European Commission

JRC TECHNICAL REPORTS

Application of the Virtual Cell Based Assay for Simulation of *in vitro* Chemical fate following Acute Exposure

Proença S., Paini A., Joossens E., Sala Benito J. V., Berggren E., Worth A., and Prieto P

2017

Joint Research Centre

EUR 28694 EN

This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

Contact information

Name: Pilar Prieto
Address: TP126 I-21027 Ispra (VA), Italy
Email:pilar.prieto-peraita@ec.europa.eu
Tel.: 00390332785534

Name: Alicia Paini
Address: TP126 I-21027 Ispra (VA), Italy
Email: alicia.paini@ec.europa.eu
Tel.: 00390332783986

JRC Science Hub
https://ec.europa.eu/jrc

JRC107407

EUR 28694 EN

PDF	ISBN 978-92-79-70867-1	ISSN 1831-9424	doi:10.2760/475757
Print	ISBN 978-92-79-70866-4	ISSN 1018-5593	doi:10.2760/715415

Luxembourg: Publications Office of the European Union, 2017

© European Union, 2017

Reuse is authorised provided the source is acknowledged. The reuse policy of European Commission documents is regulated by Decision 2011/833/EU (OJ L 330, 14.12.2011, p. 39).

For any use or reproduction of photos or other material that is not under the EU copyright, permission must be sought directly from the copyright holders.

How to cite this report: Author(s), *Title*, EUR, Publisher, Publisher City, Year of Publication, ISBN, doi

All images © European Union 2017

Contents

Acknowledgements

The authors would like to thank Dr Anthony Williams (EPA, USA) and Dr Nynke Kramer (IRAS, Utrecht) for discussion of the results.

Authors

Susana Proença, Alicia Paini, Elisabeth Joossens, Jose Vicente Sala Benito, Elisabet Berggren, Andrew Worth and Pilar Prieto

Directorate General Joint Research Centre; Directorate F – Health, Consumers and Reference Materials; Chemicals Safety and Alternative Methods Unit (F.3) incorporating EURL ECVAM. Via E. Fermi, 2749. TP126 I-21027 Ispra (VA), Italy

1. Abstract

In order to reliably assess the risk of adverse systemic effects of chemicals by using *in vitro* methods, there is a need to simulate their absorption, distribution, metabolism, and excretion (ADME) *in vivo* to determine the target organ bioavailable concentration, and to compare this predicted internal concentration with an effective internal concentration. The effective concentration derived from *in vitro* toxicity studies should ideally take into account the fate of chemicals in the *in vitro* test system, since there can be significant differences between the applied nominal concentration and the *in vitro* bioavailable concentration. Whereas PBK models have been developed to simulate ADME properties *in vivo*, the Virtual Cell Based Assay (VCBA) has been developed to simulate *in vitro* fate. In this project, the VCBA model in R code, was applied to better interpret previously obtained *in vitro* acute toxicity data and study how they can be compared to results from acute toxicity *in vivo*.

For 178 chemicals previously tested *in vitro* with the 3T3 BALB/c cell line using the Neutral Red Uptake cytotoxicity assay, physicochemical parameters were retrieved and curated. Of these chemicals, 83 were run in the VCBA to simulate a 96-well microplate set up with 5% serum supplementation, and their no effect concentration (NEC) and killing rate (Kr) optimized against the experimental data. Analyses of results of partitioning of the chemicals show a strong relation with their lipophilicity, expressed here as the logarithm of the octanol/water partitioning coefficient, with highly lipophilic chemicals binding mostly to medium lipid. Among the chemicals analysed, only benzene and xylene were modelled to evaporate by more than 10 %, and these were also the chemicals with highest degradation rates during the 48 hours assay. Chemical degradation is dependent not only on the air and water degradation rates but also on the extent of binding of the chemical.

Due to the strong binding of some chemicals to medium lipids and proteins we analysed the impact of different serum supplementations (0%, 5% and 10%) on the chemical dissolved concentrations. As expected, for the more lipophilic chemicals, different serum levels result in different dissolved concentrations, with lipid and protein binding reducing chemical loss by evaporation. Still the lack of saturation modelling might mislead the 0 % supplementation since the lipids coming solely from cells exudates are able to sequester chemical to a large extent, eg. after 48 hours, 63% (1.2E-5 M) of dimethyldioctadecylammonium chloride was bound to lipid from the cells. Although highly lipophilic chemicals have a very small bioavailable fraction, cellular uptake rate is also dependent on $\log K_{ow}$, which compensates for this lack of bioavailability to some extent.

Based on the relevance of lipophilicity on *in vitro* chemical bioavailability, we have developed an alert system based on $\log K_{ow}$, creating four classes of chemicals for the experimental condition with 10% serum supplementation: $\log K_{ow}$ 5- 10 (A), $\log K_{ow}$ <5 (B), $\log K_{ow}$ <2.5 (C), and $\log K_{ow}$ <2 (D). New

chemicals from Classes A and B, which will in the future be tested *in vitro*, were run first on the VCBA, without considering toxicity (NEC and Kr set to 0). VCBA simulations indicated that these chemicals are more than 50% bound to medium proteins, lipids and plastic. Therefore, for chemicals with $logK_{ow}$ falling in these classes, special care should be taken when extrapolating the obtained *in vitro* toxic concentrations to *in vivo* relevant doses.

A comparison of the VCBA-predicted dissolved concentrations corresponding to nominal IC50 values with the available rat oral LD50 values did not improve the previously obtained correlations. This is probably because other *in vivo* kinetic processes play an important role but were not considered in this *in vitro-in vivo* extrapolation.

The comparison of the VCBA predicted IC50 dissolved concentrations with the available rat oral LD50 values, did not improve the previously obtained correlations. Nevertheless, other *in vivo* kinetic processes that are not modelled may play an important role. They should be considered in the *in vitro-in vivo* extrapolations.

A local sensitivity analysis showed the relative low impact of Molar Volume and Molecular Diffusion Volume on the final dissolved concentration, supporting the use of approximated values obtained through the herein created QSARs. The $logk_{ow}$ and Henry Law Constant showed, as expected, a high impact in partitioning. Killing rate was shown to also have a relative low impact in the final chemical concentration, indicating that although its optimization is important, finding the Kr that leads to the absolute best correlation between experimental and predicted concentration-viability curves, is not imperative.

The VCBA can be applied to virtually any chemical as long as the physicochemical data (for the fate model) and the experimental toxicity data (that include cell growth/death) are available. However, being such a generic model, several assumptions had to be made: i) no distinction of chemical classes (inorganic, polar organic chemicals), ii) no consideration of metabolism, iii) saturation kinetics and iv) external *in vitro* conditions.

The advantages of having a generic model are that the VCBA can fit several experimental set ups and should be used in an exploratory manner, to help refinement of experimental conditions. The herein obtained VCBA results should be double checked experimentally the partition with a set of chemical compounds to better understand to what extent VCBA represents chemicals of different properties.

In future developments, it would be important to reduce the uncertainties of the model such as binding-saturation and consider inclusion of other endpoints such as metabolic activity.

2. Introduction

Global production of chemicals has increased from 1 million tonnes in 1930 to 400 million tonnes in 2001, with some of these new chemicals constituting a hazard to human health and the environment. To obtain information on chemicals on the EU market, and to determine the risks they may pose, the REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) Regulation was implemented, under which all chemicals that are produced 1 tonne or more per year need to be registered[1]. This includes the requirement for manufacturers and importers to gather information on the properties of their chemical substances. Although issues around animal experimentation have already been recognized for some years with Russell and Burch elaboration of 3Rs principles (replacement, reduction and refinement) in 1959, risk assessment of chemicals has a long history of relying on animal models. REACH promotes the use of alternative tests for the generation of information on intrinsic properties of substances (article 13), and efforts have been made to develop and show the potential of alternatives to animal experimentation methods.

In vitro models have been emerging as the main animal experimental alternatives, offering the possibility of using several types of animal and human cells. Considering that toxicological events initiate mostly at a cellular level[2,3], these models are highly relevant as they can give further insights of toxic mechanisms. Nevertheless, *in vitro* toxicity data should not be directly compared to *in vivo* data due to the fact that complex biokinetic and toxicodynamic processes that occur *in vivo* resulting in a heterogeneous chemical distribution in the animal or human's body, cannot be captured as such in an *in vitro* system.

To convert the *in vitro* concentration-response curve and median inhibitory concentration (IC_{50}) into more relevant doses for human risk and safety assessment (e.g. *in* vivo median Lethal Dose (LD_{50})), *in silico* physiologically-based kinetic (PBK) modelling have been created. PBK models consist in sets of differential equations that simulate pharmacokinetic processes such as absorption, distribution, metabolism, and excretion (ADME). Thus, these *in silico* models allow both the calculation from the nominal *in vivo* dose to the target-organ bioavailable concentration and consequently the extrapolation from *in vitro* to *in vivo* (IVIVE) [4–7].

In this context, several PBK models have been developed, most of which are compiled in Lu *et al*.[8] The integration of *in vitro* toxicity data and these models has been indeed indicating a good correlation between the prediction and experimental concentrations for animal and human toxicity [9–11].

This integration of *in vitro* and *in silico* (including PBK models) was the approach used by Gubbels van Hal et al [12] to analyse a set of 10 compounds. This work showed that it was possible to decrease by 38% the number of the animals used. Still, one of the endpoints which showed to be more difficult to evaluate

without the use of animal data, was acute oral toxicity in which half of the compounds had their toxicity over-estimated.

Although integration of these modelling techniques allows accounting for the *in vivo* toxicokinetics, approximating the toxic doses obtained in the different models, *in vitro* cells are still in a different microenvironment, as it is herein illustrated:

i) Underrepresentation of the toxicological targets, some of which require multi-organ interactions, thus not captured in an *in vitro* system.

ii) Incomplete differentiation into organ specific phenotypes, making metabolism and clearance, hallmarks troublesome to represent [13–15].

iii) Although reduced, there are some pharmacokinetic processes *in vitro,* such as binding to the supplemented serum proteins and lipids [16,17]*,* binding to plastic [18] and evaporation, which do not occur *in vivo*.

iv) Frequently the dose metric used in *in vitro* systems is concentration, which does not reflect the amount of compound per number of cells. Gulden *et al*, 2001 showed that cell quantity does change the free concentration and toxic effects, with higher cell numbers in culture having higher IC_{50} values[19].

While solutions to the points i) and ii) pass through more sophisticated *in vitro* technologies such as body-on-a-chip, and more refined techniques of differentiation, the last two points are an issue of dosimetry/kinetics.

Therefore, a better approximation to *in vivo* might be obtained if these *in vitro* biokinetic processes are modulated, determining the concentration that is effectively dissolved in the exposure medium and unbound (free concentration). This simulation might reduce the gap between the *in vitro* and the *in vivo* freely available plasma concentration, especially for highly volatile and/or lipophilic chemical compounds.

Hence, several *in vitro* kinetic models have been developed as summarized in Table 1, which shows the different focus and design of these models.

