem. Biol. Interact. 142, 317–333 (2003).
??, [n+]1ccc(c2ccXcc2)cc1 X=c or [n+]. The [n+]s can be in other positions	??, bipyridinium or phenylpyridinium	paraquat, MPP+, rotenone	4.4	ALL: RedOx cycling (1e- reduction then give this e- to O2 to give the superoxide anion (ROS) and the parent compound, and so on)	2	Cohen, G. M., and M. d'Arcy Doherty. 1987. 'Free Radical Mediated Cell Toxicity by Redox Cycling Chemicals.' The British Journal of Cancer. Supplement 8 (June): 46–52.
O=C1C=CC(=O)C=C 1 / O=C1C=CC=CC1=O	quinones	1,4-benzoquinone	3.2 & 4.4	ALL: Michaël addition of proteins and DNA => adducts, impaired function / RedOx cycling	1	Bolton, Judy L., Michael A. Trush, Trevor M. Penning, Glenn Dryhurst, and Terrence J. Monks. 2000. 'Role of Quinones in Toxicology'. Chemical Research in Toxicology 13 (3): 135–60. doi:10.1021/tx9902082.
[CH]C#N	nitriles with a H in alpha	acetonitrile, other aliphatics	m4.6 & 1.1	M: cytP450 alpha-hydroxylation (cyanohydrin), then non-enzymatic degradation to give CN- and corresponding ketone/aldehyde	1	Grogan, James, Stephen C. DeVito, Robert S. Pearlman, and Kenneth R. Korzekwa. 1992. 'Modeling Cyanide Release from Nitriles: Prediction of Cytochrome P 450-Mediated Acute Nitrile Toxicity.' Chemical Research in Toxicology 5 (4): 548–52. doi:10.1021/tx00028a014.
ROC(=O)c1ccccc1C (=O)OR'	o-phthalate	dibutyl phthalate	2.1 & mC2.2	M: digestive hydolysis of one ester =>chronic=> induction of PPARalpha (peroxisome proliferator-activated receptor, precursor of liver cancer development) and endocrine disruption (if R>C4, can undergo further oxidative metabolism)	2	Adams, William J., Gregory R. Biddinger, Kenneth A. Robillard, and Joseph W. Gorsuch. 1995. 'A Summary of the Acute Toxicity of 14 Phthalate Esters to Representative Aquatic Organisms.' Environmental Toxicology and Chemistry 14 (9): 1569–74. doi:10.1002/etc.5620140916.
ClCCl	dichloromethane	dichloromethane	m4.6 & 1.1	M: metabolisation into CO with effects apparent in a few minutes to hours (depending on the dose) => fixing to haemoglobin, diminished O2 supply (coma then death) / directly CNS (not understood)	1	ATSDR. Medical Management Guidelines (MMGs): Methylene Chloride. (2014). Available at: http://www.atsdr.cdc.gov/mmg/mmg.asp?id=230&tid=42. (Accessed: 16th July 2015)

Tableau S1 : Règles structurales.

Nc1ccccc1 and n1ccccc1 substituted such that it is very acidic (pKa<0?)	acidic anilines & pyridines	2,4-dinitroaniline, pentachloropyridine	6	ALL: Oxidative Phophorylation uncoupling. Probably through the inhibition of electron transport chain? (Maybe by coordination to heme, inhibiting them?)	3	Bearden, A. P. & Schultz, T. W. Structure-activity relationships for Pimephales and Tetrahymena: A mechanism of action approach. *Environ. Toxicol. Chem.* **16,** 1311–1317 (1997). Martin, T. M. et al. Prediction of Aquatic Toxicity Mode of Action Using Linear Discriminant and Random Forest Models. J. Chem. Inf. Model. 53, 2229–2239 (2013).
ClC(Cl)(Cl)CO	2,2,2-trichloroethanol	2,2,2-trichloroethanol	m6.6 & 1.1	M: Activate GABA receptor (Cl- channel) (=metabolite of chloral=2,2,2-trichloroacetaldehyde/chloral hydrate)	1	Peoples, R. W. & Weight, F. F. Trichloroethanol potentiation of gamma-aminobutyric acid-activated chloride current in mouse hippocampal neurones. *Br. J. Pharmacol.* **113,** 555–563 (1994).
Oc1ccccc1 substituted such that pKa<8,0	phenols with pKa<8,0. e.g. with >=2Cl or nitro	2,5-dinitrophenol, pentachlorophenol	5.1	ALL: Oxidative Phosphorylation Uncoupling by inner mitochondrial membrane pH gradient suppression (or other vesicles pH gradient also). It also depends on steric factors that favor dimer formation of hide the charged oxygen to cross the membrane...	1	Schultz, T. W. 1987. 'The Use of the Ionization Constant (pKa) in Selecting Models of Toxicity in Phenols.' Ecotoxicology and Environmental Safety 14 (2): 178–83.
R1P(=O)(R2)LG or R1P(=S)(R2)LG (R1, R2= alkyl, N-alkyl or O-alkyl ; LG less restrictive than with benzyl/allyl reactivity. Can be F, can be not acidic phenoxy, can be thiol)	organophosphorous	Tabun, Sarin, disulfoton, malathion	6.1	Animals: AChE inhibition (covalent bond with P formed) (U&P: Probably inhibits other esterases)	1	Elersek, T.; Filipic, M. Organophosphorus Pesticides - Mechanisms of Their Toxicity. In Pesticides - The Impacts of Pesticides Exposure; Stoytcheva, M., Ed.; InTech, 2011. Gonçalves, A. da S.; França, T. C. C.; Wilter, A.; Figueroa-Villar, J. D. Molecular Dynamics of the Interaction of Pralidoxime and Deazapralidoxime with Acetylcholinesterase Inhibited by the Neurotoxic Agent Tabun. J. Braz. Chem. Soc. 2006, 17 (5), 968–975 DOI: 10.1590/S0103-50532006000500022. Fukuto, T. R. Mechanism of action of organophosphorus and carbamate insecticides. Environ. Health Perspect. 87, 245–254 (1990).
NC(=X)XR X=O or S, OR/SR being a good enough LG	carbamates	methomyl, aldicarb	6.1	Animals: AChE inhibition (reversible bonding to carbonyl) (A: more toxic than esters. Probably inhibits other esterases)	1	Fukuto, T. R. Mechanism of action of organophosphorus and carbamate insecticides. Environ. Health Perspect. 87, 245–254 (1990).
dioxins, polychlorodibenzo furans, PCBs	dioxins, polychlorodibenzofurans, PCBs	dioxin, TCDD, PCB 105	#uC6.7 & 1.1	induction of CYP450 (binding to AhR, but then not / very slowly metabolised by cytP450 consequently producted) => oxidative stress	1	Buckpitt, A., B. Boland, M. Isbell, D. Morin, M. Shultz, R. Baldwin, K. Chan, et al. 2002. 'Naphthalene-Induced Respiratory Tract Toxicity: Metabolic Mechanisms of Toxicity.' Drug Metabolism Reviews 34 (4): 791–820. doi:10.1081/DMR-120015694.
XC(=O)CC(=O)X X=C,O,N	beta-dicarbonyls	pentane-2,4-dione, 1-benzoylacetone	6.4	ALL: chelating Fe, degrading cytP450, also reacts slowly with arginine residues to form adducts, causes inactivaton of B vitamins	3	O'Donoghue, J. L. 'Ketones of Four or Five Carbons'. in *Patty's Toxicology* (John Wiley & Sons, Inc., 2001).
C#O	Carbon monoxide	CO	6.7	ALL: Displaces O2 from hemes of hemoglobin, preventing then the O2 supply of the body	1	Olson, Kent R. 1984. 'Carbon Monoxide Poisoning: Mechanisms, Presentation, and Controversies in Management.' The Journal of Emergency Medicine 1 (3): 233–43. doi:10.1016/0736-4679(84)90078-7.
[C-]#N	cyanide	cyanide	#up6.7	Animals: cytochrome c oxidase (complex IV) inhibition	1	Way, James L, Peter Leung, Elizabeth Cannon, Ronald Morgan, Chamille Tamulinas, Jon Leong-Way, Lynn Baxter, Amina Nagl, and Carie Chui. 2007. 'The Mechanism of Cyanide Intoxication and Its Antagonism - Ciba Foundation Symposium 140 - Cyanide Compounds in Biology - Way - Wiley Online Library.' In , 232–48. D. Evered and S. Hamett. http://onlinelibrary.wiley.com/doi/10.1002/9780470513712.ch14/summary.
OC(=O)C(O)C(O)C(=O)O	tartaric acid	tartaric acid	6.7	ALL: tartaric acid inhibits fumarase thus stopping the Krebs cycle.	2	Shaw, W. Chapter 3 Organic Acid Testing, Byproducts of Yeast and Their Relationship to Autism. In *Biological Treatments for Autism and PDD. Causes and Biomedical Therapies for Autism & PDD.* ; 2002.
c1cc(X)ccc1[CH](Z)c2ccc(Y)cc2 see publi for details of X, Y and Z nature	DDT analogs	DDT	#up6.6	Animals: prevents the closure of Na+ channel from the nerve, maintaining the signal => causes death by respiratory failure	1	Coats, J R. 1990. 'Mechanisms of Toxic Action and Structure-Activity Relationships for Organochlorine and Synthetic Pyrethroid Insecticides.' Environmental Health Perspectives 87 (July): 255–62.
chlorinated alicyclics	chlorinated alicyclics	lindane, dieldrin	#up6.6	Animals: inhibition of GABAergic Cl- channel into the nerves (binding at the picrotoxinin site), maintaining the signal => causes death by respiratory failure	1	Coats, J R. 1990. 'Mechanisms of Toxic Action and Structure-Activity Relationships for Organochlorine and Synthetic Pyrethroid Insecticides.' Environmental Health Perspectives 87 (July): 255–62.
[C]=CC(C1(C)C)C1C(=O)OR see publi for R nature	pyrethroids	resmethrin, cypermethrin, tefluthrin	#up6.6	Animals: prevents the closure of Na+ channel from the nerve, maintaining the signal / inhibits GABAergic Cl- channel into the nerve / inhibits Ca-ATPase or Ca-Mg-ATPase, disturbing Ca2+ regulation	1-2	Coats, J R. 1990. 'Mechanisms of Toxic Action and Structure-Activity Relationships for Organochlorine and Synthetic Pyrethroid Insecticides.' Environmental Health Perspectives 87 (July): 255–62.
O=C(NCC(N[C@@](C(NCC(N[C@@H](C3)C(N[C@@H](CC(N)=O)C(N5[C@@H]4C[C@@H](O)C5)=O)=O)=O)=O)([H])[C@@H](C)CC)=O)[C@H](CC1=C(S3=O)NC2=C1C=CC(O)=C2)NC([C@@]([C@@H](C)[C@@H](O)CO)([H])N[C@@]4=O)=O	alpha-amanitin	alpha-amanitin	6.7	ALL: binds to RNA polymerase II (and less strongly III) inhibiting its action, so the DNA almost can't be transcripted anymore. Death occuring in ca. 10 days in humans.	1	Gong, Xue Q., Yuri A. Nedialkov, and Zachary F. Burton. 2004. 'α-Amanitin Blocks Translocation by Human RNA Polymerase II.' Journal of Biological Chemistry 279 (26): 27422–27. doi:10.1074/jbc.M402163200.
mimics of acetylcholine (and nicotine/muscarine)	mimics of acetylcholine (and nicotine/muscarine)	muscarine, nicotine, bethanechol	#upC6.2	Animals: Agonism of Ach receptors (muscarinic of nicotinic) (with our without further action with AChE)	1	King, Michael W. 2015. 'Biochemistry of Neurotransmitters and Nerve Transmission'. September 9. http://themedicalbiochemistrypage.org/nerves.php.
XC(=O)c1c(Y)cccc1 X,Y = N, O, OR	salicylic acid-like compounds	acetylsalicylic acid, 2-hydroxybenzamide	m+f6.7 & 1.2	M: analgesic, antipyretic, anti-inflammatory, certainly mostly due to the inhibition of COX-2 (and COX-1). Main side effect at high doses is gastrointestinal bleeding (thought to be due to COX-1 inhibition) F: other enzyme-specific interactions disturbing the stress response. Other : ?	3	Vane, John R, and Regina M Botting. 1998. 'Mechanism of Action of Nonsteroidal Anti-Inflammatory Drugs'. The American Journal of Medicine 104 (3, Supplement 1): 2S – 8S. doi:10.1016/S0002-9343(97)00203-9.
c12ccccc1N(C34)C(=O)CC(C45)OCC=C(C5CC67)CN6CCC237	strychnine	strychnine	#up6.7	Animals: antagonist of glycine in the glycine receptor of nerve jonctions	1	Makarovsky, I. *et al.* Strychnine - A Killer from the Past. *IMAJ Isr. Med. Assoc. J.* **10,** 142–145 (2008).
CN1C=2N=CN(C)C2C(=O)N(C)C1=O	caffeine	caffeine	6.7 & 6.6	Animals: CNS: antagonism of adenosine at its receptors (regulation of adenylate cyclase synthesising cAMP), cyclic nucleotide (cAMP) phosphodiesterase inhibition (=> release of Ca2+), activation of Calcium channels. (Plants: Ca2+ regulation also disrupted)	1	Nehlig, A., Daval, J.-L. & Debry, G. Caffeine and the central nervous system: mechanisms of action, biochemical, metabolic and psychostimulant effects. *Brain Res. Rev.* **17,** 139–170 (1992).

Tableau S1 : Règles structurales.

NC(C)Cc1ccccc1	amphetamine	amphetamine	#up6.3	Animals: forces release of dopamine at nerve junctions while competitively inhibiting the transport of dopamine to the cell	1	Calipari, E. S. & Ferris, M. J. Amphetamine Mechanisms and Actions at the Dopamine Terminal Revisited. *J. Neurosci. Off. J. Soc. Neurosci.* **33,** 8923–8925 (2013).
c1ncccc1C2CCCN2C	nicotine	nicotine	#up6.2	Animals: agonism of nicotinic Ach receptors, stimulation of neurotransmitters production, leading to a desensitization and upregulation of nAchRs in chronic	1	Narahashi, T. *et al.* Symposium Overview: Mechanism of Action of Nicotine on Neuronal Acetylcholine Receptors, from Molecule to Behavior. *Toxicol. Sci.* **57,** 193–202 (2000).
ClC(Cl)(Cl)C(O)(c1ccc(Cl)cc1)c2ccc(Cl)cc2	dicofol (kelthane)	dicofol (kelthane)	6.7	ALL: thyroid hormone antagonist, weak estrogen agonist, potent inducer of CYP2B, weak aromatase inhibitor, some ability to bind with andogen receptors	3	Sparling, D. W., Linder, G., Bishop, C. A. & Krest, S. *Ecotoxicology of Amphibians and Reptiles, Second Edition* . (CRC Press, 2010).
??	?? (some pesticides)	rotenone, 2-acetylaminofluorene	#up6.7	Animals: inhibition of mitochondrial e- transport (rotenone => complex I)	1	Sherer, Todd B., Jason R. Richardson, Claudia M. Testa, Byoung Boo Seo, Alexander V. Panov, Takao Yagi, Akemi Matsuno-Yagi, Gary W. Miller, and J. Timothy Greenamyre. 2007. 'Mechanism of Toxicity of Pesticides Acting at Complex I: Relevance to Environmental Etiologies of Parkinson's Disease.' Journal of Neurochemistry 100 (6): 1469–79. doi:10.1111/j.1471-4159.2006.04333.x.
steroids or steroid-like compounds	steroids or steroid-like compounds	EE2, mifepristone, 17β-trenbolone	#upC6.7?	Animals: endocrine disruption (sexual differenciation/contraceptive/...)	2	IPCS. 2013. 'Global Assessment of the State-of-the-Science of Endocrine Disruptors'. http://www.who.int/ipcs/publications/new_issues/endocrine_disruptors/en/.
OH + aromatic + 3° amine e.g.Oc1ccc(CC5C3C=CC4O)c(C3(CCN5C)C4O2)c12	opioids	morphine	#up6.6	Animals: binds to opioid receptors (GPCRs) at presynaptic and postsynaptic neurons, inhibiting the neuronal signal by upregulating the K+ channels and downregulating the Ca2+ channels thereby inhibiting the release of neurotransmitter (presynaptic effect).	1	Chahl, L. A. Experimental and Clinical Pharmacology: Opioids - Mechanisms of Action. Aust. Prescr. 1996, 19 (3), 63–65 DOI: 10.18773/austprescr.1996.063.
CCCCOCCOC(=O)COc1ccc(Cl)cc1Cl	2,4-D butoxyethanol ester	2,4-D butoxyethanol ester	p2.2 & #p2.1	Plants: hydrolysed. Then the acid is an antagonist of phytohormone auxin, overstimulating plant growth, more than it can afford (because too slowly metabolised then). M: after hydrolysis, acid rapidly eliminated.	1	Song, Y. Insight into the Mode of Action of 2,4-Dichlorophenoxyacetic Acid (2,4-D) as an Herbicide. J. Integr. Plant Biol. 2014, 56 (2), 106–113 DOI: 10.1111/jipb.12131. Jervais, G.; Luukinen, B.; Buhl, K.; Stone, D. 2,4-D Technical Fact Sheet http://npic.orst.edu/factsheets/archive/2,4-DTech.html (accessed Dec 16, 2016).
OC(=O)COc1ccc(Cl)cc1Cl	2,4-dichlorophenoxyacetic acid	2,4-D	pC6.7 & 1.1	Plants: it is an antagonist of phytohormone auxin, overstimulating plant growth, more than it can afford (because too slowly metabolised then). M: rapidly eliminated.	1	Song, Y. Insight into the Mode of Action of 2,4-Dichlorophenoxyacetic Acid (2,4-D) as an Herbicide. J. Integr. Plant Biol. 2014, 56 (2), 106–113 DOI: 10.1111/jipb.12131. Jervais, G.; Luukinen, B.; Buhl, K.; Stone, D. 2,4-D Technical Fact Sheet http://npic.orst.edu/factsheets/archive/2,4-DTech.html (accessed Dec 16, 2016).
N1CCCCC1R R = Alkyl (or H)	alkylpiperidine	coniine, piperidine	#up6.2	Animals: blocks nicotinic Ach receptors. Strong effect with alkyl = propyl (or bigger)	1	National Center for Biotechnology Information. CONIINE | C8H17N - PubChem https://pubchem.ncbi.nlm.nih.gov/compound/441072 (accessed Dec 21, 2016). Committee on Acute Exposure Guideline Levels; Committee on Toxicology; Board on Environmental Studies and Toxicology; Division on Earth and Life Studies; National Research Council. Piperidine. In Acute Exposure Guideline Levels for selected Airbone Chemicals; National Academies Press (US), 2012; Vol. 13. Vetter, J. Poison Hemlock (Conium Maculatum L.). Food Chem. Toxicol. 2004, 42 (9), 1373–1382 DOI: 10.1016/j.fct.2004.04.009.
Nc(ncn1)nc1N	diaminotriazines	atrazine	p6.5 & #uc6.7	Plants: inhibits Protein D-1 at photosystem II => blocks production of O2 + ROS generated. Other: carcinogenic, reproductive and developmental toxicity (several MechoAs are postulated to explain these effects)	2	APVMA. Atrazine Toxicity: Analysis of Potential Modes of Action; APVMA, 2010; p 49. van Rensen, J. J. S. Molecular Mechanisms of Herbicide Action near Photosystem II. Physiol. Plant. 1982, 54 (4), 515–521 DOI: 10.1111/j.1399-3054.1982.tb00719.x.
OP(=O)(O)CNCC(=O)O	glyphosate	glyphosate	p6.7 & #uc6.7 / 4.1?	Plants: inhibit aromatic amino acids synthesis by 5-enolpyruvylshikimimate-3-phosphate synthase inhibition. M: genotoxic and probable carcinogen.	2	Mensah, P. K.; Palmer, C. G.; Odume, O. N. Ecotoxicology of Glyphosate and Glyphosate-Based Herbicides — Toxicity to Wildlife and Humans. In *Toxicity and Hazard of Agrochemicals* ; Larramendy, M. L., Soloneski, S., Eds.; InTech, 2015.
CCCN(CCC)c1c([N+](=O)[O-])cc(C(F)(F)F)cc1[N+](=O)[O-]	trifluralin	trifluralin	6.7	Plants: inhibit tubulin polymerisation into microtubules (playing a central role in cellular division), stops mitosis. All: disrupts Ca2+ concentration equilibria, thus disrupting a lot of processes in the cell. Very toxic to fish, and very harmless for mammals	1	Fernandes, T. C.; Pizano, M. A.; Marin-Morales, M. A. Characterization, Modes of Action and Effects of Trifluralin: A Review. In *Herbicides - Current Research and Case Studies in Use* ; Price, A. J., Kelton, J. A., Eds.; InTech, 2013.
NCCc(cn1)c2cc(O)ccc12	serotonin	serotonin	#up6.7 & #up4.1	Animals: endogenous neurotransmitter (controlling a lot of things, vomiting, locomotion, thermoregulation, feelings, ...). In reasonable exogenous quantities it is metabolised following the normal pathway (oxidation of amine into aldehyde then acid).	1	King, M. W. Biochemistry of Neurotransmitters and Nerve Transmission http://themedicalbiochemistrypage.org/nerves.php (accessed Sep 25, 2015). Frazer, A.; Hensler, J. G. Serotonin. Basic Neurochemistry: Molecular, Cellular and medical aspects.; Siegel GJ, Agranoff BW, Albers RW, et al., 1999.
Clc1ccccc1CN2OCC(C)C)C2=O	clomazone	clomazone	p6.7 & / #up6.1	Plants: inhibit carotenoid biosynthesis (essential for photosynthesis). Other: slight AChE inhibition effect	1	Van Scoy, A. R.; Tjeerdema, R. S. Environmental Fate and Toxicology of Clomazone. Rev. Environ. Contam. Toxicol. 2014, 229, 35–49 DOI: 10.1007/978-3-319-03777-6_3.
ClCC(=O)Nc1ccccc1	chloroacetanilides	metolachlor	4.1 & p6.7	Conjugation with thiol groups (including CoA, inhibiting all the processes using it), not precisely known... Plants: makes it a good herbicide. Plants & Other: detoxification with GSH.	1	Fuerst, E. P. Understanding the Mode of Action of the Chloroacetamide and Thiocarbamate Herbicides. Weed Technol. 1987, 1 (4), 270–277. Rivard, L. Environmental Fate of Metolachlor. Chemistry 2003, 51218, 45–2.

confidence factors	confidence	criteria
1	very good	widely described and accepted, inter-species physiological effects correlated
2	good	well described, physiological effects correlated for at least one species
3	poor	well described without physiological effects OR poorly descibed with physiological effects
4	very poor	described in one phrase with no details, no physiological effects reported

Abbreviations	
EDG	Electron-Donating Group
EWG	Electron-Withdrawing Group
LG	Leaving Group = halide (except F), O-ester, phenoxy (if acidic enough), N if protonated at neutral pH.

Tableau S2 : Set de validation. Partie 1.

N°	Molecule	CAS N°	SMILES Code	Literature MoA	Toxtree 2.6.6 MoA (Verhaar modified)	MechoA scheme	Literature MechoA
1	ethanoic acid (acetic acid)	000064-19-7	O=C(O)C	3	1	5.2	5.2
2	2-methylpropanoic acid (isobutyric acid)	000079-31-2	O=C(O)C(C)C	3	1	5.2	5.2
3	pentanoic acid	000109-52-4	O=C(O)CCCC	3	1	5.2	5.2
4	2-methylbutanoic acid	000116-53-0	O=C(O)C(C)CC	3	1	5.2	5.2
5	2-ethylhexanoic acid	000149-57-5	O=C(O)C(CCCC)CC	3	1	5.2	5.2
6	3,5,5-trimethylhexanoic acid	003302-10-1	O=C(O)CC(CC(C)(C)C)C	3	1	5.2	5.2
7	methacrylic acid	000079-41-4	O=C(O)C(=C)C	3	3	3.2	5.2 & 3.2 & 4.1
8	sorbic acid	000110-44-1	O=C(O)C=CC=CC	3	3	3.2	5.2
9	cyclohexane-1,2-dicarboxylic anhydride	000085-42-7	O=C(OC(=O)C1CCCC2)C12	3	3	2.1	3.1 & 2.1
10	tripropenyl succinic anhydride	092077-08-2	C(=CC)C1C(OC(C1(C=CC)C=CC)=O)=O	3	3	2.1	3.1 & 2.1
11	3-dodecenyl dihydrofuran-2,5-dione	026544-38-7	O=C1OC(=O)CC1CCCCCCCCC=CCC	3	3	2.1	3.1 & 2.1
12	methanol	000067-56-1	OC	1	1	1.1	m4.5 & 1.1 / m4.6
13	2-methylpropan-1-ol (isobutanol)	000078-83-1	OCC(C)C	1	1	1.1	1.1
14	1-pentanol	000071-41-0	OCCCCC	1	1	1.1	1.1
15	3-methylbutan-1-ol (isopentanol)	000123-51-3	OCCC(C)C	1	1	1.1	1.1
16	2-methyl-butan-1-ol	000137-32-6	OCC(CC)C	1	1	1.1	1.1
17	1-heptanol	000111-70-6	OCCCCCCC	1	1	1.1	1.1
18	2-ethylhexan-1-ol	000104-76-7	OCC(CCCC)CC	1	1	1.1	1.1
19	1-nonanol	000143-08-8	OCCCCCCCCC	1	1	1.1	1.1
20	1-decanol	000112-30-1	OCCCCCCCCCC	1	1	1.1	1.1
21	2-propylheptan-1-ol	010042-59-8	OCC(CCCCC)CCC	1	1	1.1	1.1
22	1-undecanol	000112-42-5	OCCCCCCCCCCC	1	1	1.1	1.1
23	5-cyclohexyl-2-methylpentan-1-ol (R/S)	1141487-54-8	CC(CCCC1CCCCC1)CO	1	1	1.1	1.1
24	2-butyloctan-1-ol	003913-02-8	OCC(CCCC)CCCCCC	1	1	1.1	1.1
25	isotridecanol	027458-92-0	OCCCCCCCCCCC(C)C	1	1	1.1	1.1
26	1-tetradecanol	000112-72-1	OCCCCCCCCCCCCCC	1	1	1.1	1.1
27	1-hexadecanol	036653-82-4	OCCCCCCCCCCCCCCCCC	1	5	1.1	1.1
28	1-octadecanol	000112-92-5	OCCCCCCCCCCCCCCCCCCC	1	5	1.1	1.1
29	3-methyl-2-buten-1-ol (prenol)	000556-82-1	OCC=C(C)C	1	3	1.1	1.1 / 4.3
30	3,7-dimethyl-2,6-octadien-1-ol (geraniol)	000106-24-1	OCC=C(CCC=C(C)C)C	1	3	1.1	1.1 / 4.3
31	3,7-dimethyloct-6-en-1-ol (dihydrogeraniol/citronellol)	000106-22-9/007540-51-4/026489-01-0	OCCC(CCC=C(C)C)C	1	5	1.1	1.1
32	CBI	CBI	CBI	1	5	1.1	1.1
33	sanderol (2-ethyl-4-(2,2,3-trimethylcyclopent-3-en-1-yl)but-(2E)-en-1-ol)	028219-61-6	OCC(=CCC(C(C(=C1)C)(C)C)C1)CC	1	3	1.1	1.1 / 4.3
34	2-phenylethanol	000060-12-8	OCCc1ccccc1	1	1	1.1	1.1
35	3-methyl-5-phenylpentan-1-ol (mefrosol)	055066-48-3	OCCC(CCc(cccc1)c1)C	1	1	1.1	1.1
36	CBI	CBI	CBI	1	1	1.1	1.1
37	cinnamyl alcohol (3-phenylprop-2-en-1-ol)	000104-54-1	OCC=Cc1ccccc1	1	3	1.1	1.1 / 4.3
38	2-methoxyethanol	000109-86-4	O(CCO)C	1	1	1.1	1.1 & c4.3

Code couleur : Vert = correctement classifié ; jaune = légèrement différent ; rouge = mal classifié ; gris = non classifié

Tableau S2 : Set de validation. Partie 1. (suite)

39	2-ethoxyethanol	000110-80-5	O(CCO)CC	1	1	1.1	1.1 & c4.3
40	2-propoxyethanol	002807-30-9	O(CCO)CCC	1	1	1.1	1.1 & c4.3
41	2-isopropoxyethanol	000109-59-1	O(C(C)C)CCO	1	1	1.1	1.1 & c4.3
42	3-methoxybutan-1-ol	002517-43-3	O(C(CCO)C)C	1	1	1.1	1.1
43	2-butoxyethanol	000111-76-2	O(CCCC)CCO	1	1	1.1	1.1 & c4.3
44	2-(2-methoxyethoxy)ethanol (diethylene glycol monoethyl ether) (diethylene glycol monomethyl ether) (methyl carbitol) (DEGME) (methyl diglycol ether)	000111-77-3	O(CCOC)CCO	1	5	1.1	1.1 & c4.3
45	2-(2-ethoxyethoxy)ethanol (diethylene glycol monoethyl ether)	000111-90-0	O(CCOCC)CCO	1	1	1.1	1.1 & c4.3
46	2-(hexyloxy)ethanol	000112-25-4	O(CCCCCC)CCO	1	1	1.1	1.1 & c4.3
47	2-(2-butoxyethoxy)ethanol	000112-34-5	O(CCOCCO)CCCC	1	1	1.1	1.1 & c4.3
48	2-(2-(2-methoxyethoxy)ethoxy)ethanol	000112-35-6	O(CCOCCOC)CCO	1	5	1.1	1.1 & c4.3
49	2-(2-(2-ethoxyethoxy)ethoxy)ethanol	000112-50-5	C(OCCOCC)COCCO	1	1	1.1	1.1 & c4.3
50	2-[2-(hexyloxy)ethoxy]ethanol	000112-59-4	O(CCOCCO)CCCCCC	1	1	1.1	1.1 & c4.3
51	2-[2-(2-butoxyethoxy)ethoxy]ethanol (TEGBE) (Triethylene glycol monobutyl ether) (Triethylene glycol butyl ether)	000143-22-6	O(CCOCCOCCO)CCCC	1	1	1.1	1.1 & c4.3
52	2,5,8,11-tetraoxatridecan-13-ol	023783-42-8	O(CCOCCOCCOC)CCO	1	5	1.1	1.1 & c4.3
53	2-(2-phenoxyethoxy)ethanol	000104-68-7	O(CCOc(cccc1)c1)CCO	1	1	1.1	1.1 & c4.3
54	2-(2-naphthyloxy)ethanol	000093-20-9	c12cc(OCCO)ccc1cccc2	1	1	1.1	1.1 & c4.3
55	2,2'-p-phenylenedioxydiethanol	000104-38-1	O(c(ccc(OCCO)c1)c1)CCO	1	1	1.1	1.1 & c4.3
56	nonylphenol, ethoxylated	009016-45-9	C1=CC(=CC=C1CCCCCCCCC)OCCOCCO	1	1	1.1	1.1 & c4.3
57	2,2'-isopropylidenebis(p-phenyleneoxy)diethanol	000901-44-0	O(c(ccc(c1)C(c(ccc(OCCO)c2)c2)(C)C)c1)CCO	1	1	1.1	1.1 & c4.3
58	2-(2-chloroethoxy)ethanol	000628-89-7	C(OCCCl)CO	1	3	1.1	1.1 & c4.3
59	ethylene glycol	000107-21-1	OCCO	1	5	1.1	1.1 & 4.5
60	propane-1,3- diol	000504-63-2	C(CO)CO	1	5	1.1	1.1
61	2,2-dimethylpropane-1,3-diol	000126-30-7	OCC(CO)(C)C	1	1	1.1	1.1
62	cyclohex-1,4-ylenedimethanol (cyclohexane-1,4-diyldimethanol)	000105-08-8	OCC(CCC(C1)CO)C1	1	5	1.1	1.1
63	tricyclodecanedimethanol (tricyclo[2.2.2.2^{1,4}]decane-2,5-diyldimethanol)	026896-48-0	C(C3C1C(C2C(CO)CC1C2)CC3)O	1	5	1.1	1.1
64	pentane-1,5-diol	000111-29-5	OCCCCCO	1	1	1.1	1.1
65	3-methylpentane-1,5-diol	004457-71-0	OCCC(CCO)C	1	1	1.1	1.1
66	hexane-1,6-diol	000629-11-8	OCCCCCCO	1	1	1.1	1.1
67	2-butyl-2-ethylpropane-1,3-diol	000115-84-4	C(C(CC)(CO)CO)CCC	1	1	1.1	1.1
68	diethylene glycol	000111-46-6	O(CCO)CCO	1	5	1.1	1.1 & 4.5
69	dipropylene glycol (mixture of 3 position isomers)	025265-71-8	OCCCOCCCO / CC(O)COCC(O)C / OCCCOCC(O)C	1	1	1.1	1.1
70	tripropylene glycol	024800-44-0	CC(O)COC(C)COC(C)CO	1	1	1.1	1.1
71	3,6,9-trioxaundecane-1,11-diol (2,2'-[oxybis(ethane-2,1-diyloxy)]diethanol) (tetraethylene glycol)	000112-60-7	O(CCOCCOCCO)CCO	1	5	1.1	1.1 & 4.5
72	2,2-bis(bromomethyl)propane-1,3-diol	003296-90-0	OCC(CBr)(CBr)CO	1	1	1.1	1.1
73	propylene glycol (propane-1,2-diol)	000057-55-6	OCC(O)C	1	1	1.1	1.1
74	butane-1,4-diol	000110-63-4	OCCCCO	1	1	1.1	1.1
75	pentane-1,2-diol	005343-92-0	OC(CCC)CO	1	1	1.1	1.1

76	1,2-octanediol	001117-86-8	OCC(O)CCCCCC	1	1	1.1	1.1
77	2-ethylhexane-1,3-diol	000094-96-2	OCC(C(O)CCC)CC	1	1	1.1	1.1
78	2,2,4-trimethylpentane-1,3-diol	000144-19-4	OCC(C(O)C(C)C)(C)C	1	1	1.1	1.1
79	3-chloropropane-1,2-diol	000096-24-2	OCC(O)CCl	1	3	1.1	1.1
80	2-ethyl-2-(hydroxymethyl)propane-1,3-diol, TMP (propylidynetrimethanol)	000077-99-6	OCC(CC)(CO)CO	1	1	1.1	1.1
81	2,2-bis(hydroxymethyl)-1,3-propanediol, pentaerythritol, pentaerythritol technical	000115-77-5	OCC(CO)(CO)CO	1	5	1.1	1.1
82	2,2'-[oxybis(methylene)]bis(2-ethylpropane-1,3-diol)	023235-61-2	O(CC(CC)(CO)CO)CC(CC)(CO)CO	1	1	1.1	1.1
83	2,2,2',2'-tetrakis(hydroxymethyl)-3,3'-oxydipropan-1-ol (2,2'-[oxybis(methylene)]bis[2-(hydroxymethyl)propane-1,3-diol]) (D-PE300)	000126-58-9	O(CC(CO)(CO)CO)CC(CO)(CO)CO	1	5	1.1	1.1
84	glycerol, propoxylated - UVCB (1 - 6.5 moles propoxylated)	025791-96-2	ND	1	1	1.1	1.1
85	glycerol (1,2,3-propanetriol)	000056-81-5	OCC(O)CO	1	5	1.1	1.1
86	glycerol (1,2,3-propanetriol, homopolymer)	025618-55-7	OCC(O)CO	1	5	1.1	1.1
87	(2R, 3R, 4S)-pentane-1,2,3,4,5-pentol (xylitol)	000087-99-0	OCC(O)C(O)C(O)CO	1	5	1.1	1.1
88	(3R,4S,5S,6R)-2-hexoxy-6-(hydroxymethyl)oxane-3,4,5-triol	054549-24-5	CCCCCCOC1C(C(C(C(O1)CO)O)O)O	1	1	1.1	1.1
89	4-methyl-2-pentanol	000108-11-2	OC(CC(C)C)C	1	1	1.1	1.1
90	cyclohexanol	000108-93-0	OC(CCCC1)C1	1	5	1.1	1.1
91	CBI	CBI	CBI	1	5	1.1	1.1
92	cyclododecanol	001724-39-6	OC(CCCCCCCCCC1)C1	1	5	1.1	1.1
93	2,2,4,4-tetramethylcyclobutane-1,3-diol, mixed isomers	003010-96-6	CC1(C(C(C1O)(C)C)O)C	1	5	1.1	1.1
94	4,4'-isopropylidenedicyclohexanol	000080-04-6	OC(CCC(C(C(CCC(O)C1)C1)(C)C)C2)C2	1	5	1.1	1.1
95	(3S,3aR,6R,6aR)-2,3,3a,5,6,6a-hexahydrofuro[5,4-d]furan-3,6-diol (isosorbide)	000652-67-5	[C@@H]12[C@@H]([C@@H](O)CO1)OC[C@H]2O	1	5	1.1	1.1
96	2-methylpentane-2,4-diol	000107-41-5	OC(CC(O)(C)C)C	1	1	1.1	1.1
97	(undecavertol) 4-methyl-3-decen-5-ol	081782-77-6	CCC=C(C(CCCCC)O)C	1	3	1.1	1.1
98	CBI	CBI	CBI	1	5	1.1	1.1
99	polysantol	107898-54-4	C1C=C(C)C(C)(C)C1C=CC(C)(C)C(O)C	1	5	1.1	1.1
100	(ebanol) 3-methyl-5-(2,2,3-trimethylcyclopent-3-en-1-yl)pent-4-en-2-ol	067801-20-1	OC(C(C=CC(C(C(=C1)C)(C)C)C1)C)C	1	5	1.1	1.1
101	1-methoxy-2-propanol	000107-98-2	O(CC(O)C)C	1	1	1.1	1.1
102	exo-1-methyl-4-(1-methylethyl)-7-oxabicyclo[2.2.1]heptan-2-ol	087172-89-2	CC(C)C12OC(C(O)C1)(CC2)C	1	1	1.1	1.1
103	1,1'-[propane-2,2-diylbis(4,1-phenyleneoxy)]dipropan-2-ol	000116-37-0	O(c(ccc(c1)C(c(ccc(OCC(O)C)c2)c2)(C)C)c1)CC(O)C	1	1	1.1	1.1
104	1,3-dichloropropan-2-ol	000096-23-1	OC(CCl)CCl	1	3	1.1	1.1
105	2-methyl-butan-2-ol	000075-85-4	OC(CC)(C)C	1	1	1.1	1.1
106	3,7-dimethyloctan-3-ol (tetrahydrolinalool)	000078-69-3	OC(CCCC(C)C)(CC)C	1	1	1.1	1.1
107	2-methyl-3-buten-2-ol	000115-18-4	OC(C=C)(C)C	1	3	1.1	1.1
108	terpineol (mixture of position and stereo- isomers)	008000-41-7	OC(C1CC=C(CC1)C)(C)C	1	5	1.1	1.1
109	linalool	000078-70-6	OC(C=C)(CCC=C(C)C)C	1	3	1.1	1.1
110	3,7-dimethylnona-1,6-dien-3-ol (ethyllinalool)	010339-55-6	OC(C=C)(CCC=C(CC)C)C	1	3	1.1	1.1
111	CBI	CBI	CBI	1	3	1.1	1.1
112	1-ethynylcyclohexanol	000078-27-3	OC(C#C)(CCCC1)C1	1	5	1.1	1.1
113	2,5-dimethylhexane-2,5-diol	000110-03-2	OC(CCC(O)(C)C)(C)C	1	1	1.1	1.1
114	formaldehyde	000050-00-0	O=C	3	3	3.1	3.1
115	butanal (butyraldehyde)	000123-72-8	O=CCCC	3	3	3.1	3.1
116	isobutyraldehyde (2-methylpropanal)	000078-84-2	O=CC(C)C	3	3	3.1	3.1
117	isovaleraldehyde (3-methylbutanal)	000590-86-3	O=CCC(C)C	3	3	3.1	3.1
118	pentanal (valeraldehyde)	000110-62-3	O=CCCCC	3	3	3.1	3.1