Table 1 -List of published references which characterize the fate of a chemical in *in vitro* cell lines. Legend: **Sin**- Single Exposure, **Rep**- Repeated exposure, **PHH**-Primary Human Hepatocytes; **PRH**-Primary Rat Hepatocytes; **HepaRG**- Human hepatic stem cell line; **HepG2**- Human hepatocellular carcinoma-derived cell line; **RTL-W1** and **Rtgill-W1-** Rainbow trout cell lines; **BALB 3T3**- Mouse fibroblast cell line; **HEK293**- Human embryonal kidney cell line; **A549**- Human adenocarcinoma-derived alveolar basal epithelial cell line; **MCF-7**- human breast adenocarcinoma cell line

Model		**Cell Type**	**Chemical(s)**	**Exposure**		**Dynamic endpoint: cell viability**	**Ref.**
Compartments	**Kinetic elements**			**Sin.**	**Rep.**		
Cell Membrane Medium	Lipid and Protein	HEK293T HEK293H HepG2, HCT116 ME-180	100 chemicals from neutral to ionogenic	X			20
VCBA Cell, Medium Headspace	Serum lipid Serum Protein, Plastic, Water and Air Degradation, Dissolved organic matter (cell exudates)	3T3 HepaRG HepG2 A549	Any as long the respective physical-chemical parameters are found.	X	X	X	21, 51
Cell Medium	ECM proteins; Metabolism	PHH PRH HepaRG	Ibuprofen	X	X	X	22
Cells, Medium Headspace	Plastic	RTgill-W1	Imidacloprid, Dimethoate, Carbendazim, Malathion, Cyproconazole, Propiconazole, Pentachlorophenol, Cypermethrin, 1,2,3-Trichlorobenzene, Naphtalene, Hexachlorobenzene	X			23
Cells/tissue Medium Headspace	Serum , Plastic, Water solubility, dissolved organic matter,	Any	Any	x			24
Cells Medium	Plastic; Metabolism via clearance	PRH HepaRG	Chlorpromazine	X	X		25
Cells, Medium Headspace	Protein, Plastic	BALB 3T3 RTgill-W1	Phenanthrene	X		X	18

Model		Cell Type	Chemical(s)	Exposure		Dynamic endpoint: cell viability	Ref
Compartments	**Kinetic elements**			**Sin.**	**Rep.**		
Cells Medium Headspace	Protein, Plastic	RTL-W1 RTgill-W1	Benzo(a)pyrene, 1,2-dichlorobenzene, and 1,2,4-trichlorobenzene		X	X	26
Medium, Cells/tissue	Protein	MCF-7 cells	Genistein, bisphenol A, Octylpneol	X			27
Cells, Medium		HEK293	[3H]estradiol, octylphenol.	X			28
Cells, Medium Culture Vessel		Sperm cells	Antimycin A, digitonin, thioridazine HCl, hexachlorophene 4,4'-DDE, dieldrin, pentachlorophenol, methylmercury, chloride and xylene and 1-nitronaphthalene	X			19

As an example, Heringa *et al* [28] showed to have obtained better correlations between *in vivo* and *in vitro* toxic potencies when the calculated freely available concentrations were used as measure of the cytotoxic potency instead of the nominal concentrations. The relevance of toxicokinetic modelling is further highlighted by the fact that the partition with other cell culture components has a special great impact in compounds with higher cytotoxicity potencies. The quantity of non-bioavailable compound can easily surpass the bioavailable one. In the case of less toxic compounds, the impact of serum binding may be negligible if the nominal toxic concentration exceeds the binding capacity of the serum proteins [29].

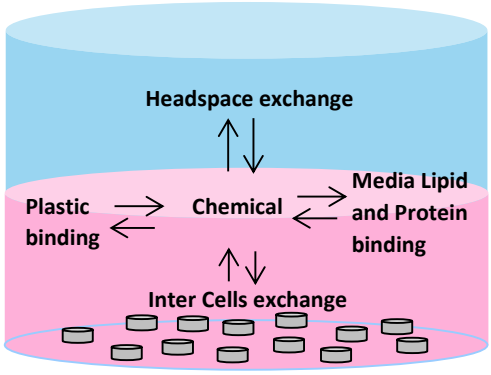

Figure 1. Schematic representation of the Fate and Transport model of the virtual cell based assay (VCBA) that simulates the kinetics of a chemical tested.

The Virtual Cell Based Assay (VCBA) is another of these *in silico* models that simulates the chemical fate *in vitro*, and was developed as part of the EU FP7 COSMOS project (http://www.cosmostox.eu) to clarify the actual bioavailable concentration required to cause perturbations in cells [30,31]. Briefly, this model is represented in Figure 1 and consists of 4 interconnected models:

[1] Fate and transport model, is based on each compound physical-chemical properties. It describes the dynamic mass balance of compound with its partition between headspace (gas exchange equations), plastic and serum lipid and protein [16,26] and compound degradation;

[2] Cell partitioning model, which accounts for cells uptake/excretion and intracellular partition between lipid, protein and aqueous fractions which depends on the chemical characteristics and cell type composition;

[3] Cell growth and division model, which is simulated through a 4-staged (G1, S, G2 and M cell cycle phases) approach using a Leslie Matrix;

[4] Toxicity and effects model, which merges the experimental *in vitro* obtained concentration-response curve with the cell growth and division model to optimize the toxicity parameters, Killing rate(Kr) and No-Effect Concentration (NEC);

Additionally, the VCBA takes into account the experimental set up, which includes the well shape and size, the volume of media and the amount of supplemented serum and, thus, protein and lipid content in the media.

The mathematical equation describing the four interconnected models of the VCBA are reported in Zaldivar et al.[21] To run the VCBA specific inputs parameters for chemicals, cell types and experimental set up are needed. Herein we aimed to analyse 178 compounds used in international projects and validation studies [NICEATM/ECVAM validation study (NIH, 2006); the PF6 EU project ACuteTox (http://www.acutetox.eu/; Prieto et al., 2013a); ECVAM validation study (Prieto et al., 2013b] where the cell line BALB/c 3T3 was used and cytotoxicity was measured with a Neutral Red Uptake (NRU) assay. In the ACuteTox project the in vitro cytotoxicity assay was complemented with specific target organ *in vitro* assays in an attempt to improve the prediction of human acute oral systemic toxicity. With regard to classification of compounds into acute oral toxicity categories according to the EU CLP Regulation (Classification, Labelling and Packaging of Substances and Mixture), the results showed difficulties in predicting the 4 toxicity categories with any of the proposed combinations. Nevertheless, substances belonging to the non-classified group (LD_{50} >2000mg/kg) were predicted relatively well, with a false negative rate lower than 5 %[32]. To rationalize the true/false predictions obtained, kinetic parameters should be considered as recommended in the EURL ECVAM strategy to replace, reduce and refine the use of animals in the assessment of acute

mammalian toxicity[33]. With this in mind, the value of the kinetic simulations obtained with the VCBA model has been explored.

The overall goal of this work was to i) clean, harmonize and evaluate the previous VCBA model code; ii) retrieve the physicochemical parameters of 178 compounds, iii) optimize and run the model, obtaining the concentration of compounds partition in the several elements/compartments of the *in vitro* assay. iv) to analyse if the calculated dissolved and unbound IC_{50} correlates better with the *in vivo* LD_{50} than the nominal IC_{50}, possibly explaining the misclassifications obtained with the 3T3 NRU cytotoxicity assay. Furthermore, we used the VCBA to help identifying compounds prone to have an *in vitro* determined toxic concentration, very discrepant from the *in vivo* one and that may offer specific difficulties *in vitro*. We also propose a simple approach, as a system based on $LogK_{ow}$, to understand chemical's fate

3. Methodology

To run the Virtual Cell Based Assay (VCBA) selected physical-chemical parameters are required: molecular weight, molecular diffusion volumes (indicates as atomic diffusion in Zaldivar et al., *in press*), molar volume, Henry law constant and degradation rate in water and air and the logarithm of octanol-water partition ($LogK_{ow}$). Although for several compounds data reported were obtained experimentally, for some others only predictions were available. In case of predictions different values were often reported, and the choice of the prediction method must be carefully addressed.

The web based chemical databases Chemical Dashboard (https://comptox.epa.gov/dashboard) and Chemspider (http://www.chemspider.com/) were used for searching these parameters. While CompTox Dashboard has its own predictive tools, Chemspider (Royal Society of Chemistry) relies on the prediction tools EPI Suite™ (**US Environmental Protection Agency's**), ACD/LABS and Chemicalize. The time frame of search was from July to October 2016.

Table 2 - Web chemicals databases/prediction tools used to retrieve each chemical parameter.

	Chemspider					**Chemical Dashboard**	
Parameter:	**Exp.**	**Pred.**				**Exp.**	**Pred.**
		EPI Suite		**ACD/Labs**	**Chemicalize**		
$LogK_{ow}$	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Henry Law Constant		**X**	**X**			**X**	**X**
Air and Water half life			**X**				
Molar Volume				**X**			
Molecular Weight						**X**	
Molecular Diffusion Volume	**Fuller Method of atomic diffusion volume increments addition**						

For each chemical parameter, Table 2 summarizes the selected database where the values were retrieved; $LogK_{ow}$ was found both on Chemspider (ACD/LABS, EPI Suite and/or Chemicalize) and Chemical Dashboard.

Water and air degradation rates (s^{-1}) were calculated from the compounds' half-life (hr) in water and air, parameters retrieved from EPIsuite, database available online through Chemspider. Molar Volume was retrieved from Chemspider as well, more specifically from ACD/Labs.

Both experimental and predicted values of Henry law constant (HLC) were available in both Chemspider (EPIsuite) and Chemical Dashboard, although in the latter the values were removed in August 2016, being reposted only after a few months. Values were converted atm*m^3/mol. to Pa*m^3/mol.

3.1. Molecular Diffusion Volumes

Molecular diffusion volumes (dimensionless) were calculated following Fuller semi-empirical method which consists in the sum of the specific atomic diffusion volumes and discounting the volume for each aromatic/heterocyclic ring [34,35].

Table 3 - Atomic Diffusion Volume increments based on Fuller, 1966 and 1969

Atomic and Structural Diffusion Volume increments					
	Fuller, 1966	Fuller, 1969		Fuller, 1966	Fuller, 1969
C	16.5	15.9	**F**	8.78	14.7
H	1.98	2.31	**Cl**	19.5	21.0
O	5.48	6.11	**Br**	33.6	21.9
N	5.69	4.54	**I**	-	29.8
Aromatic Ring	-20.2	-18.3	**S**	17	22.9
Heterocyclic Ring	-20.2	-18.3			

Although, initially the atomic increments used[21] were from Fuller *et al,* 1966, slightly different increments in Fuller *et al,* 1969 were posteriorly found. Both increments are shown in Table 3. In spite of being regarded[36] as a precise method, predicting the diffusion coefficients of organic compounds with errors of <10%, its use is limited to molecules that are solely composed by the tabled atoms. Also, it is noteworthy that the method is not as precise with inorganic compounds, where the prediction is ±30 % of the measured values[37].

For 37 compounds among the 178 would require the additional Atom Diffusion Volumes increments: B, Na, Cd, Pt, P, Cu, Fe, Hg, K, Se, Tl and Zn. The similarity to Molecular Weight (MW) was analysed by correlating both

parameters as represented in Figure 2. Simultaneously, we compared the molecular diffusion volumes calculated through the different two atomic increments.

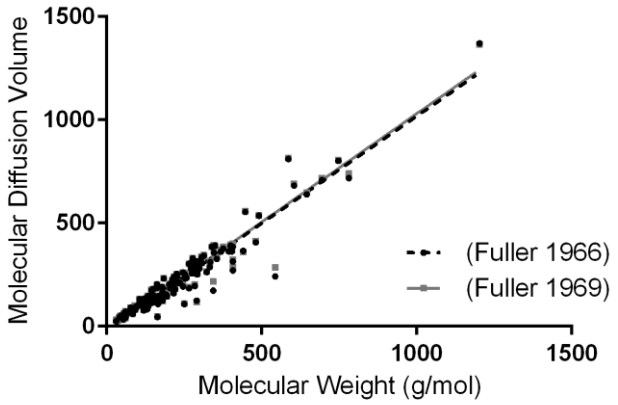

Figure 2 - Scatter plot of Molecular Weight against the respective calculated Molecular Diffusion Volume calculated through Fuller, 1966 (Black dots and continuous black line) and Fuller,1969 (grey dots and discontinuous grey line). The lines result from the regression of all presented data points.

A high linear relationship was indeed found between these two parameters, and thus for the compounds missing the molecular diffusion volume (SVcomp), it was calculated through the equation:

Fuller, 1966: $SVcomp = 1.038MW - 19.86$ R^2=0.91

Fuller, 1969: $SVcomp = 1.05MW - 19.56$ R^2=0.92

The difference between atomic increments described in Fuller, 1966 and Fuller 1969 is that for the latter ones more experimental replicates were added refine the atomic increments. Indeed, SVcomp calculated through Fuller, 1969 atomic increments have a slightly better correlation with MW. Therefore, Fuller *et al*, 1969 increments were used in this report.

3.2. Molar Volume

Molar Volume (MV) in cm^3/mol was found for 123 compounds. To understand if the missing values could as the Molecular Diffusion Volume, come from the Molecular Weight, a scatter plot was made using the found values of Molar Volume (Figure 3).

Again, a high linear correlation (R^2=0.90) was found and, therefore, the missing values for Molar Volume were calculated through the equation:

$$MV = 0.8003MW + 0.5764$$

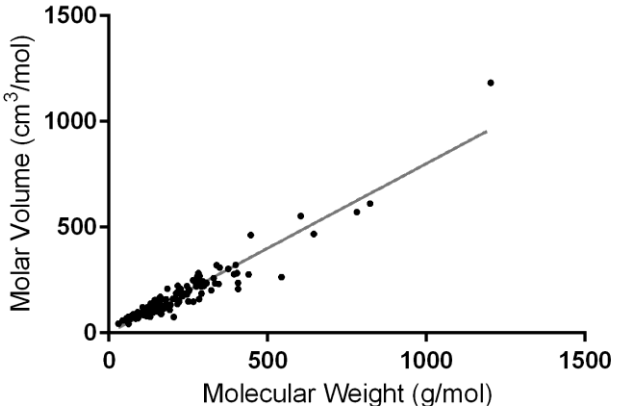

Figure 3 - Scatter plot of Molecular Weight against the respective Molar Volume. Black line is the regression line drawn through all presented data points.

3.3. LogK_{ow}

The experimental values for 109 chemical compounds were found at least in one of the web chemicals databases, and when two values were available, the average was used. For the remaining compounds, predicted values had to be used. To understand the differences between predictive tools and potentially if one of them was more reliable, an analysis between found experimental values and respective predictions was made. The linear correlation between the several sources of logK_{ow} found was very similar, as seen in Table 4:

Table 4 - Trend lines equations and correlation coefficient of experimental and predicted LogK_{ow}.

	ACD/Labs	**EPI-Suite**	**Chemicalize**	**Chemical Dashboard**
Eq.trendline	Pred= 0.9929Exp+0.0291	Pred= 0.9597Exp+0.0056	Pred= 0.9355Exp+0.0453	Pred= 0.9070Exp+0.1386
R^2	0.96	0.93	0.87	0.94

Both prediction models from EpiSuite and from Chemical Dashboard are based on the same PHYSPROP data, a collection of datasets, some coming from as early as the late 80s. However, several errors and inaccuracies have been reported and Chemical Dashboard developers have addressed this issue for some parameters such as logK_{ow}, by developing an automated curation procedure. The QSAR (quantitative structure-active relationship) resulting from this curated datasets indeed had statistically improved predictive performance [38].

Besides, Chemical Dashboard models comply with the OECD principles for QSARs[39], with unambiguous algorithms, a defined global and local applicability domain, mechanistic interpretations of the used descriptors that are reduced to the most relevant minimum, and with available information on the overall model performance. Moreover, the model is transparent, allowing access to the training and test sets from its FTP site, detailed QSAR Model Reporting Format for each model and model details for each chemical and each endpoint. All used descriptors are also free and open source (PaDEL descriptors) [Dr.Anthony Williams USA.EPA private communication].

Therefore, after the experimental data, Chemical Dashboard predictions were the ones used preferentially for following the OECD guidelines and having revised database for their prediction-model development. When these predictions were not present, the average of the other predictive tools was used.

3.4. Henry Law Constant

For 57 compounds the experimental values were found. Among the 33 compounds with experimental values found in both web databases, only 5 had different values. The largest difference was found with hexachlorobenzene with values differing 82 $Pa*m^3/mol$ between them. Hence, the average of experimental values was used. Again an analysis was made to decide upon the predictions to be used, here including different methods: Group, Bond and HENRYWIN™ method, which uses both Group and Bond method. Because HLC values were removed from Chemical Dashboard in the in middle of August 2016, for around 40 chemicals, HLC values were found only in Chemspider. Therefore, for comparison of methods/databases these 40 chemicals were excluded, and the remaining chemicals which experimental data was found, were introduced in a chart to measure the correlations between predictions and the respective experimental values. The chemical 1,1,1-trichloroethane was also excluded since its HLC is much higher than any of the other compounds and could unbalance the distribution.

The plot in Figure 4 shows relevant differences between the predictions with HENRYWIN™ showing the weakest correlation and Group Method the highest correlation (Table 5). Values retrieved from Chemical Dashboard also have a relatively weak correlation.

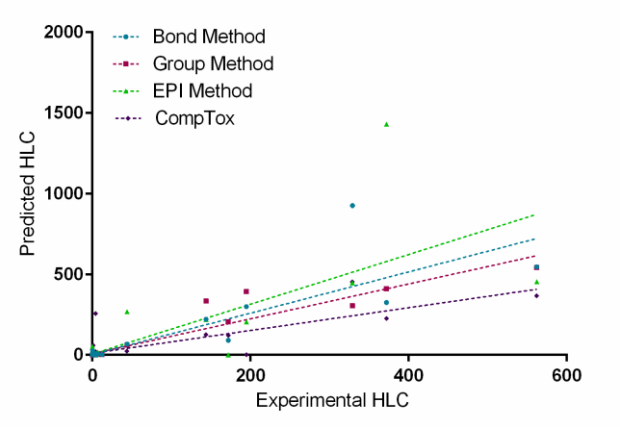

*Figure 4 - Scatter plot of Experimental Henry Law Constant against the respective Predictions obtained through Bond, Group, HENRYWIN™ (EPI method) and Chemical Dashboard (CompTox) (Units =Pa*m^3/mole).*

Since the values of Henry Law Constant spread in such a wide range, Figure 4 does not allow observation of the lower HLC values distribution. Thus, a separated analysis was additionally made using a HCL value of 1. The selection of this threshold is, nevertheless, subjective.

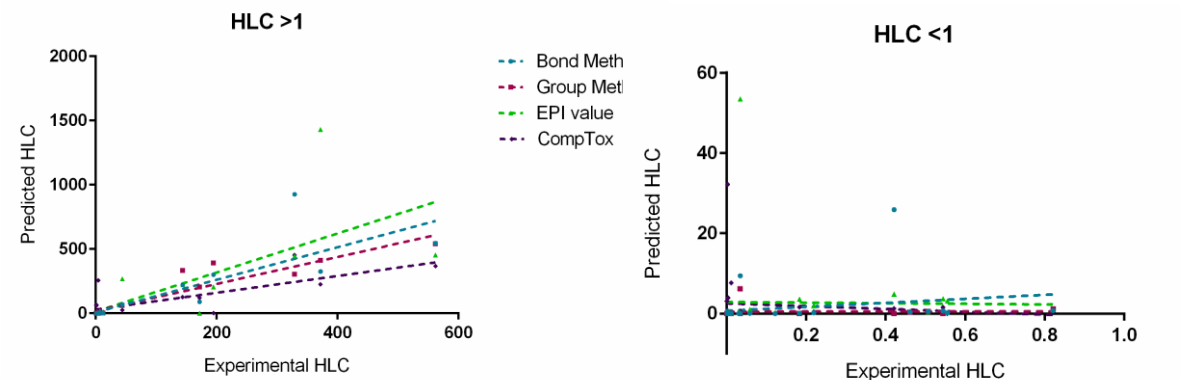

*Figure 5 - Scatter plot of Experimental Henry Law Constant, against the respective Predictions obtained through Bond, Group, HENRYWIN™ (EPI method) and Chemical Dashboard (CompTox) (Units =Pa*m^3/mole).*

Figure 5 and Table 5 show the drastic different coefficients of distribution between the two plots, with predicted HLCs having a better correlation with experimental values >1 Pa×m^3/mol than the lower ones. Hence, predictions suffer a decrease of sensitivity for small HLC values, with the Bond method showing the highest correlation. A lower sensitivity in HLC values lower than 1 might not have a significant impact as any of the predictions indicate that the compound does not evaporate. However, it is noteworthy the presence of some predictions, such as the ones for lindane and formaldehyde, where in spite of low experimental HLC values, predicted values are among ranges where significant evaporation might occur. This can lead to significant different results, ex: lindane experimental value was 4.21×10^{-1} while all predictions except the

Group Method would indicate values 1.28-25.9 and formaldehyde experimental was 3.41×10^{-2} while all the methods except Chemical Dashboard indicated values 6.22-53.4.

Table 5 - Trend lines equations and correlation coefficient of experimental and predicted Henry Law Constants

Method/ Database	**HLC Complete Range**			**HLC >1**			**HLC <1**		
	Eq. trendline	**R^2**	**N**	**Eq. trendline**	**R^2**	**N**	**Eq. trendline**	**R^2**	**N**
Bond	Y=1.28x+3.62	0.74	45	Y=1.26X+8.63	0.69	18	Y=4.94X+0.71	0.05	27
Group	Y=1.08X+8.15	0.90	35	Y=1.05X+19.56	0.88	17	Y=0.10X+0.48	< 0.001	17
HENRYWIN™	Y=1.54X+6.40	0.57	47	Y=1.52X+14.6	0.51	18	Y=-0.73X+2.88	< 0.001	27
Chemical Dashboard	Y=0.706X+10.86	0.67	47	Y=0.65X+29.85	0.60	17	Y=-3.52X+2.57	0.02	27

Although Chemical Dashboard HLC predictions are here shown with a relative low correlation with experimental data, its predictive model, as the one for $\log K_{ow}$, was based on a well curated data and follows OECD guidelines, hence being a trustworthy tool. Still, as this parameter was temporary removed from the website, EPIsuite predictions were preferably used.

In total, at least one HLC was found for 148 chemicals. Experimental data found for the compounds with higher HLC, ranged from the 1740 of 1,1,1-trichloroethane to 1.76×10^{-7} of urea. For the compounds which experimental data were not found, the priorities of predictions were: Group Method which ranged compounds with HLC = $0.3670\text{-}2.85\times10^{-15}$ (N=17); Bond Method which ranged compounds with HLC = $66.2\text{-}2.760\times10^{-37}$ (N=62) and at last CompTox Dashboard which ranged compounds with HLC =$23.41\text{-}1.880\times10^{-6}$(N=12).

3.5. VCBA Code Refinement

The VCBA model was initially created in Matlab[30,31] and more recently translated to R language to be a free toll for users and to be implemented in a KNIME environment. The differential equations describing the mass balance resulting from fate, cell dynamics and toxicodynamics are solved by the DeSolve R package. With time several versions of the VCBA code were created, all with slight modifications. Therefore, before running the chemicals a revision was made of all the versions, verifying all equations and input parameters such as cell and experimental input parameters. The code was harmonized and cleaned of redundant/duplicated equations, with the final form presented in annex 1.

3.5.1. Cell line parameters

The values herein used (Table 6) were the same as in Zaldivar *et al*[21]except the protein intracellular concentration which was corrected from 11 mol/m^3 to 4.4 mol/m^3. The value was obtained from the protein density and fraction and cell volume.