119	2-methylpentanal (2-methylvaleraldehyde)	000123-15-9	O=CC(CCC)C	3	3	3.1	3.1
120	heptanal	000111-71-7	O=CCCCCCC	3	3	3.1	3.1
121	2-ethylhexanal	000123-05-7	O=CC(CCCC)CC	3	3	3.1	3.1
122	decanal	000112-31-2	O=CCCCCCCCCC	3	3	3.1	3.1
123	2-methylundecanal	000110-41-8	O=CC(CCCCCCCCC)C	3	3	3.1	3.1
124	dodecanal	000112-54-9	O=CCCCCCCCCCCC	3	3	3.1	3.1
125	3-methyl-2-butenal	000107-86-8	O=CC=C(C)C	3	3	3.2	3.2
126	2-methyl-2-pentenal	000623-36-9	O=CC(C)=CCC	3	3	3.2	3.2
127	2-ethylhex-2-enal	000645-62-5	O=CC(=CCCC)CC	3	3	3.2	3.2
128	citral (mixture of (Z)- and (E)-3,7-dimethyl-2,6-octadienal)	005392-40-5	O=CC=C(CCC=C(C)C)C	1	3	3.2	3.2 / 4.1 / 1.1
129	CBI	CBI	CBI	3	3	3.1	3.1
130	cyclohex-3-ene-1-carbaldehyde	000100-50-5	O=CC(CCC=C1)C1	3	3	3.1	3.1
131	citronellal	000106-23-0	O=CCC(CCC=C(C)C)C	1	3	3.1	3.1 / 4.1
132	n-pentane	000109-66-0	C(CCC)C	1	1	1.1	1.1
133	methylcyclohexane	000108-87-2	C(CCCC1)(C1)C	1	1	1.1	1.1
134	bicyclo[4,4,0]decane (decahydronaphthalene)	000091-17-8	C(C(CCC1)CCC2)(C1)C2	1	1	1.1	1.1
135	1-isopropyl-4-methylcyclohexane (p-menthane)	000099-82-1	CC(C)C1CCC(C)CC1	1	1	1.1	1.1
136	n-decane	000124-18-5	C(CCCCCCCC)C	1	1	1.1	1.1
137	n-undecane	001120-21-4	C(CCCCCCCCC)C	1	1	1.1	1.1
138	cyclododecane	000294-62-2	C1CCCCCCCCCCC1	1	5	1.1	1.1
139	n-dodecane	000112-40-3	C(CCCCCCCCCC)C	1	5	1.1	1.1
140	n-tridecane	000629-50-5	CCCCCCCCCCCCCC	1	5	1.1	1.1
141	n-tetradecane	000629-59-4	CCCCCCCCCCCCCC	1	5	1.1	1.1
142	n-pentadecane	000629-62-9	CCCCCCCCCCCCCCC	1	5	1.1	1.1
143	n-hexadecane	000544-76-3	CCCCCCCCCCCCCCCC	1	5	1.1	1.1
144	n-heptadecane	000629-78-7	CCCCCCCCCCCCCCCCC	1	5	1.1	1.1
145	n-octadecane	000593-45-3	CCCCCCCCCCCCCCCCCC	1	5	1.1	1.1
146	n-nonadecane	000629-92-5	CCCCCCCCCCCCCCCCCCC	1	5	1.1	1.1
147	n-eicosane	000112-95-8	CCCCCCCCCCCCCCCCCCCC	1	5	1.1	1.1
148	1-pentene	000109-67-1	C(=C)CCC	1	1	1.1	1.1
149	2-methyl-2-butene	000513-35-9	C(=CC)(C)C	1	1	1.1	1.1
150	cyclohexene	000110-83-8	C(=CCCC1)C1	1	1	1.1	1.1
151	4-methyl-1-pentene	000691-37-2	C(=C)CC(C)C	1	1	1.1	1.1
152	1-hexene	000592-41-6	C(=C)CCCC	1	1	1.1	1.1
153	8,9,10-trinorborn-2-ene (norborn-2-ene)	000498-66-8	C(C=CC1C2)(C1)C2	1	1	1.1	1.1
154	(Z)-cyclooctene	000931-87-3	C1=CCCCCCC1	1	1	1.1	1.1
155	2,4,4-trimethylpentene (isooctene)	025167-70-8	C=C(C)CC(C)(C)C	1	1	1.1	1.1
156	1-octene	000111-66-0	C(=C)CCCCCC	1	1	1.1	1.1
157	isooctene	011071-47-9	C(=C)CCCCCC	1	1	1.1	1.1
158	CBI	CBI	CBI	1	1	1.1	1.1
159	CBI	CBI	CBI	1	1	1.1	1.1
160	camphene	000079-92-5	C=C(C12)C(C)(C)C(C1)CC2	1	1	1.1	1.1
161	3-carene (delta)	013466-78-9	C(=CCC(C12)C1(C)C)(C2)C	1	1	1.1	1.1
162	1-decene	000872-05-9	CCCCCCCCC=C	1	1	1.1	1.1
163	1-dodecene	000112-41-4	CCCCCCCCCCC=C	1	5	1.1	1.1
164	1-tetradecene	001120-36-1	CCCCCCCCCCCCC=C	1	5	1.1	1.1
165	1-hexadecene	000629-73-2	C(=C)CCCCCCCCCCCCCC	1	5	1.1	1.1

Tableau S2 : Set de validation. Partie 1. (suite)

166	allyl bromide (3-bromopropene)	000106-95-6	BrCC=C	3	3	3.1 & 4.1	3.1 & 4.1
167	1,1-dichloroethene	000075-35-4	C(=C)(Cl)Cl	1	1	1.1	1.1
168	isoprene	000078-79-5	C=C(C)C=C	1	1	1.1	1.1
169	1,3,5-cycloheptatriene	000544-25-2	C1C=CC=CC=C1	1	1	1.1	1.1
170	4-vinylcyclohexene	000100-40-3	C(=CCCC1C=C)C1	1	1	1.1	1.1
171	cycloocta-1,5-diene	000111-78-4	C1=CCCC=CCC1	1	1	1.1	1.1
172	octa-1,7-diene	003710-30-3	C=CCCCCC=C	1	1	1.1	1.1
173	5-ethylidene-2-norbornene	016219-75-3	C(C(C=CC12)C1)(=CC)C2	1	1	1.1	1.1
174	7-methylocta-1,6-diene	042152-47-6	C=CCCCC=C(C)C	1	1	1.1	1.1
175	CBI	CBI	CBI	1	1	1.1	1.1
176	p-mentha-1,4(8)-diene (terpinolene)	000586-62-9	C(=C(C)C)(CCC(=C1)C)C1	1	1	1.1	1.1
177	alpha-terpinene	000099-86-5	C(=CC=C(C1)C)(C1)C(C)C	1	1	1.1	1.1
178	gamma-terpinene	000099-85-4	C(=CCC(=C1)C)(C1)C(C)C	1	1	1.1	1.1
179	myrcene	000123-35-3	C(C=C)(=C)CCC=C(C)C	1	1	1.1	1.1
180	cyclododeca-1,5,9-triene	004904-61-4	C1C=CCCC=CCCC=CC1	1	1	1.1	1.1
181	1,3,5-trimethylbenzene (mesitylene)	000108-67-8	c(cc(cc1C)C)(c1)C	1	1	1.1	1.1
182	isopropylbenzene (cumene)	000098-82-8	c(cccc1)(c1)C(C)C	1	1	1.1	1.1
183	1,2,4-trimethylbenzene	000095-63-6	c(ccc(c1C)C)(c1)C	1	1	1.1	1.1
184	p-diethylbenzene	000105-05-5	c(ccc(c1)CC)(c1)CC	1	1	1.1	1.1
185	m-diethylbenzene	000141-93-5	c(cccc1CC)(c1)CC	1	1	1.1	1.1
186	n-pentylbenzene	000538-68-1	c(cccc1)(c1)CCCCC	1	1	1.1	1.1
187	biphenyl	000092-52-4	c(c(cccc1)c1)(cccc2)c2	1	1	1.1	1.1
188	cyclohexylbenzene	000827-52-1	c(cccc1)(c1)C(CCCC2)C2	1	1	1.1	1.1
189	1,3-diisopropylbenzene	000099-62-7	c(cccc1C(C)C)(c1)C(C)C	1	1	1.1	1.1
190	1,4-diisopropylbenzene	000100-18-5	c(ccc(c1)C(C)C)(c1)C(C)C	1	1	1.1	1.1
191	diphenylmethane	000101-81-5	c(cccc1)(c1)Cc(cccc2)c2	1	1	1.1	1.1
192	acenaphthylene	000208-96-8	c1ccc2cccc3c2c1C=C3	1	1	1.1 & m4.3	1.1 & m4.3
193	9H-fluorene (fluorene)	000086-73-7	c(c(c(c1ccc2)c2)ccc3)(c3)C1	1	1	1.1 & m4.3	1.1 & m4.3
194	anthracene	000120-12-7	c(c(ccc1)cc(c2ccc3)c3)(c1)c2	1	1	1.1 & m4.3	1.1 & m4.3
195	phenanthrene	000085-01-8	c(c(c(c(c1)ccc2)c2)ccc3)(c1)c3	1	1	1.1 & m4.3	1.1 & m4.3
196	fluoranthene	000206-44-0	c(c(ccc1)ccc2)(c1c(c3ccc4)c4)c23	1	1	1.1 & m4.3	1.1 & m4.3
197	benzofuran	000271-89-6	c1ccc2ccoc2c1	1	5	1.1	1.1 & m4.3
198	1,1'-(4-methylpent-1-ene-2,4-diyl)dibenzene	006362-80-7	c(cccc1)(c1)C(CC(c(cccc2)c2)=C)(C)C	1	1	1.1	1.1
199	o-fluorotoluene	000095-52-3	Fc(c(ccc1)C)c1	1	1	1.1	1.1
200	2-chlorotoluene	000095-49-8	c(c(ccc1)Cl)(c1)C	1	1	1.1	1.1
201	p-chlorotoluene	000106-43-4	c(ccc(c1)Cl)(c1)C	1	1	1.1	1.1
202	3,4-dimethylchlorobenzene (4-chloro-o-xylene)	000615-60-1	Clc1cc(C)c(C)cc1	1	1	1.1	1.1
203	1-bromo-4-propylbenzene	000588-93-2	c(ccc(Br)c1)(c1)CCC	1	1	1.1	1.1
204	1,3-dichlorobenzene	000541-73-1	c(cccc1Cl)(c1)Cl	1	1	1.1	1.1
205	1,2-dichlorobenzene	000095-50-1	c(c(ccc1)Cl)(c1)Cl	1	1	1.1	1.1
206	1,4-dichlorobenzene	000106-46-7	c(ccc(c1)Cl)(c1)Cl	1	1	1.1	1.1
207	2,4-dichlorotoluene	000095-73-8	c(ccc(c1Cl)C)(c1)Cl	1	1	1.1	1.1
208	3,4-dichlorotoluene	000095-75-0	c(ccc(c1Cl)Cl)(c1)C	1	1	1.1	1.1
209	4-chloro-α,α,α-trifluorotoluene	000098-56-6	FC(F)(F)c(ccc(c1)Cl)c1	1	5	1.1	1.1
210	1,2,3-trichlorobenzene	000087-61-6	c(c(c(cc1)Cl)Cl)(c1)Cl	1	1	1.1	1.1
211	1,2,4-trichlorobenzene	000120-82-1	c(ccc(c1Cl)Cl)(c1)Cl	1	1	1.1	1.1
212	methyl acetate	000079-20-9	O=C(OC)C	1	5	2.1	2.1

Tableau S2 : Set de validation. Partie 1. (suite)

213	propyl acetate	000109-60-4	O=C(OCCC)C	1	5	2.1	2.1
214	isopropyl acetate	000108-21-4	O=C(OC(C)C)C	1	5	2.1	2.1
215	n-butyl acetate	000123-86-4	O=C(OCCCC)C	1	5	2.1	2.1
216	isobutyl acetate	000110-19-0	O=C(OCC(C)C)C	1	5	2.1	2.1
217	methyl hexanoate	000106-70-7	O=C(OC)CCCCC	1	5	2.1	2.1
218	isoamyl acetate	000123-92-2	O=C(OCCC(C)C)C	1	5	2.1	2.1
219	n-hexylacetate	000142-92-7	O=C(OCCCCCC)C	1	5	2.1	2.1
220	n-butyl butyrate	000109-21-7	O=C(OCCCC)CCC	1	5	2.1	2.1
221	isobutyl butyrate	000097-85-8	O=C(OCC(C)C)C(C)C	1	5	2.1	2.1
222	amyl propionate (n-pentyl propionate)	000624-54-4	O=C(OCCCCC)CC	1	5	2.1	2.1
223	ethyl hexanoate	000123-66-0	O=C(OCC)CCCCC	1	5	2.1	2.1
224	2-methylcyclohexyl acetate	005726-19-2	CC(OC1C(C)CCCC1)=O	1	5	2.1	2.1
225	2-ethylhexyl acetate	000103-09-3	O=C(OCC(CCCC)CC)C	1	5	2.1	2.1
226	3,5,5-trimethylhexyl acetate	058430-94-7	O=C(OCCC(CC(C)(C)C)C)C	1	5	2.1	2.1
227	bornyl acetate (pichtosin)	000076-49-3/000125-12-2	O=C(OC(C(C(C1C2)(C)C)(C2)C)C1)C	1	5	2.1	2.1
228	2-tert-butylcyclohexyl acetate	020298-69-5	O=C(OC(C(C(C)(C)C)CCC1)C1)C	1	5	2.1	2.1
229	menthyl acetate	000089-48-5/016409-45-3	O=C(OC(C(CCC1C)C(C)C)C1)C	1	5	2.1	2.1
230	4-tert-butylcyclohexyl acetate (dorisyl)	032210-23-4	O=C(OC(CCC(C(C)(C)C)C1)C1)C	1	5	2.1	2.1
231	CBI	CBI	CBI	1	5	2.1	2.1
232	tetradecyl myristate	003234-85-3	O=C(OCCCCCCCCCCCCCC)CCCCCCCCCCCCCC	1	5	2.1	2.1
233	ethyl methacrylate	000097-63-2	O=C(OCC)C(=C)C	1	3	3.2	3.2 & 4.1
234	n-butyl methacrylate	000097-88-1	O=C(OCCCC)C(=C)C	1	3	3.2	3.2 & 4.1
235	i-butyl methacrylate	000097-86-9	O=C(OCC(C)C)C(=C)C	1	3	3.2	3.2 & 4.1
236	2-ethylhexyl methacrylate	000688-84-6	CCCCC(CC)COC(=O)C(=C)C	1	3	3.2	3.2 & 4.1
237	iso-decyl methacrylate	000007-21-0/029964-84-9	C=C(C)C(=O)OCCCCCCCC(C)C	1	3	3.2	3.2 & 4.1
238	2-hydroxyethyl methacrylate (HEMA)	000868-77-9	O=C(OCCO)C(=C)C	1	3	3.2	3.2 & 4.1
239	ethylene dimethacrylate	000097-90-5	CC(=C)C(=O)OCCOC(=O)C(=C)C	1	3	3.2	3.2 & 4.1
240	1-methyltrimethylene dimethacrylate	001189-08-8	O=C(OCCC(OC(=O)C(=C)C)C)C(=C)C	1	3	3.2	3.2 & 4.1
241	propylidynetrimethyl trimethacrylate	003290-92-4	O=C(OCC(CC)(COC(=O)C(=C)C)COC(=O)C(=C)C)C(=C)C	1	3	3.2	3.2 & 4.1
242	vinyl neodecanoate	051000-52-3	O=C(OC=C)CCCCCC(C)(C)C	3	5	1.1 & c4.3	4.3 / 2.2
243	3-hydroxy-2,2,4-trimethylpentyl 2-methylpropanoate (texanol)	025265-77-4	O=C(OCC(C(C(C)C)O)(C)C)C(C)C	1	1	2.1	2.1
244	propyl L-lactate (propyl (2S)-2-hydroxypropanoate)	053651-69-7	CCCOC(=O)C(C)O	1	3	2.1	2.1
245	butyl lactate	000138-22-7	O=C(OCCCC)C(O)C	1	3	2.1	2.1

Tableau S2 : Set de validation. Partie 1. (suite)

246	octyl lactate	051191-33-4	O=C(C(O)C)OCCCCCCCC	1	3	2.1	2.1
247	2-ethylhexyl 2-hydroxypropanoate (2-ethylhexyl lactate)	006283-86-9	OC(C)C(=O)OCC(CCCC)CC	1	3	2.1	2.1
248	propanoic acid, 2-hydroxy-, (1R,2S,5R)-5-methyl-2-(1-methylethyl)cyclohexyl ester, (2S)- (L-menthyl lactate)	061597-98-6	CC1CCC(C(C1)OC(=O)C(C)O)C(C)C	1	3	2.1	2.1
249	diketene (4-methyleneoxetan-2-one / but-3-en-3-olide)	000674-82-8	O=C(OC1=C)C1	1	5	2.1	2.1
250	γ-butyrolactone (oxolan-2-one)	000096-48-0	O=C(OCC1)C1	1	5	2.1	2.1
251	tetrahydro-2H-pyran-2-one (δ-valerolactone)	000542-28-9	O=C(OC1)CCC1	1	5	2.1	2.1
252	decan-5-olide	000705-86-2	O=C(O1)CCCC1CCCCC	1	5	2.1	2.1
253	ε-caprolactone (hexan-6-olide)	000502-44-3	O=C(OC1)CCCC1	1	5	2.1	2.1
254	α-hydroxy-β,β-dimethyl-γ-butyrolactone	000599-04-2	O=C(OCC1(C)C)C1O	1	5	2.1	2.1
255	(1R, 4R, 5S) 4-hydroxy-6,6-dimethyl-3-Oxabicyclo[3.1.0]hexan-2-one	073611-02-6	OC2C1C(C1C(O2)=O)(C)C	1	5	2.1	2.1
256	3-hexyl-4-hydroxy-6-undecyl-tetrahydro-pyran-2-one	136890-12-5	C(CCCCC)C1C(CC(OC1=O)CCCCCCCCCCC)O	1	5	2.1	2.1
257	CBI	CBI	CBI	4	5	2.1	2.1 / 6.7?
258	(4E)-4-(C13-C17)alkylidene-3-(C12-C16)alkyloxetan-2-one	084989-41-3	O=C1OC(CCCCCCCCCCCCC)C1=CCCCCCCCCCCC	1	5	3.2	2.1
259	(1R,4R,5S)-6,6-dimethyl-4-(tribromomethyl)-3-oxabicyclo(3.1.0)hexan-2-one	087305-03-1	CC1(C2C(OC(C12)C(Br)(Br)Br)=O)C	4	5	2.1	2.1 / 2.2? / 6.7?
260	α-acetyl-γ-butyrolactone	000517-23-7	CC(C1C(OCC1)=O)=O	1	1	6.4	2.1 / 6.4
261	10,13-dimethyl-1,8,9,10,11,12,13,14,15,16-decahydro-3'H-spiro[cyclopenta[a]phenanthrene-17,2'-furan]-3,5'(2H,4'H)-dione (canrenone)	000976-71-6	C1C(=O)C=C2C=CC3C4CCC5(OC(=O)CC5)C4(C)CCC3C2(C)C1	4	5	2.1	2.1 & c6.7
262	1,4-dioxacyclohexadecane-5,16-dione	054982-83-1	O=C(OCCOC(=O)CCCCCCCCC1)C1	1	5	2.1	2.1
263	1,4-dioxacycloheptadecane-5,17-dione (astrotone)	000105-95-3	O=C(OCCOC(=O)CCCCCCCCCCC1)C1	1	5	2.1	2.1
264	firascone - ((1RS, 2SR) (+/-) methyl trans 2,6,6,-trimethyl-3-cyclohexene-1-carboxylate)	ND	O(C(=O)C1C(C=CCC1(C)C)C)C	1	5	2.1	2.1
265	α-terpinyl acetate	000080-26-2	O=C(OC(C(CCC(=C1)C)C1)(C)C)C	1	5	2.1	2.1
266	methyl undec-10-enoate	000111-81-9	O=C(OC)CCCCCCCCC=C	1	5	2.1	2.1
267	3a,4,5,6,7,7a-hexahydro-4,7-methano-1H-indenyl propionate (cyclaprop)	068912-13-0	C3C1C2C(C=CC2C3(CC1))OC(=O)CC	1	5	3.1 & 4.1	2.1
268	nopyl acetate	000128-51-8	O=C(OCCC(C(CC1C2)C1(C)C)=C2)C	1	5	2.1	2.1
269	linalyl acetate	000115-95-7	O=C(OC(C=C)(CCC=C(C)C)C)C	1	5	3.1 & 4.1	2.1
270	1-phenylethyl acetate	000093-92-5	O=C(OC(c(cccc1)c1)C)C	1	5	3.1 & 4.1	2.1
271	2-methoxy-1-methylethyl acetate (propylene glycol methyl ether acetate)	000108-65-6	COCC(OC(=O)C)C	1	1	2.1	2.1
272	CBI	CBI	CBI	1	1	2.1	2.1?
273	CBI	CBI	CBI	1	1	2.1	2.1
274	methyl 3-oxo-2-pentylcyclopentaneacetate (hedione)	024851-98-7	O=C(OC)CC(C(C(=O)C1)CCCCC)C1	1	1	2.1	2.1
275	dimethyl succinate	000106-65-0	O=C(OC)CCC(=O)OC	3	5	2.1	2.1
276	dimethyl glutarate	001119-40-0	O=C(OC)CCCC(=O)OC	3	5	2.1	2.1
277	dimethyl 2-acetylsuccinate	010420-33-4	O=C(OC)C(C(=O)C)CC(=O)OC	1	3	6.4	2.1?
278	triacetin	000102-76-1	O=C(OCC(OC(=O)C)COC(=O)C)C	3	5	2.1	2.1
279	fructalate (diethyl 1,4-cyclohexanedicarboxylate)	072903-27-6	O=C(C1CCC(CC1)C(=O)OCC)OCC	3	5	2.1	2.1
280	dimethyl sebacate	000106-79-6	O=C(OC)CCCCCCCCC(=O)OC	3	5	2.1	2.1

Tableau S2 : Set de validation. Partie 1. (suite)

281	diisopropyl adipate	006938-94-9	O=C(OC(C)C)CCCCC(=O)OC(C)C	3	5	2.1	2.1
282	dibutyl adipate	000105-99-7	O=C(OCCCC)CCCCC(=O)OCCCC	3	5	2.1	2.1
283	romandolide	236391-76-7	CCC(=O)(OCC(=O)(OC(C)C1CCCC(C)(C)C1))	3	5	2.1	2.1
284	1-isopropyl-2,2-dimethyltrimethylene diisobutyrate (Eastman TXIB TM)	006846-50-0	O=C(OCC(C)(C)C(OC(=O)C(C)C)C(C)C)C(C)C	3	5	2.1	2.1
285	dihexyl adipate	000110-33-8	O=C(OCCCCCC)CCCCC(=O)OCCCCCC	3	5	2.1	2.1
286	dibutyl sebacate	000109-43-3	O=C(OCCCC)CCCCCCCCC(=O)OCCCCC	3	5	2.1	2.1
287	bis(2-ethylhexyl) succinate	002915-57-3	O=C(OCC(CCCC)CC)CCC(=O)OCC(CCCC)CC	3	5	2.1	2.1
288	bis(2-ethylhexyl) adipate	000103-23-1	O=C(OCC(CCCC)CC)CCCCC(=O)OCC(CCCC)CC	3	5	2.1	2.1
289	bis(2-ethylhexyl) sebacate	000122-62-3	O=C(OCC(CCCC)CC)CCCCCCCCC(=O)OCC(CCCC)CC	3	5	2.1	2.1
290	diisononyl adipate	000151-32-6	O=C(OCCCCCCCCC)CCCCC(=O)OCCCCCCCCC	3	5	2.1	2.1
291	ditridecyl adipate (UVCB)	016958-92-2	O=C(OCCCCCCCCCCCCC)CCCCC(=O)OCCCCCCCCCCCCC	3	5	2.1	2.1
292	diisotridecyl adipate	026401-35-4	CC(C)CCCCCCCCCCOC(=O)CCCCC(=O)OCCCCCCCCCCC(C)C	3	5	2.1	2.1
293	dibutyl fumarate (dibutyl (E)-but-2-enedioate)	000105-75-9	O=C(OCCCC)C=CC(=O)OCCCC	3	3	3.2	3.2
294	dibutyl maleate (dibutyl (Z)-but-2-enedioate)	000105-76-0	O=C(OCCCC)C=CC(=O)OCCCC	3	3	3.2	2.1?
295	2-(acryloyloxy)ethyl hydrogen succinate (MAES)	050940-49-3	O=C(C=C)OCCOC(=O)CCC(=O)O	3	3	3.2	3.2
296	bis(2,2,6,6-tetramethylpiperidin-4-yl) sebacate (Tinuvin 770)	052829-07-9	O=C(OC(CC(NC1(C)C)(C)C)C1)CCCCCCCC(=O)OC(CC(NC2(C)C)(C)C)C2	3	5	2.1	2.1
297	bis[2-(2-butoxyethoxy)ethyl] adipate	000141-17-3	O=C(OCCOCCOCCCC)CCCCC(=O)OCCOCCOCCCC	3	1	2.1	2.1
298	dimethyl 2,5-dioxocyclohexane-1,4-dicarboxylate	006289-46-9	O=C(OC)C(C(=O)CC(C1=O)C(=O)OC)C1	1	5	6.4	2.1?
299	2,2-dimethylpropane-1,3-diyl dibenzoate	004196-89-8	O=C(OCC(C)(C)COC(=O)c(cccc1)c1)c(cccc2)c2	3	5	2.1	2.1
300	oxydipropyl dibenzoate	027138-31-4	O=C(c1ccccc1)OCC(OC(COC(=O)c2ccccc2)C)C	3	1	2.1	2.1
301	dibutyl phthalate	000084-74-2	O=C(OCCCC)c(c(ccc1)C(=O)OCCCC)c1	1	5	2.1 & mC2.2	2.1 & mC2.2
302	diisobutyl phthalate	000084-69-5/003126-90-7	O=C(OCC(C)C)c(c(ccc1)C(=O)OCC(C)C)c1	1	5	2.1 & mC2.2	2.1 & mC2.2
303	dibutyl terephthalate	001962-75-0	CCCCOC(=O)c1ccc(cc1)C(=O)OCCCC	1	5	2.1	2.1
304	butyl benzyl phthalate	000085-68-7	O=C(OCc(cccc1)c1)c(c(ccc2)C(=O)OCCCC)c2	1	5	2.1 & mC2.2	2.1 & mC2.2
305	dicyclohexyl phthalate	000084-61-7	O=C(OC(CCCC1)C1)c(c(ccc2)C(=O)OC(CCCC3)C3)c2	1	5	2.1 & mC2.2	2.1 & mC2.2
306	bis(2-ethylhexyl) phthalate	000117-81-7	O=C(OCC(CCCC)CC)c(c(ccc1)C(=O)OCC(CCCC)CC)c1	1	5	2.1 & mC2.2	2.1 & mC2.2
307	bis(2-ethylhexyl) terephthalate	006422-86-2	O=C(OCC(CCCC)CC)c(ccc(c1)C(=O)OCC(CCCC)CC)c1	1	5	2.1	2.1
308	2-(2-hydroxyethoxy)ethyl 2-hydroxypropyl 3,4,5,6-tetrabromophthalate	020566-35-2	O=C(OCCOCCO)c(c(c(c(c1Br)Br)Br)C(=O)OCC(O)C)c1Br	1	5	2.1 & mC2.2	2.1 & mC2.2
309	dimethyl 2-nitroterephthalate	005292-45-5	O=C(OC)c(ccc(c1N(=O)=O)C(=O)OC)c1	3	5	2.1	2.1
310	2-methoxy-2-methylbutane (*tert*-amyl methyl ether)	000994-05-8	COC(C)(C)CC	1	1	1.1	1.1
311	ethyl t-butyl ether	000637-92-3	O(C(C)(C)C)CC	1	1	1.1	1.1
312	di-n-butyl ether	000142-96-1	O(CCCC)CCCC	1	1	1.1	1.1
313	cyclopentyl methyl ether (NZE)	005614-37-9	C1CCCC1OC	1	1	1.1	1.1
314	herbal heptane (5-ethyl-2-methoxybicyclo[2.2.1]heptane)	122795-41-9	COC1CC2C(CC)CC1C2	1	1	1.1	1.1

Tableau S2 : Set de validation. Partie 1. (suite)

315	2,3-dihydrobenzofuran	000496-16-2	O(c(c(ccc1)C2)c1)C2	1	1	1.1	1.1
316	p-methylanisole	000104-93-8	O(c(ccc(c1)C)c1)C	1	1	1.1	1.1
317	benzylmethyl ether	000538-86-3	O(Cc(cccc1)c1)C	1	1	1.1	1.1
318	2-phenoxyethanol	000122-99-6	O(c(cccc1)c1)CCO	1	1	1.1	1.1
319	anethole	000104-46-1	O(c(ccc(c1)C=CC)c1)C	1	1	1.1	1.1
320	5-{[2-(2-butoxyethoxy)ethoxy]methyl}-6-propyl-1,3-benzodioxole (piperonyl butoxide)	000051-03-6	O(c(c(O1)cc(c2CCC)COCCOCCOCCCC)c2)C1	1	1	1.1	1.1
321	dibenzyl ether	000103-50-4	O(Cc(cccc1)c1)Cc(cccc2)c2	1	1	1.1	1.1
322	PBDE (pentabromo)	032534-81-9	Brc1cc(c(cc1Oc2c(cc(cc2)Br)Br)Br)Br	4	1	1.1	6.7
323	propylene oxide	000075-56-9	O(C1C)C1	3	3	3.1	3.1
324	1,2-epoxybutane (butylene oxide)	000106-88-7	CCC1OC1	3	3	3.1	3.1
325	2-propyloxirane (pentylene oxide)	001003-14-1	O(C1)C1CCC	3	3	3.1	3.1
326	2-octyloxirane (decylene oxide)	002404-44-6	O(C1CCCCCCCC)C1	3	3	3.1	3.1
327	allyl 2,3-epoxypropyl ether	000106-92-3	O(C1COCC=C)C1	3	3	3.1	3.1
328	2,3-epoxypropan-1-ol	000556-52-5	O(C1CO)C1	3	5	3.1	3.1
329	(epoxyethyl)benzene	000096-09-3	O(C1c(cccc2)c2)C1	3	3	3.1	3.1
330	2,3-epoxypropyl phenyl ether	000122-60-1	O(C1COc(cccc2)c2)C1	3	3	3.1	3.1
331	2,3-epoxypropyl o-tolyl ether	002210-79-9	O(C1COc(c(ccc2)C)c2)C1	3	3	3.1	3.1
332	p-tert-butylphenyl 1-(2,3-epoxy)propyl ether	003101-60-8	O(C1COc(ccc(c2)C(C)(C)C)c2)C1	3	3	3.1	3.1
333	ethyl 2,3-epoxy-3-phenylbutyrate	000077-83-8	O=C(OCC)C(O1)C1(c(cccc2)c2)C	3	3	3.1	3.1 / 2.1
334	oxiran-2-ylmethyl 2-ethyl-2,5-dimethylhexanoate	026761-45-5	O=C(OCC(O1)C1)C(CC)(C)CC(C)C	3	3	3.1	3.1 / 2.1
335	bis(oxiran-2-ylmethyl) terephthalate	007195-44-0	O1C(COC(=O)c2ccc(C(=O)OCC3OC3)cc2)C1	3	3	3.1	3.1 / 2.1
336	tris(oxiran-2-ylmethyl) benzene-1,2,4-tricarboxylate	007237-83-4	O=C(OCC(O1)C1)c(ccc(c2C(=O)OCC(O3)C3)C(=O)OCC(O4)C4)c2	3	3	3.1	3.1 / 2.1
337	2,3-epoxypropyl methacrylate	000106-91-2	O=C(OCC(O1)C1)C(=C)C	3	3	3.2	3.2 & 3.1
338	1-chloro-2,3-epoxypropane (epichlorohydrin)	000106-89-8	O(C1CCl)C1	3	3	3.1	3.1
339	1,4-butanediol, reaction product with 1-chloro-2,3-epoxypropane - UVCB	002425-79-8	O(C1COCCCCOCC(O2)C2)C1	3	3	3.1	3.1
340	reaction products of hexane-1,6-diol with 2-(chloromethyl)oxirane (1:2) - 1,6-bis(2,3-epoxypropoxy)hexane - UVCB	016096-31-4	O1C(COCCCCCCOCC2OC2)C1	3	3	3.1	3.1
341	3-(oxiran-2-ylmethoxy)-N,N-bis(oxiran-2-ylmethyl)aniline	071604-74-5	O(C1CN(c(cc(OCC(O2)C2)cc3)c3)CC(O4)C4)C1	3	3	3.1	3.1
342	4-(oxiran-2-ylmethoxy)-N,N-bis(oxiran-2-ylmethyl)aniline	005026-74-4	O(C1CN(c(ccc(OCC(O2)C2)c3)c3)CC(O4)C4)C1	3	3	3.1	3.1
343	2-methyltetrahydrofuran	000096-47-9	O(C(CC1)C)C1	1	1	1.1	1.1
344	CBI	CBI	CBI	1	1	1.1	1.1
345	isobutyl vinyl ether	000109-53-5	O(C=C)CC(C)C	1	1	1.1 & c4.3	1.1 & c4.3
346	bis(2-chloroethyl)ether	000111-44-4	O(CCCl)CCCl	1	3	1.1	1.1
347	diethoxymethane	000462-95-3	CCOCOCC	1	1	1.1	1.1
348	2-ethoxyethylether	000112-36-7	CCOCCOCCOCC	1	1	1.1	1.1

Tableau S2 : Set de validation. Partie 1. (suite)

349	p-(dimethoxymethyl)anisole	002186-92-7	O(C(OC)c(ccc(OC)c1)c1)C	1	1	1.1	1.1
350	bis(2-chloroethoxy)methane	000111-91-1	O(CCCl)COCCCl	1	3	1.1	1.1
351	1,3-dioxepane	000505-65-7	O(CCCCO1)C1	1	1	1.1	1.1
352	methylchloride	000074-87-3	ClC	1	1	1.1	1.1 & c4.3
353	2-chloropropane	000075-29-6	C(Cl)(C)C	1	1	1.1	1.1
354	2-chloro-2-methylpropane	000507-20-0	ClC(C)(C)C	1	1	1.1	1.1
355	1-chlorobutane	000109-69-3	ClCCCC	1	1	1.1	1.1
356	2-chlorobutane	000078-86-4	C(Cl)(CC)C	1	1	1.1	1.1
357	methyl bromide	000074-83-9	BrC	3	1	1.1	3.1 & c4.3
358	1-chlorohexane	000544-10-5	ClCCCCCC	1	1	1.1	1.1
359	1-chlorooctane	000111-85-3	ClCCCCCCCC	1	1	1.1	1.1
360	1,2-dichloroethane	000107-06-2	ClCCCl	1	1	1.1	1.1
361	1,1-dichloroethane	000075-34-3	C(Cl)(Cl)C	1	1	1.1	1.1
362	1,3-dichloropropane	000142-28-9	ClCCCCl	1	1	1.1	1.1
363	1,2-dichloropropane	000078-87-5	ClCC(Cl)C	1	1	1.1	1.1
364	chloroform	000067-66-3	C(Cl)(Cl)Cl	1	1	1.1	1.1 & m4.3
365	1,4-dichlorobutane	000110-56-5	ClCCCCCl	1	1	1.1	1.1
366	1,1,2-trichloroethane	000079-00-5	ClCC(Cl)Cl	1	1	1.1	1.1
367	1,5-dichloropentane	000628-76-2	ClCCCCCCl	1	1	1.1	1.1
368	1,2-dichlorocyclohexane (trans)	000822-86-6	ClC1C(Cl)CCCC1	1	1	1.1	1.1
369	1,6-dichlorohexane	002163-00-0	ClCCCCCCCl	1	1	1.1	1.1
370	1,1,2,2-tetrachloroethane	000079-34-5	C(C(Cl)Cl)(Cl)Cl	1	1	1.1	1.1
371	1,3-dibromopropane	000109-64-8	BrCCCBr	3	1	1.1	3.1 / c4.3
372	pentachloroethane	000076-01-7	C(C(Cl)Cl)(Cl)(Cl)Cl	1	1	1.1	1.1
373	cyclopentanone	000120-92-3	O=C(CCC1)C1	1	1	1.1	1.1
374	3-pentanone	000096-22-0	O=C(CC)CC	1	1	1.1	1.1
375	2-pentanone	000107-87-9	O=C(CCC)C	1	1	1.1	1.1
376	cyclohexanone	000108-94-1	O=C(CCCC1)C1	1	1	1.1	1.1
377	4-methyl-2-pentanone	000108-10-1	O=C(CC(C)C)C	1	1	1.1	1.1
378	2-heptanone	000110-43-0	O=C(CCCCC)C	1	1	1.1	1.1
379	5-methyl-2-hexanone	000110-12-3	O=C(CCC(C)C)C	1	1	1.1	1.1
380	5-methyl-3-heptanone	000541-85-5	CCC(=O)CC(C)CC	1	1	1.1	1.1
381	4-propylcyclohexanone	040649-36-3	O=C1CCC(CCC)CC1	1	1	1.1	1.1
382	camphor (bornan-2-one)	000076-22-2	O=C(C(C(C1C2)(C)C)(C2)C)C1	1	1	1.1	1.1
383	4-tert-butylcyclohexanone	000098-53-3	O=C(CCC(C(C)(C)C)C1)C1	1	1	1.1	1.1
384	2,5,6-trimethyl-2-cyclohexen-1-one	020030-30-2	O=C(C(=CCC1C)C)C1C	1	5	1.1	1.1
385	(R)-(-)-5-isopropenyl-2-methylcyclohex-2-en-1-one ((R)-(-)-carvone)	006485-40-1	O=C(C(=CCC1C(=C)C)C)C1	1	5	1.1	1.1 / 6.6
386	4-phenylbut-3-en-2-one (benzylideneacetone)	000122-57-6	O=C(C=Cc(cccc1)c1)C	1	3	1.1	1.1 / 6.7
387	pyruvaldehyde (methyl glyoxal)	000078-98-8	O=CC(=O)C	1	3	3.1	4.1 & 3.1
388	CBI	CBI	CBI	1	1	1.1	1.1
389	2'-acetonaphthone	000093-08-3	O=C(c(ccc(c1ccc2)c2)c1)C	1	5	1.1	1.1
390	cyclohexyl phenyl ketone	000712-50-5	C1CCCCC1C(=O)c1ccccc1	1	5	1.1	1.1
391	benzophenone	000119-61-9	O=C(c(cccc1)c1)c(cccc2)c2	1	5	1.1	1.1
392	(tonalid) 6-acetyl-1,1,2,4,4,7-hexamethyltetraline	001506-02-1	O=C(c(c(cc(c1C(CC2C)(C)C)C2(C)C)c1)C	1	5	1.1	1.1
393	p-hydroxyacetophenone	000099-93-4	O=C(c(ccc(O)c1)c1)C	2	5	1.2	1.2
394	2-butyryl-3-hydroxy-5-thiocyclohexan-3-yl-cyclohex-2-en-1-one (625-triketon)	094723-86-1	C1SCCCC(C2CC(O)=C(C(=O)CCC)C(=O)C2)1	1	5	6.4	1.1 / C6.7
395	lactonitrile	000078-97-7	N#CC(O)C	4	3	4.5	3.1
396	but-3-enenitrile (allylnitrile)	000109-75-1	N#CCC=C	4	3	4.5	4.5
397	2-methyl-3-butenenitrile	016529-56-9	N#CC(C=C)C	1	3	4.5	1.1 / 4.5 ?
398	2-nitrotoluene	000088-72-2	[O-][N+](=O)c(c(ccc1)C)c1	1	2	1.1 & m4.3	1.1 & mC4.3

Tableau S2 : Set de validation. Partie 1. (suite)

399	1-chloro-2-nitrobenzene	000088-73-3	[O-][N+](=O)c(c(ccc1)Cl)c1	1	2	1.1 & m4.3	1.1 & mC4.3
400	*p*-ethylphenol	000123-07-9	Oc1ccc(CC)cc1	2	2	1.2	1.2
401	2,6-xylenol	000576-26-1	Oc(c(ccc1)C)c1C	2	2	1.2	1.2
402	p-propylphenol	000645-56-7	Oc1ccc(CCC)cc1	2	2	1.2	1.2
403	2,3,6-trimethylphenol	002416-94-6	Oc1c(C)c(C)ccc1C	2	2	1.2	1.2
404	o-tert-butylphenol	000088-18-6	Oc(c(ccc1)C(C)(C)C)c1	2	2	1.2	1.2
405	2-sec-butylphenol	000089-72-5	Oc(c(ccc1)C(CC)C)c1	2	2	1.2	1.2
406	p-tert-butylphenol	000098-54-4	Oc(ccc(c1)C(C)(C)C)c1	2	2	1.2	1.2
407	5-methyl-2-(1-methylethyl)-phenol (thymol)	000089-83-8	Oc(c(ccc1C)C(C)C)c1	2	2	1.2	1.2
408	CBI	CBI	CBI	2	2	1.2	1.2
409	4-tert-pentylphenol	000080-46-6	Oc(ccc(c1)C(CC)(C)C)c1	2	2	1.2	1.2
410	2-tert-butyl-4-methylphenol	002409-55-4	Oc(c(cc(c1)C)C(C)(C)C)c1	2	2	1.2	1.2
411	2-tert-butyl-5-methylphenol	000088-60-8	Oc(c(ccc1C)C(C)(C)C)c1	2	2	1.2	1.2
412	2-tert-butyl-4,6-dimethylphenol	001879-09-0	Oc(c(cc(c1)C)C(C)(C)C)c1C	2	2	1.2	1.2
413	4-tert-octylphenol	000140-66-9	Oc(ccc(c1)C(CC(C)(C)C)(C)C)c1	2	2	1.2	1.2 / 6.7
414	di-sec-butylphenol, mixed isomers (2,6-di-sec-butylphenol for EpiSuite)	005510-99-6	CCC(C)c1cccc(C(C)CC)c1O	2	2	1.2	1.2
415	2,4-di-tert-butylphenol	000096-76-4	Oc(c(cc(c1)C(C)(C)C)C(C)(C)C)c1	2	2	1.2	1.2
416	2,6-di-tert-butylphenol	000128-39-2	Oc(c(ccc1)C(C)(C)C)c1C(C)(C)C	2	2	1.2	1.2
417	p-nonylphenol	000104-40-5	Oc(ccc(c1)CCCCCCCCC)c1	2	2	1.2	1.2 & #up6.7
418	2,6-di-tert-butyl-p-cresol (BHT)	000128-37-0	Oc(c(cc(c1)C)C(C)(C)C)c1C(C)(C)C	2	2	1.2	1.2 / 4.3
419	2,4,6-tri-sec-butylphenol	005892-47-7	Oc(c(cc(c1)C(CC)C)C(CC)C)c1C(CC)C	2	2	1.2	1.2 / 4.3?
420	2-allylphenol	001745-81-9	Oc(c(ccc1)CC=C)c1	2	2	1.2	1.2
421	biphenyl-2-ol (o-phenylphenol)	000090-43-7	Oc(c(c(cccc1)c1)ccc2)c2	2	2	1.2	1.2
422	biphenyl-4-ol (p-phenylphenol)	000092-69-3	Oc(ccc(c(cccc1)c1)c2)c2	2	2	1.2	1.2
423	4-(1-methyl-1-phenylethyl)phenol	000599-64-4	Oc(ccc(c1)C(c(cccc2)c2)(C)C)c1	2	2	1.2	1.2 & c6.7
424	methyl salicylate	000119-36-8	O=C(OC)c(c(O)ccc1)c1	2	5	m+f6.7 & 2.1	m+f6.7 & 1.2
425	amyl salicylate	002050-08-0	O=C(OCCCCC)c(c(O)ccc1)c1	2	5	m+f6.7 & 2.1	m+f6.7 & 1.2
426	hexyl salicylate	006259-76-3	O=C(OCCCCCC)c(c(O)ccc1)c1	2	5	m+f6.7 & 2.1	m+f6.7 & 1.2
427	cyclohexyl salicylate	025485-88-5	O=C(c1ccccc1O)OC2CCCCC2	2	5	m+f6.7 & 2.1	m+f6.7 & 1.2
428	phenyl salicylate	000118-55-8	O=C(Oc(cccc1)c1)c(c(O)ccc2)c2	2	5	m+f6.7 & 2.1	m+f6.7 & 1.2
429	benzyl salicylate	000118-58-1	O=C(OCc(cccc1)c1)c(c(O)ccc2)c2	2	5	m+f6.7 & 2.1	m+f6.7 & 1.2
430	2-ethylhexyl salicylate	000118-60-5	O=C(OCC(CCCC)CC)c(c(O)ccc1)c1	2	5	m+f6.7 & 2.1	m+f6.7 & 1.2
431	(Z) hex-3-en-1-yl salicylate	065405-77-8	O=C(OCCC=CCC)c(c(O)ccc1)c1	2	5	m+f6.7 & 2.1	m+f6.7 & 1.2
432	p-(2-methoxyethyl)phenol	056718-71-9	Oc1ccc(CCOC)cc1	2	1	1.2	1.2
433	o-methoxyphenol (guaiacol)	000090-05-1	O(c(c(O)ccc1)c1)C	2	1	1.2 & m4.3	1.2
434	tert-butyl-4-methoxyphenol	025013-16-5	COc1ccc(O)c(c1)C(C)(C)C	3	1	1.2 & m4.3	4.3
435	4-allyl-2-methoxyphenol (eugenol)	000097-53-0	O(c(c(O)ccc1CC=C)c1)C	3	1	1.2 & m4.3	4.3
436	2,4-dichloro-3-ethylphenol	121518-45-4	CCc1c(ccc(c1Cl)O)Cl	4	2	5.1	5.1

Tableau S2 : Set de validation. Partie 1. (suite)

437	resorcinol	000108-46-3	Oc(cccc1O)c1	4	5	1.2	1.2 / 6.7
438	4-tert-butylbenzene-1,2-diol (4-tert-butylpyrocatechol)	000098-29-3	Oc(c(O)cc(c1)C(C)(C)C)c1	3	5	4.3 & 4.4	4.3 & 4.4
439	4-(4-hydroxyphenyl)phenol (biphenyl-4,4'-diol)	000092-88-6	Oc(ccc(c(ccc(O)c1)c1)c2)c2	2	2	1.2	1.2 & #up6.7
440	4-(1-phenylethyl)-benzene-1,3-diol	000085-27-8	CC(c1ccccc1)c2ccc(O)cc2O	2	5	1.2	1.2 / 4.3
441	bisphenol A (2,2-bis(4-hydroxyphenyl)propane) (4,4'-isopropylidenediphenol)	000080-05-7	Oc(ccc(c1)C(c(ccc(O)c2)c2)(C)C)c1	2	2	1.2	1.2 & #up6.7 & m4.3
442	4-[1-(4-hydroxy-3-methylphenyl)cyclohexyl]-2-methylphenol	002362-14-3	Oc(c(cc(c1)C(c(ccc(O)c2C)c2)(CCCC3)C3)C)c1	2	2	1.2	1.2
443	4,4',4''-(ethan-1,1,1-triyl)triphenol	027955-94-8	Oc1ccc(cc1)C(c2ccc(cc2)O)(c3ccc(cc3)O)C	2	2	1.2	1.2
444	4,4'-methylenebis(2,6-di-tert-butylphenol)	000118-82-1	Oc(c(cc(c1)Cc(cc(c(O)c2C(C)(C)C)C(C)(C)C)c2)C(C)(C)C)c1C(C)(C)C	2	5	1.2	1.2
445	(hexachlorophene) 2,2'-methylenebis[3,4,6-trichlorophenol]	000070-30-4	Oc(c(c(c(c1)Cl)Cl)Cc(c(c(cc2Cl)Cl)Cl)c2O)c1Cl	4	5	5.1	5.1
446	1-propanethiol	000107-03-9	SCCC	3	5	4.4	4.4
447	1-hexanethiol	000111-31-9	SCCCCCC	3	5	4.4	4.4
448	1-octanethiol (octyl mercaptan)	000111-88-6	SCCCCCCCC	3	5	4.4	4.4
449	1-dodecanethiol	000112-55-0	SCCCCCCCCCCCC	3	5	4.4	4.4
450	2-mercaptoethanol	000060-24-2	SCCO	3	5	4.4	4.4
451	cyclohexanethiol	001569-69-3	C1CCCCC1S	3	5	4.4	4.4
452	tert-butylmercaptan	000075-66-1	SC(C)(C)C	1	5	1.1	1.1
453	2,4,5-trichlorophenol	000095-95-4	Oc1c(Cl)cc(Cl)c(Cl)c1	4	2	5.1	5.1
454	2,3,4-trichlorophenol	015950-66-0	Oc1c(Cl)c(Cl)c(Cl)cc1	4	2	5.1	5.1
455	2,3-dichlorophenol	000576-24-9	Oc1c(Cl)c(Cl)ccc1	4	2	5.1	5.1
456	2,4-dichloro-6-methylphenol	001570-65-6	Oc1c(Cl)cc(Cl)cc1C	4	2	5.1	5.1
457	phenobarbital	000050-06-6	c1ccccc1C2(CC)C(=O)NC(=O)NC2=O	4	5	2.1	6.6
458	hyoscine (scopolamine)	000051-34-3	c1ccccc1C(CO)C(=O)OC2CC(N3C)C(O4)C4C3C2	4	3	3.1	#up6.2
459	diethyl stilbestrol	000056-53-1	Oc(cc1)ccc1C(CC)=C(CC)c(cc2)ccc2O	4	2	1.2	#up6.7
460	ethynylestradiol (EE2)	000057-63-6	Oc1cc2CCC3C4CCC(O)(C#C)C4(C)CCC3c2cc1	4	5	#upC6.7?	#up6.7

Tableau S2 : Set de validation. Partie 1. (suite)

461	d-tubocurarine	000057-94-3	Oc7ccc1cc7Oc5cc6C(Cc4ccc(Oc2c3C(C1)[N+](C)(C)CCc3cc(OC)c2O)cc4)[N+](C)(C)CCc6cc5OC	4	5	#up6.6	#up6.2
462	codeine	000076-57-3	CN1CCC23C4C1CC5=C2C(=C(C=C5)OC)OC3C(C=C4)O	4	3	#up6.6	#up6.6
463	methadone	000076-99-3	CCC(=O)C(CC(C)N(C)C)(c1ccccc1)c2ccccc2	4	3	#up6.2	#up6.6
464	2,6-dichlorophenol	000087-65-0	Oc1c(Cl)cccc1Cl	4	2	5.1	5.1
465	soman	000096-64-0	CP(=O)(F)OC(C)C(C)(C)C	4	5	6.1	6.1
466	simazine	000122-34-9	CCNc(nc(Cl)n1)nc1NCC	4	5	#uC6.7 & p6.5	#uC6.7 & p6.5
467	methyl-parathion	000298-00-0	COP(=S)(OC)Oc1ccc([N+](=O)[O-])cc1	4	4	6.1	6.1
468	diazinon	000333-41-5	CCOP(=S)(OCC)Oc1nc(C(C)C)nc(C)c1	4	4	6.1	6.1
469	diazepam	000439-14-5	c1ccccc1C2c(cc3Cl)c(cc3)N(C)C(=O)CN=2	4	5	#up6.2	6.6
470	cytisine	000485-35-8	O=C1C=CC=C2C(C3)CNCC3CN12	4	5	3.2	#up6.2
471	2-methyl-4,6-dinitrophenol	000534-52-1	Oc1c([N+](=O)[O-])cc([N+](=O)[O-])cc1C	4	5	5.1	5.1
472	norgestrel (levo)	000797-63-7	O=C1C=C2CCC3C4CCC(O)(C#C)C4(CC)CCC3C2CC1	4	5	#upC6.7?	#up6.7
473	2,4-D 2-ethylhexyl ester	001928-43-4	Clc1cc(Cl)ccc1OCC(=O)OCC(CC)CCCC	1	1	#p2.1 & p2.2	p2.2 & #p2.1
474	isoprocarb (2-Isopropylphenyl methylcarbamate)	002631-40-5	CNC(=O)Oc1c(C(C)C)cccc1	4	4	6.1	6.1
475	chlorpyrifos	002921-88-2	CCOP(=S)(OCC)Oc1c(Cl)cc(Cl)c(Cl)n1	4	4	6.1	6.1
476	picrotoxinin	017617-45-7	O=C1OC2C(C3C(=C)C)OC(=O)C3C(O)(C2C)CC(O4)C14	4	3	3.1	6.6
477	MPTP	028289-54-5	CN(CC1)CC=C1c2ccccc2	3	5	1.1 & 5.2	4.4
478	acetochlor	034256-82-1	ClCC(=O)N(COCC)c1c(CC)cccc1C	1	3	4.1 & p6.7	4.1 & p6.7
479	hexazinone	051235-04-2	CN(C)C1=NC(=O)N(C2CCCCC2)C(=O)N1C	1	5	p6.5 & 1.1	p6.5
480	cypermethrin	052315-07-8	ClC(Cl)=CC(C1(C)C)C1C(=O)OC(C#N)c2cc(ccc2)Oc3ccccc3	4	4	#up6.6	#up6.6
481	carbosulfan	055285-14-8	CCCCN(CCCC)SN(C)C(=O)Oc1c(OC2(C)C)c(C2)ccc1	4	4	6.1	6.1
482	MCPB methyl ester	057153-18-1	c1cc(Cl)cc(C)c1OCCCC(=O)OC	1	1	2.1	p2.2 & #p2.1
483	tefluthrin	079538-32-2	FC(F)(F)C(Cl)=CC(C1(C)C)C1C(=O)OCc2c(F)c(F)c(C)c(F)c2F	4	4	#up6.6	#up6.6
484	imazapyr	081334-34-1	OC(=O)c1cccnc1C2=NC(C)(C(C)C)C(=O)N2	1	5	5.2	p6.7
485	raloxifene	084449-90-1	C1CCCCN1CCOc(cc2)ccc2C(=O)c3c4ccc(O)cc4sc3c(cc5)ccc5O	4	5	1.2	#up6.7
486	epibatidine	140111-52-0	Clc(nc1)ccc1C(C2)C(N3)CCC23	4	5	#up6.2	#up6.2
487	N,N,N-Trimethyl-3-[(methylcarbamoyl)oxy]anilinium	ND	CNC(=O)Oc1cc([N+](C)(C)C)ccc1	4	4	6.1	6.1
488	n-butylamine	000109-73-9	NCCCC	2	2	1.2	1.2
489	butylethylamine	013360-63-9	CCNCCCC	2	1	1.2	1.2
490	tripropylamine	000102-69-2	CCCN(CCC)CCC	2	1	1.1	1.2
491	octylamine	000111-86-4	NCCCCCCCC	2	2	1.2	1.2
492	N,N-dimethylbenzylamine	000103-83-3	CN(C)Cc1ccccc1	2	5	1.1	1.2
493	decyldimethylamine	001120-24-7	CN(C)CCCCCCCCCC	2	1	1.1	1.2

Tableau S2 : Set de validation. Partie 2.

N°	MechoA detail	Reference
1	H+ dessicates cells (consumes water to make hydronium ion) = corrosion? ; acetic acid may enhance lipid solubility into water, disrupting the membrane integrity	Barker, R. M. The Bactericidal Action of Low Molecular Weight Compounds on Mycobacterium tuberculosis. J. Appl. Bacteriol. 27, 213–220 (1964). https://toxnet.nlm.nih.gov/cgi-bin/sis/search/a?dbs+hsdb:@term+@DOCNO+40
2	transport of protons into the cytosol, acidifying the cell.	Barker, R. M. The Bactericidal Action of Low Molecular Weight Compounds on Mycobacterium tuberculosis. J. Appl. Bacteriol. 27, 213–220 (1964).
3	transport of protons into the cytosol, acidifying the cell.	Barker, R. M. The Bactericidal Action of Low Molecular Weight Compounds on Mycobacterium tuberculosis. J. Appl. Bacteriol. 27, 213–220 (1964).
4	transport of protons into the cytosol, acidifying the cell.	Barker, R. M. The Bactericidal Action of Low Molecular Weight Compounds on Mycobacterium tuberculosis. J. Appl. Bacteriol. 27, 213–220 (1964).
5	transport of protons into the cytosol, acidifying the cell.	Barker, R. M. The Bactericidal Action of Low Molecular Weight Compounds on Mycobacterium tuberculosis. J. Appl. Bacteriol. 27, 213–220 (1964).
6	transport of protons into the cytosol, acidifying the cell.	Barker, R. M. The Bactericidal Action of Low Molecular Weight Compounds on Mycobacterium tuberculosis. J. Appl. Bacteriol. 27, 213–220 (1964).
7	transport of protons into the cytosol, acidifying the cell. M: Easy metabolism/Michaël.	Methacrylic acid [MAK Value Documentation, 2010]. in The MAK-Collection for Occupational Health and Safety (Wiley-VCH Verlag GmbH & Co. KGaA, 2002). Barker, R. M. The Bactericidal Action of Low Molecular Weight Compounds on Mycobacterium tuberculosis. J. Appl. Bacteriol. 27, 213–220 (1964).
8	transport of protons into the cytosol, acidifying the cell.	Plumridge, A. et al. The Weak Acid Preservative Sorbic Acid Inhibits Conidial Germination and Mycelial Growth of Aspergillus niger through Intracellular Acidification. Appl. Environ. Microbiol. 70, 3506–3511 (2004). Barker, R. M. The Bactericidal Action of Low Molecular Weight Compounds on Mycobacterium tuberculosis. J. Appl. Bacteriol. 27, 213–220 (1964).
9	reacts with lysine to form adducts, or with water to generate the diacid	Venables, K. M. Low Molecular Weight Chemicals, Hypersensitivity, and Direct Toxicity: The Acid Anhydrides. Br. J. Ind. Med. 46, 222–232 (1989).
10	reacts with lysine to form adducts, or with water to generate the diacid	Venables, K. M. Low Molecular Weight Chemicals, Hypersensitivity, and Direct Toxicity: The Acid Anhydrides. Br. J. Ind. Med. 46, 222–232 (1989).
11	reacts with lysine to form adducts, or with water to generate the diacid	Venables, K. M. Low Molecular Weight Chemicals, Hypersensitivity, and Direct Toxicity: The Acid Anhydrides. Br. J. Ind. Med. 46, 222–232 (1989).
12	metabolism into formic acid, creating metabolic acidosis. And formate ion inhibit mitochondrial cytochrome oxidase. (The metabolism and toxicity of methanol differs markedly between species (Jacobsen & Martin, 1986). In primates and human beings, methanol causes metabolic acidosis and ocular damage. In non-primate laboratory animals, however, it acts as a CNS-depressant (Gosselin et al., 1984; Roe, 1982))	Bozza-Marrubini, M., Brucato, A., Locatelli, C. & Ruggeroni, M. L. METHANOL (PIM 335). (2002). Available at: http://www.inchem.org/documents/pims/chemical/pim335.htm#7.1. (Accessed: 9th August 2016) Liesivuori, J. & Savolainen, A. H. Methanol and Formic Acid Toxicity: Biochemical Mechanisms. Pharmacol. Toxicol. 69, 157–163 (1991).
13		KREATiS internal experimental database and QSARs.
14		KREATiS internal experimental database and QSARs.
15		KREATiS internal experimental database and QSARs.
16		KREATiS internal experimental database and QSARs.
17		KREATiS internal experimental database and QSARs.
18		KREATiS internal experimental database and QSARs.
19		KREATiS internal experimental database and QSARs.
20		KREATiS internal experimental database and QSARs.
21		KREATiS internal experimental database and QSARs.
22		KREATiS internal experimental database and QSARs.
23		KREATiS internal experimental database and QSARs.
24		KREATiS internal experimental database and QSARs.
25		KREATiS internal experimental database and QSARs.
26		KREATiS internal experimental database and QSARs.
27		KREATiS internal experimental database and QSARs.
28		KREATiS internal experimental database and QSARs.
29	oxidation into aldehyde and adducts	Belsito, D. et al. A safety assessment of non-cyclic alcohols with unsaturated branched chain when used as fragrance ingredients: The RIFM expert panel. Food Chem. Toxicol. 48, Supplement 3, S1–S42 (2010).
30	oxidation into aldehyde and adducts	Belsito, D. et al. A safety assessment of non-cyclic alcohols with unsaturated branched chain when used as fragrance ingredients: The RIFM expert panel. Food Chem. Toxicol. 48, Supplement 3, S1–S42 (2010).
31		KREATiS internal experimental database and QSARs.
32		KREATiS internal experimental database and QSARs.
33	oxidation into aldehyde and adducts	Belsito, D. et al. A safety assessment of non-cyclic alcohols with unsaturated branched chain when used as fragrance ingredients: The RIFM expert panel. Food Chem. Toxicol. 48, Supplement 3, S1–S42 (2010).
34		KREATiS internal experimental database and QSARs.
35		KREATiS internal experimental database and QSARs.
36		KREATiS internal experimental database and QSARs.
37		KREATiS internal experimental database and QSARs.
38	oxidation into correponding acid, generating acidosis, and giving developmental / reproductive toxicity	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012). Bagchi, G. & Waxman, D. J. Toxicity of Ethylene Glycol Monomethyl Ether: Impact on Testicular Gene Expression. Int. J. Androl. 31, 269–274 (2008).

Tableau S2 : Set de validation. Partie 2. (suite)

39	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
40	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
41	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
42		KREATiS internal experimental database and QSARs.
43	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
44	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
45	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
46	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
47	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
48	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
49	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
50	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
51	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
52	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
53	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
54	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
55	ecotox apparent MoA1	KREATiS internal experimental database and QSARs.
56	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
57		KREATiS internal experimental database and QSARs.
58	oxidation into correponding acid, generating acidosis	Pohl, H. R., Ruiz, P., Scinicariello, F. & Mumtaz, M. M. Joint toxicity of alkoxyethanol mixtures: Contribution of in silico applications. Regul. Toxicol. Pharmacol. 64, 134–142 (2012).
59	metabolised into glycoaldehyde, glycolate, glyoxylate and oxalate. "Theses metabolites are OPU, inhibit glucose and serotonin metabolism, protein synthesis, DNA replication, ribosomal RNA formation." These effecs may all be due to acidosis generated, and oxalate crystallising with calcium, leading to deposits blocking circulation etc. and eventually hypocalcemia)	ATSDR. Ethylene Glycol and Propylene Glycol Toxicity: What Are the Stages of Ethylene Glycol Intoxication? | ATSDR - Environmental Medicine & Environmental Health Education - CSEM. (2007). Available at: http://www.atsdr.cdc.gov/csem/csem.asp?csem=12&po=9. (Accessed: 15th July 2015)
60		KREATiS internal experimental database and QSARs.
61		KREATiS internal experimental database and QSARs.
62		KREATiS internal experimental database and QSARs.
63		KREATiS internal experimental database and QSARs.
64		KREATiS internal experimental database and QSARs.
65		KREATiS internal experimental database and QSARs.
66		KREATiS internal experimental database and QSARs.
67		KREATiS internal experimental database and QSARs.
68	metabolised into more toxic compound such as oxalic acid, ethylene dioxyacetic acid, and possibly ethylene glycol (leading to glycoaldehyde, glycolate, glyoxylate and oxalate. "Theses metabolites are OPU, inhibit glucose and serotonin metabolism, protein synthesis, DNA replication, ribosomal RNA formation." These effecs may all be due to acidosis generated, and oxalate crystallising with calcium, leading to deposits blocking circulation etc. and eventually hypocalcemia))	ATSDR. Ethylene Glycol and Propylene Glycol Toxicity: What Are the Stages of Ethylene Glycol Intoxication? | ATSDR - Environmental Medicine & Environmental Health Education - CSEM. (2007). Available at: http://www.atsdr.cdc.gov/csem/csem.asp?csem=12&po=9. (Accessed: 15th July 2015)
69		KREATiS internal experimental database and QSARs.
70		KREATiS internal experimental database and QSARs.
71	metabolised into more toxic compound such as oxalic acid, ethylene dioxyacetic acid, and possibly ethylene glycol (leading to glycoaldehyde, glycolate, glyoxylate and oxalate. "Theses metabolites are OPU, inhibit glucose and serotonin metabolism, protein synthesis, DNA replication, ribosomal RNA formation." These effecs may all be due to acidosis generated, and oxalate crystallising with calcium, leading to deposits blocking circulation etc. and eventually hypocalcemia))	ATSDR. Ethylene Glycol and Propylene Glycol Toxicity: What Are the Stages of Ethylene Glycol Intoxication? | ATSDR - Environmental Medicine & Environmental Health Education - CSEM. (2007). Available at: http://www.atsdr.cdc.gov/csem/csem.asp?csem=12&po=9. (Accessed: 15th July 2015)
72		KREATiS internal experimental database and QSARs.
73		KREATiS internal experimental database and QSARs.
74		KREATiS internal experimental database and QSARs.
75		KREATiS internal experimental database and QSARs.

Tableau S2 : Set de validation. Partie 2. (suite)

76		KREATiS internal experimental database and QSARs.
77		KREATiS internal experimental database and QSARs.
78		KREATiS internal experimental database and QSARs.
79		KREATiS internal experimental database and QSARs.
80	not toxic at all for all aquatic species	KREATiS internal experimental database and QSARs.
81		KREATiS internal experimental database and QSARs.
82		KREATiS internal experimental database and QSARs.
83		KREATiS internal experimental database and QSARs.
84		KREATiS internal experimental database and QSARs.
85		KREATiS internal experimental database and QSARs.
86		KREATiS internal experimental database and QSARs.
87		KREATiS internal experimental database and QSARs.
88		KREATiS internal experimental database and QSARs.
89		KREATiS internal experimental database and QSARs.
90		KREATiS internal experimental database and QSARs.
91		KREATiS internal experimental database and QSARs.
92		KREATiS internal experimental database and QSARs.
93		KREATiS internal experimental database and QSARs.
94		KREATiS internal experimental database and QSARs.
95		KREATiS internal experimental database and QSARs.
96		KREATiS internal experimental database and QSARs.
97		KREATiS internal experimental database and QSARs.
98		KREATiS internal experimental database and QSARs.
99		KREATiS internal experimental database and QSARs.
100		KREATiS internal experimental database and QSARs.
101		KREATiS internal experimental database and QSARs.
102		KREATiS internal experimental database and QSARs.
103	ecotox apparent MoA1	KREATiS internal experimental database and QSARs.
104	ecotox apparent MoA1	KREATiS internal experimental database and QSARs.
105		KREATiS internal experimental database and QSARs.
106		KREATiS internal experimental database and QSARs.
107		KREATiS internal experimental database and QSARs.
108		KREATiS internal experimental database and QSARs.
109		KREATiS internal experimental database and QSARs.
110		KREATiS internal experimental database and QSARs.
111		KREATiS internal experimental database and QSARs.
112		KREATiS internal experimental database and QSARs.
113		KREATiS internal experimental database and QSARs.
114		A.O. Aptula, D.W. Roberts. 'Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity'. Chem. Res. Toxicol., 19 (2006), pp. 1097–1105. KREATiS internal experimental database and QSARs.
115		A.O. Aptula, D.W. Roberts. 'Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity'. Chem. Res. Toxicol., 19 (2006), pp. 1097–1105. KREATiS internal experimental database and QSARs.
116		A.O. Aptula, D.W. Roberts. 'Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity'. Chem. Res. Toxicol., 19 (2006), pp. 1097–1105. KREATiS internal experimental database and QSARs.
117		A.O. Aptula, D.W. Roberts. 'Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity'. Chem. Res. Toxicol., 19 (2006), pp. 1097–1105. KREATiS internal experimental database and QSARs.
118		A.O. Aptula, D.W. Roberts. 'Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity'. Chem. Res. Toxicol., 19 (2006), pp. 1097–1105. KREATiS internal experimental database and QSARs.

119		A.O. Aptula, D.W. Roberts. 'Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity'. Chem. Res. Toxicol., 19 (2006), pp. 1097–1105. KREATiS internal experimental database and QSARs.
120		A.O. Aptula, D.W. Roberts. 'Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity'. Chem. Res. Toxicol., 19 (2006), pp. 1097–1105. KREATiS internal experimental database and QSARs.
121		A.O. Aptula, D.W. Roberts. 'Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity'. Chem. Res. Toxicol., 19 (2006), pp. 1097–1105. KREATiS internal experimental database and QSARs.
122		A.O. Aptula, D.W. Roberts. 'Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity'. Chem. Res. Toxicol., 19 (2006), pp. 1097–1105. KREATiS internal experimental database and QSARs.
123		A.O. Aptula, D.W. Roberts. 'Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity'. Chem. Res. Toxicol., 19 (2006), pp. 1097–1105. KREATiS internal experimental database and QSARs.
124		A.O. Aptula, D.W. Roberts. 'Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity'. Chem. Res. Toxicol., 19 (2006), pp. 1097–1105. KREATiS internal experimental database and QSARs.
125		Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037. KREATiS internal experimental database and QSARs.
126		Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037. KREATiS internal experimental database and QSARs.
127		Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037. KREATiS internal experimental database and QSARs.
128	ecotox apparent MoA1	Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037. KREATiS internal experimental database and QSARs.
129		A.O. Aptula, D.W. Roberts. 'Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity'. Chem. Res. Toxicol., 19 (2006), pp. 1097–1105. KREATiS internal experimental database and QSARs.
130		A.O. Aptula, D.W. Roberts. 'Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity'. Chem. Res. Toxicol., 19 (2006), pp. 1097–1105. KREATiS internal experimental database and QSARs.
131	ecotox less toxic than MoA1, has the same toxicity as citronellol...	A.O. Aptula, D.W. Roberts. 'Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity'. Chem. Res. Toxicol., 19 (2006), pp. 1097–1105. KREATiS internal experimental database and QSARs.
132		KREATiS internal experimental database and QSARs.
133		KREATiS internal experimental database and QSARs.
134		KREATiS internal experimental database and QSARs.
135		KREATiS internal experimental database and QSARs.
136		KREATiS internal experimental database and QSARs.
137		KREATiS internal experimental database and QSARs.
138		KREATiS internal experimental database and QSARs.
139		KREATiS internal experimental database and QSARs.
140		KREATiS internal experimental database and QSARs.
141		KREATiS internal experimental database and QSARs.
142		KREATiS internal experimental database and QSARs.
143		KREATiS internal experimental database and QSARs.
144		KREATiS internal experimental database and QSARs.
145		KREATiS internal experimental database and QSARs.
146		KREATiS internal experimental database and QSARs.
147		KREATiS internal experimental database and QSARs.
148		KREATiS internal experimental database and QSARs.
149		KREATiS internal experimental database and QSARs.
150		KREATiS internal experimental database and QSARs.
151		KREATiS internal experimental database and QSARs.
152		KREATiS internal experimental database and QSARs.
153		KREATiS internal experimental database and QSARs.
154		KREATiS internal experimental database and QSARs.
155		KREATiS internal experimental database and QSARs.
156		KREATiS internal experimental database and QSARs.
157		KREATiS internal experimental database and QSARs.
158		KREATiS internal experimental database and QSARs.
159		KREATiS internal experimental database and QSARs.
160		KREATiS internal experimental database and QSARs.
161		KREATiS internal experimental database and QSARs.
162		KREATiS internal experimental database and QSARs.
163		KREATiS internal experimental database and QSARs.
164		KREATiS internal experimental database and QSARs.
165		KREATiS internal experimental database and QSARs.