In order to allow inputs of different initial cell numbers, the initial cell number per cell cycle phases must be in fraction/percentage, then multiplied by the overall initial cell number, rather than a fixed value. However, care should be taken with the initial cell number input as 3T3 BALB/C cells have their growth inhibited when confluent (50,000 cells/cm^2) and, therefore, the initial cell number must allow growth during 48 hours without reaching this confluence. Higher initials cell numbers would require another type of fecundity functions.

Table 6– Cell line 3T3 Balb/c defined parameters to run the VCBA model.

3T3 Cell Parameters				
Aqueous Fraction (% weight)	0.614			
Protein Fraction	0.244			
Lipid Fraction	0.142			
Protein Concentration (mol/m^3)	4.4			
Lipid Concentration (kg/m^3)	170.7			
Initial Cell number (per well)	1680			
Cell Cycle phase	G1	S	G2	M
Duration (H)	9.63	3.65	3.45	2.26
Mortality (h^{-1})	0.005	0.005	0.04	0.04
Volume (m^3)	1.73E-15	2.4E-15	2.4E-15	2.4E-15
Mass (g)	2.08E-9	2.4E-9	2.4E-9	2.4E-9
Initial Cell Population (%)	50.7	19.2	18.18	11.92
Cell Division Rate (h^{-1})	1.026			

3.5.2. Experimental set up

Likewise to cell type descriptors, experimental parameters had to be revised and harmonized with the experimental protocol, such as the 48 hours of duration of the assay, the percentage of supplemented serum of 5 %, which consists in 0.0234 mol/m^3 (protein content) and 0.08 kg/m^3 (lipid content) and the 100 µL of media in a 96 well-plate well. Experimental data are available for 8 concentrations for which cells were exposed at the beginning of the 48 hours.

Table 7 – Experimental set up according to Neutral Red Uptake protocol.

Water density (g/L)	1000
Protein density (g/L)	1350
Lipid density (g/L)	900
Assay time (h)	48
% Supplemented serum	5
Protein in Medium (mol/m^3)	0.0234
Lipid in Medium (kg/m^3)	0.08
Volume Medium (m^3)	1E-7
Headspace volume (m^3)	2.68E-07
Cell assay surface (m^2)	3.31E-05
Plastic surface (m^2)	9.39E-05

3.6. Optimization and running the VCBA

Optimization of NEC and Kr was made with the chemicals input parameters and experimental concentration-response curve, consisting in 8 concentrations and a value linked to the control response (which is included as a response of 100% at 0 µM).

The code was run on R, with cell growth in hours and differential equations solved using a discretization in seconds as shown in Figure 6. The time run can be adjusted to run in minutes or other time endpoints as long as the cell growth rates are set to the correct time units. However independent of the time run because compound distribution is represented by differential equations and cell growth is not, the two processes are not synchronized in the model. For example in time run herein used for every second in one hour the distribution is calculated with the initial cell number in that hour. At the end of the hour, the model computes the cell growth/death that occurred during that hour with the compound distribution given in the end of the hour and not during all the seconds. This specific time was chosen to make optimization a faster process, as cell growth/death in minutes makes VCBA run more cycles slowing down the optimization process.

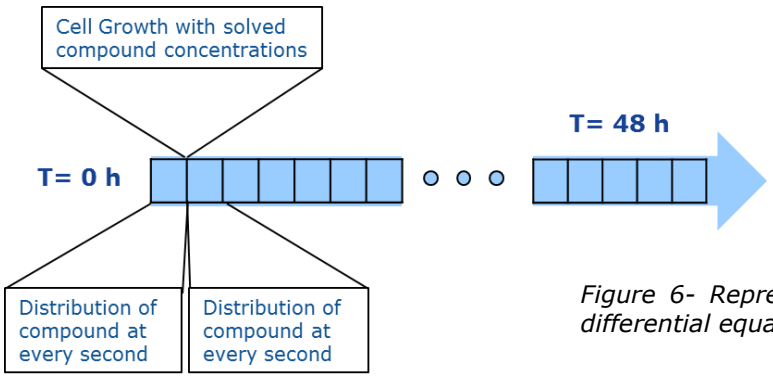

Figure 6- Representation of the VCBA model's differential equations and "for cycle" solving.

Implementation of the VCBA in KNIME for use as a web tool

The VCBA R code was implemented as an open source tool into the KNIME platform. KNIME is a user-friendly graphical workbench for data analysis (http://www.KNIME.org/) and R is a language and environment for statistical computing and graphics (http://www.r-project.org/). KNIME consists of a series of pieces of program code called nodes that can be connected in such way that the input of one node is the output of the previous one. Each node has a dialog box that accepts the user input.

This VCBA KNIME represented in Figure 7, can be divided in to three separate zones: input, core and output, this version of the VCBA is only for single exposure simulation.

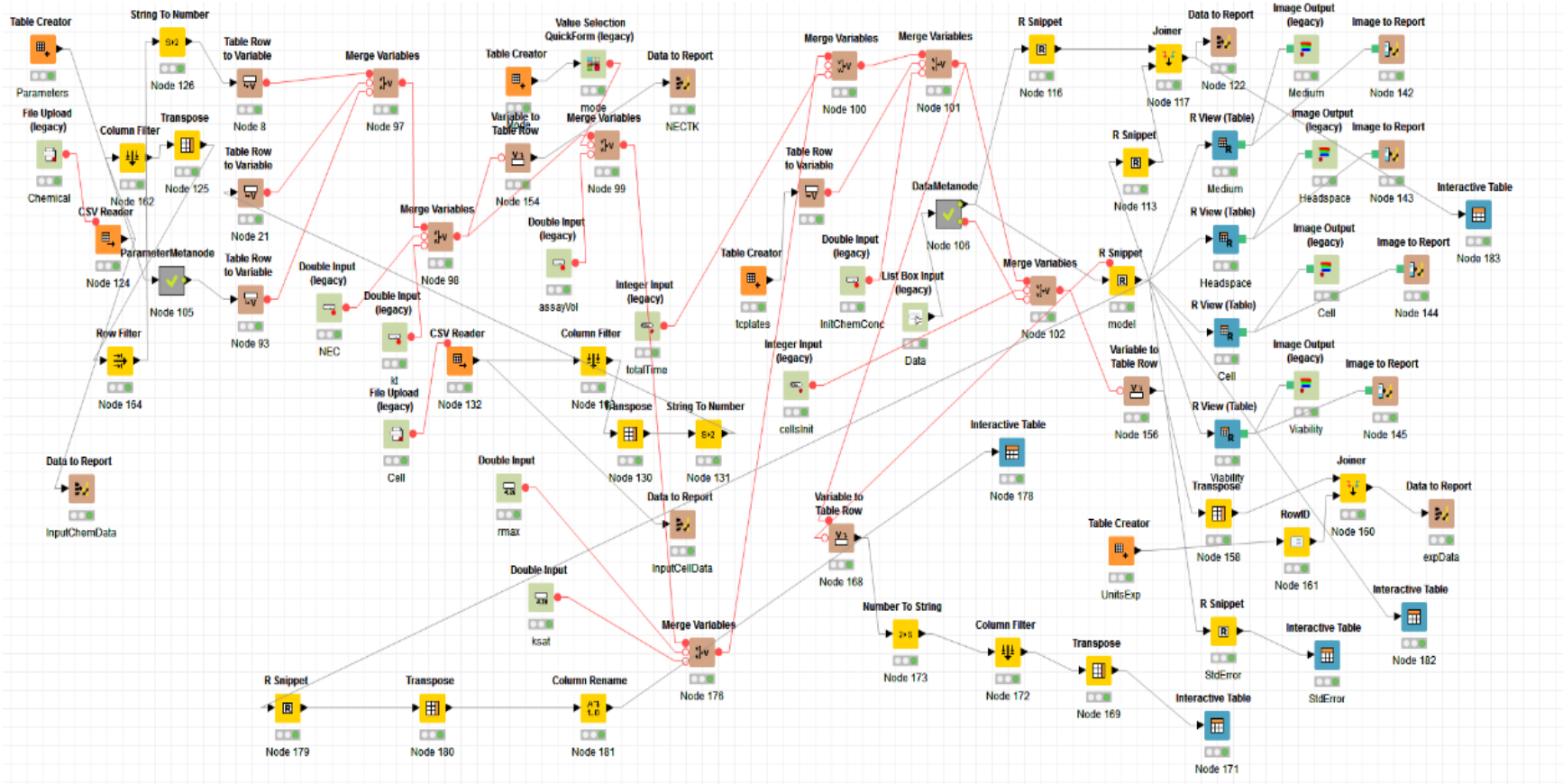

Figure 7 - KNIME workflow for the current VCBA where three zones are defined: input, model (core), and output.

3.7. Sensitivity analysis

Local sensitivity analysis was made for the impact of the $\log K_{ow}$, MV, SVcomp, Kr and HLC parameters on the dissolved concentration (M) for different compounds (caffeine, benzene, xylene, ochratoxin A and dimethyldioctadecylammonium chloride) with the chemical IC_{50} calculated from the concentration-response curves and NEC and Kr calculated for the initial input parameters.

For each parameter its original value was changed to $\pm$ 10 %, maintaining other parameters constant[40]. The normalised sensitivity coefficient (SC) was calculated using the equation:

$$SC = \frac{D' - D}{P' - P} \times \frac{P}{D}$$

where D is the initial outcome of the model, which in this case is the dissolved concentration (M) and D is the output of the model after the 10% parameter change. P is the initial parameter value, and P' is the parameter value modified by an increase/decrease of 10%. The sensitivity analysis was conducted for 5 % serum, 48 hours and the previously obtained respective IC_{50}.

3.8. Statistical analysis

GraphPad Prism 3.0 (San Diego, CA) was used for plotting and analyzing the data, except for Figure 6 which was made directly in R console.

4. Results and discussion

For 35 out of the initial 178 compounds, we could not find either the HLC or the air and water half-lives, and for other 9 compounds (di-isodecyl phthalate, tris(nonylphenyl)phosphite, hexachlorobenzene, 2-ethylhexyl acrylate, 1,2-dichlorobenzene, 1,1,1-trichloroethane, aconitine, 1,2-benzenedicarboxylic acid, malononitrile) the concentration-response curves were considered not to be suitable to optimize the VCBA. It is notable that 5 of these compounds (1,1,1-trichloroethane HLC=1740, 1,2-dichlorobenzene HLC=195, hexachlorobenzene HLC=131, tris(nonylphenyl)phosphite HLC=66.2, 2-ethylhexyl acrylate 43.8) are among the 11 compounds with the highest HLC values and, thus, a prevalence of evaporation might have a role in the difficulties found in the *in vitro* assays.
Therefore 83 compounds were optimized and run using the VCBA. The optimization was done after harmonization of the VCBA code, and was performed by applying the available *in vitro* concentration response curves. The VCBA values that were optimized (NEC and Kr) are presented in annex II. In Figure 8 it is exemplified for some compounds how the VCBA with the optimized parameters can modulate concentrations throughout the time in culture and its effect on cells growth/death.

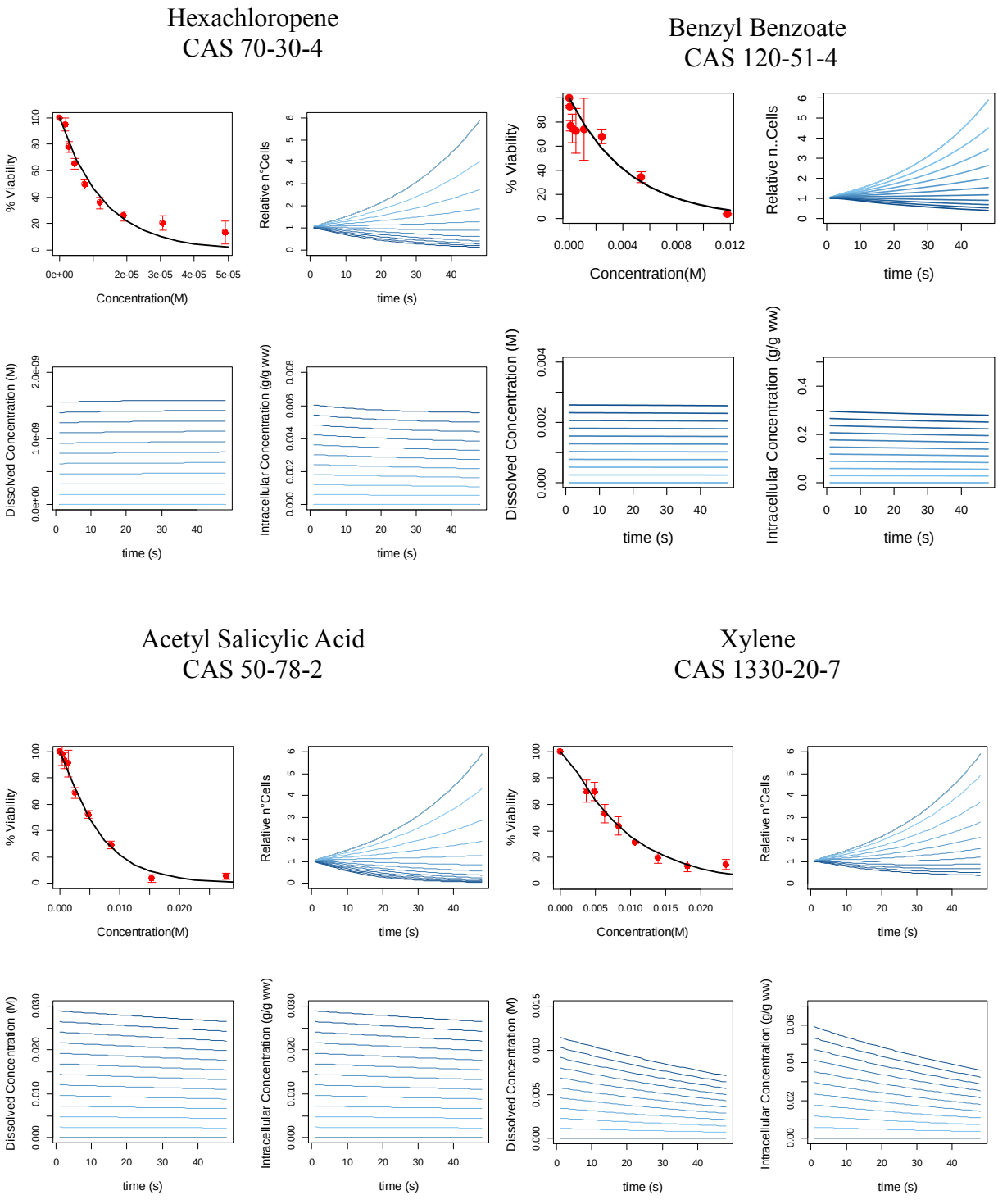

Figure 8-Plots for chemical compounds concentration and effect in cell number with the optimized NEC and Kr: Percentage of viability at 48 hours with the nominal concentration: black line-prediction; red dots-experimental values; Relative number of cells, Dissolved Concentration and Intracellular Concentration throughout the 48 hours of the culture for at least 10 concentrations of each compound, represented with a gradient blue color from the lower concentrations=light color to higher concentrations=darker color (Hexachlorophene 5×10^{-6} to 5×10^{-5} (M), Benzyl Benzoate 0.0012 to 0.012 (M), Acetyl Salicylic Acid : 0.0028 to 0.028 (M) and Xylene: 0.0026 to 0.026 (M))

4.1. Influence of $logK_{ow}$ and HLC in chemical partitioning

The $logK_{ow}$ has a high impact on description of the chemical partitioning with lipid, plastic, protein and cellular uptake. Therefore, is expected that the obtained chemical compounds partition among these several compartments has a strong correlation with $logK_{ow}$. Indeed, sorting the chemical compounds from the highest to the lowest $logK_{ow}$ and plotting their partition in Figure 9 corroborates the strong influence $logK_{ow}$ has on the chemical fate partition. Even though the model requires other parameters which widely vary among the compounds, the lipophilicity, here represented by the $logK_{ow}$, seems determinant: for the 14 chemicals with $logK_{ow}$ higher or equal to 4.46 (phenantrene has $logK_{ow}$ of 4.46) less than 10% of the chemical was dissolved in the medium. Still, for different $logK_{ow}$ ranges, different compartments of the *in vitro* system will be the strongest sequesters of the chemical. Observing from bottom to top of Figure 9 from $logK_{ow}$ 1.88 (triethylene glycol dimethacrylate) the dissolved concentration starts to decrease in detriment to protein bound chemical until $logK_{ow}$ 3.83 (endosulfan) in which lipid gradually binds to more chemical, eventually sequestering chemical even from protein binding. Although there is a significant binding to plastic it is never higher than 18 %

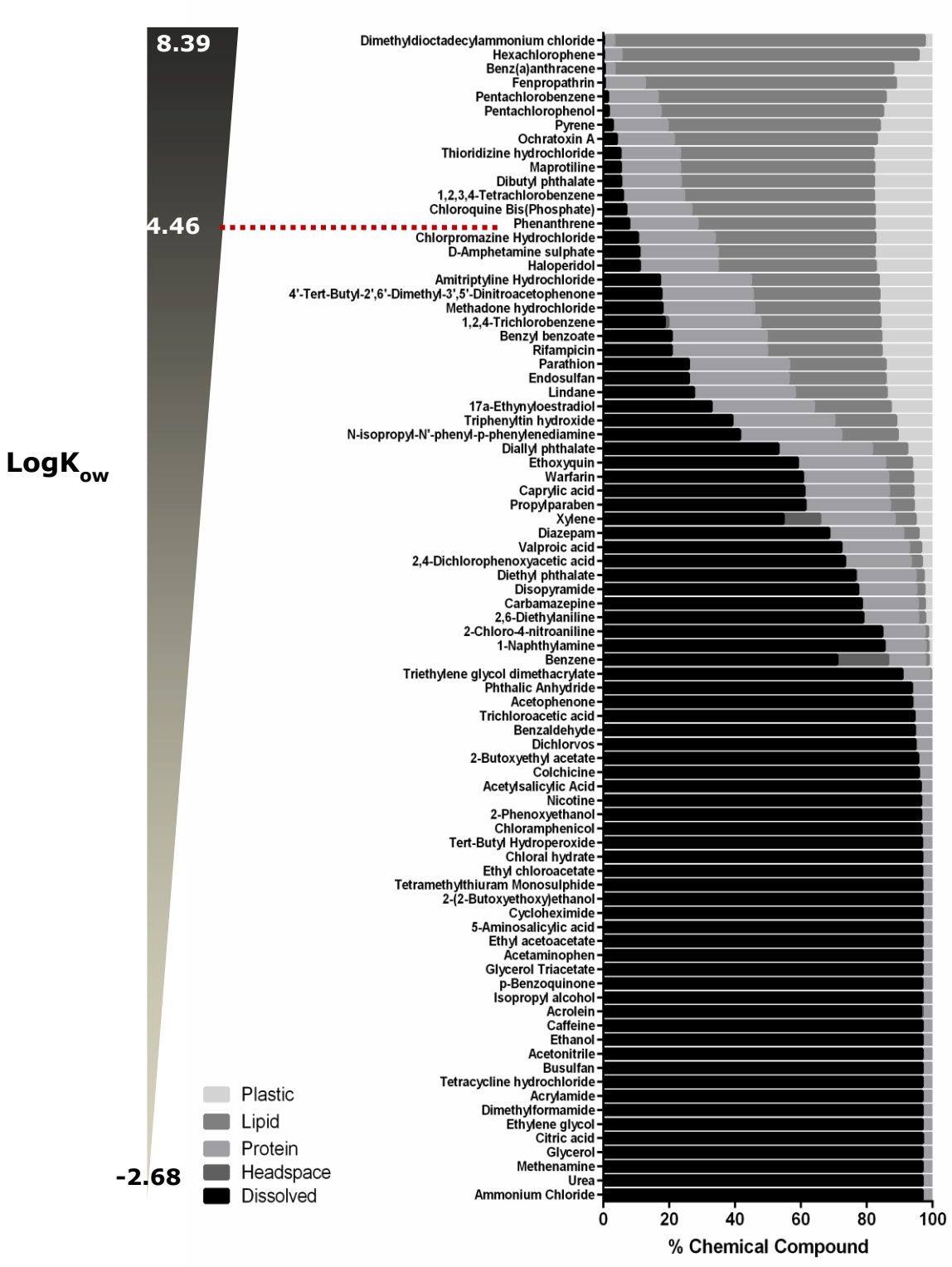

Figure 9- Partition in the several components of the in vitro assay for the 83 compounds optimized and run in the VCBA model. Represented in % to total final concentration

As for the evaporation, the VCBA model indicates that just for the two chemical compounds with higher HLC, benzene (HLC= 562 Pa×m^3/mole) and xylene

(HLC=525 Pa×m^3/mole), the percentage of compound that evaporates is higher than 10 %. For 1,2,4-trichlorobenzene that has a HLC of 144 Pa×m^3/mole, just 1% evaporates.

Although chemical partition has a very high impact on its free dissolved concentration after 48 hours, attention should also be paid to degradation hallmark which is considered in the VCBA model. Herein, the difference between total initial nominal concentration and total final concentration was considered as an indication of degradation of chemicals. Degradation and its descriptors are shown in Table 8. Comparing the final concentration with the initial one of each chemical, xylene stands out with the most degradation as its final concentration is approximately half of the initial one. Benzene follows with 30 % degradation. Both chemicals do not have relatively high degradation rates. But in general air degradation rate is higher than water, hence the higher the percentage of chemical in the headspace the fastest the degradation of the chemical, highlighting the importance of the HLC as an indicator of chemical kinetics. All remaining chemicals were not predicted to have more than 15 % degradation. It is noteworthy that overall degradation of a chemical does not depend solely of specific degradation rates but also on its partition. Comparing the ratio of total final concentration/initial concentration with air degradation there is no correlation while for water there is a slight trend (R^2=0.37). Still, these degradation rates will only affect chemicals that are either in the aqueous or gas phase. Chemical bound to protein, lipid or plastic is not considered in the VCBA to be protected from degradation. Therefore, comparing the total final concentration/initial concentration with the percentage of chemical in headspace and dissolved in medium, there is a slight inverted correlation (R^2=3.22). By analyzing all these parameters together (Table 8) with a colour scale pattern it can better perceived how degradation is a multifactorial event.

Table 8 – Fraction of chemical degraded ([Final]/[Initial] and the parameters describing it, Air and water degradation rate (s-1) and amount of unbound chemical, meaning chemical available for degradation at 48 hours. The colour coding is a gradient from the lower values coloured red to the highest values coloured green relative to each column/parameter.