166	Adduct formation with SH groups of proteins	Rand, G. M. in Fundamentals Of Aquatic Toxicology: Effects, Environmental Fate And Risk Assessment 637 (CRC Press, 1995).
167		KREATiS internal experimental database and QSARs.
168		KREATiS internal experimental database and QSARs.
169		KREATiS internal experimental database and QSARs.
170		KREATiS internal experimental database and QSARs.
171		KREATiS internal experimental database and QSARs.
172		KREATiS internal experimental database and QSARs.
173		KREATiS internal experimental database and QSARs.
174		KREATiS internal experimental database and QSARs.
175		KREATiS internal experimental database and QSARs.
176		KREATiS internal experimental database and QSARs.
177		KREATiS internal experimental database and QSARs.
178		KREATiS internal experimental database and QSARs.
179		KREATiS internal experimental database and QSARs.
180		KREATiS internal experimental database and QSARs.
181		KREATiS internal experimental database and QSARs.
182		KREATiS internal experimental database and QSARs.
183		KREATiS internal experimental database and QSARs.
184		KREATiS internal experimental database and QSARs.
185		KREATiS internal experimental database and QSARs.
186		KREATiS internal experimental database and QSARs.
187		KREATiS internal experimental database and QSARs.
188		KREATiS internal experimental database and QSARs.
189		KREATiS internal experimental database and QSARs.
190		KREATiS internal experimental database and QSARs.
191		KREATiS internal experimental database and QSARs.
192		Tsuruda, L. S., Lamé, M. W. & Jones, A. D. Formation of epoxide and quinone protein adducts in B6C3F1 mice treated with naphthalene, sulfate conjugate of 1,4-dihydroxynaphthalene and 1,4-naphthoquinone. Arch. Toxicol. 69, 362–367 (1995). Buckpitt, A. et al. Naphthalene-induced respiratory tract toxicity: metabolic mechanisms of toxicity. Drug Metab. Rev. 34, 791–820 (2002). KREATiS internal experimental database and QSARs.
193		Tsuruda, L. S., Lamé, M. W. & Jones, A. D. Formation of epoxide and quinone protein adducts in B6C3F1 mice treated with naphthalene, sulfate conjugate of 1,4-dihydroxynaphthalene and 1,4-naphthoquinone. Arch. Toxicol. 69, 362–367 (1995). Buckpitt, A. et al. Naphthalene-induced respiratory tract toxicity: metabolic mechanisms of toxicity. Drug Metab. Rev. 34, 791–820 (2002). KREATiS internal experimental database and QSARs.
194		Tsuruda, L. S., Lamé, M. W. & Jones, A. D. Formation of epoxide and quinone protein adducts in B6C3F1 mice treated with naphthalene, sulfate conjugate of 1,4-dihydroxynaphthalene and 1,4-naphthoquinone. Arch. Toxicol. 69, 362–367 (1995). Buckpitt, A. et al. Naphthalene-induced respiratory tract toxicity: metabolic mechanisms of toxicity. Drug Metab. Rev. 34, 791–820 (2002). KREATiS internal experimental database and QSARs.
195		Tsuruda, L. S., Lamé, M. W. & Jones, A. D. Formation of epoxide and quinone protein adducts in B6C3F1 mice treated with naphthalene, sulfate conjugate of 1,4-dihydroxynaphthalene and 1,4-naphthoquinone. Arch. Toxicol. 69, 362–367 (1995). Buckpitt, A. et al. Naphthalene-induced respiratory tract toxicity: metabolic mechanisms of toxicity. Drug Metab. Rev. 34, 791–820 (2002). KREATiS internal experimental database and QSARs.
196		Tsuruda, L. S., Lamé, M. W. & Jones, A. D. Formation of epoxide and quinone protein adducts in B6C3F1 mice treated with naphthalene, sulfate conjugate of 1,4-dihydroxynaphthalene and 1,4-naphthoquinone. Arch. Toxicol. 69, 362–367 (1995). Buckpitt, A. et al. Naphthalene-induced respiratory tract toxicity: metabolic mechanisms of toxicity. Drug Metab. Rev. 34, 791–820 (2002). KREATiS internal experimental database and QSARs.
197	ecotox apparent MoA1	Tsuruda, L. S., Lamé, M. W. & Jones, A. D. Formation of epoxide and quinone protein adducts in B6C3F1 mice treated with naphthalene, sulfate conjugate of 1,4-dihydroxynaphthalene and 1,4-naphthoquinone. Arch. Toxicol. 69, 362–367 (1995). Buckpitt, A. et al. Naphthalene-induced respiratory tract toxicity: metabolic mechanisms of toxicity. Drug Metab. Rev. 34, 791–820 (2002). KREATiS internal experimental database and QSARs.
198		KREATiS internal experimental database and QSARs.
199		KREATiS internal experimental database and QSARs.
200		KREATiS internal experimental database and QSARs.
201		KREATiS internal experimental database and QSARs.
202		KREATiS internal experimental database and QSARs.
203		KREATiS internal experimental database and QSARs.
204		KREATiS internal experimental database and QSARs.
205		KREATiS internal experimental database and QSARs.
206		KREATiS internal experimental database and QSARs.
207		KREATiS internal experimental database and QSARs.
208		KREATiS internal experimental database and QSARs.
209		KREATiS internal experimental database and QSARs.
210		KREATiS internal experimental database and QSARs.
211		KREATiS internal experimental database and QSARs.
212		KREATiS internal experimental database and QSARs.

213		KREATiS internal experimental database and QSARs.
214		KREATiS internal experimental database and QSARs.
215		KREATiS internal experimental database and QSARs.
216		KREATiS internal experimental database and QSARs.
217		KREATiS internal experimental database and QSARs.
218		KREATiS internal experimental database and QSARs.
219		KREATiS internal experimental database and QSARs.
220		KREATiS internal experimental database and QSARs.
221		KREATiS internal experimental database and QSARs.
222		KREATiS internal experimental database and QSARs.
223		KREATiS internal experimental database and QSARs.
224		KREATiS internal experimental database and QSARs.
225		KREATiS internal experimental database and QSARs.
226		KREATiS internal experimental database and QSARs.
227		KREATiS internal experimental database and QSARs.
228		KREATiS internal experimental database and QSARs.
229		KREATiS internal experimental database and QSARs.
230		KREATiS internal experimental database and QSARs.
231		KREATiS internal experimental database and QSARs.
232		KREATiS internal experimental database and QSARs.
233	methacrylates ecotox apparent MoA1	Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037. KREATiS internal experimental database and QSARs.
234	methacrylates ecotox apparent MoA1	Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037. KREATiS internal experimental database and QSARs.
235	methacrylates ecotox apparent MoA1	Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037. KREATiS internal experimental database and QSARs.
236	methacrylates ecotox apparent MoA1	Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037. KREATiS internal experimental database and QSARs.
237	methacrylates ecotox apparent MoA1	Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037. KREATiS internal experimental database and QSARs.
238	methacrylates ecotox apparent MoA1	Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037. KREATiS internal experimental database and QSARs.
239	methacrylates ecotox apparent MoA1	Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037. KREATiS internal experimental database and QSARs.
240	methacrylates ecotox apparent MoA1	Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037. KREATiS internal experimental database and QSARs.
241	methacrylates ecotox apparent MoA1	Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037. KREATiS internal experimental database and QSARs.
242	vinyl esters are much more toxic than normal esters in ecotox (is it by oxidation of vinyl to epoxide, or by ester hydrolysis, generating by tautomerisation an aldehyde?)	KREATiS internal experimental database and QSARs.
243		Jaworska, J. S.; Hunter, R. S.; Schultz, T. W. Quantitative Structure-Toxicity Relationships and Volume Fraction Analyses for Selected Esters. Arch. Environ. Contam. Toxicol. 1995, 29 (1), 86–93 DOI: 10.1007/BF00213091. KREATiS internal experimental database and QSARs.
244	+ fast metabolism into Krebs cycle	IPCS Inchem. Ethyl Lactate. WHO Food Additives Series. 1980. / Adeva-Andany, M.; López-Ojén, M.; Funcasta-Calderón, R.; Ameneiros-Rodríguez, E.; Donapetry-García, C.; Vila-Altesor, M.; Rodríguez-Seijas, J. Comprehensive Review on Lactate Metabolism in Human Health. Mitochondrion 2014, 17, 76–100 DOI: 10.1016/j.mito.2014.05.007. KREATiS internal experimental database and QSARs.
245	+ fast metabolism into Krebs cycle	IPCS Inchem. Ethyl Lactate. WHO Food Additives Series. 1980. / Adeva-Andany, M.; López-Ojén, M.; Funcasta-Calderón, R.; Ameneiros-Rodríguez, E.; Donapetry-García, C.; Vila-Altesor, M.; Rodríguez-Seijas, J. Comprehensive Review on Lactate Metabolism in Human Health. Mitochondrion 2014, 17, 76–100 DOI: 10.1016/j.mito.2014.05.007. KREATiS internal experimental database and QSARs.

246	+ fast metabolism into Krebs cycle	IPCS Inchem. Ethyl Lactate. WHO Food Additives Series. 1980. / Adeva-Andany, M.; López-Ojén, M.; Funcasta-Calderón, R.; Ameneiros-Rodríguez, E.; Donapetry-García, C.; Vila-Altesor, M.; Rodríguez-Seijas, J. Comprehensive Review on Lactate Metabolism in Human Health. Mitochondrion 2014, 17, 76–100 DOI: 10.1016/j.mito.2014.05.007. KREATiS internal experimental database and QSARs.
247	+ fast metabolism into Krebs cycle	IPCS Inchem. Ethyl Lactate. WHO Food Additives Series. 1980. / Adeva-Andany, M.; López-Ojén, M.; Funcasta-Calderón, R.; Ameneiros-Rodríguez, E.; Donapetry-García, C.; Vila-Altesor, M.; Rodríguez-Seijas, J. Comprehensive Review on Lactate Metabolism in Human Health. Mitochondrion 2014, 17, 76–100 DOI: 10.1016/j.mito.2014.05.007. KREATiS internal experimental database and QSARs.
248	+ fast metabolism into Krebs cycle	IPCS Inchem. Ethyl Lactate. WHO Food Additives Series. 1980. / Adeva-Andany, M.; López-Ojén, M.; Funcasta-Calderón, R.; Ameneiros-Rodríguez, E.; Donapetry-García, C.; Vila-Altesor, M.; Rodríguez-Seijas, J. Comprehensive Review on Lactate Metabolism in Human Health. Mitochondrion 2014, 17, 76–100 DOI: 10.1016/j.mito.2014.05.007. KREATiS internal experimental database and QSARs.
249	hydrolysed abiotically, generating acetoacetic acid, a compound normally appearing in mammals during fatty acids metabolism. However, it competes with lactic acid in the monocarboxylate transporters, which can pose a problem in high acute doses.	Gómez-Bombarelli, R.; González-Pérez, M.; Pérez-Prior, M. T.; Manso, J. A.; Calle, E.; Casado, J. Kinetic Study of the Neutral and Base Hydrolysis of Diketene. J. Phys. Org. Chem. 2009, 22 (5), 438–442 DOI: 10.1002/poc.1483. National Center for Biotechnology Information. Acetoacetic acid https://pubchem.ncbi.nlm.nih.gov/compound/acetoacetic_acid#section=Top (accessed Oct 4, 2016).
250		KREATiS internal experimental database and QSARs.
251		KREATiS internal experimental database and QSARs.
252		KREATiS internal experimental database and QSARs.
253		KREATiS internal experimental database and QSARs.
254	F: normal ester toxicity	KREATiS internal experimental database and QSARs.
255	hydrolysis gives an hemiacetal, in equilibirum with aldehyde in water. ecotox apparent MoA ester	KREATiS internal experimental database and QSARs.
256		KREATiS internal experimental database and QSARs.
257	10 times more toxic to fish and daphnia than ester toxicity, 2 times for algae	KREATiS internal experimental database and QSARs.
258	Ecotox: all not toxic above solubility limit	KREATiS internal experimental database and QSARs.
259	10 times more toxic to fish and daphnia than ester toxicity, 2 times for algae	KREATiS internal experimental database and QSARs.
260	Ecotox: normal ester toxicity	KREATiS internal experimental database and QSARs.
261	Acute: 2 times more toxic to fish, 4 times to daphnia, 10 times to algae. Probably only because of experimental issues… M: hallucinatory, causes forced sleep. It is a mineralocorticoid receptor antagonist (VS aldosterone = normal substrate) and glucocorticoid recpetor antagonist, and it also inhibit the biosynthesis of aldosterone.	KREATiS internal experimental database and QSARs.
262	10 times more toxic to fish, normal ester toxicity to daphnia and algae	KREATiS internal experimental database and QSARs.
263		KREATiS internal experimental database and QSARs.
264	ecotox: normal ester toxicity	KREATiS internal experimental database and QSARs.
265		KREATiS internal experimental database and QSARs.
266		KREATiS internal experimental database and QSARs.
267	ecotox: normal ester toxicity	KREATiS internal experimental database and QSARs.
268		KREATiS internal experimental database and QSARs.
269	ecotox: normal ester toxicity	KREATiS internal experimental database and QSARs.
270	Fish: normal ester toxicity	KREATiS internal experimental database and QSARs.
271	ecotox: normal ester toxicity	KREATiS internal experimental database and QSARs.
272	2 times more toxic to fish and daphnia than normal ester, not toxic at EC50(pred)*2 for algae	KREATiS internal experimental database and QSARs.
273	ecotox: normal ester toxicity	KREATiS internal experimental database and QSARs.
274	ecotox: normal ester toxicity	KREATiS internal experimental database and QSARs.
275	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.
276	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.
277	fish: less toxic than ester narcosis…	KREATiS internal experimental database and QSARs.
278	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.
279	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.
280	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.

281	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.
282	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.
283	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.
284	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.
285	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.
286	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.
287	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.
288	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.
289	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.
290	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.
291	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.
292	diesters are more toxic than esters, probably because 2 esters are hydrolysed, generating twice as much acidity as for monoesters	KREATiS internal experimental database and QSARs.
293	ecotox: 10 times more toxic than normal diester toxicity	KREATiS internal experimental database and QSARs.
294	because of C=C being (Z), esters (at least one) are not conjugated, hence it behaves like normal diesters? Ecotox results gives weight to this hypothesis.	KREATiS internal experimental database and QSARs.
295	ecotox: 4 times more toxic than normal diester toxicity	KREATiS internal experimental database and QSARs.
296	ecotox: normal diester toxicity. It is an antioxidant (Oxygen radical trapping).	KREATiS internal experimental database and QSARs.
297	fish: normal diester toxicity	KREATiS internal experimental database and QSARs.
298	ecotox: less toxic than ester narcosis…	KREATiS internal experimental database and QSARs.
299		KREATiS internal experimental database and QSARs.
300	ecotox: normal diester toxicity	KREATiS internal experimental database and QSARs.
301	M: digestive hydolysis of one ester =>chronic=> induction of PPARalpha (peroxisome proliferator-activated receptor, precursor of liver cancer development) and endocrine disruption (if R>C4, can undergo further oxidative metabolism)	Adams, William J., Gregory R. Biddinger, Kenneth A. Robillard, and Joseph W. Gorsuch. 1995. 'A Summary of the Acute Toxicity of 14 Phthalate Esters to Representative Aquatic Organisms.' Environmental Toxicology and Chemistry 14 (9): 1569–74. doi:10.1002/etc.5620140916.
302	M: digestive hydolysis of one ester =>chronic=> induction of PPARalpha (peroxisome proliferator-activated receptor, precursor of liver cancer development) and endocrine disruption (if R>C4, can undergo further oxidative metabolism)	Adams, William J., Gregory R. Biddinger, Kenneth A. Robillard, and Joseph W. Gorsuch. 1995. 'A Summary of the Acute Toxicity of 14 Phthalate Esters to Representative Aquatic Organisms.' Environmental Toxicology and Chemistry 14 (9): 1569–74. doi:10.1002/etc.5620140916.
303	ecotox: EC50 > solubility limit	KREATiS internal experimental database and QSARs.
304	M: digestive hydolysis of one ester =>chronic=> induction of PPARalpha (peroxisome proliferator-activated receptor, precursor of liver cancer development) and endocrine disruption (if R>C4, can undergo further oxidative metabolism)	Adams, William J., Gregory R. Biddinger, Kenneth A. Robillard, and Joseph W. Gorsuch. 1995. 'A Summary of the Acute Toxicity of 14 Phthalate Esters to Representative Aquatic Organisms.' Environmental Toxicology and Chemistry 14 (9): 1569–74. doi:10.1002/etc.5620140916.
305	M: digestive hydolysis of one ester =>chronic=> induction of PPARalpha (peroxisome proliferator-activated receptor, precursor of liver cancer development) and endocrine disruption (if R>C4, can undergo further oxidative metabolism)	Adams, William J., Gregory R. Biddinger, Kenneth A. Robillard, and Joseph W. Gorsuch. 1995. 'A Summary of the Acute Toxicity of 14 Phthalate Esters to Representative Aquatic Organisms.' Environmental Toxicology and Chemistry 14 (9): 1569–74. doi:10.1002/etc.5620140916.
306	M: digestive hydolysis of one ester =>chronic=> induction of PPARalpha (peroxisome proliferator-activated receptor, precursor of liver cancer development) and endocrine disruption (if R>C4, can undergo further oxidative metabolism)	Adams, William J., Gregory R. Biddinger, Kenneth A. Robillard, and Joseph W. Gorsuch. 1995. 'A Summary of the Acute Toxicity of 14 Phthalate Esters to Representative Aquatic Organisms.' Environmental Toxicology and Chemistry 14 (9): 1569–74. doi:10.1002/etc.5620140916.
307	ecotox: EC50 > solubility limit	KREATiS internal experimental database and QSARs.
308	M: digestive hydolysis of one ester =>chronic=> induction of PPARalpha (peroxisome proliferator-activated receptor, precursor of liver cancer development) and endocrine disruption (if R>C4, can undergo further oxidative metabolism)	Adams, William J., Gregory R. Biddinger, Kenneth A. Robillard, and Joseph W. Gorsuch. 1995. 'A Summary of the Acute Toxicity of 14 Phthalate Esters to Representative Aquatic Organisms.' Environmental Toxicology and Chemistry 14 (9): 1569–74. doi:10.1002/etc.5620140916.
309	Fish: 1,5 to 4 times more toxic than normal diester toxicity	KREATiS internal experimental database and QSARs.
310		KREATiS internal experimental database and QSARs.
311		KREATiS internal experimental database and QSARs.
312		KREATiS internal experimental database and QSARs.
313		KREATiS internal experimental database and QSARs.
314		KREATiS internal experimental database and QSARs.

315	Fish: baseline toxicity	KREATiS internal experimental database and QSARs.
316		KREATiS internal experimental database and QSARs.
317		KREATiS internal experimental database and QSARs.
318		KREATiS internal experimental database and QSARs.
319		KREATiS internal experimental database and QSARs.
320	ecotox: baseline toxicity	KREATiS internal experimental database and QSARs.
321	ecotox: baseline toxicity	KREATiS internal experimental database and QSARs.
322	similar to thyroid hormones, it disturbs their balance in the organism (thyroid hormones influences a lot of processes), binds to AhR, mimics ER, binding to its receptors with strong affinity, it is carcinogen also…	Siddiqi, M. A., Laessig, R. H. & Reed, K. D. Polybrominated Diphenyl Ethers (PBDEs): New Pollutants-Old Diseases. Clin. Med. Res. 1, 281–290 (2003).
323	ecotox: reactive	Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
324		Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
325		Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
326		Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
327		Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
328	ecotox: reactive	Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
329	ecotox: reactive	Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
330		Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
331		Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
332		Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
333	reactive toxicity for fish, ester toxicity for daphnids and algae…	Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
334	ecotox: between reactive and ester toxicity for all species	Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
335		Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
336		Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
337	ecotox: very toxic / reactive	Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. Schwöbel, J. A. H.; Madden, J. C.; Cronin, M. T. D. Application of a Computational Model for Michael Addition Reactivity in the Prediction of Toxicity to Tetrahymena Pyriformis. Chemosphere 2011, 85 (6), 1066–1074 DOI: 10.1016/j.chemosphere.2011.07.037. KREATiS internal experimental database and QSARs.
338	slow chemical hydrolysis in water, faster with acids. Attacks almost everything! Alkylating.	INRS. 1-Chloro-2,3-époxypropane - Fiche toxicologique n° 187. INRS February 2016. KREATiS internal experimental database and QSARs.
339	ecotox: reactive	Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
340		Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
341		Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
342		Kazius J., McGuire R. and Roberta B. (2005), Derivation and Validation of Toxicophores for Mutagenicity Prediction. J. Med. Chem., 48: 312-320. KREATiS internal experimental database and QSARs.
343		KREATiS internal experimental database and QSARs.
344		KREATiS internal experimental database and QSARs.
345	ecotox: baseline toxicity	KREATiS internal experimental database and QSARs.
346	ecotox: baseline toxicity	KREATiS internal experimental database and QSARs.
347	ecotox: baseline toxicity	KREATiS internal experimental database and QSARs.
348		KREATiS internal experimental database and QSARs.

349	ecotox: baseline toxicity	KREATiS internal experimental database and QSARs.
350	ecotox: baseline toxicity	KREATiS internal experimental database and QSARs.
351	ecotox: baseline toxicity	KREATiS internal experimental database and QSARs.
352	M: conjugation with GSH, then further reduction (thus oxidative stress) to generate methanethiol, which produces CNS injury. Fish: baseline toxicity	ATSDR. Toxicological Profile: Chloromethane https://www.atsdr.cdc.gov/toxprofiles/TP.asp?id=587&tid=109 (accessed Oct 25, 2016). KREATiS internal experimental database and QSARs.
353		KREATiS internal experimental database and QSARs.
354		KREATiS internal experimental database and QSARs.
355		KREATiS internal experimental database and QSARs.
356		KREATiS internal experimental database and QSARs.
357	M: Alkylation through SN2 with thiol and amino residues. Protection with free cysteine and GSH. The conjugation with GSH and further metabolism may be the main cause for chronic toxicity as for methyl chloride. Ecotox: 100x more toxic than baseline.	Yang, R. S. H.; Witt, K. L.; Alden, C. J.; Cockerham, L. G. Toxicology of Methyl Bromide. In Reviews of Environmental Contamination and Toxicology; Ware, G. W., Ed.; Reviews of Environmental Contamination and Toxicology; Springer New York, 1995; pp 65–85. KREATiS internal experimental database and QSARs.
358		KREATiS internal experimental database and QSARs.
359		KREATiS internal experimental database and QSARs.
360	ecotox: baseline toxicity	KREATiS internal experimental database and QSARs.
361	Algae: baseline toxicity	KREATiS internal experimental database and QSARs.
362		KREATiS internal experimental database and QSARs.
363		KREATiS internal experimental database and QSARs.
364	H/M: metabolised by CYP450 into phosgene, highly reative (generates HCl in the medium and binds to proteins to make adducts). Ecotox: baseline toxicity	ATSDR. Toxicological Profile: Chloroform https://www.atsdr.cdc.gov/toxprofiles/tp.asp?id=53&tid=16 (accessed Oct 26, 2016). KREATiS internal experimental database and QSARs.
365		KREATiS internal experimental database and QSARs.
366		KREATiS internal experimental database and QSARs.
367		KREATiS internal experimental database and QSARs.
368	Fish: baseline toxicity	KREATiS internal experimental database and QSARs.
369	ecotox: baseline toxicity	KREATiS internal experimental database and QSARs.
370		KREATiS internal experimental database and QSARs.
371	Fish: 20x more toxic than baseline	KREATiS internal experimental database and QSARs.
372	Ecotox: between 1,5 and 2 times more toxic than baseline. M: metabolised into trichloroethanol, trichloroethylene, no surprising or high toxic effects.	Pubchem, citing [Bingham, E.; Cohrssen, B.; Powell, C.H.; Patty's Toxicology Volumes 1-9 5th ed. John Wiley & Sons. New York, N.Y. (2001)., p. 144]. KREATiS internal experimental database and QSARs.
373		KREATiS internal experimental database and QSARs.
374		KREATiS internal experimental database and QSARs.
375		KREATiS internal experimental database and QSARs.
376		KREATiS internal experimental database and QSARs.
377		KREATiS internal experimental database and QSARs.
378	ecotox: baseline toxicity	KREATiS internal experimental database and QSARs.
379		KREATiS internal experimental database and QSARs.
380		KREATiS internal experimental database and QSARs.
381		KREATiS internal experimental database and QSARs.
382	M: activates and desensitizes TRP channels, thus having an analgesic effect.	Xu, H.; Blair, N. T.; Clapham, D. E. Camphor Activates and Strongly Desensitizes the Transient Receptor Potential Vanilloid Subtype 1 Channel in a Vanilloid-Independent Mechanism. J. Neurosci. Off. J. Soc. Neurosci. 2005, 25 (39), 8924–8937 DOI: 10.1523/JNEUROSCI.2574-05.2005. KREATiS internal experimental database and QSARs.
383		KREATiS internal experimental database and QSARs.
384	ecotox: baseline toxicity	KREATiS internal experimental database and QSARs.
385	Ecotox: 7x more toxic to fish, baseline for daphnia, 2x more toxic for algae. M: calcium channel blocker in intestine (inhibits contraction).	Souza, F. V. M.; da Rocha, M. B.; de Souza, D. P.; Marçal, R. M. (−)-Carvone: Antispasmodic Effect and Mode of Action. Fitoterapia 2013, 85, 20–24 DOI: 10.1016/j.fitote.2012.10.012. KREATiS internal experimental database and QSARs.
386	Fish: 8x more toxic than baseline. (weak) inhibition of tyrosinase	KREATiS internal experimental database and QSARs. Liu, X.; Jia, Y.; Chen, J.; Liang, G.; Guo, H.; Hu, Y.; Shi, Y.; Zhou, H.-T.; Chen, Q.-X. Inhibition Effects of Benzylideneacetone, Benzylacetone, and 4-Phenyl-2-Butanol on the Activity of Mushroom Tyrosinase. J. Biosci. Bioeng. 2015, 119 (3), 275–279 DOI: 10.1016/j.jbiosc.2014.08.014.
387	endogenous compound involved in different endogenous pathways. Ecotox: between aldehyde and baseline toxicity.	KREATiS internal experimental database and QSARs.
388	ecotox: baseline toxicity	KREATiS internal experimental database and QSARs.
389	ecotox: baseline toxicity	KREATiS internal experimental database and QSARs.
390	ecotox: baseline toxicity	KREATiS internal experimental database and QSARs.
391	Fish and Daphnid: baseline. Algae: 9x more toxic.	KREATiS internal experimental database and QSARs.
392		KREATiS internal experimental database and QSARs.
393	Ecotox: polar narcosis toxicity	KREATiS internal experimental database and QSARs.
394	Reprotoxic. Ecotox: less toxic than baseline…	KREATiS internal experimental database and QSARs.
395	spontaneously decomposes giving CN- and acetaldehyde. Very high toxicity to aquatic life and mammals.	Grogan, J.; DeVito, S. C.; Pearlman, R. S.; Korzekwa, K. R. Modeling Cyanide Release from Nitriles: Prediction of Cytochrome P 450-Mediated Acute Nitrile Toxicity. *Chem. Res. Toxicol.* **1992**, *5* (4), 548–552 DOI: 10.1021/tx00028a014.
396	Specific study on this compound showed metabolism by CYP2E1 generating CN-, correlated with toxicity. In addition, epoxide is formed by CYP2A5. Fish: ca. 10x more toxic than baseline.	Boadas-Vaello, P.; Jover, E.; Saldaña-Ruíz, S.; Soler-Martín, C.; Chabbert, C.; Bayona, J. M.; Llorens, J. Allylnitrile Metabolism by CYP2E1 and Other CYPs Leads to Distinct Lethal and Vestibulotoxic Effects in the Mouse. *Toxicol. Sci. Off. J. Soc. Toxicol.* **2009**, *107* (2), 461–472 DOI: 10.1093/toxsci/kfn233. KREATiS internal experimental database and QSARs.
397	no H/M data. Ecotox seems much less toxic than baseline.	KREATiS internal experimental database and QSARs.
398	ecotox: baseline toxicity. M: probably genotoxic. Reduction into nitroso cause DNA and protein damage. Induces oxidative stress, generating methemoglobinemia.	IARC. 2010. 'General Discussion of Common Mechanisms for Aromatic Amines.' IARC Monographs on the Evaluation of Carcinogenic Risks to Humans 99: 41–54. KREATiS internal experimental database and QSARs.

399	ecotox: baseline toxicity. M: probably genotoxic. Reduction into nitroso cause DNA and protein damage. Induces oxidative stress, generating methemoglobinemia.	IARC. 2010. 'General Discussion of Common Mechanisms for Aromatic Amines.' IARC Monographs on the Evaluation of Carcinogenic Risks to Humans 99: 41–54. KREATiS internal experimental database and QSARs.
400	Ecotox: polar narcosis toxicity	KREATiS internal experimental database and QSARs.
401	Ecotox: polar narcosis toxicity	KREATiS internal experimental database and QSARs.
402		KREATiS internal experimental database and QSARs.
403		KREATiS internal experimental database and QSARs.
404		KREATiS internal experimental database and QSARs.
405		KREATiS internal experimental database and QSARs.
406		KREATiS internal experimental database and QSARs.
407	Ecotox/M: polar narcosis.	Ferreira, J. V. N.; Capello, T. M.; Siqueira, L. J. A.; Lago, J. H. G.; Caseli, L. Mechanism of Action of Thymol on Cell Membranes Investigated through Lipid Langmuir Monolayers at the Air–Water Interface and Molecular Simulation. Langmuir 2016, 32 (13), 3234–3241 DOI: 10.1021/acs.langmuir.6b00600. KREATiS internal experimental database and QSARs.
408	polar narcosis	KREATiS internal experimental database and QSARs.
409		KREATiS internal experimental database and QSARs.
410		KREATiS internal experimental database and QSARs.
411		KREATiS internal experimental database and QSARs.
412		KREATiS internal experimental database and QSARs.
413		KREATiS internal experimental database and QSARs.
414		KREATiS internal experimental database and QSARs.
415		KREATiS internal experimental database and QSARs.
416	antioxidant (radical trapping).	Rahimtula, A. D.; Jernström, B.; Dock, L.; Moldeus, P. Effects of Dietary and in Vitro 2(3)-t-Butyl-4-Hydroxy-Anisole and Other Phenols on Hepatic Enzyme Activities in Mice. *Br. J. Cancer* **1982**, *45* (6), 935–944. KREATiS internal experimental database and QSARs.
417	it is an estrogenic compound (3 to 6 orders of magnitude less potent than estradiol), but the corresponding effects can only be seen at high doses.	OEHHA. Toxicological Profile for Nonylphenol. September 2009. KREATiS internal experimental database and QSARs.
418	antioxidant (radical trapping). Creates adducts after metabolisation into a quinone methide in the lung.	Shearn, C. T.; Fritz, K. S.; Meier, B. W.; Kirichenko, O. V.; Thompson, J. A. Carbonyl Reductase Inactivation May Contribute to Mouse Lung Tumor Promotion by Electrophilic Metabolites of Butylated Hydroxytoluene: Protein Alkylation in Vivo and in Vitro. *Chem. Res. Toxicol.* **2008**, *21* (8), 1631–1641 DOI: 10.1021/tx800162p. KREATiS internal experimental database and QSARs.
419		Shearn, C. T.; Fritz, K. S.; Meier, B. W.; Kirichenko, O. V.; Thompson, J. A. Carbonyl Reductase Inactivation May Contribute to Mouse Lung Tumor Promotion by Electrophilic Metabolites of Butylated Hydroxytoluene: Protein Alkylation in Vivo and in Vitro. *Chem. Res. Toxicol.* **2008**, *21* (8), 1631–1641 DOI: 10.1021/tx800162p. KREATiS internal experimental database and QSARs.
420	Ecotox: polar narcosis toxicity	KREATiS internal experimental database and QSARs.
421	M: minor pathway is oxidation into phenylhydroquinone, and then redox cycle between phenylhydroquinone and phenylbenzoquinone.	IARC. Monograph of Ortho-Pehylphenol and Its Sodium Salt; 1999; Vol. 73, p 451. KREATiS internal experimental database and QSARs.
422	M: minor pathway is oxidation into phenylcatechol, and then redox cycle between phenylcatechol and phenylbenzoquinone.	IARC. Monograph of Ortho-Pehylphenol and Its Sodium Salt; 1999; Vol. 73, p 451. KREATiS internal experimental database and QSARs.
423	Binds to Estrogen-Related Receptor (ERR) Gamma (like Bisphenol A which has a hydroxy group in addition)	Matsushima, A.; Teramoto, T.; Okada, H.; Liu, X.; Tokunaga, T.; Kakuta, Y.; Shimohigashi, Y. ERRγ Tethers Strongly Bisphenol A and 4-α-Cumylphenol in an Induced-Fit Manner. *Biochem. Biophys. Res. Commun.* **2008**, *373* (3), 408–413 DOI: 10.1016/j.bbrc.2008.06.050. KREATiS internal experimental database and QSARs.
424	all esters of salicylic acid present the same effects than salicylic acid. Inhibition of COX-2 and -1. Ecotox: salicylates stand between esters and phenols (most are bad data)	Vane, John R, and Regina M Botting. 1998. 'Mechanism of Action of Nonsteroidal Anti-Inflammatory Drugs'. The American Journal of Medicine 104 (3, Supplement 1): 2S – 8S. doi:10.1016/S0002-9343(97)00203-9. KREATiS internal experimental database and QSARs.
425	all esters of salicylic acid present the same effects than salicylic acid. Inhibition of COX-2 and -1. Ecotox: salicylates stand between esters and phenols (most are bad data)	Vane, John R, and Regina M Botting. 1998. 'Mechanism of Action of Nonsteroidal Anti-Inflammatory Drugs'. The American Journal of Medicine 104 (3, Supplement 1): 2S – 8S. doi:10.1016/S0002-9343(97)00203-9. KREATiS internal experimental database and QSARs.
426	all esters of salicylic acid present the same effects than salicylic acid. Inhibition of COX-2 and -1. Ecotox: salicylates stand between esters and phenols (most are bad data)	Vane, John R, and Regina M Botting. 1998. 'Mechanism of Action of Nonsteroidal Anti-Inflammatory Drugs'. The American Journal of Medicine 104 (3, Supplement 1): 2S – 8S. doi:10.1016/S0002-9343(97)00203-9. KREATiS internal experimental database and QSARs.
427	all esters of salicylic acid present the same effects than salicylic acid. Inhibition of COX-2 and -1. Ecotox: salicylates stand between esters and phenols (most are bad data)	Vane, John R, and Regina M Botting. 1998. 'Mechanism of Action of Nonsteroidal Anti-Inflammatory Drugs'. The American Journal of Medicine 104 (3, Supplement 1): 2S – 8S. doi:10.1016/S0002-9343(97)00203-9. KREATiS internal experimental database and QSARs.
428	all esters of salicylic acid present the same effects than salicylic acid. Inhibition of COX-2 and -1. Ecotox: salicylates stand between esters and phenols (most are bad data)	Vane, John R, and Regina M Botting. 1998. 'Mechanism of Action of Nonsteroidal Anti-Inflammatory Drugs'. The American Journal of Medicine 104 (3, Supplement 1): 2S – 8S. doi:10.1016/S0002-9343(97)00203-9. KREATiS internal experimental database and QSARs.
429	all esters of salicylic acid present the same effects than salicylic acid. Inhibition of COX-2 and -1. Ecotox: salicylates stand between esters and phenols (most are bad data)	Vane, John R, and Regina M Botting. 1998. 'Mechanism of Action of Nonsteroidal Anti-Inflammatory Drugs'. The American Journal of Medicine 104 (3, Supplement 1): 2S – 8S. doi:10.1016/S0002-9343(97)00203-9. KREATiS internal experimental database and QSARs.
430	all esters of salicylic acid present the same effects than salicylic acid. Inhibition of COX-2 and -1. Ecotox: salicylates stand between esters and phenols (most are bad data)	Vane, John R, and Regina M Botting. 1998. 'Mechanism of Action of Nonsteroidal Anti-Inflammatory Drugs'. The American Journal of Medicine 104 (3, Supplement 1): 2S – 8S. doi:10.1016/S0002-9343(97)00203-9. KREATiS internal experimental database and QSARs.
431	all esters of salicylic acid present the same effects than salicylic acid. Inhibition of COX-2 and -1. Ecotox: salicylates stand between esters and phenols (most are bad data)	Vane, John R, and Regina M Botting. 1998. 'Mechanism of Action of Nonsteroidal Anti-Inflammatory Drugs'. The American Journal of Medicine 104 (3, Supplement 1): 2S – 8S. doi:10.1016/S0002-9343(97)00203-9. KREATiS internal experimental database and QSARs.
432	Ecotox: polar narcosis toxicity	KREATiS internal experimental database and QSARs.
433	Visible effects in vivo seem not much different than another phenol. Ecotox: polar narcosis toxicity	Riley, P. A. Mechanism of Pigment-Cell Toxicity Produced by Hydroxyanisole. J. Pathol. 1970, 101 (2), 163–169 DOI: 10.1002/path.1711010211. KREATiS internal experimental database and QSARs.
434	M: Dealkoxylation, then RedOx cycle between reduced and oxidized quinone. Ecotox: approx. 3x more toxic than polar narcosis. Oxidised by melanocytes tyrosinase.	Riley, P. A. Mechanism of Pigment-Cell Toxicity Produced by Hydroxyanisole. J. Pathol. 1970, 101 (2), 163–169 DOI: 10.1002/path.1711010211. KREATiS internal experimental database and QSARs.
435	M: Dealkoxylation, then RedOx cycle between reduced and oxidized quinone. Ecotox: approx. 3x more toxic than polar narcosis.	Riley, P. A. Mechanism of Pigment-Cell Toxicity Produced by Hydroxyanisole. J. Pathol. 1970, 101 (2), 163–169 DOI: 10.1002/path.1711010211. KREATiS internal experimental database and QSARs.
436	OPU (H+ transport), seen in ecotox KTS	KREATiS internal experimental database and QSARs.