	[Final]/[Initial]	Air Deg. rate s-1	Water Deg. rate (s-1)	Non-binding % (Headspace+ Dissolved)
Xylene	0.52	1.03E-05	5.35E-07	73.91
Benzene	0.74	9.21E-07	2.14E-07	102.11
Diethyl phthalate	0.85	2.25E-05	9.26E-07	77.01
Ethanol	0.86	2.45E-06	9.26E-07	96.90
Citric acid	0.86	5.26E-06	9.26E-07	96.88
Ethylene glycol	0.86	5.78E-06	9.26E-07	96.88
Glycerol	0.86	1.41E-05	9.26E-07	96.88
2-(2-Butoxyethoxy)ethanol	0.86	5.58E-05	9.26E-07	96.84
2-Butoxyethyl acetate	0.86	1.59E-05	9.26E-07	95.52
Acrolein	0.88	1.52E-05	5.35E-07	97.34
p-Benzoquinone	0.91	5.91E-06	5.35E-07	97.05
Caprylic acid	0.91	6.27E-06	9.26E-07	60.93
Benzaldehyde	0.91	9.68E-06	5.35E-07	94.64
Ethyl chloroacetate	0.91	9.08E-07	5.35E-07	96.93
Tert-Butyl Hydroperoxide	0.91	2.25E-06	5.35E-07	96.78
Dimethylformamide	0.91	1.32E-05	5.35E-07	96.90
Isopropyl alcohol	0.91	3.81E-06	5.35E-07	96.90
Acetonitrile	0.91	1.25E-07	5.35E-07	96.92
Ethyl acetoacetate	0.91	1.48E-06	5.35E-07	96.85
Ammonium Chloride	0.91	1.93E-09	5.35E-07	96.88
Caffeine	0.91	1.46E-05	5.35E-07	96.88
Glycerol Triacetate	0.91	6.38E-06	5.35E-07	96.86
Acrylamide	0.91	9.63E-06	5.35E-07	96.88
Acetaminophen	0.91	1.33E-05	5.35E-07	96.85
Urea	0.91	1.50E-06	5.35E-07	96.89
5-Aminosalicylic acid	0.91	1.57E-05	5.35E-07	96.85
2-Phenoxyethanol	0.91	2.45E-05	5.35E-07	96.40
Acetylsalicylic Acid	0.91	9.82E-07	5.35E-07	96.26
Acetophenone	0.92	2.05E-06	5.35E-07	93.80
Phthalic Anhydride	0.92	5.61E-07	5.35E-07	93.57
Triethylene glycol dimethacrylate	0.92	6.69E-05	5.35E-07	90.72
Tetramethylthiuram Monosulphide	0.93	1.04E-04	2.14E-07	96.86
Valproic acid	0.94	6.13E-06	5.35E-07	72.16
Propylparaben	0.95	1.06E-05	5.35E-07	61.34
Diallyl phthalate	0.95	5.85E-05	5.35E-07	53.09
Cycloheximide	0.96	4.91E-05	2.14E-07	96.84
Busulfan	0.96	3.53E-06	2.14E-07	96.88
Chloral hydrate	0.97	1.44E-06	2.14E-07	96.76
Nicotine	0.97	6.83E-05	2.14E-07	96.35
Dichlorvos	0.97	7.08E-06	2.14E-07	94.73
Trichloroacetic acid	0.97	3.90E-07	2.14E-07	94.36
1-Naphthylamine	0.97	1.50E-04	2.14E-07	85.30
2-Chloro-4-nitroaniline	0.97	2.94E-06	2.14E-07	84.58
2,6-Diethylaniline	0.97	1.22E-04	2.14E-07	78.85
Carbamazepine	0.97	2.37E-04	2.14E-07	78.33
2,4-Dichlorophenoxyacetic acid	0.97	4.98E-06	2.14E-07	73.25
Diazepam	0.98	7.43E-06	2.14E-07	68.51
Chloramphenicol	0.98	2.32E-05	1.34E-07	96.50
Methenamine	0.98	3.81E-04	1.34E-07	96.88
Tetracycline hydrochloride	0.98	1.54E-04	1.34E-07	96.88
Colchicine	0.98	6.98E-04	1.34E-07	95.67
Warfarin	0.98	1.36E-04	2.14E-07	60.52
Ethoxyquin	0.98	1.93E-04	2.14E-07	58.99
1,2,4-Trichlorobenzene	0.98	4.12E-07	1.34E-07	19.82
Benzyl benzoate	0.99	5.23E-06	5.35E-07	20.52
N-isopropyl-N'-phenyl-p-phenylenediamine	0.99	1.63E-04	2.14E-07	41.31
Triphenyltin hydroxide	0.99	4.49E-06	2.14E-07	39.05
Disopyramide	0.99	9.39E-05	4.46E-08	77.32
Parathion	0.99	6.90E-05	2.14E-07	25.84
17α-Ethynyloestradiol	1.00	9.39E-05	1.34E-07	32.72
Dibutyl phthalate	1.00	6.95E-06	9.26E-07	5.32
Amitriptyline Hydrochloride	1.00	1.23E-03	2.14E-07	17.06
Endosulfan	1.00	6.80E-06	4.46E-08	25.85
1,2,3,4-Tetrachlorobenzene	1.00	6.17E-08	1.34E-07	5.96
D-Amphetamine sulphate	1.00	4.25E-05	2.14E-07	10.81
Thioridizine hydrochloride	1.00	1.93E-04	1.34E-07	5.06
Methadone hydrochloride	1.00	2.94E-05	1.34E-07	17.87
4'-Tert-Butyl-2',6'-Dimethyl-3',5'-Dinitroacetophenone	1.00	1.19E-06	1.34E-07	17.50
Pentachlorophenol	1.00	4.13E-07	4.46E-08	1.52
Lindane	1.00	1.05E-07	4.46E-08	27.45
Maprotiline	1.00	7.05E-05	2.14E-07	5.19
Rifampicin	1.00	6.59E-04	4.46E-08	20.59
Phenanthrene	1.01	9.77E-06	1.34E-07	7.77
Chlorpromazine Hydrochloride	1.01	1.34E-04	1.34E-07	10.39
Ochratoxin A	1.01	1.87E-05	2.14E-07	4.00
Haloperidol	1.01	8.71E-05	4.46E-08	10.89
Chloroquine Bis(Phosphate)	1.01	6.42E-05	4.46E-08	6.83
Pyrene	1.01	3.75E-05	1.34E-07	2.77
Pentachlorobenzene	1.01	4.34E-08	4.46E-08	1.31
Benz(a)anthracene	1.01	3.75E-05	1.34E-07	0.29
Fenpropathrin	1.01	1.34E-05	1.34E-07	0.31
Hexachlorophene	1.01	1.63E-06	4.46E-08	0.00
Dimethyldioctadecylammonium chloride	1.02	5.18E-05	2.14E-07	0.00

4.2. Influence of the experimental set up on chemical partitioning

With external lipid and protein playing such a relevant role in chemicals partition, we further explored how different percentage of supplemented serum could influence the partition of compounds. Sixteen compounds were selected and simulations run with 0, 5, and 10% of serum by changing the initial protein and lipid concentrations. The compounds were chosen to spread across the range of $logK_{ow}$ (from dimethyldioctadecylammonium chloride with 8.392 to ammonium chloride with -2.68) including the two compounds with significant evaporation (benzene and xylene).

Figure 10, indicates substantial differences between 0% and the other two plots. No chemical fraction is bound to protein under 0% serum, and although there is an increase in the dissolved concentration, most is halted by lipid binding. In the absence of serum, binding to lipids is high despite the fact that proteins and lipids come solely from exudes from cell death. In the absence of serum also it has been described the relevance of plastic binding[23]. Observing the equations that describe these partitions, the distribution of chemical in the several elements is easily perceived:

Plastic partition constant: $Kp=10^{(0.97*logkow-6.94)}$

Protein partition constant : $Ks<-10^{(vals-1.178)}$

$$vals=\begin{cases} -1.31, if\ logKow < 1.09 \\ 0.57 \times logKow + 0.69, if\ 1.09 \leq logKow\ \leq 4.6 \\ logKow - 1.3, if\ logKow > 4.6 \end{cases}$$

Lipid partition constant : $Kl<-10^{(1.25*logkow-3.70)}$

$LogK_{ow}$ has a higher impact on lipid partition than on plastic for all range of values while for proteins the relation $logK_{ow}$ protein-binding partition depends on the $logK_{ow}$ values. For $logK_{ow}$ values lower than 1.09 and higher than 3.7 proteins exhibit the highest chemical binding. Partition also depends on the concentration of lipids and proteins and the surface area of the plastic. It should be noticed that although in the absence of serum the lipids in medium after 48 hours would be one eight of the lipid content in 5 % serum supplementation, it still has a high binding ability as observed in the Figure 8. Indeed, two facts contribute for the 0 % serum plot to be far from reality. Firstly, actually at 0 % serum most cell lines do not grow and thus the lipids in the medium would even be a smaller fraction. Secondly, saturation is herein not being included. Experimentally, plastic partition does seem to change with nominal concentration[26]. As for lipids, saturation is also expected although it would be a difficult parameter to modulate as the interactions of lipids with chemicals occurs more in the form of aggregates than individually[41].

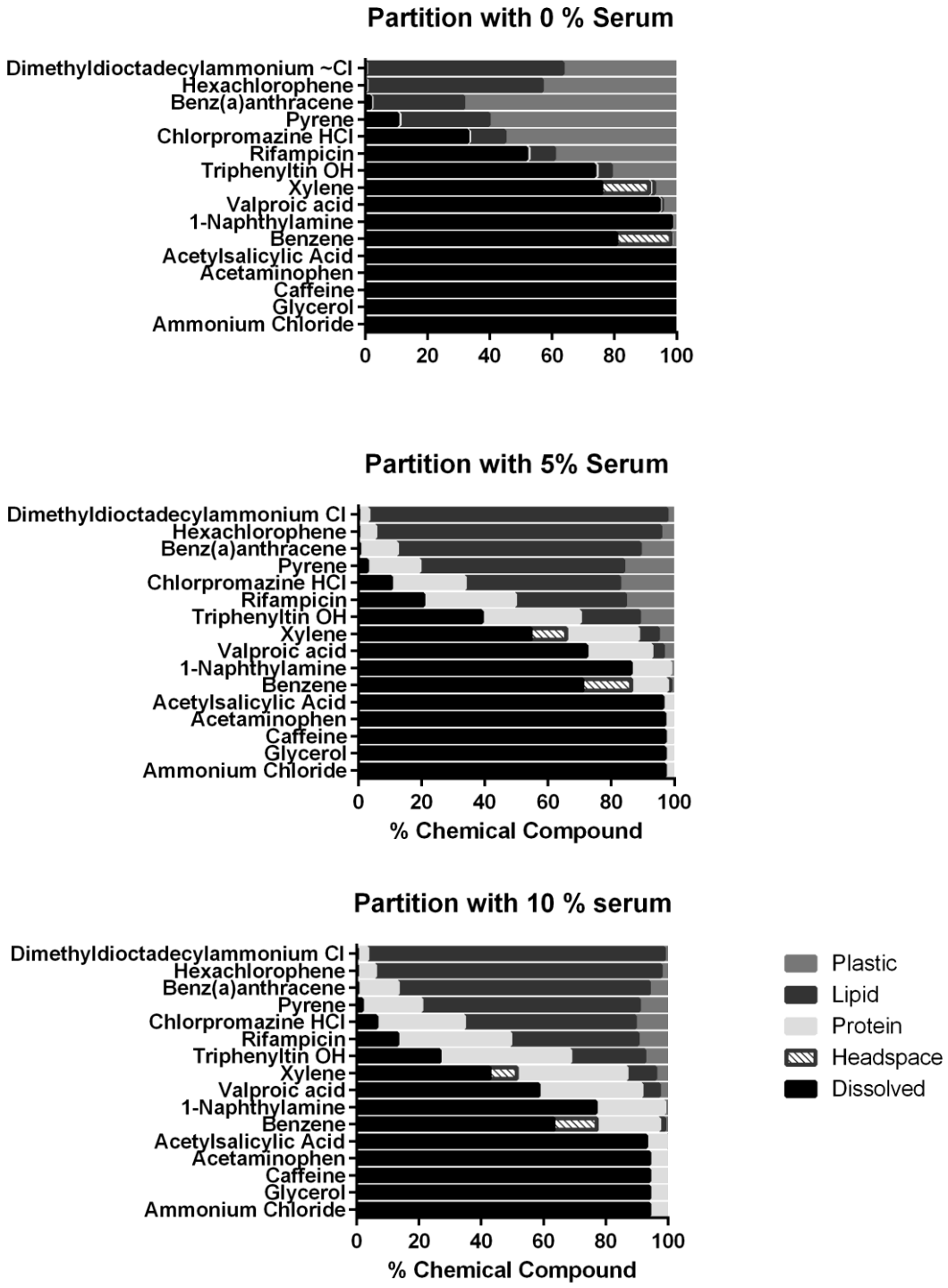

Figure 10- Partition in the several components of the in vitro assay for the 16 compounds optimized and run in the VCBA model with 0, 5 and 10 % of supplemented serum (FBS) for 48 hours with the previously optimized NEC and Kr.