437	Ecotox: 5x more toxic to fish, 70x more toxic for daphnia, 10x more toxic to algae than polar narcosis. M: in high doses, it inhibits thyroid peroxidases thus disrupting the depending hormone system.	Welsch, F. Routes and Modes of Administration of Resorcinol and Their Relationship to Potential Manifestations of Thyroid Gland Toxicity in Animals and Man. *Int. J. Toxicol.* **2008**, *27* (1), 59–63 DOI: 10.1080/10915810701876687. KREATiS internal experimental database and QSARs.
438	Ecotox: 100x more toxic to fish, 20x to daphnia, 3x to algae than polar narcosis. Oxidised by tyrosinase.	Riley, P. A. Mechanism of Pigment-Cell Toxicity Produced by Hydroxyanisole. J. Pathol. 1970, 101 (2), 163–169 DOI: 10.1002/path.1711010211. KREATiS internal experimental database and QSARs.
439	Ecotox: not more toxic to fish, 5x to daphnia and algae than polar narcosis. M: agonist of Estrogen Receptor.	Paris, F.; Balaguer, P.; Térouanne, B.; Servant, N.; Lacoste, C.; Cravedi, J.-P.; Nicolas, J.-C.; Sultan, C. Phenylphenols, Biphenols, Bisphenol-A and 4-Tert-Octylphenol Exhibit α and β Estrogen Activities and Antiandrogen Activity in Reporter Cell Lines. Mol. Cell. Endocrinol. 2002, 193 (1–2), 43–49 DOI: 10.1016/S0303-7207(02)00094-1. KREATiS internal experimental database and QSARs.
440	Ecotox: not more toxic to fish, 7x to daphnia and 10x to algae than polar narcosis.	KREATiS internal experimental database and QSARs.
441	Fish: polar narcosis toxicity. Estrogen Receptor agonist, induce oxidative stress (metabolism to catechol moieties?) (and adducts to proteins). A lot of other effects, probably caused by these 2?	Rezg, R.; El-Fazaa, S.; Gharbi, N.; Mornagui, B. Bisphenol A and Human Chronic Diseases: Current Evidences, Possible Mechanisms, and Future Perspectives. Environ. Int. 2014, 64, 83–90 DOI: 10.1016/j.envint.2013.12.007. KREATiS internal experimental database and QSARs.
442	certainly good antioxidant properties.	KREATiS internal experimental database and QSARs.
443	Algae: 2,4x more toxic than polar narcosis. no H/M data.	KREATiS internal experimental database and QSARs.
444	so hydrophobic that it is not bioavailable, not toxic at solubility limit in aquatic toxicity, not especially toxic in plants and soil organisms.	KREATiS internal experimental database and QSARs.
445	B,M: Inihibits electron transport (lactate dehydrogenase / succinate dehydrogenase in complex II) in the membranes.	Lokanatha, V.; Sailaja, P.; Rajendra, W. In Vitro Kinetics of the Rat Brain Succinate Dehydrogenase Inhibition by Hexachlorophene. *J. Biochem. Mol. Toxicol.* **1999**, *13* (6), 303–306 DOI: 10.1002/(SICI)1099-0461(1999)13:6<303::AID-JBT3>3.0.CO;2-3.
446	1° & 2° thiols, in their anionic form, are oxidized into thiyl radical and then enter a redox cycle between the thiyl radical and the disulfides. 3° are not reactive enough due to steric hindrance. Aromatics are more reactive (conjugation stabilization), EWG are stabilizing the anion, therefore decreasing the reactivity, while EDG will stabilize the thiyl radical and favor the oxidation.	Munday, R. Toxicity of Thiols and Disulphides: Involvement of Free-Radical Species. Free Radic. Biol. Med. 1989, 7 (6), 659–673 DOI: 10.1016/0891-5849(89)90147-0. KREATiS internal experimental database and QSARs.
447	1° & 2° thiols, in their anionic form, are oxidized into thiyl radical and then enter a redox cycle between the thiyl radical and the disulfides. 3° are not reactive enough due to steric hindrance. Aromatics are more reactive (conjugation stabilization), EWG are stabilizing the anion, therefore decreasing the reactivity, while EDG will stabilize the thiyl radical and favor the oxidation.	Munday, R. Toxicity of Thiols and Disulphides: Involvement of Free-Radical Species. Free Radic. Biol. Med. 1989, 7 (6), 659–673 DOI: 10.1016/0891-5849(89)90147-0. KREATiS internal experimental database and QSARs.
448	1° & 2° thiols, in their anionic form, are oxidized into thiyl radical and then enter a redox cycle between the thiyl radical and the disulfides. 3° are not reactive enough due to steric hindrance. Aromatics are more reactive (conjugation stabilization), EWG are stabilizing the anion, therefore decreasing the reactivity, while EDG will stabilize the thiyl radical and favor the oxidation.	Munday, R. Toxicity of Thiols and Disulphides: Involvement of Free-Radical Species. Free Radic. Biol. Med. 1989, 7 (6), 659–673 DOI: 10.1016/0891-5849(89)90147-0. KREATiS internal experimental database and QSARs.
449	1° & 2° thiols, in their anionic form, are oxidized into thiyl radical and then enter a redox cycle between the thiyl radical and the disulfides. 3° are not reactive enough due to steric hindrance. Aromatics are more reactive (conjugation stabilization), EWG are stabilizing the anion, therefore decreasing the reactivity, while EDG will stabilize the thiyl radical and favor the oxidation.	Munday, R. Toxicity of Thiols and Disulphides: Involvement of Free-Radical Species. Free Radic. Biol. Med. 1989, 7 (6), 659–673 DOI: 10.1016/0891-5849(89)90147-0. KREATiS internal experimental database and QSARs.
450	1° & 2° thiols, in their anionic form, are oxidized into thiyl radical and then enter a redox cycle between the thiyl radical and the disulfides. 3° are not reactive enough due to steric hindrance. Aromatics are more reactive (conjugation stabilization), EWG are stabilizing the anion, therefore decreasing the reactivity, while EDG will stabilize the thiyl radical and favor the oxidation.	Munday, R. Toxicity of Thiols and Disulphides: Involvement of Free-Radical Species. Free Radic. Biol. Med. 1989, 7 (6), 659–673 DOI: 10.1016/0891-5849(89)90147-0. KREATiS internal experimental database and QSARs.
451	1° & 2° thiols, in their anionic form, are oxidized into thiyl radical and then enter a redox cycle between the thiyl radical and the disulfides. 3° are not reactive enough due to steric hindrance. Aromatics are more reactive (conjugation stabilization), EWG are stabilizing the anion, therefore decreasing the reactivity, while EDG will stabilize the thiyl radical and favor the oxidation.	Munday, R. Toxicity of Thiols and Disulphides: Involvement of Free-Radical Species. Free Radic. Biol. Med. 1989, 7 (6), 659–673 DOI: 10.1016/0891-5849(89)90147-0. KREATiS internal experimental database and QSARs.
452	Baseline toxicity for fish and algae, 5x more toxic for daphnia	KREATiS internal experimental database and QSARs.
453	OPU (H+ transport), seen in ecotox KTS	KREATiS internal experimental database and QSARs.
454	OPU (H+ transport), seen in ecotox KTS	KREATiS internal experimental database and QSARs.
455	OPU (H+ transport), seen in ecotox KTS	KREATiS internal experimental database and QSARs.
456	OPU (H+ transport), seen in ecotox KTS	KREATiS internal experimental database and QSARs.
457	binds to barbiturate site of GABAa receptors, increasing the activity of GABA.	Olsen, R. W.; DeLorey, T. M. GABA Receptor Physiology and Pharmacology. 1999.
458	muscarinic antagonist	King, M. W. Biochemistry of Neurotransmitters and Nerve Transmission http://themedicalbiochemistrypage.org/nerves.php (accessed Sep 25, 2015).
459	strong agonist of estrogen receptor	Shi, L.; Tong, W.; Fang, H.; Xie, Q.; Hong, H.; Perkins, R.; Wu, J.; Tu§, M.; Blair, R. M.; Branham, W. S.; Waller, C.; Walker, J.; Sheehan, D. M. An Integrated "4-Phase" Approach for Setting Endocrine Disruption Screening Priorities–phase I and II Predictions of Estrogen Receptor Binding Affinity. *SAR QSAR Environ. Res.* **2002**, *13* (1), 69–88 DOI: 10.1080/10629360290002235.
460	strong agonist of estrogen receptor	Shi, L.; Tong, W.; Fang, H.; Xie, Q.; Hong, H.; Perkins, R.; Wu, J.; Tu§, M.; Blair, R. M.; Branham, W. S.; Waller, C.; Walker, J.; Sheehan, D. M. An Integrated "4-Phase" Approach for Setting Endocrine Disruption Screening Priorities–phase I and II Predictions of Estrogen Receptor Binding Affinity. *SAR QSAR Environ. Res.* **2002**, *13* (1), 69–88 DOI: 10.1080/10629360290002235.

461	nicotinic agonist	King, M. W. Biochemistry of Neurotransmitters and Nerve Transmission http://themedicalbiochemistrypage.org/nerves.php (accessed Sep 25, 2015).
462	agonist of morphine at μ-opioid receptor	Pathan, H.; Williams, J. Basic Opioid Pharmacology: An Update. Br. J. Pain 2012, 6 (1), 11–16 DOI: 10.1177/2049463712438493. Chahl, L. A. Experimental and Clinical Pharmacology: Opioids - Mechanisms of Action. Aust. Prescr. 1996, 19 (3), 63–65 DOI: 10.18773/austprescr.1996.063.
463	agonist of morphine at μ-opioid receptor	Pathan, H.; Williams, J. Basic Opioid Pharmacology: An Update. Br. J. Pain 2012, 6 (1), 11–16 DOI: 10.1177/2049463712438493. Chahl, L. A. Experimental and Clinical Pharmacology: Opioids - Mechanisms of Action. Aust. Prescr. 1996, 19 (3), 63–65 DOI: 10.18773/austprescr.1996.063.
464	OPU (H+ transport), seen in ecotox KTS	KREATiS internal experimental database and QSARs.
465	AChE inhibitor	Elersek, T.; Filipic, M. Organophosphorus Pesticides - Mechanisms of Their Toxicity. In *Pesticides - The Impacts of Pesticides Exposure* ; Stoytcheva, M., Ed.; InTech, 2011.
466	Plants: inhibits Protein D-1 at photosystem II => blocks production of O2 + ROS generated. Other: carcinogenic, reproductive and developmental toxicity (several MechoAs are postulated to explain these effects)	Breckenridge, C. B.; Werner, C.; Stevens, J. T.; Sumner, D. D. Chapter 25 - Hazard Assessment for Selected Symmetrical and Asymmetrical Triazine Herbicides. In The Triazine Herbicides; Elsevier: San Diego, 2008; pp 387–398. APVMA. *Atrazine Toxicity: Analysis of Potential Modes of Action* ; APVMA, 2010; p 49.
467	AChE inhibitor	Fukuto, T. R. Mechanism of Action of Organophosphorus and Carbamate Insecticides. *Environ. Health Perspect.* **1990**, *87* , 245–254.
468	AChE inhibitor	Fukuto, T. R. Mechanism of Action of Organophosphorus and Carbamate Insecticides. *Environ. Health Perspect.* **1990**, *87* , 245–254.
469	binds to benzodiazepine site of GABAa receptors, increasing the activity of GABA.	Olsen, R. W.; DeLorey, T. M. GABA Receptor Physiology and Pharmacology. 1999.
470	nicotinic agonist	Tzankova, V.; Danchev, N. Cytisine—from Ethomedical Use to the Development as a Natural Alternative for Smoking Cessation. *Biotechnol. Biotechnol. Equip.* **2007**, *21* (2), 151–160 DOI: 10.1080/13102818.2007.10817436.
471	OPU (H+ transport), seen in ecotox KTS	KREATiS internal experimental database and QSARs.
472	binds to progesterone and estrogen receptors, inhibiting ovulation.	Kahlenborn, C.; Peck, R.; Severs, W. B. Mechanism of Action of Levonorgestrel Emergency Contraception. *Linacre Q.* **2015**, *82* (1), 18–33 DOI: 10.1179/2050854914Y.0000000026.
473	Plants: hydrolysed. Then the acid is an antagonist of phytohormone auxin, overstimulating plant growth, more than it can afford (because too slowly metabolised then). M: after hydrolysis, acid rapidly eliminated.	Song, Y. Insight into the Mode of Action of 2,4-Dichlorophenoxyacetic Acid (2,4-D) as an Herbicide. J. Integr. Plant Biol. 2014, 56 (2), 106–113 DOI: 10.1111/jipb.12131. Jervais, G.; Luukinen, B.; Buhl, K.; Stone, D. 2,4-D Technical Fact Sheet http://npic.orst.edu/factsheets/archive/2,4-DTech.html (accessed Dec 16, 2016).
474	AChE inhibitor	Fukuto, T. R. Mechanism of Action of Organophosphorus and Carbamate Insecticides. *Environ. Health Perspect.* **1990**, *87* , 245–254.
475	AChE inhibitor	Fukuto, T. R. Mechanism of Action of Organophosphorus and Carbamate Insecticides. *Environ. Health Perspect.* **1990**, *87* , 245–254.
476	binds to picrotoxin site of GABAa receptors, blocking the channel. Very toxic for mammals and fish.	Olsen, R. W.; DeLorey, T. M. GABA Receptor Physiology and Pharmacology. 1999.
477	oxidized by metabolism to generate MPP+, RedOx cycler. It is inducing Parkinson decease, because MPP+ is generated in astrocytes in the brain, and then actively transported to dopaminergic neurons by dopamine transporters, to attack mitochondria inside these neurons.	Langston, J. W. The MPTP Story. J. Park. Dis. 7 (Suppl 1), S11–S22 DOI: 10.3233/JPD-179006. Adams, J. D.; Chang, M. L.; Klaidman, L. Parkinson's Disease--Redox Mechanisms. Curr. Med. Chem. 2001, 8 (7), 809–814.
478	Conjugation with thiol groups (including CoA, inhibiting all the processes using it), not precisely known… Plants: makes it a good herbicide. Plants & Other: detoxification with GSH.	Fuerst, E. P. Understanding the Mode of Action of the Chloroacetamide and Thiocarbamate Herbicides. *Weed Technol.* **1987**, *1* (4), 270–277.
479	Plants: inhibits Protein D-1 at photosystem II => blocks production of O2 + ROS generated. Other: slightly to not toxic, in acute and in chronic.	Breckenridge, C. B.; Werner, C.; Stevens, J. T.; Sumner, D. D. Chapter 25 - Hazard Assessment for Selected Symmetrical and Asymmetrical Triazine Herbicides. In The Triazine Herbicides; Elsevier: San Diego, 2008; pp 387–398. APVMA. *Atrazine Toxicity: Analysis of Potential Modes of Action* ; APVMA, 2010; p 49.
480	prevents the closure of Na+ channel from the nerve, maintaining the signal / inhibits GABAergic Cl- channel into the nerve / inhibits Ca-ATPase or Ca-Mg-ATPase, disturbing Ca2+ regulation	Coats, J R. 1990. 'Mechanisms of Toxic Action and Structure-Activity Relationships for Organochlorine and Synthetic Pyrethroid Insecticides.' Environmental Health Perspectives 87 (July): 255–62.
481	AChE inhibitor	Fukuto, T. R. Mechanism of Action of Organophosphorus and Carbamate Insecticides. *Environ. Health Perspect.* **1990**, *87* , 245–254.
482	Plants: hydrolysed. Then, beta-oxidation to obtain the phenoxyacetic acid which is an antagonist of phytohormone auxin, overstimulating plant growth, more than it can afford (because too slowly metabolised then). M: after hydrolysis, acid rapidly eliminated.	Arteca, R. N. Chapter 13 Weed Control. In *Plant Growth Substances: Principles and Applications* ; Springer Science & Business Media, 1996; p 282.
483	prevents the closure of Na+ channel from the nerve, maintaining the signal / inhibits GABAergic Cl- channel into the nerve / inhibits Ca-ATPase or Ca-Mg-ATPase, disturbing Ca2+ regulation	Coats, J R. 1990. 'Mechanisms of Toxic Action and Structure-Activity Relationships for Organochlorine and Synthetic Pyrethroid Insecticides.' Environmental Health Perspectives 87 (July): 255–62.
484	inhibit acetolactate synthase, an enzyme involved in the biosynthesis of valine, leucine and isoleucine in plants, but not in animals.	Tu, M.; Hurd, C.; Randall, J. M. Imazapyr. In *Weed Control Methods Handbook: tools & techniques for use in natural areas* ; The Nature Conservancy, 2001.
485	selective estrogen receptor modulator. Agonist in bone and in lipi metabolism, antagonist in uterine endometrium and breast.	Shi, L.; Tong, W.; Fang, H.; Xie, Q.; Hong, H.; Perkins, R.; Wu, J.; Tu§, M.; Blair, R. M.; Branham, W. S.; Waller, C.; Walker, J.; Sheehan, D. M. An Integrated "4-Phase" Approach for Setting Endocrine Disruption Screening Priorities–phase I and II Predictions of Estrogen Receptor Binding Affinity. *SAR QSAR Environ. Res.* **2002**, *13* (1), 69–88 DOI: 10.1080/10629360290002235.
486	nicotinic agonist	Changgeng Qian; Tongchuan Li; Shen, T. Y.; Libertine-Garahan, L.; Eckman, J.; Biftu, T.; Ip, S. Epibatidine Is a Nicotinic Analgesic. *Eur. J. Pharmacol.* **1993**, *250* (3), R13–R14 DOI: 10.1016/0014-2999(93)90043-H.
487	AChE inhibitor	Fukuto, T. R. Mechanism of Action of Organophosphorus and Carbamate Insecticides. *Environ. Health Perspect.* **1990**, *87* , 245–254.
488	fish: polar narcosis toxicity (KREATiS)	KREATiS internal experimental database and QSARs.
489	fish: polar narcosis toxicity (KREATiS)	KREATiS internal experimental database and QSARs.
490	fish: polar narcosis toxicity (KREATiS)	KREATiS internal experimental database and QSARs.
491	fish: polar narcosis toxicity (KREATiS)	KREATiS internal experimental database and QSARs.
492	fish: polar narcosis toxicity (KREATiS)	KREATiS internal experimental database and QSARs.
493	fish: polar narcosis toxicity (KREATiS)	KREATiS internal experimental database and QSARs.

A.3) lignes directrices OCDE pour les tests en laboratoire

OECD GUIDELINE FOR THE TESTING OF CHEMICALS

Adopted by the Council on 27th July 1995

Partition Coefficient (n-octanol/water): Shake Flask Method

INTRODUCTION

1. This guideline is a revised version of the original Guideline 107 which was adopted in 1981. The only change is one of format. The substantive content of the original guideline was not changed.

INITIAL CONSIDERATIONS

2. P_{OW} values in the range log P_{OW} between -2 and 4 (occasionally up to 5) can be experimentally determined by the method set out in this guideline (1)(2). P_{OW} values in the range log P_{OW} between 0 and 6 can be estimated using high performance liquid chromatography as is set out in Guideline 117. Before deciding on what procedure to use, a preliminary estimate of the P_{OW} should be obtained from calculation (see the annex to Guideline 117), or where appropriate from the ratio of the solubilities of the test substance in the pure solvents (3).

3. Before determining P_{OW} it is useful to have preliminary information on the structural formula, dissociation constant, water solubility, hydrolysis, n-octanol solubility and surface tension of the substance.

4. The shake flask method is impossible to use with surface-active materials (for these, a P_{OW} can be calculated from individual solubilities in water and n-octanol).

DEFINITIONS AND UNITS

5. The partition coefficient (P) is defined as the ratio of the equilibrium concentrations of a dissolved substance in a two-phase system consisting of two largely immiscible solvents. In the case of n-octanol and water:

$$P_{ow} = \frac{c_{n\text{-}octanol}}{c_{water}}$$

The partition coefficient, being the quotient of two concentrations, or the quotient of the fractions of the test substance in the two phases multiplied by a fixed volume ratio, is dimensionless and is usually given in the form of its logarithm to base ten.

REFERENCE SUBSTANCES

6. Reference substances do not need to be employed. They serve primarily to check the performance of the method from time to time and to allow comparison with results from other methods.

PRINCIPLE OF THE METHOD

7. The Nernst partition law applies at constant temperature, pressure and pH for dilute solutions. It strictly applies to a pure substance dispersed between two pure solvents and when the concentration of the solute in either phase is not more than 0.01 mol per litre. If several different solutes occur in one or both phases at the same time, this may affect the results. Dissociation or association of the dissolved molecules result in deviations from the partition law. Such deviations are indicated by the fact that the partition coefficient becomes dependent upon the concentration. Measurements should be made on ionizable substances only in their non-ionized form (free acid or free base) produced by the use of an appropriate buffer with a pH of at least one unit below (free acid) or above (free base) the pK.

DESCRIPTION OF THE METHOD

Test substance and solvents

8. n-Octanol of analytical grade and distilled or double-distilled water should be used. Water taken directly from an ion exchanger should not be used. A stock solution of known concentration of the test substance in n-octanol, pre-saturated with water, is prepared. This solution should be stored under conditions which ensure its stability.

9. Before a partition coefficient is determined, the two solvents are mutually saturated at the temperature of the experiment. To do this, it is practical to shake two large stock bottles, one containing n-octanol and a sufficient quantity of water, and the other containing water and a sufficient quantity of n-octanol, for 24 hours on a mechanical shaker, and then to let them stand long enough to allow the phases to separate.

Test conditions

10. The test should be done at a temperature in the range 20 to 25°C, kept constant at ± 1°C.

11. For a first run, a volume ratio of n-octanol to water and a quantity of substance are chosen with regard given to:

- the preliminary estimation of the partition coefficient,
- the minimum concentration of test substance in each phase required for the analytical procedure,
- a maximum concentration of the test substance in each phase of 0,01 mol per litre.

For a second run, the originally chosen volume ratio is divided by two, and, for a third run, it is multiplied by two. In the second and third run the quantity of substance added may have to be different from that used in the first run in order to fulfill the above criteria.

12. Duplicate vessels containing accurately measured amounts of the two solvents and stock solution are used in all three runs.

Establishment of the partition equilibrium

13. The two-phase system should nearly fill the entire volume of the test vessels. This will help prevent loss of material due to volatilisation.

14. The test vessels are placed in a mechanical shaker or are shaken by hand. When using a centrifuge tube as the test vessel, it is recommended to rotate the tube through 180° about its transverse axis (approximately a hundred times during five minutes), allowing the trapped air to rise through the two phases.

Phase separation

15. The separation of the two phases, in general, is achieved by centrifugation. This should preferably be done at the test temperature. If a centrifuge without temperature control is used, the centrifuge tubes should be kept for equilibration at the test temperature for at least one hour before analysis.

Analysis

16. It is necessary to determine the concentrations of the test substance in both phases. For the measurement, substance-specific methods are preferred. Analytical methods which may be appropriate are: photometry, gas chromatography and high performance liquid chromatography. The total quantity of substance present in both phases should be calculated and compared with the quantity originally introduced.

17. The aqueous phase should be sampled by a procedure that minimizes the risk of including traces of n-octanol. This can be done using a syringe with a removable needle. The syringe should initially be partially filled with air. Air should be gently expelled while passing the n-octanol layer. The adequate volume of the aqueous solution is withdrawn. The syringe is quickly withdrawn and the needle is detached.

DATA AND REPORTING

Data

18. A P_{OW} value is calculated from the data of each run. Altogether six values are obtained since the test conditions are: three duplicate runs with different solvent ratios, and possibly also different quantities of test substance. The six log P_{OW} values should fall within a range of ± 0.3 units.

Test Report

19. The test report must include the following information:

- chemical identity and impurities;
- the results of the preliminary estimation (when the shake flask method is not applicable, e.g. surface active material, a calculated value or an estimate based on the individual n-octanol and water solubilities should be provided);
- all information relevant for the interpretation of the results, especially with regard to impurities and physical state of the substance;
- the test conditions: temperature, the amount of test substance introduced in the test vessels, the volume of each phase in each vessel and the calculated total amount of test substance based on the analytical data;
- pH of the water used and of the aqueous phase during the experiment;

- justification for the use of buffers; composition, concentration and pH of the buffers; pH of the aqueous phase before and after the experiment;
- duration and speed of centrifugation, if used;
- the analytical procedures;
- the concentrations measured in each run (a total of 12 concentrations);
- P_{OW} values and their mean for each set of test conditions and the overall mean (if there is the suggestion of concentration dependence of the partition coefficient, this should be noted);
- the standard deviation of individual P_{OW} values about their mean;
- the overall mean expressed as its logarithm to base 10;
- the theoretical P_{OW} when it has been calculated or when the measured value is above 10^4.

LITERATURE

(1) NF T 20-043 AFNOR (1985). Chemical products for industrial use - Determination of partition coefficient - Shake flask method

(2) 40 CFR, 796.1550, Office of the Federal Register (1989)

(3) Jübermann, O. (1958). in Houben-Weyl, Methoden der Organischen Chemie, Band I/1, 223-339, Georg Thieme Verlag, Stuttgart

OECD GUIDELINE FOR THE TESTING OF CHEMICALS

Adopted by the Council on 27th July 1995

Water Solubility

INTRODUCTION

1. This guideline is a revised version of the original Guideline 105 which was adopted in 1981. There are no differences of substance between the current version and that from 1981. Mainly the format has been changed. The revision was based on the EC method "Water Solubility" (1).

INITIAL CONSIDERATIONS

2. The water solubility of a substance can be considerably affected by the presence of impurities. This guideline addresses the determination of the solubility in water of essentially pure substances which are stable in water and not volatile. Before determining water solubility, it is useful to have some preliminary information on the substance, like structural formula, vapour pressure, dissociation constant and hydrolysis as a function of pH.

3. Two methods, the column elution method and the flask method which cover respectively solubilities below and above 10^{-2} g/l are described in this guideline. A simple preliminary test is also described. It allows to determine approximately the appropriate amount of sample to be used in the final test, as well as the time necessary to achieve saturation.

DEFINITIONS AND UNITS

4. The water solubility of a substance is the saturation mass concentration of the substance in water at a given temperature.

5. Water solubility is expressed in mass of solute per volume of solution. The SI unit is kg/m^3 but g/l is commonly used.

REFERENCE SUBSTANCES

6. Reference substances do not need to be employed when investigating a substance.

DESCRIPTION OF THE METHODS

Test conditions

7. The test is preferably run at 20 ± 0,5 °C. The chosen temperature should be kept constant in all relevant parts of the equipment.

Preliminary test

8. In a stepwise procedure, increasing volumes of water are added at room temperature to approximately 0.1 g of the sample (solid substances must be pulverized) in a 10 ml glass-stoppered measuring cylinder. After each addition of an amount of water, the mixture is shaken for 10 minutes and is visually checked for any undissolved parts of the sample. If, after addition of 10 ml of water, the sample or parts of it remain undissolved, the experiment is continued in a 100 ml measuring cylinder. The approximate solubility is given in Table 1 below under that volume of water in which complete dissolution of the sample occurs. When the solubility is low, a long time may be required to dissolve a substance and at least 24 hours should be allowed. If, after 24 hours, the substance is still not dissolved, more time (up to 96 hours) should be allowed or further dilution should be attempted to ascertain whether the column elution method or flask method should be used.

Table 1

ml of water for 0.1 g soluble	0.1	0.5	1	2	10	100	> 100
approximate solubility in g/l	> 1000	1000 to 200	200 to 100	100 to 50	50 to 10	10 to 1	< 1

Column elution method

Principle

9. This method is based on the elution of a test substance with water from a micro-column which is charged with an inert support material, previously coated with an excess of the test substance (2). The water solubility is given by the mass concentration of the eluate when this has reached a plateau as a function of time.

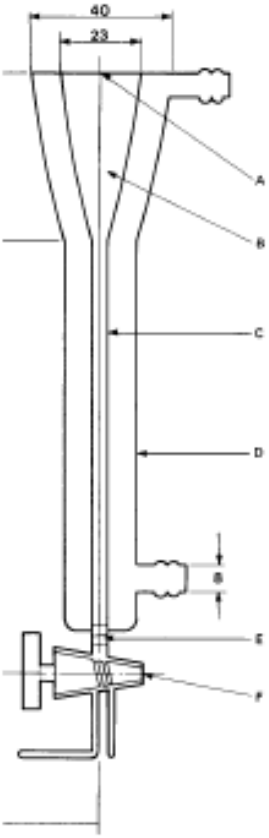
Apparatus

10. The apparatus consists of a microcolumn (Figure 1), maintained at constant temperature. It is connected either to a recirculating pump (Figure 2) or to a levelling vessel (Figure 3). The microcolumn contains an inert support held in place by a small plug of glasswool which also serves to filter out particles. Possible materials which can be employed for the support are glass beads, diatomaceous earth, or other inert materials.

11. The microcolumn shown in figure 1 is suitable for the set-up with recirculating pump. It has a head space providing for five bed volumes (discarded at the start of the experiment) and the volume of five samples (withdrawn for analysis during the experiment). Alternatively, the size can be reduced if water can be added to the system during the experiment to replace the initial five bed volumes removed with impurities. The column is connected with tubing made of an inert material to the recirculating pump, capable of delivering approximately 25 ml/h. The recirculating pump can be, for example, a peristaltic or membrane pump. Care must be taken that no contamination and/or adsorption occurs with the tube material.

12. A schematic arrangement using a levelling vessel is shown in figure 3. In this arrangement the microcolumn is fitted with a one way stopcock. The connection to the levelling vessel consists of a ground glass joint and tubing made of an inert material. Theflow rate from the levelling vessel should be approximately 25 ml/h.

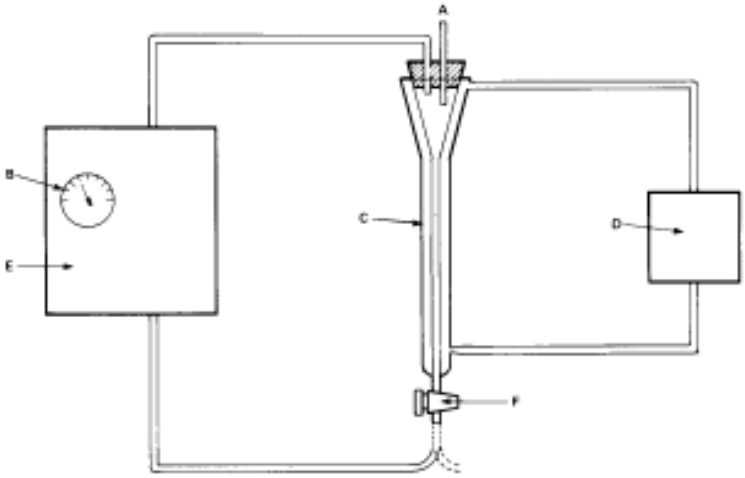
Figure 1



Dimensions in mm

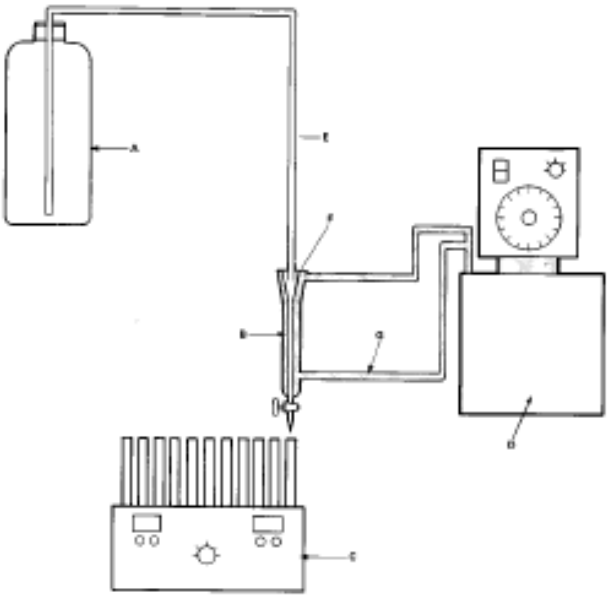
A. Connection for ground glass joint
B. Headspace
C. Interior 5
D. Exterior 19
E. Plug of glass wool
F. Stopcock

Figure 2



A. Atmospheric equilibration
B. Flowmeter
C. Microcolumn
D. Thermostatically controlled circulating pump
E. Recirculating pump
F. Two-way valve for sampling

Figure 3



A. Levelling vessel (e.g. 2.5 litres chemical flask)
B. Column
C. Fraction accumulator
D. Thermostat
E. Teflon tubing
F. Ground glass joint
G. Water line (between thermostat and column, inner diameter approximately 8 mm)

Loading of the support

13. Approximately 600 mg of support material is transferred to a 50 ml round-bottom flask. A suitable amount of test substance is dissolved in a volatile solvent of analytical reagent quality and an appropriate amount of this solution is added to the support material. The solvent is completely evaporated, e.g. using a rotary evaporator, as otherwise water saturation of the support will not be achieved during the elution step because of partitioning on the surface. The loaded support material is soaked for two hours in approximately 5 ml of water and the suspension is poured into the microcolumn. Alternatively, dry loaded support material may be poured into the water-filled microcolumn and two hours are allowed for equilibrating.

14. The loading of the support material may cause problems, leading to erroneous results, e.g. when the test substance is deposited as an oil. These problems should be examined and the details reported.

Procedure using a recirculating pump

15. The flow through the column is started. It is recommended that a flow rate of approximately 25 ml/h, corresponding to 10 bed volumes per hour for the column described, be used. At least the first five bed volumes are discarded to remove water soluble impurities. Following this, the pump is allowed to run until equilibrium is established, as defined by five successive samples whose concentrations do not differ by more than ± 30% in a random fashion. These samples should be separated from each other by time intervals corresponding to the passage of at least ten bed volumes. Depending on the analytical method used, it may be preferable to establish a concentration/time curve to show that equilibrium is reached.

Procedure using a levelling vessel

16. Successive eluate fractions should be collected and analyzed by the chosen method. Fractions from the middle eluate range, where the concentrations are constant within ± 30% in at least five consecutive fractions, are used to determine the solubility.

Notes applicable to both procedures

17. Double distilled water is the preferred eluent. Deionized water with a resistivity above 10 megohms/cm and a total organic carbon content below 0.01% can also be used.

18. Under both procedures, a second run is performed at half the flow rate of the first. If the results of the two runs are in agreement, the test is satisfactory. If the measured solubility is higher with the lower flow rate, then the halving of the flow rate must continue until two successive runs give the same solubility.

19. Under both procedures, the fractions should be checked for the presence of colloidal matter by examination of the Tyndall effect. The presence of particles invalidates the test and the test should be repeated after improvement of the filtering action of the column.

20. The pH of each sample should be measured, preferably by using special indicator strips.

Flask method

Principle

21. The substance (solids must be pulverized) is dissolved in water at a temperature somewhat above the test temperature. When saturation is achieved, the mixture is cooled and kept at the test temperature. Alternatively, and if it is assured by appropriate sampling that the saturation equilibrium is reached, the measurement can be performed directly at the test temperature. Subsequently, the mass concentration of the substance in the aqueous solution, which must not contain any undissolved particles, is determined by a suitable analytical method (3).

Apparatus

22. Following material is needed:

- normal laboratory glassware and instrumentation;
- a device for the agitation of solutions under controlled constant temperature;
- if required for emulsions, a centrifuge (preferably thermostatted); and
- analytical equipment.

Procedure

23. The quantity of test substance necessary to saturate the desired volume of water is estimated from the preliminary test. About five times thatquantity is weighed into each of three glass vessels fitted with glass stoppers (e.g. centrifuge tubes, flasks). A volume of water, chosen in function of the analytical method and solubility range, is added to each vessel. The vessels are tightly stoppered and then agitated at 30 °C. A shaking or stirring device capable of operating at constant temperature should be used, e.g. magnetic stirring in a thermostatted water bath. After one day, one of the vessels is equilibrated for 24 hours at the test temperature with occasional shaking. The contents of the vessel are then centrifuged at the test temperature and the concentration of the test substance in the clear aqueous phase is determined by a suitable analytical method. The other two flasks are treated similarly after initial equilibration at 30 °C for two and three days respectively. If the concentrations measured in at least the two last vessels do not differ by more than 15%, the test is satisfactory. If the results from vessels 1, 2 and 3 show a tendency of increasing values, the whole test should be repeated using longer equilibration times.

24. The test can also be performed without preincubation at 30 °C. In order to estimate the rate of establishment of the saturation equilibrium, samples are taken until the stirring time no longer influences the concentrations measured.

25. The pH of each sample should be measured, preferably by using special indicator strips.

Analytical determinations

26. A substance-specific method is preferred since small amounts of soluble impurities can cause large errors in the measured solubility. Examples of such methods are: gas or liquid chromatography, titration, photometry, voltametry.

DATA AND REPORTING

Data

Column elution method

27. For each run, the mean value and standard deviation from at least five consecutive samples taken from the saturation plateau should be calculated. The mean values obtained from two tests with different flows should not differ by more than 30%.

Flask method

28. The individual results from each of the three flasks, which should not differ by more than 15%, are averaged.

Test Report

Column elution method

29. The test report must include the following information:

- the results of the preliminary test;
- chemical identity and impurities (preliminary purification step, if any);
- the concentrations, flow rates and pH for each sample;
- the means and standard deviations from at least five samples from the saturation plateau of each run;
- the average of at least two successive runs;
- the temperature of the water during the saturation process;
- the method of analysis;
- the nature of the support material;
- loading of the support material;
- solvent used;
- evidence of any chemical instability of the substance during the test;
- all information relevant for the interpretation of the results, in particular with regard to impurities and physical state of the substance.

Flask method

30. The test report must include the following information:

- the results of the preliminary test;

- chemical identity and impurities (preliminary purification step, if any);
- the individual analytical determinations and the average where more than one value was determined for each flask;
- the pH of each sample;
- the average of the values for different flasks which were in agreement;
- the test temperature;
- the analytical method;
- evidence of any chemical instability of the substance during the test;
- all information relevant for the interpretation of the results, in particular with regard to impurities and physical state of the substance.

LITERATURE

(1) Official Journal of the European Communities L 383 A, 54-62 (1992)

(2) NF T 20-045 (AFNOR) (September 1985). Chemical products for industrial use - Determination of water solubility of solids and liquids with low solubility - Column elution method

(3) NF T 20-046 (AFNOR) (September 1985). Chemical products for industrial use - Determination of water solubility of solids and liquids with high solubility - Flask method

Adopted :
23 March 2006

OECD GUIDELINES FOR THE TESTING OF CHEMICALS

Partition Coefficient (1-Octanol/Water): Slow-Stirring Method

INTRODUCTION

1. 1-octanol/water partition coefficient (P_{OW}) values up to a log P_{OW} of 8.2 have been accurately determined by the slow-stirring method (1). Therefore it is a suitable experimental approach for the direct determination of P_{OW} of highly hydrophobic substances.

2. Other guidelines describing methods for the determination of the 1-octanol/water partition coefficient (P_{OW}) are OECD guidelines 107 and 117 (2) (3). The former refers to the 'shake-flask' method, the latter describes the determination of the P_{OW} from reversed phase HPLC-retention behavior. The 'shake-flask' method is prone to artifacts due to transfer of octanol microdroplets into the aqueous phase. With increasing values of P_{OW} the presence of these droplets in the aqueous phase leads to an increasing overestimation of the concentration of the test substance in the water. Therefore, its use is limited to substances with log $P_{OW} < 4$. The second method relies on solid data of directly determined P_{OW} values to calibrate the relationship between HPLC-retention behavior and measured values of P_{OW}. In addition, a draft OECD guideline is available for determining 1-octanol/water partition coefficients of ionisable substances (4).

3. This guideline has been developed in the Netherlands. The precision of the methods in this guideline has been validated and optimized in a ring-test validation study in which 15 laboratories participated (5).

INITIAL CONSIDERATIONS

Significance and use

4. For inert organic substances highly significant relationships have been found between 1-octanol/water partition coefficients (P_{OW}) and their bioaccumulation in fish. Moreover, P_{OW} has been demonstrated to be correlated to fish toxicity as well as to sorption of chemicals to solids such as soils and sediments. An extensive overview of the relationships has been given in reference (6).

5. A wide variety of relationships between the 1-octanol/water partition coefficient and other substance properties of relevance to environmental toxicology and chemistry have been established. As a consequence, the 1-octanol/water partition coefficient has evolved as a key parameter in the assessment of the environmental risk of chemicals as well as in the prediction of fate of chemicals in the environment.

Scope

6. The slow-stirring experiment is thought to reduce the formation of microdroplets from 1-octanol droplets in the water phase. As a consequence, overestimation of the aqueous concentration due to test substance molecules associated to such droplets does not occur. Therefore, the slow-stirring method is particularly suitable for the determination of P_{OW} for substances with expected log P_{OW} values of 5 and higher, for which the shake-flask method (2) is prone to yield erroneous results.

DEFINITION AND UNITS

7. The partition coefficient of a substance between water and a lipophilic solvent (1-octanol) characterizes the equilibrium distribution of the chemical between the two phases. The partition coefficient between water and 1-octanol (P_{OW}) is defined as the ratio of the equilibrium concentrations of the test substance in 1-octanol saturated with water (C_O) and water saturated with 1-octanol (C_W).

$$P_{OW} = C_O / C_W$$

As a ratio of concentrations it is dimensionless. Most frequently it is given as the logarithm to the base 10 (log P_{OW}). P_{OW} is temperature dependent and reported data should include the temperature of the measurement.

PRINCIPLE OF THE METHOD

8. In order to determine the partitioning coefficient, water, 1-octanol, and the test substance are equilibrated with each other at constant temperature. Then the concentrations of the test substance in the two phases are determined.

9. The experimental difficulties associated with the formation of microdroplets during the shake-flask experiment can be reduced in the slow-stirring experiment proposed here. In the slow-stirring experiment, water, 1-octanol and the test substance are equilibrated in a thermostated stirred reactor. Exchange between the phases is accelerated by stirring. The stirring introduces limited turbulence which enhances the exchange between 1-octanol and water without microdroplets being formed (1).

APPLICABILITY OF THE TEST

10. Since the presence of substances other than the test substance might influence the activity coefficient of the test substance, the test substance should be tested as a pure substance. The highest purity commercially available should be employed for the 1-octanol/water partition experiment.

11. The present method applies to pure substances that do not dissociate or associate and that do not display significant interfacial activity. It can be applied to determine the 1-octanol/water partition ratio of such substances and of mixtures. When the method is used for mixtures, the 1-octanol/water partition ratios determined are conditional and depend on the chemical composition of the mixture tested and on the electrolyte composition employed as aqueous phase. Provided additional steps are taken, the method is also applicable to dissociating or associating compounds (paragraph 12).

12. Due to the multiple equilibria in water and 1-octanol involved in the 1-octanol/water partitioning of dissociating substances such as organic acids and phenols, organic bases, and organometallic substances, the 1-octanol/water partition ratio is a conditional constant strongly dependent on electrolyte composition (7)(8). Determination of the 1-octanol/water partition ratio therefore requires that pH and electrolyte composition be controlled in the experiment and reported. Expert judgement has to be employed in the evaluation of these partition ratios. Using the value of dissociation constant(s), suitable pH-values need to be selected, such that a partitioning ratio is determined for each ionization state. Non-complexing buffers must be used when testing organometallic compounds (8). Taking the existing knowledge on the aqueous chemistry (complexation constants, dissociation constants) into account, the experimental conditions should be chosen in such a manner that the speciation of the test substance in the aqueous phase can be estimated. The ionic strength should be identical in all experiments by employing a background electrolyte.

13. Difficulties in the test may arise in conducting the test for substances with low water solubility or high P_{OW}, due to the fact that the concentrations in the water become very low such that their accurate determination is difficult. This Guideline provides guidance on how to deal with this problem.

INFORMATION ON THE TEST SUBSTANCE

14. Chemical reagents should be of analytical grade or of higher purity. The use of non-labelled test substances with known chemical composition and preferably at least 99% purity, or of radiolabelled test substances with known chemical composition and radiochemical purity, is recommended. In the case of short half-life tracers, decay corrections should be applied. In the case of radiolabelled test substances, a chemical specific analytical method should be employed to ensure that the measured radioactivity is directly related to the test substance.

15. An estimate of log P_{OW} may be obtained by using commercially available software for estimation of log P_{OW}, or by using the ratio of the solubilities in both solvents.

16. Before carrying out a slow-stirring experiment for determination of P_{OW}, the following information on the test substance should be available:

(a) structural formula;
(b) suitable analytical methods for determination of the concentration of the substance in water and 1-octanol;
(c) dissociation constant(s) of ionisable substances [OECD Guideline 112] (9);
(d) aqueous solubility [OECD Guideline 105] (10);
(e) abiotic hydrolysis [OECD Guideline 111] (11)
(f) ready biodegradability [OECD Guideline 301] (12);
(g) vapour pressure [OECD Guideline 104] (13).

DESCRIPTION OF THE METHOD

Equipment and apparatus

17. Standard laboratory equipment is required, in particular, the following:

- magnetic stirrers and Teflon coated magnetic stir bars are employed to stir the water phase;
- analytical instrumentation, suitable for determination of the concentration of the test substance at the expected concentrations;
- stirring-vessel with a tap at the bottom. Dependent on the estimate of log P_{OW} and the Limit of Detection (LOD) of the test compound, the use of a reaction vessel of the same geometry larger than one litre has to be considered, so that a sufficient volume of water can be obtained for chemical extraction and analysis. This will result in higher concentrations in the water extract and thus a more reliable analytical determination. A table giving estimates of the minimum volume needed, the LOD of the compound, its estimated log P_{OW} and its water solubility is given in Annex 1. The table is based on the relationship between log P_{OW} and the ratio between the solubilities in octanol and water, as presented by Pinsuwan et al. (14):

  $\log P_{OW} = 0.88 \log SR + 0.41$
  where
  $SR = S_{oct}/S_w$ (in molarity);

- and the relationship given by Lyman (15) for predicting water solubility. Water solubilities calculated with the equation given in Annex 1 must be seen as a first estimate. It should be noted that the user is free to generate an estimate of water solubility by means of any relationship that is considered to better represent the relationship between hydrophobicity and solubility. For solid compounds, inclusion of melting point in the prediction of solubility is for instance recommended. In case a modified equation is used, it should be ascertained that the equation for calculation of solubility in octanol is still valid. A schematic drawing of a glass-jacketed stirring-vessel with a volume of ca. one litre is given in Annex 2. The proportions of the vessel shown in Annex 2 have proven favorable and should be maintained when apparatus of a different size is used;

- a means for keeping the temperature constant during the slow-stirring experiment is essential.

18. Vessels should be made from inert material such that adsorption to vessel surfaces is negligible.

Preparation of the test solutions

19. The P_{OW} determination should be carried out with the highest purity 1-octanol that is commercially available (at least +99 %). Purification of 1-octanol by extraction with acid, base and water and subsequent drying is recommended. In addition, distillation can be used to purify 1-octanol. Purified 1-octanol is to be used to prepare standard solutions of the test substances. Water to be used in the P_{OW} determination should be glass or quartz distilled, or obtained from a purification system, or HPLC-grade water may be used. Filtration through a 0.22 µm filter is required for distilled water, and blanks should be included to check that no impurities are in the concentrated extracts that may interfere with the test substance. If a glass fiber filter is used, it should be cleaned by baking for at least three hours at 400 °C.

20. Both solvents are mutually saturated prior to the experiment by equilibrating them in a sufficiently large vessel. This is accomplished by slow-stirring the two-phase system for two days.

21. An appropriate concentration of test substance is selected and dissolved in 1-octanol (saturated with water). The 1-octanol/water partition coefficient needs to be determined in dilute solutions in 1-octanol and water. Therefore the concentration of the test substance should not exceed 70 % of its solubility with a maximum concentration of 0.1 M in either phase (1). The 1-octanol solutions used for the experiment must be devoid of suspended solid test substance.

22. The appropriate amount of test substance is dissolved in 1-octanol (saturated with water). If the estimate of log P_{OW} exceeds five, care has to be taken that the 1-octanol solutions used for the experiment are devoid of suspended solid test substance. To that end, the following procedure for chemicals with an estimated value of log $P_{OW} > 5$ is followed:

- the test substance is dissolved in 1-octanol (saturated with water);

- the solution is given sufficient time for the suspended solid substance to settle out. During the settling period, the concentration of the test substance is monitored;

- after the measured concentrations in the 1-octanol-solution have attained stable values, the stock solution is diluted with an appropriate volume of 1-octanol;

- the concentration of the diluted stock solution is measured. If the measured concentration is consistent with the dilution, the diluted stock solution can be employed in the slow-stirring experiment.

Extraction and analysis of samples

23. A validated analytical method should be used for the assay of test substance. The investigators have to provide evidence that the concentrations in the water saturated 1-octanol as well as in the 1-octanol saturated water phase during the experiment are above the method limit of quantification of the analytical procedures employed. Analytical recoveries of the test substance from the water phase and from the 1-octanol phase need to be established prior to the experiment in those cases for which extraction methods are necessary. The analytical signal needs to be corrected for blanks and care should be taken that no carry-over of analyte from one sample to another can occur.

24. Extraction of the water phase with an organic solvent and preconcentration of extract are likely to be required prior to analysis, due to rather low concentrations of hydrophobic test substances in the water phase. For the same reason it is necessary to reduce eventual blank concentrations. To that end, it is necessary to employ high purity solvents, preferably solvents for residue analysis. Moreover, working with carefully pre-cleaned (e.g. solvent washing or baking at elevated temperature) glassware can help to avoid cross-contamination.

25. An estimate of log P_{OW} may be obtained from an estimation program or by expert judgment. If the value is higher than six then blank corrections and analyte carry-over need to be monitored closely. Similarly, if the estimate of log P_{OW} exceeds six, the use of a surrogate standard for recovery correction is mandatory, so that high preconcentration factors can be reached. A number of software programs for the estimation of log P_{OW} are commercially available[1], e.g., Clog P (16), KOWWIN (17), ProLogP (18) and ACD log P (19). Descriptions of the estimation approaches can be found in references (20-22).

26. The limits of quantification (LOQ) for determination of the test substance in 1-octanol and water are established using accepted methods. As a rule of thumb, the method limit of quantification can be determined as the concentration in water or 1-octanol that produces a signal to noise ratio of ten. A suitable extraction and pre-concentration method should be selected and analytical recoveries should also be specified. A suitable pre-concentration factor is selected in order to obtain a signal of the required size upon analytical determination.

27. On the basis of the parameters of the analytical method and the expected concentrations, an approximate sample size required for an accurate determination of the compound concentration is determined. The use of water samples that are too small to obtain a sufficient analytical signal should be avoided. Also, the use of excessively large water samples should be avoided, since otherwise there might be too little water left for the minimum number of analyses required (n = 5). In Annex 1, the minimum sample volume is indicated as a function of the vessel volume, the LOD of the test substance and the solubility of the test substance.

28. Quantification of the test substances occurs by comparison with calibration curves of the respective compound. The concentrations in the samples analyzed must be bracketed by concentrations of standards.

[1] This information is only given for the convenience of users. Other equivalent computer programmes may be used if they can be shown to produce the same results.

29. For test compounds with a log P_{OW} estimate higher than six a surrogate standard has to be spiked to the water sample prior to extraction in order to register losses occurring during extraction and pre-concentration of the water samples. For accurate recovery correction, the surrogates must have properties that are very close to, or identical with, those of the test substance. Preferably, (stable) isotopically-labelled analogs of the substances of interest (for example, perdeuterated or ^{13}C-labelled) are used for this purpose. If the use of labelled stable isotopes i.e., ^{13}C or ^{2}H, is not possible it should be demonstrated from reliable data in the literature that the physical-chemical properties of the surrogate are very close to those of the test substance. During liquid-liquid extraction of the water phase emulsions can form. They can be reduced by addition of salt and allowing the emulsion to settle overnight. Methods used for extracting and pre-concentrating the samples need to be reported.

30. Samples withdrawn from the 1-octanol phase may, if necessary, be diluted with a suitable solvent prior to analysis. Moreover, the use of surrogate standards for recovery correction is recommended for substances for which the recovery experiments demonstrated a high degree of variation in the recovery experiments (relative standard deviation > 10%).

31. The details of the analytical method need to be reported. This includes the method of extraction, pre-concentration and dilution factors, instrument parameters, calibration routine, calibration range, analytical recovery of the test substance from water, addition of surrogate standards for recovery correction, blank values, detection limits and limits of quantification.

Performance of the Test

Optimal 1-octanol/water volume ratios

32. When choosing the water and 1-octanol volumes, the LOQ in 1-octanol and water, the pre-concentration factors applied to the water samples, the volumes sampled in 1-octanol and water, and the expected concentrations should be considered. For experimental reasons, the volume of 1-octanol in the slow-stirring system should be chosen such that the 1-octanol layer is sufficiently thick (>0.5 cm) in order to allow for sampling of the 1-octanol phase without disturbing it.

33. Typical phase ratios used for the determinations of compounds with log P_{OW} of 4.5 and higher are 20 to 50 ml of 1-octanol and 950 to 980 ml of water in a one litre vessel.

Test conditions

34. During the test the reaction vessel is thermostated to reduce temperature variation to below 1 °C. The assay should be performed at 25 °C.

35. The experimental system should be protected from daylight by either performing the experiment in a dark room or by covering the reaction vessel with aluminum foil.

36. The experiment should be performed in a dust-free (as far as possible) environment.

37. The 1-octanol-water system is stirred until equilibrium is attained. In a pilot experiment the length of the equilibration period is assessed by performing a slow-stirring experiment and sampling water and 1-octanol periodically. The sampling time points should be interspersed by a minimum period of five hours.

38. Each P_{OW} determination has to be performed employing at least three independent slow-stirring experiments.

Determination of the equilibration time

39. It is assumed that the equilibrium is achieved when a regression of the 1-octanol/water concentration ratio against time over a time span of four time points yields a slope that is not significantly different from zero at a p-level of 0.05. The minimum equilibration time is one day before sampling can be started. As a rule of thumb, sampling of substances with a log P_{OW} estimate of less than five can take place during days two and three. The equilibration might have to be extended for more hydrophobic compounds. For a compound with log P_{OW} of 8.23 (decachlorobiphenyl) 144 hours were sufficient for equilibration. Equilibrium is assessed by means of repeated sampling of a single vessel.

Starting the experiment

40. At the start of the experiment the reaction vessel is filled with 1-octanol-saturated water. Sufficient time should be allowed to reach the theremostated temperature.

41. The desired amount of test substance (dissolved in the required volume of 1-octanol saturated with water) is carefully added to the reaction vessel. This is a crucial step in the experiment, since turbulent mixing of the two phases has to be avoided. To that end, the 1-octanol phase can be pipetted slowly against the wall of the experimental vessel, close to the water surface. It will subsequently flow along the glass wall and form a film above the water phase. The decantation of 1-octanol directly into the flask should always be avoided; drops of 1-octanol should not be allowed to fall directly into the water.

42. After starting the stirring, the stirring rate should be increased slowly. If the stirring motors cannot be appropriately adjusted the use of a transformer should be considered. The stirring rate should be adjusted so that a vortex at the interface between water and 1-octanol of 0.5 to maximally 2.5 cm depth is created. The stirring rate should be reduced if the vortex depth of 2.5 cm is exceeded; otherwise microdroplets may be formed from 1-octanol droplets in the water phase, leading to an overestimation of the concentration of the test substance in the water. The maximum stirring rate of 2.5 cm is recommended on the basis of the findings in the ring-test validation study (5). It is a compromise between achieving a rapid rate of equilibration, while limiting the formation of 1-octanol microdroplets.

Sampling and Sample Treatment

43. The stirrer should be turned off prior to sampling and the liquids should be allowed to stop moving. After sampling is completed, the stirrer is started again slowly, as described above, and then the stirring rate is increased gradually.

44. The water phase is sampled from a stopcock at the bottom of the reaction vessel. Always discard the dead volume of water contained in the taps (approximately 5 ml in the vessel shown in the Annex 2). The water in the taps is not stirred and therefore not in equilibrium with the bulk. Note the volume of the water samples, and make sure that the amount of test substance present in the discarded water is taken into account when setting up a mass balance. Evaporative losses should be minimized by allowing the water to flow quiescently into the separatory funnel, such that there is no disturbance of the water/1-octanol layer.

45. 1-Octanol samples are obtained by withdrawing a small aliquot (ca. 100 µl) from the 1-octanol layer with a 100 microliter all glass-metal syringes. Care should be taken not to disturb the boundary. The volume of the sampled liquid is recorded. A small aliquot is sufficient, since the 1-octanol sample will be diluted.

46. Unnecessary sample transfer steps should be avoided. To that end the sample volume should be determined gravimetrically. In case of water samples this can be achieved by collecting the water sample in a separatory funnel that contains already the required volume of solvent.

DATA AND REPORTING

47. According to the present guideline, P_{OW} is determined by performing three slow-stirring experiments (three experimental units) with the compound under investigation employing identical conditions. The regression used to demonstrate attainment of equilibrium should be based on the results of at least four determinations of C_O / C_W at consecutive time points. This allows for calculating variance as a measure of the uncertainty of the average value obtained by each experimental unit.

48. The P_{OW} can be characterized by the variance in the data obtained for each experimental unit. This information is employed to calculate the P_{OW} as the weighted average of the results of the individual experimental units. To do so, the inverse of the variance of the results of the experimental units is employed as weight. As a result, data with a large variation (expressed as the variance) and thus with lower reliability have less influence on the result than data with a low variance.

49. Analogously, the weighted standard deviation is calculated. It characterizes the repeatability of the P_{OW} measurement. A low value of the weighted standard deviation indicates that the P_{OW} determination was very repeatable within one laboratory. The formal statistical treatment of the data is outlined below.

Treatment of the results

Demonstration of attainment of equilibrium

50. The logarithm of the ratio of the concentration of the test compound in 1-octanol and water (log (C_O/C_w)) is calculated for each sampling time. Achievement of chemical equilibrium is demonstrated by plotting this ratio against time. A plateau in this plot that is based on at least four consecutive time points indicates that equilibrium has been attained, and that the compound is truly dissolved in 1-octanol. If not, the test needs to be continued until four successive time points yield a slope that is not significantly different from 0 at a p-level of 0.05, indicating that log C_o/C_w is independent of time.

Log P_{OW}-calculation

51. The value of log P_{OW} of the experimental unit is calculated as the weighted average value of log C_o/C_w for the part of the curve of log C_o/C_w vs. time, for which equilibrium has been demonstrated. The weighted average is calculated by weighting the data with the inverse of the variance so that the influence of the data on the final result is inversely proportional to the uncertainty in the data.

Average log P_{OW}

52. The average value of log P_{OW} of different experimental units is calculated as the average of the results of the individual experimental units weighted with their respective variances.

The calculation is performed as follows:

$$\log P_{OW,Av} = (\Sigma w_i \times \log P_{OW,i}) \times (\Sigma w_i)^{-1}$$

where

$\log P_{OW,i}$ = the $\log P_{OW}$ value of the individual experimental unit i;
$\log P_{OW,Av}$ = the weighted average value of the individual $\log P_{OW}$ determinations;
w_i = the statistical weight assigned to the $\log P_{OW}$ value of the experimental unit i.

The reciprocal of the variance of $\log P_{OW,i}$ is employed as w_i ($w_i = var(\log P_{OW,i})^{-1}$).

53. The error of the average of $\log P_{OW}$ is estimated as the repeatability of $\log C_o/C_w$ determined during the equilibrium phase in the individual experimental units. It is expressed as the weighted standard deviation of $\log P_{OW,Av}$ ($\sigma_{\log Pow,Av}$) which in turn is a measure of the error associated with $\log P_{OW,Av}$. The weighted standard deviation can be computed from the weighted variance ($var_{\log Pow,Av}$) as follows:

$$var_{\log Pow,Av} = (\Sigma w_i \times (\log P_{OW,i} - \log P_{OW,Av})^2) \times (\Sigma w_i \times (n-1))^{-1}$$
$$\sigma_{\log Pow,Av} = (var_{\log Pow,Av})^{0.5}$$

The symbol n stands for the number of experimental units.

Test Report

54. The test report should include the following information:

Test substance:
- common name, chemical name, CAS number, structural formula (indicating position of label when radiolabelled substance is used) and relevant physical-chemical properties (see paragraph 17);
- purity (impurities) of test substance;
- label purity of labelled chemicals and molar activity (where appropriate);
- the preliminary estimate of $\log P_{ow}$, as well as the method used to derive the value.

Test conditions:
- dates of the performance of the studies;
- temperature during the experiment;
- volumes of 1-octanol and water at the beginning of the test;
- volumes of withdrawn 1-octanol and water samples;
- volumes of 1-octanol and water remaining in the test vessels;
- description of the test vessels and stirring conditions (geometry of the stirring bar and of the test vessel, vortex height in mm, and when available: stirring rate) used;
- analytical methods used to determine the test substance and the method limit of quantification;
- sampling times;
- the aqueous phase pH and the buffers used, when pH is adjusted for ionizable molecules;
- number of replicates.

Results:
- repeatability and sensitivity of the analytical methods used;
- determined concentrations of the test substance in 1-octanol and water as a function of time;
- demonstration of mass balance;
- temperature and standard deviation or the range of temperature during the experiment;
- the regression of concentration ratio against time;
- the average value $\log P_{ow,Av}$ and its standard error;
- discussion and interpretation of the results;

- examples of raw data figures of representative analysis (all raw data have to be stored in accordance with GLP standards), including recoveries of surrogates, and the number of levels used in the calibration (along with the criteria for the correlation coefficient of the calibration curve), and results of QA/QC;
- when available: validation report of the assay procedure (to be indicated among references).

LITERATURE

(1) De Bruijn JHM, Busser F, Seinen W, Hermens J. (1989). Determination of octanol/water partition coefficients with the 'slow-stirring' method. *Environ. Toxicol. Chem.* 8: 499-512.

(2) OECD (1995) OECD Guidelines for the Testing of Chemicals: 107 Partition Coefficients. Paris.

(3) OECD (2004) OECD Guidelines for the Testing of Chemicals: 117 Partition Coefficients by Reversed Phase Chromatography. Paris.

(4) OECD (2000) OECD Draft Guideline for the Testing of Chemicals: 122 Partition Coefficient (n-Octanol/Water): pH-Metric Method for Ionisable Substances. Paris.

(5) Tolls J. (2002). Partition Coefficient 1-Octanol/Water (Pow) Slow-Stirring Method for Highly Hydrophobic Chemicals, Validation Report. RIVM contract-Nrs 602730 M/602700/01.

(6) Boethling R.S., Mackay D. (eds.) (2000). Handbook of property estimation methods for chemicals. Lewis Publishers Boca Raton, FL, USA.

(7) Schwarzenbach RP, Gschwend PM, Imboden DM. (1993). Environmental Organic Chemistry. Wiley, New York, NY.

(8) Arnold CG, Widenhaupt A, David MM, Müller SR, Haderlein SB, Schwarzenbach RP. (1997). Aqueous speciation and 1-octanol-water partitioning of tributyl- and triphenyltin: effect of pH and ion composition. *Environ. Sci. Technol.* 31: 2596-2602.

(9) OECD (1981) OECD Guidelines for the Testing of Chemicals: 112 Dissociation Constants in Water. Paris.

(10) OECD (1995) OECD Guidelines for the Testing of Chemicals: 105 Water Solubility. Paris.

(11) OECD (2004) OECD Guidelines for the Testing of Chemicals: 111 Hydrolysis as a function of pH. Paris.

(12) OECD (1992) OECD Guidelines for the Testing of Chemicals: 301 Ready Biodegradability. Paris.

(13) OECD (1995) OECD Guidelines for the Testing of Chemicals: 104 Vapour Pressure. Paris.

(14) Pinsuwan, S., Li, A. and Yalkowsky, S.H. (1995). Correlation of octanol/water solubility ratios and partition coefficients, *J. Chem. Eng. Data.*, 40: 623-626.

(15) Lyman WJ. (1990). Solubility in water. In: Handbook of Chemical Property Estimation Methods: Environmental Behavior of Organic Compounds, Lyman WJ, Reehl WF, Rosenblatt DH, Eds. American Chemical Society, Washington, DC, 2-1 to 2-52.

(16) Leo A, Weininger D. (1989). Medchem Software Manual. Daylight Chemical Information Systems, Irvine, CA.

(17) Meylan W. (1993). SRC-LOGKOW for Windows. SRC, Syracuse, N.Y.

(18) Compudrug L. (1992). ProLogP. Compudrug, Ltd, Budapest.

(19) ACD. ACD logP; Advanced Chemistry Development: Toronto, Ontario M5H 3V9, Canada, 2001.

(20) Lyman WJ. (1990). Octanol/water partition coefficient. In Lyman WJ, Reehl WF, Rosenblatt DH, eds, *Handbook of chemical property estimation*, American Chemical Society, Washington, D.C.

(21) Rekker RF, de Kort HM. (1979). The hydrophobic fragmental constant: An extension to a 1000 data point set. *Eur. J. Med. Chem. Chim. Ther.* 14: 479-488.

(22) Jübermann O. (1958). Houben-Weyl, ed, *Methoden der Organischen Chemie*, 386-390.

ANNEX 1

SPREADSHEET FOR COMPUTATION OF MINIMUM VOLUMES OF WATER REQUIRED FOR DETECTION OF TEST COMPOUNDS OF DIFFERENT LOG P_{OW} VALUES IN AQUEOUS PHASE

Assumptions:

- Maximum volume of individual aliquots = 10 % of total volume; 5 aliquots = 50% of total volume.
- Concentration of test substances = 0.7 × solubility in either phase. In case of lower concentrations, larger volumes would be required.
- Volume used for LOD determination = 100 ml.
- log P_{ow} vs. log S_w and log P_{ow} vs. SR (S_{oct}/S_w) are reasonable representations of relationships for test compounds.

Estimation of S_w

log P_{ow}	equation	log S_w	S_w (mg/L)
4	(-)0.922*log P_{ow} + 4.184	0.496	3.133E+00
4.5	(-)0.922*log P_{ow} + 4.184	0.035	1.084E+00
5	(-)0.922*log P_{ow} + 4.184	-0.426	3.750E-01
5.5	(-)0.922*log P_{ow} + 4.184	-0.887	1.297E-01
6	(-)0.922*log P_{ow} + 4.184	-1.348	4.487E-02
6.5	(-)0.922*log P_{ow} + 4.184	-1.809	1.552E-02
7	(-)0.922*log P_{ow} + 4.184	-2.270	5.370E-03
7.5	(-)0.922*log P_{ow} + 4.184	-2.731	1.858E-03
8	(-)0.922*log P_{ow} + 4.184	-3.192	6.427E-04

Estimation of S_{oct}

log P_{ow}	equation	S_{oct} (mg/L)
4	log P_{ow} = 0.88log SR + 0.41	3.763E+04
4.5	log P_{ow} = 0.88log SR + 0.42	4.816E+04
5	log P_{ow} = 0.88log SR + 0.43	6.165E+04
5.5	log P_{ow} = 0.88log SR + 0.44	7.890E+04
6	log P_{ow} = 0.88log SR + 0.45	1.010E+05
6.5	log P_{ow} = 0.88log SR + 0.46	1.293E+05
7	log P_{ow} = 0.88log SR + 0.47	1.654E+05
7.5	log P_{ow} = 0.88log SR + 0.48	2.117E+05
8	log P_{ow} = 0.88log SR + 0.49	2.710E+05

Total Mass test compound (mg)	$Mass_{oct}/Mass_{water}$	$Mass_{H2O}$ (mg)	$Conc_{H2O}$ (mg/L)	$Mass_{oct}$ (mg)	$Conc_{oct}$ (mg/L)
1319	526	2.5017	2.6333	1 317	26 333
1686	1664	1.0127	1.0660	1 685	33 709
2158	5263	0.4099	0.4315	2 157	43 149
2762	16644	0.1659	0.1747	2 762	55 230
3535	52632	0.0672	0.0707	3 535	70 691
4524	166436	0.0272	0.0286	4 524	90 480
5790	526316	0.0110	0.0116	5 790	115 807
7411	1664357	0.0045	0.0047	7 411	148 223
9486	5263158	0.0018	0.0019	9 486	189 713

Computation of volumes

Minimum volume required for H_2O phase at each LOD concentration

log K_{ow}	LOD (micrograms/L)→	0.001	0.01	0.10	1.00	10
4		0.04	0.38	3.80	38	380
4.5		0.09	0.94	9.38	94	938
5		0.23	2.32	23.18	232	2 318
5.5		0.57	5.73	57.26	573	5 726
6		1.41	14.15	141	1 415	14 146
6.5		3.50	34.95	350	3 495	34 950
7		8.64	86.35	864	8 635	86 351
7.5		21.33	213	2 133	21 335	213 346
8		52.71	527	5 271	52 711	527 111
Volume used for LOD (L)	0.1					

Key to Computations

Represents <10% of total volume of aqueous phase, 1 liter equilibration vessel.

Represents <10% of total volume of aqueous phase, 2 liter equilibration vessel.

Represents <10% of total volume of aqueous phase, 5 liter equilibration vessel.

Represents <10% of total volume of aqueous phase, 10 liter equilibration vessel.

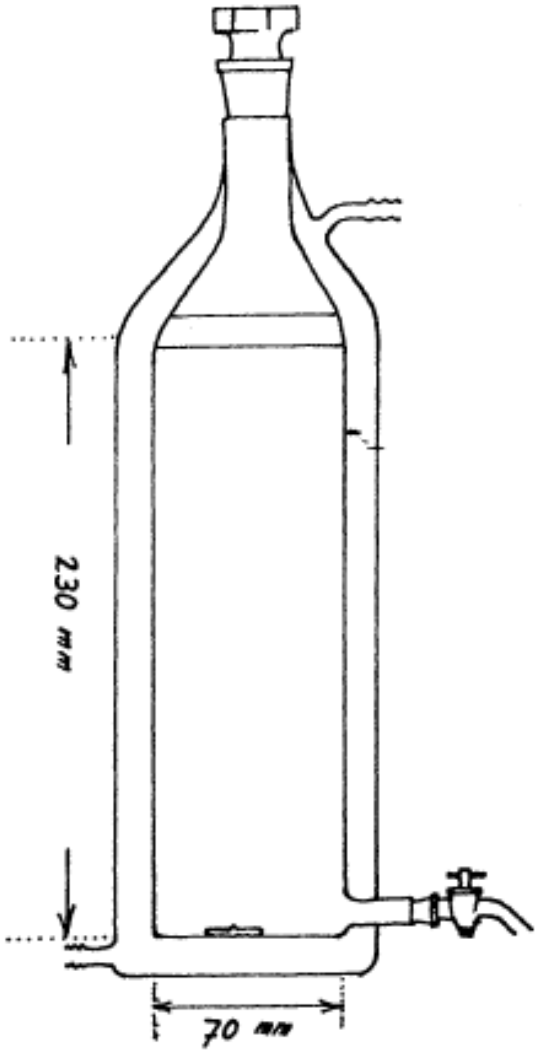
Exceeds 10% of even the 10 liter equilibration vessel.

Overview of volumes required, as a function of water solubility and Log P_{ow}

		Minimum volume required for H_2O phase at each LOD concentration (ml)					
log P_{ow}	S_w (mg/L)	LOD (micrograms/L)→	0.001	0.01	0.10	1.00	10
4	10		0.01	0.12	1.19	11.90	118.99
	5		0.02	0.24	2.38	23.80	237.97
	3		0.04	0.40	3.97	39.66	396.62
	1		0.12	1.19	11.90	118.99	1189.86
4.5	5		0.02	0.20	2.03	20.34	203.37
	2		0.05	0.51	5.08	50.84	508.42
	1		0.10	1.02	10.17	101.68	1016.83
	0.5		0.20	2.03	20.34	203.37	2033.67
5	1		0.09	0.87	8.69	86.90	869.01
	0.5		0.17	1.74	17.38	173.80	1738.02
	0.375		0.23	2.32	23.18	231.75	2317.53
	0.2		0.43	4.35	43.45	434.51	4345.05
5.5	0.4		0.19	1.86	18.57	185.68	1856.79
	0.2		0.37	3.71	37.14	371.36	3713.59
	0.1		0.74	7.43	74.27	742.72	7427.17
	0.05		1.49	14.85	148.54	1485.43	14854.35
6	0.1		0.63	6.35	63.48	634.80	6347.95
	0.05		1.27	12.70	126.96	1269.59	12695.91
	0.025		2.54	25.39	253.92	2539.18	25391.82
	0.0125		5.08	50.78	507.84	5078.36	50783.64
6.5	0.025		2.17	21.70	217.02	2170.25	21702.46
	0.0125		4.34	43.40	434.05	4340.49	43404.93
	0.006		9.04	90.43	904.27	9042.69	90426.93
	0.003		18.09	180.85	1808.54	18085.39	180853.86
7	0.006		7.73	77.29	772.89	7728.85	77288.50
	0.003		15.46	154.58	1545.77	15457.70	154577.01
	0.0015		23.19	231.87	2318.66	23186.55	231865.51
	0.001		46.37	463.73	4637.31	46373.10	463731.03
7.5	0.002		19.82	198.18	1981.77	19817.73	198177.33
	0.001		39.64	396.35	3963.55	39635.47	396354.66
	0.0005		79.27	792.71	7927.09	79270.93	792709.32
	0.00025		158.54	1585.42	15854.19	158541.86	1585418.63
8	0.001		33.88	338.77	3387.68	33876.77	338767.72
	0.0005		67.75	677.54	6775.35	67753.54	677535.44
	0.00025		135.51	1355.07	13550.71	135507.09	1355070.89
	0.000125		271.01	2710.14	27101.42	271014.18	2710141.77
Volume used for LOD (L)		0.1					

ANNEX 2

AN EXAMPLE OF GLASS-JACKETED TEST VESSEL FOR THE SLOW-STIRRING EXPERIMENT FOR DETERMINATION OF P_{OW}.



15/15

Adopted :
13 April 2004

OECD GUIDELINE FOR TESTING OF CHEMICALS

Daphnia sp., Acute Immobilisation Test

INTRODUCTION

1. OECD Guidelines for the Testing of chemicals are periodically reviewed in the light of scientific progress. Guideline 202 on "*Daphnia* sp., Acute Immobilisation Test and Reproduction Test", adopted in April 1984, included two parts: Part I - the 24h EC_{50} acute immobilisation test and Part II - the reproduction test (at least 14 days). Revision of the reproduction test has resulted in the adoption and publication of Test Guideline 211 on "*Daphnia magna* Reproduction Test" in September 1998. Consequently, the new version of Guideline 202 is restricted to the acute immobilisation test.