Saturation modelling could furthermore increase the differences among the several percentages of serum supplementation. Although the lipid binding extent is something that should alert for a careful consideration of the chemical's kinetics, a high lipid binding ability might also indicate a mode of toxicity[24] based on narcosis which can also occur by external cell effects, without cell uptake.

Although VCBA simulates for very highly lipophilic chemicals very low dissolved concentration, uptake and bioaccumulation also depend on lipophilicity and, therefore, compensates to some extent for the lack of bioavailable chemical.

Indeed, a comparison of the intracellular concentration with the dissolved concentration shows lack of correlation (Figure 11). In fact, segmenting the chemicals between the ones with $logK_{ow}$ lower than 2.5 and the ones higher (red) resulted in a much more significant correlation for the ones with lower $logK_{ow}$ values.

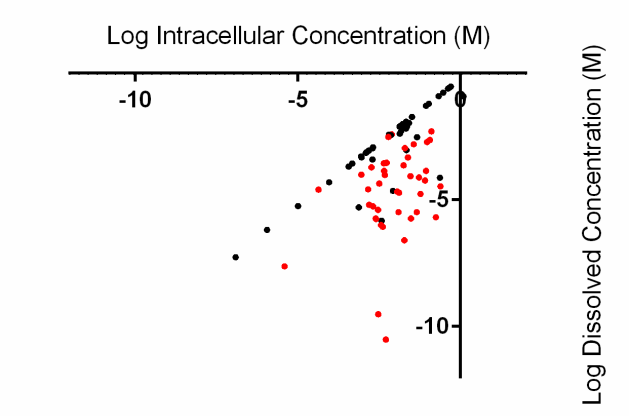

Figure 11- Comparison of the logarithms of intracellular concentration and dissolved one for the 83 simulated chemicals with 5 % serum supplementation. Red dots correspond to the chemicals with $logK_{ow}$ higher than 2.5.

Although initially serum seems to have a major role in sequestering chemicals from the dissolved phase, lipids originated in cells and plastic act as the major sequesters in the absence of serum. Also for the two compounds for which evaporation was predicted by VCBA, xylene and benzene, evaporation increases in the absence of serum which is on agreement with Kramer N. *et al* 2009[18], which experimentally verified that serum retained phenantrene in solution.

In the same article, Kramer *et al* also report a possible deficiency in the model in capturing evaporation, for it uses as a proxy of evaporation the HLC which is obtained/predicted for 20-25 °C, while experiments often use higher temperatures such as 37 °C. And HLC, being dependent of the chemical solubility and vapour pressure, does increase with temperature. Ten Hulscher[42] reported that for temperatures increases from 20 to 30 °C, HLC increased ~50 % for chlorobenzene, chlorobiphenyls and polycyclic aromatic hydrocarbons.

Evaporation of phenantrene and 1,2,4-trichlorobenzene was experimentally verified[18], while our model only points to 0.01%, 1.05 % of the respective chemical in the headspace (at 5 % supplemented serum).

Stadnicka-Michalak J. *et al* [43]al also verified substantial evaporation from 1,2,3 trichlorobenzene, naphthalene and hexachlorobenzene.

Similarly, several *reports* [44,45] have indicated that compounds with 1 $Pa.m^3$ /mol may be already prone to evaporate, while with 100 $Pa.m^3$ /mol (which would include 1,2,4-Trichlorobenzene) are even considered highly volatile. Therefore, for the compounds 1,2,3,4-tetrachlorobenzene, pentachlorobenzene, diethyl phthalate, p-benzoquinone, endosulfan, phenanthrene, ethyl chloroacetate, acetonitrile, benzaldehyde, tetramethylthiuram monosulphide, tert-Butyl hydroperoxide, benz(a)anthracene, pyrene, acetophenone we should have obtained higher percentages in the headspace while, in fact, none was modelled to evaporate more than 0.2%.

4.3. Proposal of $LogK_{ow}$ thresholds to rank chemicals

One of the purposes of this work was to possibly create a system that alerts for chemicals for which the dissolved concentration is expected to differ substantially from the nominal one, hindering an accurate extrapolation to *in vivo*. Moreover, indications of the chemical fates *in vivo* can hopefully help refining experiments depending of the chemical properties. Therefore, to put in practice this alert system, the results described in Figure 8 were used to create a system with 4 classes (A, B, C, D) based on $logK_{ow}$ thresholds for conditions of 10 % serum.

These classes and respective thresholds are represented in Figure 12.

Chemicals with $logK_{ow}$ below 2 are predicted by the VCBA to be at least 90% dissolved, constituting class D. Class C is composed by chemicals with $logK_{ow}$ between 2-2.5 which are predicted to be 80% dissolved. Compounds with a logKow values up to 5 (class B) have dissolved concentrations in water from 5 to 80%, showing a very strong correlation with $logK_{ow}$ and eventually shifting the main chemical partition from proteins to lipids. Finally, for a $logK_{ow}$ value between 5 and 10 (class A), most of the chemicals will be bound to lipids, migrated to plastic or bound to proteins.

With the alerting system created we carried out an additional exercise to estimate the partition of chemicals not previously analyzed here in the VCBA, relying only on the $logK_{ow}$ value. The chemicals chosen are a set that will be in the future tested for toxicity in an *in vitro* test system supplemented with 10 % serum. Hopefully, the information herein gained will help exploring the results of this new project. Based on Figure 12 we collected the $logK_{ow}$ for these additional 35 chemicals, and we ranked them based on their $logK_{ow}$ (Table 9).

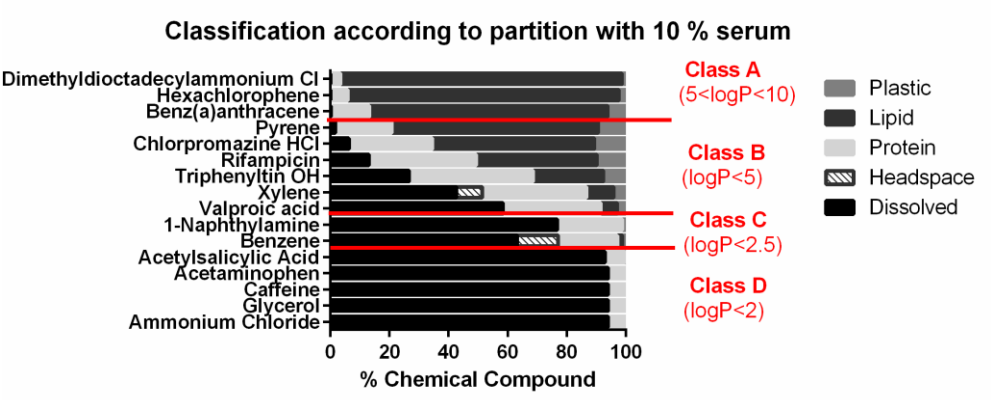

Figure 12- Proposed classes for chemical fate, delimited by thresholds based on logK_{ow}.

Table 9- List of chemicals, their respective logK_{ow} and class according to alert system for partitioning.

	Chemical Name	LogP	VCBA_cutoff
1	Bis(2-ethylhexyl) terephthalate	9.55	A There could be a potential impact on the fate of the chemical to the well / serum / experimental set up
2	Isopropyl palmitate	8.49	A
3	Benzo(a)pyrene	6.4	A
4	(-)-Ambroxide	5.4	A
5	Tween 80	5.37	A
6	Piperonyl butoxide	4.75	B
7	Rotenone	4.65	B
8	(-)Bronyl acetate	4.3	B
9	Sodium diclofenac	4.26	B
10	17-B-oestradiol	4.01	B
11	2-Nitrofluorene	3.89	B
12	Carbazole	3.72	B
13	Disulfiram	3.67	B
14	1,4-Dichlorobenzene	3.44	B
15	Dimethyl 2,6-naphthalenedicarboxylate	3.4	B
16	Cyclosporin A	3.35	B
17	Ethyl cinnamate	2.99	B
18	Valproic acid	2.75	B
19	Rifampicin	2.7	B
20	Benzyl propanoate	2.46	C
21	(+/-)-alpha-Methylbenzyl acetate	2.28	C
22	1-Chloro-2,4-dinitrobenzene (DNCB)	2.06	C
23	Clonidine hydrochloride	1.41	D more than 90%should go to the water fraction
24	Vanillin	1.21	D
25	Alosetron hydrochloride	0.88	D
26	2,5-Diaminotoluene sulphate (PTD)	0.74	D
27	Hydroquinone	0.59	D
28	Acesulfame potassium	0.47	D
29	Acetominophen	0.46	D
30	Aflatoxin B1	0.45	D
31	Formaldehyde	0.35	D
32	4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanone	0.09	D
33	1,4-Phenylenediamine	-0.3	D
34	Glutaraldehyde	-0.34	D
35	Allura Red C.I.16035	-0.55	D

Pay more attention to the experimental set up

This resulted into 5 chemicals being of concern, thus we further investigated these chemicals by means of VCBA simulations, firstly collecting the other physical-chemical properties for these 5 chemicals. No cytotoxicity data were used and, therefore, the model was run with NEC and Kr set to 0, no cell death occurred and the impact it has on concentrations of the chemicals was tested. The initial concentrations used for all these chemicals were 0.1 and 100 μM. Both concentrations induced the same % of partition. The model was run for 24 and 72 hours but not substantial differences were found on the outputs.

The partition of the chemicals belonging to Class A is represented in Figure 13. Indeed, most of the amount of chemicals is bound to lipids, proteins and/or plastic. For chemicals in Class B, although dissolved fraction varied greatly it was always less than half of the chemical total concentration after the 48 hours assay

For these compounds special care should be taken when manipulating it in plastic eppendorfs in the laboratory, and with supplementation that should be done with the exact same batch of serum.

Possible toxic mechanisms for these chemicals should be identified to find to what extent the dissolved concentration is responsible for cytotoxicity or if it is related to their lipid binding ability, destabilizing cellular membranes.

Furthermore, for these chemicals it is advised to use used other dosimetry than the nominal concentration. As indicated here, nominal concentration does not represent the concentration the cells are indeed exposed to. This is especially important in the eventuality that the results generated *in vitro* with these compounds need to be extrapolated to *in vivo* doses. In such as case, the partition data shown here should be taken into account.