2. This guideline describes an acute toxicity test to assess effects of chemicals towards daphnids. Existing test methods were used to the extent possible (1)(2)(3). The main differences in comparison with the earlier version are the extension of the test duration to 48 hours, the provision for more information on recommended culture and test media, and the introduction of a limit test at 100 mg/l of test substance.

PRINCIPLE OF THE TEST

3. Young daphnids, aged less than 24 hours at the start of the test, are exposed to the test substance at a range of concentrations for a period of 48 hours. Immobilisation is recorded at 24 hours and 48 hours and compared with control values. The results are analysed in order to calculate the EC_{50} at 48h (see Annex 1 for definitions). Determination of the EC_{50} at 24h is optional.

INFORMATION ON THE TEST SUBSTANCE

4. The water solubility and the vapour pressure of the test substance should be known and a reliable analytical method for the quantification of the substance in the test solutions with reported recovery efficiency, and limit of determination should be available. Useful information includes the structural formula, purity of the substance, stability in water and light, P_{ow} and results of a test for ready biodegradability (see Guideline 301).

Note: Guidance for testing substances with physical chemical properties that make them difficult to test is provided in a separate document (4).

REFERENCE SUBSTANCES

5. A reference substance may be tested for EC_{50} as a means of assuring that the test conditions are reliable. Toxicants used in international ring-tests (1)(5) are recommended for this purpose[1]. Test(s) with a reference substance should be done preferably every month and at least twice a year.

VALIDITY OF THE TEST

6. For a test to be valid, the following performance criteria apply:

- In the control, including the control containing the solubilising agent, not more that 10 per cent of the daphnids should have been immobilised;
- The dissolved oxygen concentration at the end of the test should be ≥ 3 mg/l in control and test vessels.

Note: For the first criterion, not more than 10 percent of the control daphnids should show immobilisation or other signs of disease or stress, for example, discoloration or unusual behaviour such as trapping at surface of water.

DESCRIPTION OF THE METHOD

Apparatus

7. Test vessels and other apparatus that will come into contact with the test solutions should be made entirely of glass or other chemically inert material. Test vessels will normally be glass test tubes or beakers; they should be cleaned before each use using standard laboratory procedures. Test vessels should be loosely covered to reduce the loss of water due to evaporation and to avoid the entry of dust into the solutions. Volatile substances should be tested in completely filled closed vessels large enough to prevent oxygen becoming limiting or too low (see paragraphs 6 and 22).

8. In addition some or all of the following equipment will be used: oxygen-meter (with microelectrode or other suitable equipment for measuring dissolved oxygen in low volumes samples); pH-meter; adequate apparatus for temperature control; equipment for the determination of total organic carbon concentration (TOC); equipment for the determination of chemical oxygen demand (COD); equipment for the determination of hardness, etc.

Test organism

9. *Daphnia magna* Straus is the preferred test species although other suitable *Daphnia* species can be used in this test (e.g. *Daphnia pulex*). At the start of the test, the animals should be less than 24 hours old and, to reduce variability, it is strongly recommended they are not first brood progeny. They should be derived from a healthy stock (i.e. showing no signs of stress such as high mortality, presence of males and ephippia, delay in the production of the first brood, discoloured animals, etc.). All organisms used for a

[1] The results of these inter laboratory tests and a Technical Corrigendum to ISO 6341 give an EC_{50}-24 h of the potassium dichromate ($K_2Cr_2O_7$) within the range 0.6 mg/l to 2.1 mg/l

particular test should have originated from cultures established from the same stock of daphnids. The stock animals must be maintained in culture conditions (light, temperature, medium) similar to those to be used in the test. If the daphnids culture medium to be used in the test is different from that used for routine daphnids culture, it is good practice to include a pre-test acclimation period. For that, brood daphnids should be maintained in dilution water at the test temperature for at least 48 hours prior to the start of the test.

Holding and dilution water

10. Natural water (surface or ground water), reconstituted water or dechlorinated tap water are acceptable as holding and dilution water if daphnids will survive in it for the duration of the culturing, acclimation and testing without showing signs of stress. Any water which conforms to the chemical characteristics of an acceptable dilution water as listed in Annex 2 is suitable as a test water. It should be of constant quality during the period of the test. Reconstituted water can be made up by adding specific amounts of reagents of recognised analytical grade to deionised or distilled water. Examples of reconstituted water are given in (1)(6) and in Annex 3. Note that media containing known chelating agents, such as M4 and M7 media in Annex 3, should be avoided for testing substances containing metals. The pH should be in the range of 6 to 9. Hardness between 140 and 250 mg/l (as $CaCO_3$) is recommended for *Daphnia magna*, while lower hardness may be also appropriate for other *Daphnia* species. The dilution water may be aerated prior to use for the test so that the dissolved oxygen concentration has reached saturation.

11. If natural water is used, the quality parameters should be measured at least twice a year or whenever it is suspected that these characteristics may have changed significantly (see paragraph 10 and Annex 2). Measurements of heavy metals (e.g. Cu, Pb, Zn, Hg, Cd, Ni) should also be made. If dechlorinated tap water is used, daily chlorine analysis is desirable. If the dilution water is from a surface or ground water source, conductivity and total organic carbon (TOC) or chemical oxygen demand (COD) should be measured.

Test solutions

12. Test solutions of the chosen concentrations are usually prepared by dilution of a stock solution. Stock solutions should preferably be prepared by dissolving the test substance in the dilution water. As far as possible, the use of solvents, emulsifiers or dispersants should be avoided. However, such compounds may be required in some cases in order to produce a suitably concentrated stock solution. Guidance for suitable solvents, emulsifiers and dispersants is given in (4). In any case, the test substance in the test solutions should not exceed the limit of solubility in the dilution water.

13. The test should be carried out without the adjustment of pH. If the pH does not remain in the range 6-9, then a second test could be carried out, adjusting the pH of the stock solution to that of the dilution water before addition of the test substance. The pH adjustment should be made in such a way that the stock solution concentration is not changed to any significant extent and that no chemical reaction or precipitation of the test substance is caused. HCl and NaOH are preferred.

PROCEDURE

Conditions of exposure

Test groups and controls

14. Test vessels are filled with appropriate volumes of dilution water and solutions of test substance. Ratio of air/water volume in the vessel should be identical for test and control groups. Daphnids are then placed into test vessels. At least 20 animals, preferably divided into four groups of five animals each, should be used at each test concentration and for the controls. At least 2 ml of test solution should be provided for each animal (i.e. a volume of 10 ml for five daphnids per test vessel). The test may be carried out using semi-static renewal or flow-through system when the concentration of the test substance is not stable.

15. One dilution-water control series and also, if relevant, one control series containing the solubilising agent (solvent control) at the level used in treatments must be run in addition to the treatment series.

Test concentrations

16. A range-finding test may be conducted to determine the range of concentrations for the definitive test unless information on toxicity of the test substance is available. For this purpose, the daphnids are exposed to a series of widely spaced concentrations of the test substance. Five daphnids should be exposed to each test concentration for 48 hours or less, and no replicates are necessary. The exposure period may be shortened (e.g. 24 hours or less) if data suitable for the purpose of the range-finding test can be obtained in less time.

17. At least five test concentrations should be used. They should be arranged in a geometric series with a separation factor preferably not exceeding 2.2. Justification should be provided if fewer than five concentrations are used. The highest concentration tested should preferably result in 100 per cent immobilisation, and the lowest concentration tested should preferably give no observable effect.

Incubation conditions

18. The temperature should be within the range of 18°C and 22°C, and for each single test it should be constant within ± 1°C. A 16-hour light and 8-hour dark cycle is recommended. Complete darkness is also acceptable, especially for test substances unstable in light.

19. The test vessels must not be aerated during the test. The test is carried out without adjustment of pH. The daphnids should not be fed during the test.

Duration

20. The test duration is 48 hours.

Observations

21. Each test vessel should be checked for immobilised daphnids at 24 and 48 hours after the beginning of the test. (see Annex 1 for definitions). In addition to immobility, any abnormal behaviour or appearance should be reported.

Analytical measurements

22. The dissolved oxygen and pH are measured at the beginning and end of the test in the control(s) and in the highest test substance concentration. The dissolved oxygen concentration in controls should be in compliance with the validity criterion (see paragraph 6). The pH should normally not vary by more than 1.5 units in any one test. The temperature is usually measured in control vessels or in ambient air and it should be recorded preferably continuously during the test or, as a minimum, at the beginning and end of the test.

23. The concentration of the test substance should be measured, as a minimum, at the highest and lowest test concentration, at the beginning and end of the test (4). It is recommended that results be based on measured concentrations. However, if evidence is available to demonstrate that the concentration of the test substance has been satisfactorily maintained within ± 20 per cent of the nominal or measured initial concentration throughout the test, then the results can be based on nominal or measured initial values.

LIMIT TEST

24. Using the procedures described in this Guideline, a limit test may be performed at 100 mg/l of test substance or up to its limit of solubility in the test medium (whichever is the lower) in order to demonstrate that the EC50 is greater than this concentration. The limit test should be performed using 20 daphnids (preferably divided into four groups of five), with the same number in the control(s). If the percentage of immobilisation exceeds 10% at the end of the test, a full study should be conducted. Any observed abnormal behaviour should be recorded.

DATA AND REPORTING

Data

25. Data should be summarised in tabular form, showing for each treatment group and control, the number of daphnids used, and immobilisation at each observation. The percentages immobilised at 24 hours and 48 hours are plotted against test concentrations. Data are analysed by appropriate statistical methods (e.g. probit analysis, etc.) to calculate the slopes of the curves and the EC_{50} with 95% confidence limits (p = 0.95) (7)(8).

26. Where the standard methods of calculating the EC_{50} are not applicable to the data obtained, the highest concentration causing no immobility and the lowest concentration producing 100 per cent immobility should be used as an approximation for the EC_{50} (this being considered the geometric mean of these two concentrations).

Test report

27. The test report must include the following:

Test substance:

- physical nature and relevant physical-chemical properties;
- chemical identification data, including purity.

Test species:

- source and species of *Daphnia*, supplier of source (if known) and the culture conditions used (including source, kind and amount of food, feeding frequency).

Test conditions:

- description of test vessels: type and volume of vessels, volume of solution, number of daphnids per test vessel, number of test vessels (replicates) per concentration;
- methods of preparation of stock and test solutions including the use of any solvent or dispersants, concentrations used;
- details of dilution water: source and water quality characteristics (pH, hardness, Ca/Mg ratio, Na/K ratio, alkalinity, conductivity, etc.); composition of reconstituted water if used;
- incubation conditions: temperature, light intensity and periodicity, dissolved oxygen, pH, etc.

Results:

- the number and percentage of daphnids that were immobilised or showed any adverse effects (including abnormal behaviour) in the controls and in each treatment group, at each observation time and a description of the nature of the effects observed;
- results and date of test performed with reference substance, if available;
- the nominal test concentrations and the result of all analyses to determine the concentration of the test substance in the test vessels; the recovery efficiency of the method and the limit of determination should also be reported;
- all physical-chemical measurements of temperature, pH and dissolved oxygen made during the test;
- the EC_{50} at 48h for immobilisation with confidence intervals and graphs of the fitted model used for their calculation, the slopes of the dose-response curves and their standard error; statistical procedures used for determination of EC_{50}; (these data items for immobilisation at 24h should also be reported when they were measured.)
- explanation for any deviation from the Test Guideline and whether the deviation affected the test results.

LITERATURE

(1) ISO 6341. (1996). Water quality - Determination of the inhibition of the mobility of *Daphnia magna* Straus (Cladocera, Crustacea) - Acute toxicity test. Third edition, 1996.

(2) EPA OPPTS 850.1010. (1996). Ecological Effects Test Guidelines - Aquatic Invertebrate Acute Toxicity Test, Freshwater Daphnids.

(3) Environment Canada. (1996) Biological test method. Acute Lethality Test Using *Daphnia* spp. EPS 1/RM/11. Environment Canada, Ottawa, Ontario, Canada.

(4) Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures. OECD Environmental Health and Safety Publication. Series on Testing and Assessment. No. 23. Paris 2000.

(5) Commission of the European Communities. Study D8369. (1979). Inter-laboratory Test Programme concerning the study of the ecotoxicity of a chemical substance with respect to *Daphnia*.

(6) OECD Guidelines for the Testing of Chemicals. Guideline 211: *Daphnia magna* Reproduction Test, adopted September 1998.

(7) Stephan C.E. (1977). Methods for calculating an LC50. In Aquatic Toxicology and Hazard Evaluation (edited by F.I. Mayer and J.L. Hamelink). ASTM STP 634 - American Society for Testing and Materials. Pp65-84

(8) Finney D.J. (1978). Statistical Methods in Biological Assay. 3rd ed. London. Griffin, Weycombe, UK.

ANNEX 1

DEFINITIONS

In the context of this guideline, the following definitions are used:

EC_{50} is the concentration estimated to immobilise 50 per cent of the daphnids within a stated exposure period. If another definition is used, this must be reported, together with its reference.

Immobilisation: Those animals that are not able to swim within 15 seconds, after gentle agitation of the test vessel are considered to be immobilised (even if they can still move their antennae).

ANNEX 2

SOME CHEMICAL CHARACTERISTICS OF AN ACCEPTABLE DILUTION WATER

Substance	Concentration
Particulate matter	<20 mg/l
Total organic carbon	< 2 mg/l
Unionised ammonia	< 1 µg/l
Residual chlorine	<10 µg/l
Total organophosphorus pesticides	<50 ng/l
Total organochorine pesticides plus polychlorinated biphenyls	<50 ng/l
Total organic chlorine	<25 ng/l

ANNEX 3

EXAMPLES OF SUITABLE RECONSTITUTED TEST WATER

ISO Test water (1)

Stock solutions (single substance)		To prepare the reconstituted water, add the following volumes of stock solutions to 1 litre water*
Substance	Amount added to 1 litre water*	
Calcium chloride $CaCl_2, 2H_2O$	11.76 g	25 ml
Magnesium sulfate $MgSO_4, 7H_2O$	4.93 g	25 ml
Sodium bicarbonate $NaHCO_3$	2.59 g	25 ml
Potassium chloride KCl	0.23 g	25 ml

* Water of suitable purity, for example deionised, distilled or reverse osmosis with conductivity preferably not exceeding 10 $\mu S.cm^{-1}$.

ANNEX 3(Cont.)

Elendt M7 and M4 medium

Acclimation to Elendt M4 and M7 medium

Some laboratories have experienced difficulty in directly transferring Daphnia to M4 and M7 media. However, some success has been achieved with gradual acclimation, i.e. moving from own medium to 30% Elendt, then to 60% Elendt and then to 100% Elendt. The acclimation periods may need to be as long as one month.

Preparation

Trace element

Separate stock solutions (I) of individual trace elements are first prepared in water of suitable purity, for example deionised, distilled or reverse osmosis. From these different stock solutions (I) a second single stock solution (II) is prepared, which contains all trace elements (combined solution), i.e.:

Stock solution(s) I (single substance)	Amount added to water (mg/l)	Concentration (related to medium M4)	To prepare the combined stock solution II, add the following amount of stock solution I to water (ml/l)	
			M4	M7
H_3BO_3	57 190	20 000-fold	1.0	0.25
$MnCl_2 \bullet 4H_2O$	7 210	20 000-fold	1.0	0.25
LiCl	6 120	20 000-fold	1.0	0.25
RbCl	1 420	20 000-fold	1.0	0.25
$SrCl_2 \bullet 6H_2O$	3 040	20 000-fold	1.0	0.25
NaBr	320	20 000-fold	1.0	0.25
$Na_2MoO_4 \bullet 2H_2O$	1 230	20 000-fold	1.0	0.25
$CuCl_2 \bullet 2H_2O$	335	20 000-fold	1.0	0.25
$ZnCl_2$	260	20 000-fold	1.0	1.0
$CoCl_2 \bullet 6H_2O$	200	20 000-fold	1.0	1.0
KI	65	20 000-fold	1.0	1.0
Na_2SeO_3	43.8	20 000-fold	1.0	1.0
NH_4VO_3	11.5	20 000-fold	1.0	1.0
$Na_2EDTA \bullet 2H_2O$	5 000	2 000-fold	-	-
$FeSO_4 \bullet 7H_2O$	1991	2 000-fold	-	-
Both Na_2EDTA and $FeSO_4$ solutions are prepared singly, poured together and autoclaved immediately. This gives:				
21 Fe-EDTA solution		1 000-fold	20.0	5.0

M4 and M7 media

M4 and M7 media are prepared using stock solution II, the macro-nutrients and vitamin as follows:

	Amount added to water (mg/l)	Concentration (related to medium M4)	Amount of stock solution II added to prepare medium (ml/l)	
			M4	M7
Stock solution II (combined trace elements)		20-fold	50	50
Macro nutrient stock solutions (single substance)				
$CaCl_2 \bullet 2H_2O$	293 800	1 000-fold	1.0	1.0
$MgSO_4 \bullet 7H_2O$	246 600	2 000-fold	0.5	0.5
KCl	58 000	10 000-fold	0.1	0.1
$NaHCO_3$	64 800	1 000-fold	1.0	1.0
$Na_2SiO_3 \bullet 9H_2O$	50 000	5 000-fold	0.2	0.2
$NaNO_3$	2 740	10 000-fold	0.1	0.1
KH_2PO_4	1 430	10 000-fold	0.1	0.1
K_2HPO_4	1 840	10 000-fold	0.1	0.1
Combined Vitamin stock	-	10 000-fold	0.1	0.1
The combined vitamin stock solution is prepared by adding the 3 vitamin to 1 litre water, as shown below:				
Thiamine hydrochloride	750	10 000-fold		
Cyanocobalamine (B_{12})	10	10 000-fold		
Biotine	7.5	10 000-fold		

The combined vitamin stock is stored frozen in small aliquots. Vitamins are added to the media shortly before use.

N.B: To avoid precipitation of salts when preparing the complete media, add the aliquots of stock solutions to about 500 – 800 ml deionised water and then fill up to 1 litre.

N.N.B: The first publication of the M4 medium can be found in Elendt, B. P. (1990). Selenium deficiency in crustacea; an ultrastructual approach to antennal damage in *Daphnia magna* Straus. Protoplasma, 154, 25-33.

Adopted: 23 March 2006
Annex 5 corrected: 28 July 2011

OECD GUIDELINES FOR THE TESTING OF CHEMICALS

Freshwater Alga and Cyanobacteria, Growth Inhibition Test

INTRODUCTION

1. OECD Guidelines for Testing of Chemicals are periodically reviewed and updated in the light of scientific progress. With respect to Guideline 201, Alga, Growth Inhibition Test (adopted June 1984), the need to extend the Guideline to include additional species and update it to meet the requirements for hazard assessment and classification of chemicals has been identified. The revision has been completed on the basis of extensive practical experience, scientific progress in the field of algal toxicity studies, and extensive regulatory use, which has occurred since the original adoption.

2. Definitions used are given in Annex 1.

PRINCIPLE OF THE TEST

3. The purpose of this test is to determine the effects of a substance on the growth of freshwater microalgae and/or cyanobacteria. Exponentially growing test organisms are exposed to the test substance in batch cultures over a period of normally 72 hours. In spite of the relatively brief test duration, effects over several generations can be assessed.

4. The system response is the reduction of growth in a series of algal cultures (test units) exposed to various concentrations of a test substance. The response is evaluated as a function of the exposure concentration in comparison with the average growth of replicate, unexposed control cultures. For full expression of the system response to toxic effects (optimal sensitivity), the cultures are allowed unrestricted exponential growth under nutrient sufficient conditions and continuous light for a sufficient period of time to measure reduction of the specific growth rate.

5. Growth and growth inhibition are quantified from measurements of the algal biomass as a function of time. Algal biomass is defined as the dry weight per volume, e.g. mg algae/litre test solution. However, dry weight is difficult to measure and therefore surrogate parameters are used. Of these surrogates, cell counts are most often used. Other surrogate parameters include cell volume, fluorescence, optical density, etc. A conversion factor between the measured surrogate parameter and biomass should be known.

6. The test endpoint is inhibition of growth, expressed as the logarithmic increase in biomass (average specific growth rate) during the exposure period. From the average specific growth rates recorded in a series of test solutions, the concentration bringing about a specified x % inhibition of growth rate (e.g. 50%) is determined and expressed as the E_rC_x (e.g. E_rC_{50}).

7. An additional response variable used in this Guideline is yield, which may be needed to fulfil specific regulatory requirements in some countries. It is defined as the biomass at the end of the exposure

© OECD, (2011)
You are free to use this material for personal, non-commercial purposes without seeking prior consent from the OECD, provided the source is duly mentioned. Any commercial use of this material is subject to written permission from the OECD.

period minus the biomass at the start of the exposure period. From the yield recorded in a series of test solutions, the concentration bringing about a specified x % inhibition of yield (e.g., 50 %) is calculated and expressed as the E_yC_x (e.g. E_yC_{50}).

8. In addition, the lowest observed effect concentration (LOEC) and the no observed effect concentration (NOEC) may be statistically determined.

INFORMATION ON THE TEST SUBSTANCE

9. Information on the test substance which may be useful in establishing the test conditions includes structural formula, purity, stability in light, stability under the conditions of the test, light absorption properties, pKa, and results of studies of transformation including biodegradability in water.

10. The water solubility, octanol water partition coefficient (P_{ow}) and vapour pressure of the test substance should be known and a validated method for the quantification of the substance in the test solutions with reported recovery efficiency and limit of detection should be available.

VALIDITY OF THE TEST

11. For the test to be valid, the following performance criteria should be met:

- The biomass in the control cultures should have increased exponentially by a factor of at least 16 within the 72-hour test period. This corresponds to a specific growth rate of 0.92 day^{-1}. For the most frequently used species the growth rate is usually substantially higher (see Annex 2). This criterion may not be met when species that grow slower than those listed in Annex 2 are used. In this case, the test period should be extended to obtain at least a 16-fold growth in control cultures, while the growth has to be exponential throughout the test period. The test period may be shortened to at least 48 hours to maintain unlimited, exponential growth during the test as long as the minimum multiplication factor of 16 is reached.

- The mean coefficient of variation for section-by-section specific growth rates (days 0-1, 1-2 and 2-3, for 72-hour tests) in the control cultures (See Annex 1 under "coefficient of variation") must not exceed 35%. See paragraph 49 for the calculation of section-by-section specific growth rate. This criterion applies to the mean value of coefficients of variation calculated for replicate control cultures.

- The coefficient of variation of average specific growth rates during the whole test period in replicate control cultures must not exceed 7% in tests with *Pseudokirchneriella subcapitata* and *Desmodesmus subspicatus*. For other less frequently tested species, the value should not exceed 10%.

REFERENCE SUBSTANCE

12. Reference substance(s), such as 3,5-dichlorophenol used in the international ring test (1), may be tested as a means of checking the test procedure. Potassium dichromate can also be used as a reference substance for green algae. It is desirable to test a reference substance at least twice a year.

APPLICABILITY OF THE TEST

13. This Guideline is most easily applied to water-soluble substances which, under the conditions of the test, are likely to remain in the water. For testing of substances that are volatile, strongly adsorbing,

© OECD, (2011)

coloured, having a low solubility in water or substances that may affect the availability of nutrients or minerals in the test medium, certain modifications of the described procedure may be required (e.g., closed system, conditioning of the test vessels). Guidance on some appropriate modifications is given in (2) (3) and (4).

DESCRIPTION OF THE METHOD

Apparatus

14. Test vessels and other apparatus which will come into contact with the test solutions should be made entirely of glass or other chemically inert material. The items should be thoroughly washed to ensure that no organic or inorganic contaminants may interfere with the algal growth or composition of the test solutions.

15. The test vessels will normally be glass flasks of dimensions that allow a sufficient volume of culture for measurements during the test and a sufficient mass transfer of CO_2 from the atmosphere (see paragraph 30). Note that the liquid volume must be sufficient for analytical determinations (see paragraph 37).

16. In addition some or all of the following equipment may be required:

- Culturing apparatus: a cabinet or chamber is recommended, in which the chosen incubation temperature can be maintained at ± 2°C.

- Light measurement instruments: it is important to note that the method of measurement of light intensity, and in particular the type of receptor (collector), may affect the measured value. Measurements should preferably be made using a spherical (4 π) receptor (which responds to direct and reflected light from all angles above and below the plane of measurement), or a 2 π receptor (which responds to light from all angles above the measurement plane).

- Apparatus to determine algal biomass. Cell count, which is the most frequently used surrogate parameter for algal biomass, may be made using an electronic particle counter, a microscope with counting chamber, or a flow cytometer. Other biomass surrogates can be measured using a flow cytometer, fluorimeter, spectrophotometer or colorimeter. A conversion factor relating cell count to dry weight is useful to calculate. In order to provide useful measurements at low biomass concentrations when using a spectrophotometer, it may be necessary to use cuvettes with a light path of at least 4 cm.

Test organisms

17. Several species of non-attached microalgae and cyanobacteria may be used. The strains listed in Annex 2 have been shown to be suitable using the test procedure specified in this Guideline.

18. If other species are used, the strain and/or origin should be reported. Confirm that exponential growth of the selected test alga can be maintained throughout the test period under the prevailing conditions.

Growth medium

19. Two alternative growth media, the OECD and the AAP medium, are recommended. The compositions of these media are shown in Annex 3. Note that the initial pH value and the buffering

© OECD, (2011)

capacity (regulating pH increase) of the two media are different. Therefore the results of the tests may be different depending on the medium used, particularly when testing ionising substances.

20. Modification of the growth media may be necessary for certain purposes, e.g. when testing metals and chelating agents or testing at different pH values. Use of a modified medium should be described in detail and justified (3) (4).

Initial biomass concentration

21. The initial biomass in the test cultures must be the same in all test cultures and sufficiently low to allow exponential growth throughout the incubation period without risk of nutrient depletion. The initial biomass should not exceed 0.5 mg/L as dry weight. The following initial cell concentrations are recommended:

Pseudokirchneriella subcapitata:	5×10^3 - 10^4	cells/mL
Desmodesmus subspicatus	2-5 $\times 10^3$	cells/mL
Navicula pelliculosa	10^4	cells/mL
Anabaena flos-aquae	10^4	cells/mL
Synechococcus leopoliensis	5×10^4 - 10^5	cells/mL

Concentrations of test substance

22. The concentration range in which effects are likely to occur may be determined on the basis of results from range-finding tests. For the final definitive test at least five concentrations, arranged in a geometric series with a factor not exceeding 3.2, should be selected. For test substances showing a flat concentration response curve a higher factor may be justified. The concentration series should preferably cover the range causing 5-75 % inhibition of algal growth rate.

Replicates and controls

23. The test design should include three replicates at each test concentration. If determination of the NOEC is not required, the test design may be altered to increase the number of concentrations and reduce the number of replicates per concentration. The number of control replicates must be at least three, and ideally should be twice the number of replicates used for each test concentration.

24. A separate set of test solutions may be prepared for analytical determinations of test substance concentrations (see paragraphs 36 and 38).

25. When a solvent is used to solubilise the test substance, additional controls containing the solvent at the same concentration as used in the test cultures must be included in the test design.

Preparation of inoculum culture

26. In order to adapt the test alga to the test conditions and ensure that the algae are in the exponential growth phase when used to inoculate the test solutions, an inoculum culture in the test medium is prepared 2-4 days before start of the test. The algal biomass should be adjusted in order to allow exponential growth to prevail in the inoculum culture until the test starts. Incubate the inoculum culture under the same conditions as the test cultures. Measure the increase in biomass in the inoculum culture to ensure that growth is within the normal range for the test strain under the culturing conditions. An example of the procedure for algal culturing is described in Annex 4. To avoid synchronous cell divisions during the test a second propagation step of the inoculum culture may be required.

© OECD, (2011)

Preparation of test solutions

27. All test solutions must contain the same concentrations of growth medium and initial biomass of test alga. Test solutions of the chosen concentrations are usually prepared by mixing a stock solution of the test substance with growth medium and inoculum culture. Stock solutions are normally prepared by dissolving the substance in test medium.

28. Solvents, e.g. acetone, t-butyl alcohol and dimethyl formamide, may be used as carriers to add substances of low water solubility to the test medium (2)(3). The concentration of solvent should not exceed 100 µl/L, and the same concentration of solvent should be added to all cultures (including controls) in the test series.

Incubation

29. Cap the test vessels with air-permeable stoppers. The vessels are shaken and placed in the culturing apparatus. During the test it is necessary to keep the algae in suspension and to facilitate transfer of CO_2. To this end constant shaking or stirring should be used. The cultures should be maintained at a temperature in the range of 21 to 24°C, controlled at ± 2°C. For species other than those listed in Annex 2, e.g. tropical species, higher temperatures may be appropriate, providing that the validity criteria can be fulfilled. It is recommended to place the flasks randomly and to reposition them daily in the incubator.

30. The pH of the control medium should not increase by more than 1.5 units during the test. For metals and compounds that partly ionise at a pH around the test pH, it may be necessary to limit the pH drift to obtain reproducible and well defined results. A drift of < 0.5 pH units is technically feasible and can be achieved by ensuring an adequate CO_2 mass transfer rate from the surrounding air to the test solution, e.g. by increasing the shaking rate. Another possibility is to reduce the demand for CO_2 by reducing the initial biomass or the test duration.

31. The surface where the cultures are incubated should receive continuous, uniform fluorescent illumination e.g. of «cool-white» or «daylight» type. Strains of algae and cyanobacteria vary in their light requirements. The light intensity should be selected to suit the test organism used. For the recommended species of green algae, select the light intensity at the level of the test solutions from the range of 60-120 $\mu E \cdot m^{-2} s^{-1}$ when measured in the photosynthetically effective wavelength range of 400-700 nm using an appropriate receptor. Some species, in particular *Anabaena flos-aquae*, grow well at lower light intensities and may be damaged at high intensities. For such species an average light intensity in the range 40-60 $\mu E \cdot m^{-2} \cdot s^{-1}$ should be selected. (For light-measuring instruments calibrated in lux, an equivalent range of 4440 – 8880 lux for cool white light corresponds approximately to the recommended light intensity 60-120 $\mu E \cdot m^{-2} \cdot s^{-1}$). Maintain the light intensity within ±15% from the average light intensity over the incubation area.

Test duration

32. Test duration is normally 72 hours. However, shorter or longer test durations may be used provided that all validity criteria in paragraph 11 can be met.

Measurements and analytical determinations

© OECD, (2011)

33. The algal biomass in each flask is determined at least daily during the test period. If measurements are made on small volumes removed from the test solution by pipette, these should not be replaced.

34. Measurement of biomass is done by manual cell counting by microscope or an electronic particle counter (by cell counts and/or biovolume). Alternative techniques, e.g. flow cytometry, *in vitro* or *in vivo* chlorophyll fluorescence (5) (6), or optical density can be used if a satisfactory correlation with biomass can be demonstrated over the range of biomass occurring in the test.

35. Measure the pH of the solutions at the beginning and at the end of the test.

36. Provided an analytical procedure for determination of the test substance in the concentration range used is available, the test solutions should be analysed to verify the initial concentrations and maintenance of the exposure concentrations during the test.

37. Analysis of the concentration of the test substance at the start and end of the test of a low and high test concentration and a concentration around the expected EC_{50} may be sufficient where it is likely that exposure concentrations will vary less than 20% from nominal values during the test. Analysis of all test concentrations at the beginning and at the end of the test is recommended where concentrations are unlikely to remain within 80-120 % of nominal. For volatile, unstable or strongly adsorbing test substances, additional samplings for analysis at 24 hour intervals during the exposure period are recommended in order to better define loss of the test substance. For these substances, extra replicates may be needed. In all cases, determination of test substance concentrations need only be performed on one replicate vessel at each test concentration (or the contents of the vessels pooled by replicate).

38. The test media prepared specifically for analysis of exposure concentrations during the test should be treated identically to those used for testing, i.e. they should be inoculated with algae and incubated under identical conditions. If analysis of the dissolved test substance concentration is required, it may be necessary to separate algae from the medium. Separation should preferably be made by centrifugation at a low g-force, sufficient to settle the algae.

39. If there is evidence that the concentration of the substance being tested has been satisfactorily maintained within ± 20 % of the nominal or measured initial concentration throughout the test, analysis of the results can be based on nominal or measured initial values. If the deviation from the nominal or measured initial concentration is not within the range of ± 20 %, analysis of the results should be based on geometric mean concentration during exposure or on models describing the decline of the concentration of the test substance (3) (7).

40. The alga growth inhibition test is a more dynamic test system than most other short-term aquatic toxicity tests. As a consequence, the actual exposure concentrations may be difficult to define, especially for adsorbing substances tested at low concentrations. In such cases, disappearance of the test substance from solution by adsorption to the increasing algal biomass does not mean that it is lost from the test system. When the result of the test is analysed, it should be checked whether a decrease in concentration of the test substance in the course of the test is accompanied by a decrease in growth inhibition. If this is the case, application of a suitable model describing the decline of the concentration of the test substance (7) may be considered. If not, it may be appropriate to base the analysis of the results on the initial (nominal or measured) concentrations.

© OECD, (2011)

Other observations

41. Microscopic observation should be performed to verify a normal and healthy appearance of the inoculum culture and to observe any abnormal appearance of the algae (as may be caused by the exposure to the test substance) at the end of the test.

Limit test

42. Under some circumstances, e.g. when a preliminary test indicates that the test substance has no toxic effects at concentrations up to 100 mg/L or up to its limit of solubility in the test medium (whichever is the lower), a limit test involving a comparison of responses in a control group and one treatment group (100 mg/L or a concentration equal to the limit of solubility), may be undertaken. It is strongly recommended that this be supported by analysis of the exposure concentration. All previously described test conditions and validity criteria apply to a limit test, with the exception that the number of treatment replicates should be at least six. The response variables in the control and treatment group may be analysed using a statistical test to compare means, e.g. a Student's t-test. If variances of the two groups are unequal, a t-test adjusted for unequal variances should be performed

DATA AND REPORTING

Plotting growth curves

43. The biomass in the test vessels may be expressed in units of the surrogate parameter used for measurement (e.g. cell number, fluorescence).

44. Tabulate the estimated biomass concentration in test cultures and controls together with the concentrations of test material and the times of measurement, recorded with a resolution of at least whole hours, to produce plots of growth curves. Both logarithmic scales and linear scales can be useful at this first stage, but logarithmic scales are mandatory and generally give a better presentation of variations in growth pattern during the test period. Note that exponential growth produces a straight line when plotted on a logarithmic scale, and inclination of the line (slope) indicates the specific growth rate.

45. Using the plots, examine whether control cultures grow exponentially at the expected rate throughout the test. Examine all data points and the appearance of the graphs critically and check raw data and procedures for possible errors. Check in particular any data point that seems to deviate by a systematic error. If it is obvious that procedural mistakes can be identified and/or considered highly likely, the specific data point is marked as an outlier and not included in subsequent statistical analysis. (A zero algal concentration in one out of two or three replicate vessels may indicate the vessel was not inoculated correctly, or was improperly cleaned). State reasons for rejection of a data point as an outlier clearly in the test report. Accepted reasons are only (rare) procedural mistakes and not just bad precision. Statistical procedures for outlier identification are of limited use for this type of problem and cannot replace expert judgement. Outliers (marked as such) should preferably be retained among the data points shown in any subsequent graphical or tabular data presentation.

Response variables

46. The purpose of the test is to determine the effects of the test substance on the growth of algae. This Guideline describes two response variables, as member countries have different preferences and regulatory needs. In order for the test results to be acceptable in all member countries, the effects should be evaluated using both response variables (a) and (b) described below.

© OECD, (2011)

(a) Average specific growth rate: this response variable is calculated on the basis of the logarithmic increase of biomass during the test period, expressed per day

(b) Yield: this response variable is the biomass at the end of the test minus the starting biomass.

47. It should be noted that toxicity values calculated by using these two response variables are not comparable and this difference must be recognised when using the results of the test. EC_x values based upon average specific growth rate (E_rC_x) will generally be higher than results based upon yield (E_yC_x) if the test conditions of this Guideline are adhered to, due to the mathematical basis of the respective approaches. This should not be interpreted as a difference in sensitivity between the two response variables, simply that the values are different mathematically. The concept of average specific growth rate is based on the general exponential growth pattern of algae in non-limited cultures, where toxicity is estimated on the basis of the effects on the growth rate, without being dependent on the absolute level of the specific growth rate of the control, slope of the concentration-response curve or on test duration. In contrast, results based upon the yield response variable are dependent upon all these other variables. E_yC_x is dependent on the specific growth rate of the algal species used in each test and on the maximum specific growth rate that can vary between species and even different algal strains. This response variable should not be used for comparing the sensitivity to toxicants among algal species or even different strains. While the use of average specific growth rate for estimating toxicity is scientifically preferred, toxicity estimates based on yield are also included in this Guideline to satisfy current regulatory requirements in some countries.

Average growth rate

48. The average specific growth rate for a specific period is calculated as the logarithmic increase in the biomass from the equation for each single vessel of controls and treatments [1]:

$$\mu_{i-j} = \frac{\ln X_j - \ln X_i}{t_j - t_i} \text{ (day}^{-1}\text{)} \quad \text{----- [1],}$$

where:
μ_{i-j} is the average specific growth rate from time i to j;
X_i is the biomass at time i;
X_j is the biomass at time j

For each treatment group and control group, calculate a mean value for growth rate along with variance estimates.

49. Calculate the average specific growth rate over the entire test duration (normally days 0-3), using the nominally inoculated biomass as the starting value rather than a measured starting value, because in this way greater precision is normally obtained. If the equipment used for biomass measurement allows sufficiently precise determination of the low inoculum biomass (e.g. flow cytometer) then the measured initial biomass concentration can be used. Assess also the section-by-section growth rate, calculated as the specific growth rates for each day during the course of the test (days 0-1, 1-2 and 2-3) and examine whether the control growth rate remains constant (see validity criteria, paragraph 11). A significantly lower specific growth rate on day one than the total average specific growth rate may indicate a lag phase. While a lag phase can be minimised and practically eliminated in control cultures by proper propagation of the pre-culture, a lag phase in exposed cultures may indicate recovery after initial toxic stress or reduced

© OECD, (2011)

exposure due to loss of test substance (including sorption onto the algal biomass) after initial exposure. Hence the section-by-section growth rate may be assessed in order to evaluate effects of the test substance occurring during the exposure period. Substantial differences between the section-by-section growth rate and the average growth rate indicate deviation from constant exponential growth and that close examination of the growth curves is warranted.

50. Calculate the percent inhibition of growth rate for each treatment replicate from equation [2]:

$$\%I_r = \frac{\mu_C - \mu_T}{\mu_C} \times 100 \text{ ----- [2]},$$

where:

$\%I_r$: percent inhibition in average specific growth rate;
μ_C mean value for average specific growth rate (μ) in the control group;
μ_T average specific growth rate for the treatment replicate.

51. When solvents are used to prepare the test solutions, the solvent controls rather than the controls without solvents should be used in calculation of percent inhibition.

Yield

52. Yield is calculated as the biomass at the end of the test minus the starting biomass for each single vessel of controls and treatments. For each test concentration and control, calculate a mean value for yield along with variance estimates. The percent inhibition in yield ($\%I_y$) may be calculated for each treatment replicate as follows:

$$\%I_y = \frac{(Y_c - Y_T)}{Y_c} \times 100 \text{ ----- [3]}$$

where:

$\%I_y$: percent inhibition of yield;
- Y_C: mean value for yield in the control group;
- Y_T: value for yield for the treatment replicate.

Plotting concentration response curve

53. Plot the percentage of inhibition against the logarithm of the test substance concentration and examine the plot closely, disregarding any such data point that was singled out as an outlier in the first phase. Fit a smooth line through the data points by eye or by computerised interpolation to get a first impression of the concentration-response relationship, and then proceed with a more detailed method, preferably a computerised statistical method. Depending on the intended usage of data; the quality (precision) and amount of data as well as the availability of data analysis tools, it may be decided (and sometimes well justified) to stop the data analysis at this stage and simply read the key figures EC_{50} and EC_{10} (and/or EC_{20}) from the eye fitted curve (see also section below on stimulatory effects). Valid reasons for not using a statistical method may include:

- Data are not appropriate for computerised methods to produce any more reliable results than can be obtained by expert judgement - in such situations some computer programs may even fail to produce a reliable solution (iterations may not converge etc.)

- Stimulatory growth responses cannot be handled adequately using available computer programs (see below).

© OECD, (2011)

Statistical procedures

54. The aim is to obtain a quantitative concentration-response relationship by regression analysis. It is possible to use a weighted linear regression after having performed a linearising transformation of the response data - for instance into probit or logit or Weibull units (8), but non-linear regression procedures are preferred techniques that better handle unavoidable data irregularities and deviations from smooth distributions. Approaching either zero or total inhibition, such irregularities may be magnified by the transformation, interfering with the analysis (8). It should be noted that standard methods of analysis using probit, logit, or Weibull transforms are intended for use on quantal (e.g. mortality or survival) data, and must be modified to accommodate growth or biomass data. Specific procedures for determination of EC_x values from continuous data can be found in (9) (10) and (11). The use of non-linear regression analysis is further detailed in Annex 5.

55. For each response variable to be analysed, use the concentration-response relationship to calculate point estimates of EC_x values. When possible, the 95% confidence limits for each estimate should be determined. Goodness of fit of the response data to the regression model should be assessed either graphically or statistically. Regression analysis should be performed using individual replicate responses, not treatment group means. If, however nonlinear curve fitting is difficult or fails because of too great scatter in the data, the problem may be circumvented by performing the regression on group means as a practical way of reducing the influence of suspected outliers. Use of this option should be identified in the test report as a deviation from normal procedure because curve fits with individual replicates did not produce a good result.

56. EC_{50} estimates and confidence limits may also be obtained using linear interpolation with bootstrapping (13), if available regression models/methods are unsuitable for the data.

57. For estimation of the LOEC and hence the NOEC, for effects of the test substance on growth rate, it is necessary to compare treatment means using analysis of variance (ANOVA) techniques. The mean for each concentration must then be compared with the control mean using an appropriate multiple comparison or trend test method. Dunnett's or Williams' test may be useful (12)(14)(15)(16)(17). It is necessary to assess whether the ANOVA assumption of homogeneity of variance holds. This assessment may be performed graphically or by a formal test (17). Suitable tests are Levene's or Bartlett's. Failure to meet the assumption of homogeneity of variances can sometimes be corrected by logarithmic transformation of the data. If heterogeneity of variance is extreme and cannot be corrected by transformation, analysis by methods such as step-down Jonkheere trend tests should be considered. Additional guidance on determining the NOEC can be found in (11).

58. Recent scientific developments have led to a recommendation of abandoning the concept of NOEC and replacing it with regression based point estimates EC_x. An appropriate value for x has not been established for this algal test. A range of 10 to 20 % appears to be appropriate (depending on the response variable chosen), and preferably both the EC_{10} and EC_{20} should be reported.

Growth stimulation

59. Growth stimulation (negative inhibition) at low concentrations is sometimes observed. This can result from either hormesis ("toxic stimulation") or from addition of stimulating growth factors with the test material to the minimal medium used. Note that the addition of inorganic nutrients should not have any direct effect because the test medium should maintain a surplus of nutrients throughout the test. Low dose stimulation can usually be ignored in EC_{50} calculations unless it is extreme. However, if it is extreme, or an EC_x value for low x is to be calculated, special procedures may be needed. Deletion of stimulatory responses from the data analysis should be avoided if possible, and if available curve fitting software

© OECD, (2011)

cannot accept minor stimulation, linear interpolation with bootstrapping can be used. If stimulation is extreme, use of a hormesis model may be considered (18).

Non toxic growth inhibition

60. Light absorbing test materials may give rise to a growth rate reduction because shading reduces the amount of available light. Such physical types of effects should be separated from toxic effects by modifying the test conditions and the former should be reported separately. Guidance may be found in (2) and (3).

Test report

61. The test report must include the following:

Test substance:
- physical nature and relevant physical-chemical properties, including water solubility limit;
- chemical identification data (e.g., CAS Number), including purity (impurities).

Test species:
- the strain, supplier or source and the culture conditions used.

Test conditions:
- date of start of the test and its duration;
- description of test design: test vessels, culture volumes, biomass density at the beginning of the test;
- composition of the medium;
- test concentrations and replicates (e.g., number of replicates, number of test concentrations and geometric progression used);
- description of the preparation of test solutions, including use of solvents etc.
- culturing apparatus;
- light intensity and quality (source, homogeneity);
- temperature;
- concentrations tested: the nominal test concentrations and any results of analyses to determine the concentration of the test substance in the test vessels. The recovery efficiency of the method and the limit of quantification in the test matrix should be reported.;
- all deviations from this Guideline;
- method for determination of biomass and evidence of correlation between the measured parameter and dry weight;

Results:
- pH values at the beginning and at the end of the test at all treatments;
- biomass for each flask at each measuring point and method for measuring biomass;
- growth curves (plot of biomass versus time);
- calculated response variables for each treatment replicate, with mean values and coefficient of variation for replicates;
- graphical presentation of the concentration/effect relationship;
- estimates of toxicity for response variables e.g., EC_{50}, EC_{10}, EC_{20} and associated confidence intervals. If calculated, LOEC and NOEC and the statistical methods used for their determination;

© OECD, (2011)

- if ANOVA has been used, the size of the effect which can be detected (e.g. the least significant difference);
- any stimulation of growth found in any treatment;
- any other observed effects, e.g. morphological changes of the algae;
- discussion of the results, including any influence on the outcome of the test resulting from deviations from this Guideline.

LITERATURE

(1) International Organisation for Standardisation (1993). ISO 8692 Water quality – Algal growth inhibition test.

(2) International Organisation for Standardisation (1998). ISO/DIS 14442. Water quality – Guidelines for algal growth inhibition tests with poorly soluble materials, volatile compounds, metals and waster water.

(3) OECD (2000). Guidance Document on Aquatic Toxicity Testing of Difficult Substances and mixtures. Environmental Health and Safety Publications. Series on Testing and Assessment, no. 23. Organisation for Economic Co-operation and Development, Paris.

(4) International Organisation for Standardisation (1998). ISO 5667-16 Water quality – Sampling – Part 16: Guidance on Biotesting of Samples.

(5) Mayer, P., Cuhel, R. and Nyholm, N. (1997). A simple in vitro fluorescence method for biomass measurements in algal growth inhibition tests. *Water Research* 31: 2525-2531.

(6) Slovacey, R.E. and Hanna, P.J. (1997). In vivo fluorescence determinations of phytoplancton chlorophyll, Limnology & Oceanography 22, 5, pp.919-925

(7) Simpson, S.L., Roland, M.G.E., Stauber, J.L. and Batley, G.E. (2003) Effect of declining toxicant concentrations on algal bioassay endpoints. *Environ. Toxicol. Chem.* 22, 2073-2079.

(8) Christensen, E.R., Nyholm, N. (1984). Ecotoxicological Assays with Algae: Weibull Dose-Response Curves. *Env. Sci. Technol.* 19, 713-718.

(9) Nyholm, N. Sørensen, P.S., Kusk, K.O. and Christensen, E.R. (1992): Statistical treatment of data from microbial toxicity tests. *Environ. Toxicol. Chem.* 11, 157-167.

(10) Bruce, R.D.,and Versteeg, D.J. (1992). A statistical procedure for modelling continuous toxicity data. *Environ. Toxicol. Chem.* 11:1485-1494.

(11) OECD. (2005). Current Approaches in the Statistical Analysis of Ecotoxicity Data: A Guidance to Application. Organisation for Economic Co-operation and Development, Paris.

(12) Dunnett, C.W. (1955) A multiple comparisons procedure for comparing several treatments with a control. J. Amer. Statist. Assoc. 50: 1096-1121

(13) Norberg-King T.J. (1988) An interpolation estimate for chronic toxicity: The ICp approach. National Effluent Toxicity Assessment Center Technical Report 05-88. US EPA, Duluth, MN.

© OECD, (2011)

(14) Dunnett, C.W. (1964) New tables for multiple comparisons with a control. Biometrics 20: 482-491.

(15) Williams, D.A. (1971). A test for differences between treatment means when several dose levels are compared with a zero dose control. Biometrics 27: 103-117.

(16) Williams, D.A. (1972). The comparison of several dose levels with a zero dose control. Biometrics 28: 519-531.

(17) Draper, N.R. and Smith, H. (1981). Applied Regression Analysis, second edition. Wiley, New York.

(18) Brain, P. and Cousens, R. (1989). An equation to describe dose-responses where there is stimulation of growth at low doses. *Weed Research*, 29, 93-96.

© OECD, (2011)

ANNEX 1

DEFINITIONS

The following definitions and abbreviations are used for the purposes of this Guideline:

Biomass is the dry weight of living matter present in a population expressed in terms of a given volume; e.g., mg algae/litre test solution. Usually "biomass" is defined as a mass, but in this test this word is used to refer to mass per volume. Also in this test, surrogates for biomass, such as cell counts, fluorescence, etc. are typically measured and the use of the term "biomass" thus refers to these surrogate measures as well.

Coefficient of variation is a dimensionless measure of the variability of a parameter, defined as the ratio of the standard deviation to the mean. This can also be expressed as a percent value. Mean coefficient of variation of average specific growth rate in replicate control cultures should be calculated as follows:

1. Calculate % CV of average specific growth rate out of the daily/section by section growth rates for the respective replicate;
2. Calculate the mean value out of all values calculated under point 1 to get the mean coefficient of variation of the daily/section by section specific growth rate in replicate control cultures.

EC_x is the concentration of the test substance dissolved in test medium that results in an x % (e.g. 50%) reduction in growth of the test organism within a stated exposure period (to be mentioned explicitly if deviating from full or normal test duration). To unambiguously denote an EC value deriving from growth rate or yield the symbol "E_rC" is used for growth rate and "E_yC" is used for yield.

Growth medium is the complete synthetic culture medium in which test algae grow when exposed to the test substance. The test substance will normally be dissolved in the test medium.

Growth rate (average specific growth rate) is the logarithmic increase in biomass during the exposure period.

Lowest Observed Effect Concentration (LOEC) is the lowest tested concentration at which the substance is observed to have a statistically significant reducing effect on growth (at $p < 0.05$) when compared with the control, within a given exposure time. However, all test concentrations above the LOEC must have a harmful effect equal to or greater than those observed at the LOEC. When these two conditions cannot be satisfied, a full explanation must be given for how the LOEC (and hence the NOEC) has been selected.

No Observed Effect Concentration (NOEC) is the test concentration immediately below the LOEC.

Response variable is a variable for the estimation of toxicity derived from any measured parameters describing biomass by different methods of calculation. For this guideline growth rates and yield are response variables derived from measuring biomass directly or any of the surrogates mentioned.

Specific growth rate is a response variable defined as quotient of the difference of the natural logarithms of a parameter of observation (in this Guideline, biomass) and the respective time period

Yield is the value of a measurement variable at the end of the exposure period minus the measurement variable's value at the start of the exposure period to express biomass increase during the test.

© OECD, (2011)

ANNEX 2

STRAINS SHOWN TO BE SUITABLE FOR THE TEST

Green algae

- *Pseudokirchneriella subcapitata,* (formerly known as *Selenastrum capricornutum*) , ATCC 22662, CCAP 278/4, 61.81 SAG
- *Desmodesmus subspicatus* (formerly known as *Scenedesmus subspicatus*) 86.81 SAG

Diatoms

- *Navicula pelliculosa,* UTEX 664

Cyanobacteria

- *Anabaena flos-aquae,* UTEX 1444, ATCC 29413, CCAP 1403/13A
- *Synechococcus leopoliensis,* UTEX 625, CCAP 1405/1

Sources of Strains

The strains recommended are available in unialgal cultures from the following collections (in alphabetical order):

ATCC: American Type Culture Collection
10801 University Boulevard
Manassas, Virginia 20110-2209
USA

CCAP, Culture Collection of Algae and Protozoa
Institute of Freshwater Ecology,
Windermere Laboratory
Far Sawrey, Amblerside
Cumbria LA22 0LP
UK

SAG: Collection of Algal Cultures
Inst. Plant Physiology
University of Göttingen
Nicholausberger Weg 18
D-3400 Göttingen
GERMANY

UTEX Culture Collection of Algae
Section of Molecular, Cellular and Developmental Biology
School of Biological Sciences
the University of Texas at Austin
Austin, Texas 78712
USA.

© OECD, (2011)

Appearance and characteristics of recommended species

	P. subcapitata	*D. subspicatus*	*N. pelliculosa*	*A. flos-aquae*	*S. leopoliensis*
Appearance	Curved, twisted single cells	Oval, mostly single cells	Rods	Chains of oval cells	Rods
Size (L x W) µm	8-14 x 2-3	7-15 x 3-12	7.1 x 3.7	4.5 x 3	6 x 1
Cell volume (μm^3/cell)	40-60[1]	60-80[1]	40-50[1]	30-40[1]	2.5[2]
Cell dry weight (mg/cell)	2-3 x 10^{-8}	3-4 x 10^{-8}	3-4 x 10^{-8}	1-2 x 10^{-8}	2-3 x 10^{-9}
Growth rate[3] (day^{-1})	1.5 -1.7	1.2-1.5	1.4	1.1-1.4	2.0 - 2.4

[1] Measured with electronic particle counter
[2] Calculated from size
[3] Most frequently observed growth rate in OECD medium at light intensity approx. 70 µE m^{-2} s^{-1} and 21 °C

Specific Recommendations on Culturing and Handling of Recommended Test Species

Pseudokirchneriella subcapitata and *Desmodesmus subspicatus*

These green algae are generally easy to maintain in various culture media. Information on suitable media is available from the culture collections. The cells are normally solitary, and cell density measurements can easily be performed using an electronic particle counter or microscope.

Anabaena flos-aquae

Various growth media may be used for keeping a stock culture. It is particularly important to avoid allowing the batch culture to go past log phase growth when renewing, recovery is difficult at this point.

Anabaena flos-aquae develops aggregates of nested chains of cells. The size of these aggregates may vary with culturing conditions. It may be necessary to break up these aggregates when microscope counting or an electronic particle counter is used for determination of biomass.

Sonication of sub-samples may be used to break up chains to reduce count variability. Longer sonication than required for breaking up chains into shorter lengths may destroy the cells. Sonication intensity and duration must be identical for each treatment.

Count enough fields on the hemocytometer (at least 400 cells) to help compensate for variability. This will improve reliability of microscopic density determinations.

An electronic particle counter can be used for determination of total cell volume of *Anabaena* after breaking up the cell chains by careful sonification. The sonification energy has to be adjusted to avoid disruption of the cells.

© OECD, (2011)

Use a vortex mixer or similar appropriate method to make sure the algae suspension used to inoculate test vessels is well mixed and homogeneous.

Test vessels should be placed on an orbital or reciprocate shaker table at about 150 revolutions per minute. Alternatively, intermittent agitation may be used to reduce the tendency of *Anabaena* to form clumps. If clumping occurs, care must be taken to achieve representative samples for biomass measurements. Vigorous agitation before sampling may be necessary to disintegrate algal clumps.

Synechococcus leopoliensis

Various growth media may be used for keeping a stock culture. Information on suitable media is available from the culture collections.

Synechococcus leopoliensis grows as solitary rod-shaped cells. The cells are very small, which complicates the use of microscope counting for biomass measurements. Electronic particle counters equipped for counting particles down to a size of approximately 1 µm are useful. *In vitro* fluorometric measurements are also applicable.

Navicula pelliculosa

Various growth media may be used for keeping a stock culture. Information on suitable media is available from the culture collections. Note that silicate is required in the medium.

Navicula pelliculosa may form aggregates under certain growth conditions. Due to production of lipids the algal cells sometimes tend to accumulate in the surface film. Under those circumstances special measures have to be taken when sub-samples are taken for biomass determination in order to obtain representative samples. Vigorous shaking, e.g. using a vortex mixer may be required.

© OECD, (2011)

ANNEX 3

GROWTH MEDIA

One of the following two growth media may be used:

OECD medium: Original medium of OECD TG 201, also according to ISO 8692
US. EPA medium AAP also according to ASTM.

When preparing these media, reagent or analytical-grade chemicals should be used and deionised water.

Composition of The AAP-medium (US. EPA) and the OECD TG 201 medium.

Component	AAP		OECD	
	mg/L	mM	mg/L	mM
$NaHCO_3$	15.0	0.179	50.0	0.595
$NaNO_3$	25.5	0.300		
NH_4Cl			15.0	0.280
$MgCl_2 \cdot 6(H_2O)$	12.16	0.0598	12.0	0.0590
$CaCl_2 \cdot 2(H_2O)$	4.41	0.0300	18.0	0.122
$MgSO_4 \cdot 7(H_2O)$	14.6	0.0592	15.0	0.0609
K_2HPO_4	1.044	0.00599		
KH_2PO_4			1.60	0.00919
$FeCl_3 \cdot 6(H_2O)$	0.160	0.000591	0.0640	0.000237
$Na_2EDTA \cdot 2(H_2O)$	0.300	0.000806	0.100	0.000269*
H_3BO_3	0.186	0.00300	0.185	0.00299
$MnCl_2 \cdot 4(H_2O)$	0.415	0.00201	0.415	0.00210
$ZnCl_2$	0.00327	0.000024	0.00300	0.0000220
$CoCl_2 \cdot 6(H_2O)$	0.00143	0.000006	0.00150	0.00000630
$Na_2MoO_4 \cdot 2(H_2O)$	0.00726	0.000030	0.00700	0.0000289
$CuCl_2 . 2(H_2O)$	0.000012	0.00000007	0.00001	0.00000006
pH	7.5		8.1	

- The molar ratio of EDTA to iron slightly exceed unity. This prevents iron precipitation and at the same time, chelation of heavy metal ions is minimised.

In test with the diatom *Navicula pelliculosa* both media must be supplemented with $Na_2SiO_3 \cdot 9H_20$ to obtain a concentration of 1.4 mg Si/L.

© OECD, (2011)

The pH of the medium is obtained at equilibrium between the carbonate system of the medium and the partial pressure of CO_2 in atmospheric air. An approximate relationship between pH at 25 °C and the molar bicarbonate concentration is:

$$pH_{eq} = 11.30 + \log[HCO_3]$$

With 15 mg $NaHCO_3$/L, pH_{eq} = 7.5 (U.S. EPA medium) and with 50 mg $NaHCO_3$/L, pH_{eq} = 8.1 (OECD medium).

Element composition of test media

Element	AAP mg/L	OECD mg/L
C	2.144	7.148
N	4.202	3.927
P	0.186	0.285
K	0.469	0.459
Na	11.044	13.704
Ca	1.202	4.905
Mg	2.909	2.913
Fe	0.033	0.017
Mn	0.115	0.115

© OECD, (2011)

Preparation of OECD medium

Nutrient	Concentration in stock solution
Stock solution 1: macro nutrients	
NH_4Cl	1.5 g/L
$MgCl_2 \cdot 6H_2O$	1.2 g/L
$CaCl_2 \cdot 2H_2O$	1.8 g/L
$MgSO_4 \cdot 7H_2O$	1.5 g/L
KH_2PO_4	0.16 g/L
Stock solution 2: iron	
$FeCl_3 \cdot 6H_2O$	64 mg/L
$Na_2EDTA \cdot 2H_2O$	100 mg/L
Stock solution 3: trace elements	
H_3BO_3	185 mg/L
$MnCl_2 \cdot 4H_2O$	415 mg/L
$ZnCl_2$	3 mg/L
$CoCl_2 \cdot 6H_2O$	1.5 mg/L
$CuCl_2 \cdot 2H_2O$	0.01 mg/L
$Na_2MoO_4 \cdot 2H_2O$	7 mg/L
Stock solution 4: bicarbonate	
$NaHCO_3$	50 g/L
$Na_2SiO_3 \cdot 9H_20$	

Sterilize the stock solutions by membrane filtration (mean pore diameter 0.2 µm) or by autoclaving (120 °C, 15 min). Store the solutions in the dark at 4 °C.
Do not autoclave stock solutions 2 and 4, but sterilise them by membrane filtration.

Prepare a growth medium by adding an appropriate volume of the stock solutions 1-4 to water:

Add to 500 ml of sterilised water:

- 10 ml of stock solution 1
- 1 ml of stock solution 2
- 1 ml of stock solution 3
- 1 ml of stock solution 4

Make up to 1 000 mL with sterilised water.

Allow sufficient time for equilibrating the medium with the atmospheric CO_2, if necessary by bubbling with sterile, filtered air for some hours.

© OECD, (2011)

Preparation of U.S. EPA medium

1. Add 1 mL of each stock solution in *2.1–2.7* to approximately 900 mL of deionized or distilled water and then dilute to 1 litre.

2. Macronutrient stock solutions are made by dissolving the following into 500 mL of deionised or distilled water. Reagents *2.1, 2.2, 2.3,* and *2.4* can be combined into one stock solution.

2.1 $NaNO_3$—12.750 g.
2.2 $MgCl_2 \cdot 6H_2O$—6.082 g.
2.3 $CaCl_2 \cdot 2H_2O$—2.205 g.
2.4 *Micronutrient Stock Solution*—(see 3).
2.5 $MgSO_4 \cdot 7H_2O$—7.350 g.
2.6 K_2HPO_4—0.522 g.
2.7 $NaHCO_3$—7.500 g.
2.8 $Na_2SiO_3 \cdot 9H_2O$—See Note 1.

NOTE 1: Use for diatom test species only. May be added directly (202.4 mg) or by way of stock solution to give 20 mg/L Si final concentration in medium.

3. The micronutrient stock solution is made by dissolving the following into 500 mL of deionised or distilled water:

3.1 H_3BO_3—92.760 mg.
3.2 $MnCl_2 \cdot 4H_2O$—207.690 mg.
3.3 $ZnCl_2$—1.635 mg.
3.4 $FeCl_3 \cdot 6H_2O$—79.880 mg.
3.5 $CoCl_2 \cdot 6H_2O$—0.714 mg.
3.6 $Na_2MoO_4 \cdot 2H_2O$—3.630 mg.
3.7 $CuCl_2 \cdot 2H_2O$—0.006 mg.
3.8 $Na_2EDTA \cdot 2H_2O$—150.000 mg. [Disodium (Ethylenedinitrilo) tetraacetate].
3.9 $Na_2SeO_4 \cdot 5H_2O$—0.005 mg See Note 2.

NOTE 2: Use only in medium for stock cultures of diatom species.

4. Adjust pH to 7.5± 0.1 with 0.1 *N* or 1.0 *N* NaOH or HCl.

5. Filter the media into a sterile container through either a 0.22 µm membrane filter if a particle counter is to be used or a 0.45 µm filter if a particle counter is not to be used.

6. Store medium in the dark at approximately 4°C until use.

© OECD, (2011)

ANNEX 4

EXAMPLE OF A PROCEDURE FOR THE CULTURING OF ALGAE

General observations

The purpose of culturing on the basis of the following procedure is to obtain algal cultures for toxicity tests.

Use suitable methods to ensure that the algal cultures are not infected with bacteria. Axenic cultures may be desirable but unialgal cultures must be established and used.

All operations must be carried out under sterile conditions in order to avoid contamination with bacteria and other algae.

Equipment and materials

See under Test Guideline: Apparatus.

Procedures for obtaining algal cultures

Preparation of nutrient solutions (media):

All nutrient salts of the medium are prepared as concentrated stock solutions and stored dark and cold. These solutions are sterilised by filtration or by autoclaving.

The medium is prepared by adding the correct amount of stock solution to sterile distilled water, taking care that no infection occurs. For solid medium 0.8 per cent of agar is added.

Stock culture:

The stock cultures are small algal cultures that are regularly transferred to fresh medium to act as initial test material. If the cultures are not used regularly they are streaked out on sloped agar tubes. These are transferred to fresh medium at least once every two months.

The stock cultures are grown in conical flasks containing the appropriate medium (volume about 100 ml). When the algae are incubated at 20°C with continuous illumination, a weekly transfer is required.

During transfer an amount of "old" culture is transferred with sterile pipettes into a flask of fresh medium, so that with the fast-growing species the initial concentration is about 100 times smaller than in the old culture.

The growth rate of a species can be determined from the growth curve. If this is known, it is possible to estimate the density at which the culture should be transferred to new medium. This must be done before the culture reaches the death phase.

Pre-culture:

The pre-culture is intended to give an amount of algae suitable for the inoculation of test cultures. The pre-culture is incubated under the conditions of the test and used when still exponentially growing, normally after an incubation period of 2 to 4 days. When the algal cultures contain deformed or abnormal cells, they must be discarded.

© OECD, (2011)

ANNEX 5

DATA ANALYSIS BY NONLINEAR REGRESSION

General considerations

The response in algal tests and other microbial growth tests - growth of biomass - is by nature a continuous or metric variable – a process rate if growth rate is used and its integral over time if biomass is selected. Both are referenced to the corresponding mean response of replicate non-exposed controls showing maximum response for the conditions imposed - with light and temperature as primary determining factors in the algal test. The system is distributed or homogenous and the biomass can be viewed as a continuum without consideration of individual cells. The variance distribution of the type of response for a such system relate solely to experimental factors (described typically by the log-normal or normal distributions of error). This is by contrast to typical bioassay responses with quantal data for which the tolerance (typically binomially distributed) of individual organisms are often assumed to be the dominant variance component. Control responses are here zero or background level.

In the uncomplicated situation, the normalized or relative response, r, decreases monotonically from 1 (zero inhibition) to 0 (100 per cent inhibition). Note, that all responses have an error associated and that apparent negative inhibitions can be calculated as a result of random error only.

Regression analysis

Models

A regression analysis aims at quantitatively describing the concentration response curve in the form of a mathematical regression function $Y = f(C)$ or more frequently $F(Z)$ where $Z = \log C$. Used inversely $C = f^{-1}(Y)$ allows the calculation of, EC_x figures, including the EC_{50}, EC_{10} and EC_{20}, and their 95% confidence limits. Several simple mathematical functional forms have proved to successfully describe concentration - response relationships obtained in algal growth inhibition tests. Functions include for instance the logistic equation, the nonsymmetrical Weibul equation and the log normal distribution function, which are all sigmoid curves asymptotically approaching zero for $C \rightarrow 0$ and one for $C \rightarrow$ infinity.

The use of continuous threshold function models (e.g. the Kooijman model "for inhibition of population growth" Kooijman et al. 1996) is a recently proposed or alternative to asymptotic models. This model assumes no effects at concentrations below a certain threshold EC_0+ that is estimated by extrapolation of the response concentration relationship to intercept the concentration axis using a simple continuous function that is not differentiable in the starting point.

Note that the analysis can be a simple minimization of sums of residual squares (assuming constant variance) or weighted squares if variance heterogeneity is compensated

Procedure

The procedure can be outlined as follows: Select an appropriate functional equation, $Y = f(C)$, and fit it to the data by non-linear regression. Use preferably the measurements from each individual flask rather than means of replicates, in order to extract as much information from the data as possible. If the variance is high, on the other hand, practical experience suggests that means of replicates may provide a more robust mathematical estimation less influenced by systematic errors in the data, than with each individual data point retained.

Plot the fitted curve and the measured data and examine whether the curve fit is appropriate. Analysis of residuals may be a particular helpful tool for this purpose. If the chosen functional relationship to fit the concentration response does not describe well the whole curve or some essential part of it, such as the response at low concentrations, choose another curve fit option - e.g., a non-symmetrical curve like the Weibul function instead of a symmetrical one. Negative inhibitions may be a problem with for instance the

© OECD, (2011)

log - normal distribution function likewise demanding an alternative regression function. It is not recommended to assign a zero or small positive value to such negative values because this distorts the error distribution. It may be appropriate to make separate curve fits on parts of the curve such as the low inhibition part to estimate $EC_{low\ x}$ figures. Calculate from the fitted equation (by "inverse estimation", $C = f^{-1}(Y)$), characteristic point estimates EC_x's, and report as a minimum the EC_{50} and one or two $EC_{low\ x}$ estimates. Experience from practical testing has shown that the precision of the algal test normally allows a reasonably accurate estimation at the 10 % inhibition level if data points are sufficient - unless stimulation occurs at low concentrations as a confounding factor. The precision of an EC_{20} estimate is often considerably better than that of an EC_{10}, because the EC_{20} is usually positioned on the approximately linear part of the central concentration response curve. Sometimes EC_{10} can be difficult to interpret because of growth stimulation. So while the EC_{10} is normally obtainable with a sufficient accuracy it is recommended to report always also the EC_{20}.

Weighting factors

The experimental variance generally is not constant and typically includes a proportional component, and a weighted regression is therefore advantageously carried out routinely. Weighting factors for a such analysis are normally assumed inversely proportional to the variance:

$$W_i = 1/Var(r_i)$$

Many regression programs allow the option of weighted regression analysis with weighting factors listed in a table. Conveniently weighting factors should be normalized by multiplying them by $n/\Sigma\ w_i$ (n is the number of datapoints) so their sum be one.

Normalizing responses

Normalizing by the mean control response gives some principle problems and gives rise to a rather complicated variance structure. Dividing the responses by the mean control response for obtaining the percentage of inhibition, one introduces an additional error caused by the error on the control mean. Unless this error is negligibly small, weighting factors in the regression and confidence limits must be corrected for the covariance with the control (Draper and Smith, 1981). Note that high precision on the estimated mean control response is important in order to minimize the overall variance for the relative response. This variance is as follows:

(Subscript i refers to concentration level i and subscript 0 to the controls)

$$Yi = \text{Relative response} = r_i/r_0 = 1 - I = f(C_i)$$

with a variance $Var(Y_i) = Var(r_i/r_0) \cong (\partial Yi / \partial r_i)^2 \cdot Var(r_i) + ((\partial Y_i/\partial r_0)^2 \cdot Var(r_0)$

and since $(\partial Y_i/\partial r_i) = 1/r_0$ and $(\partial Y_i/\partial r_0) = r_i/r_0^{\ 2}$

with normally distributed data and m_i and m_0 replicates: $Var(r_i) = \sigma^2/m_i$

the total variance of the relative response Y_i thus becomes

$Var(Y_i) = \sigma^2/(r_0^{\ 2} \cdot m_i) + r_i^{\ 2} \cdot \sigma^2/r_0^{\ 4} \cdot m_0$

The error on the control mean is inversely proportional to the square root of the number of control replicates averaged, and sometimes it can be justified to include historic data and in this way greatly reduce the error. An alternative procedure is not to normalize the data and fit the absolute responses including the control response data but introducing the control response value as an additional parameter to be fitted by non linear regression. With a usual 2 parameter regression equation, this method necessitates the fitting of 3 parameters, and therefore demands more data points than non-linear regression on data that are normalized using a pre-set control response .

Inverse confidence intervals

The calculation of non-linear regression confidence intervals by inverse estimation is rather complex and not an available standard option in ordinary statistical computer program packages. Approximate confidence limits may be obtained with standard non-linear regression programs with re-parameterisation (Bruce and Versteeg, 1992), which involves rewriting the mathematical equation with the desired point estimates, e.g. the EC_{10} and the EC_{50} as the parameters to be estimated. (Let the function be $I = f(\alpha, \beta,$

Concentration) and utilize the definition relationships f (α, β, EC_{10}) = 0.1 and f (α, β, EC_{50}) = 0.5 to substitute f (α, β, concentration) with an equivalent function g(EC_{10}, EC_{50}, concentration).
A more direct calculation (Andersen et al, 1998) is performed by retaining the original equation and using a Taylor expansion around the means of r_i and r_0.
Recently "boot strap methods" have become popular. Such methods use the measured data and a random number generator directed frequent re-sampling to estimate an empirical variance distribution.

Kooijman, S.A.L.M.; Hanstveit, A.O.; Nyholm, N. (1996): No-effect concentrations in algal growth inhibition tests. *Water Research*, 30, 1625-1632.

Draper, N.R. and Smith, H. (1981). Applied Regression Analysis, second edition. Wiley, New York.

Bruce, R..D. and Versteeg,, D.J.(1992) A Statistical Procedure for Modelling Continuous Ecotoxicity Data. *Environ. Toxicol. Chem.*11, 1485-1494

Andersen, J.S., Holst, H., Spliid, H., Andersen, H., Baun, A. & Nyholm, N. (1998): Continuous ecotoxicological data evaluated relative to a control response. *Journal of Agricultural, Biological and Environmental Statistics*, 3, 405-420.

© OECD, (2011)

Franklin BAUER

Une meilleure caractérisation des mécanismes d'action toxique à partir de la structure moléculaire.

Résumé

Connaître le mécanisme d'action (MechoA) des substances est une première étape clé pour l'évaluation de risque (éco)toxicologique. Les MechoAs désignent les interactions moléculaires entre les xénobiotiques et le matériel biologique, entraînant des effets toxiques.

Des informations de MechoA ont été collectées pour des centaines de molécules chez différentes espèces, principalement le rat et le poisson. Une première méthode pour prédire les MechoAs à partir de la structure a été développée à l'aide de calculs quantiques. Elle a résulté en 70% de classifications correctes.

Par la suite, une deuxième méthode de prédiction utilisant des alertes structurales a été développée, et utilise 6 MechoAs généraux incluant 23 MechoAs détaillés. Cette méthode a obtenu 92% de classifications correctes, tout en étant simple et rapide à mettre en œuvre. Nous l'avons comparée avec d'autres méthodes disponibles et notre arbre de décision a obtenu les meilleures statistiques. Cette méthode est en cours d'implémentation dans un logiciel, et sera disponible gratuitement.

Des expériences en laboratoire ont aussi été effectuées et ont permis de consolider les données pour des substances mal caractérisées, en mesurant le logP, la solubilité aqueuse et la toxicité aigüe pour les daphnies.

Mots-clés : toxicité, écotoxicité, mécanisme d'action, prédiction

Résumé en anglais

A knowledge of the mechanism of action (MechoA) of substances is a crucial first step in (eco)toxicological risk assessment approaches. MechoAs refer to molecular interactions between xenobiotics and biological material, leading to an adverse outcome.

MechoA information has been gathered for hundreds of molecules in several species, mainly rats and fish. A first method to predict MechoAs from the structure has been developed using quantum calculations. This resulted in 70% correct classifications.

A second method to predict MechoAs using structural alerts was developed, using a classification with 6 general MechoAs including 23 detailed MechoAs. This method achieved 92% correct classifications, while being simple and fast to use. We compared this method with other available methods and our decision tree showed the best statistics. This method is currently being implemented into a software, and it will be made freely available as we consider it as a useful support in risk assessment.

Some laboratory studies have also been performed and have consolidated data for substances which were not well characterised, measuring logKOW, water solubility and acute toxicity to daphnids.

Keywords : toxicity, ecotoxicity, mechanism of action, prediction