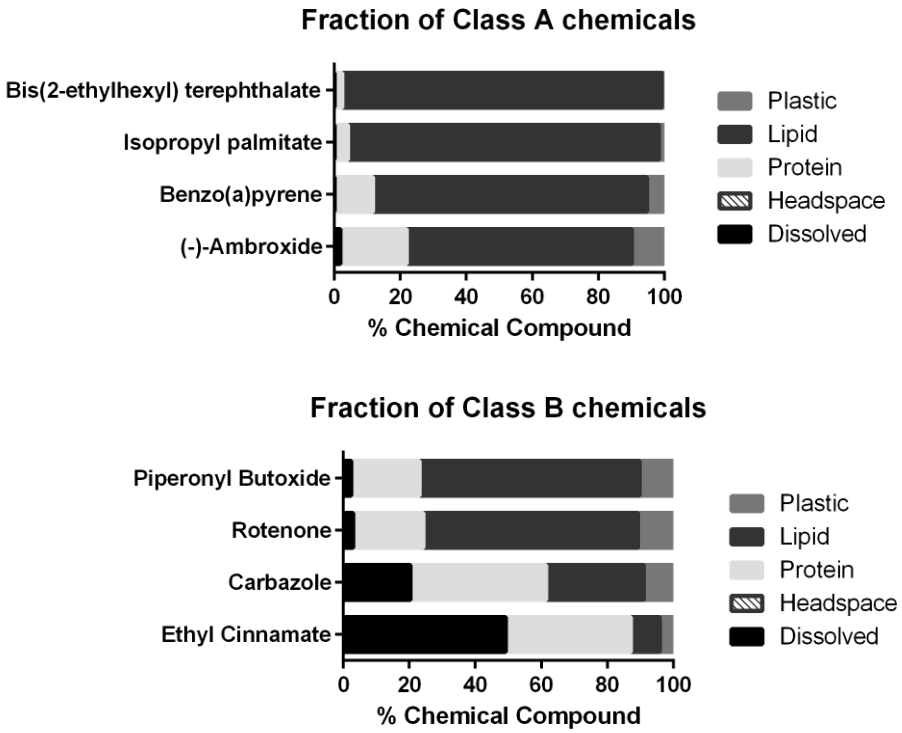

Figure 13- Partition of the Class A and Class B chemicals through the several in vitro components in medium supplemented with 10 % serum and not accounting for cell death. Results from running the chemical at 0.1 μM for a 24 hours experiment.

4.4. Impact of chemical partitioning on toxicity prediction

The chemicals used were primarily tested in BALB/c 3T3 NRU *in vitro* assays to further evaluate the predictive power of the *in vitro* method, by comparing them to the acute rat oral *in vivo* test (mean LD_{50} values). This comparison was made through the EU CLP system for acute oral toxicity, which is based on the cut-off of 2000 mg/kg to assign chemicals to the classified (mean LD_{50}< 2000 mg/kg) or non-classified group (mean LD_{50}> 2000 mg/kg). *In vitro* IC_{50} were converted to LD_{50} and its resulting classification was compared to the *in vivo* rat oral classification (also based on mean LD_{50}). The *in vitro* correctly classified chemicals were designated True (Positive and Negative), while the incorrectly classified were called False Positive (i.e. LD_{50} predicted *in vitro* ≤ 2000 mg/kg while *in vivo* the observed value was > 2000 mg/kg) or Negative (i.e. LD_{50} predicted *in vitro* > 2000 mg/kg while *in vivo* the observed value was ≤ 2000 mg/kg). With the premise that the partition results would help understanding these false predictions of this binary classification system, and considering the high impact that $logK_{ow}$ has on partitioning, it would be expected that the compounds which toxic class was falsely predicted, would have $logK_{ow}$ values falling onto specific ranges.

However as shown in Figure 14 this was not observed. The distribution of values of $logK_{ow}$ of the compounds predicted as false positives and false negatives is not significantly different from the compounds with true predictions.

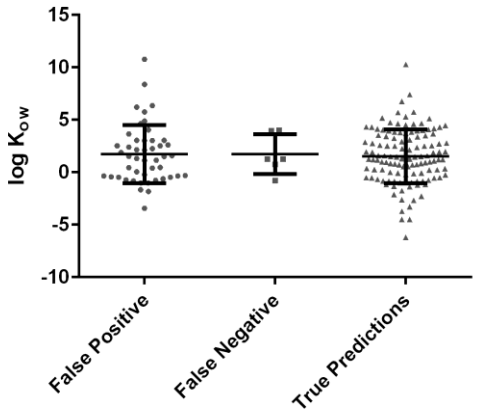

Figure 14- Distribution of $logK_{ow}$ of the chemical compounds which toxicity was correctly (True Predictions) or not correctly classified, either because in vitro indicated toxicity only at concentrations > 2000 mg/kg while in vivo studies shown toxic effect at lower concentrations (False Negative) or the opposite (False Positive).

After converting the nominal and predicted dissolved IC_{50s} from molar concentration (M) to mg/L (mg/kg), they were plotted with the *in vivo* LD_{50}, Figure 15, to analyze which concentration would correlate best with the *in vivo* one. The following trendline equations were obtained:

$\log Nominal\ 50 = 0.7625 \times LD50 - 0.1983$, R^2=0.31

$\log Dissolved\ 50 = 0.8734 \times LD50 - 1.076$, R^2=0.20

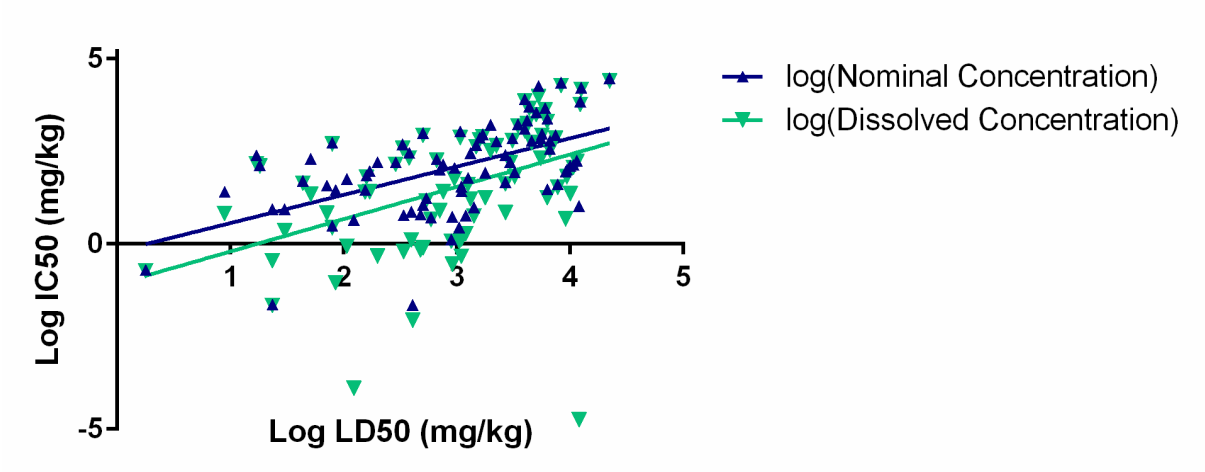

Figure 15- Comparison of the in vivo LD_{50} (mg/kg) with the in vitro nominal and the predicted dissolved IC_{50} (mg/kg).

The logarithm of the *in vivo* rat oral LD_{50} values correlated with the logarithm of dissolved IC_{50} values has a lower coefficient of determination than with the nominal IC_{50} values. The partitioning and evaporation are events that have been broadly described *in vitro*, hence using the nominal or the actual bioavailable concentration for *in vitro-in vivo* extrapolations does make a difference. However, besides the uncertainties of the VCBA model, the correct way of integrating the predicted concentrations in this extrapolation is still being studied, e.g. can we assume that the dissolved concentration reflects the bioavailable one? Furthermore, the correlations should be interpreted in the light of the inherent high variability of the LD_{50} data, as shown by Hoffmann et al (2010). Moreover, the *in vivo* dose here being used is the nominal oral dose and chemical kinetics *in vivo* comprise more hallmarks such as metabolism and absorption through the gastrointestinal tract, which are being ignored in these comparisons. The cell line used in this assays indeed has no metabolic ability so it will not predict correctly the compounds that might be bio-activated or more easily cleared/excreted through metabolism. It is noteworthy that also the

higher protein-binding ability of lipophilic compound includes higher affinity with xenobiotic-metabolism enzymes. Thus these compounds tend to be more metabolized than polar compounds[46] which further explain the lack of trend on false predictions and high $logK_{ow}$. However, to make a more accurate comparison of free unbound blood plasma concentrations *in vivo* and dissolved *in vitro* concentrations it would require PBK modelling all the 83 compounds, a task which extends beyond the scope of this specific work but hopefully will be explored in the future.

As part of the ACuteTox project, kinetics transformations were used to estimate the oral dose from the nominal concentrations obtained *in vitro* (IC_{50}) using a set of algorithms that took into account lipophilicity, metabolic clearance and protein binding and intestinal permeability using Caco-2 cells (http://www.acutetox.eu/WP5.pdf). The calculations were only possible for a limited set of compounds for which the kinetic input data were obtained and not clear conclusions were drawn other than recommending further evaluation (Prieto et al., 2013a).

Efforts are still needed to prove that the VCBA simulations are relevant to predict acute oral toxicity for different regulatory contexts. In this context, an initial step could be to check experimentally the simulations obtained with the VCBA using a set of compounds. Table 10 shows the proposed compounds, chosen to represent a wide range of physicochemical properties and acute oral toxicity categories estimated *in vivo* and predicted *in vitro*.

Table 10 –Suggested chemical compounds to be tested *in vitro,* with their physical-chemical characteristics and the result of the comparison of the toxicity EU-CLP classification derived from the mean *in vitro* predicted and *in vivo* LD_{50}s.[1]

	MW	$logK_{ow}$	HLC	Acute oral toxicity prediction
Benzyl benzoate	212	3,97	$2,37\times10^{-02}$	FN
Isoniazid	137	-0,78	$1,23\times10^{-09}$	FN
Dimethyldioctadecylammonium chloride	586	8,39	$6,45\times10^{-03}$	FP
Hexachlorobenzene	285	5,73	$1,31\times10^{02}$	FP
Xylene	106	3,04	$5,25\times10^{02}$	FP
Ethanol	46	-0,31	$5,07\times10^{-01}$	FP
Benzene	78	2,13	$5,62\times10^{02}$	TN
Disulfoton	274	4,02	$2,19\times10^{-09}$	TP
Hexachlorophene	407	7,40	$5,55\times10^{-08}$	TP
Caffeine	194	-0,07	$3,63\times10^{-06}$	TP

[1] *In this classification system, chemicals' toxicity with LD50≤2000 mg/kg is considered classified and with LD50s> 2000 mg/kg, non-classified:

FN (False Negative) – in vitro toxicity was non-classified while in vivo it was classified.

FP (False Positive) – in vitro toxicity was classified while in vivo it was non-classified.

TN (True Negative) – both in vitro and in vivo were non-classified

TP (True Positive) – both in vitro and in vivo were classified. Still among classified toxicity, categories : Category 1- LD50≤5 mg/mL; Category 2- 5<LD50≤50 mg/kg; Category 3- 50<LD50≤300 mg/kg; Category 4- 300>LD50≤2000 mg/kg . Even in a chemical is considered TP, reflecting a correct prediction of classified toxicity, its category can be mispredicted; while Hexachlorophene category was correctly predicted, Caffeine was underpredicted for 1 category and Disulfuton for 3 categories

4.5. Sensitivity analysis of several input parameters

A local sensitivity analysis was performed to determine how influential some of the input parameters are on the dissolved concentration. In Figure 16 shows the influence of parameters Kr, HLC and $logK_{ow}$ on the output of the dissolved concentration.

As expected the $logK_{ow}$ is the parameter with higher impact in the dissolved concentration, having sensitivity coefficients higher than 0.1.

The higher the $LogK_{ow}$, the higher the absolute value of sensitivity coefficient, showing a high impact on 4 chemicals ($logK_{ow}$ =2.31-8.39) but not on Caffeine which has the lower $logK_{ow}$ (-0.07).

A negative sensitivity coefficient indicates inversed relation with the output, which indeed is true for $logK_{ow}$ as the higher lipophilicity, the less bioavailable the chemical is, e.g. for the dimethyldioctadecylammonium chloride a variation of 10 % on $logK_{ow}$ value induces a change of one decimal unit in the dissolved concentration. Likewise for HLC, the higher its value the more the chemical evaporates, hence less chemical is bioavailable.

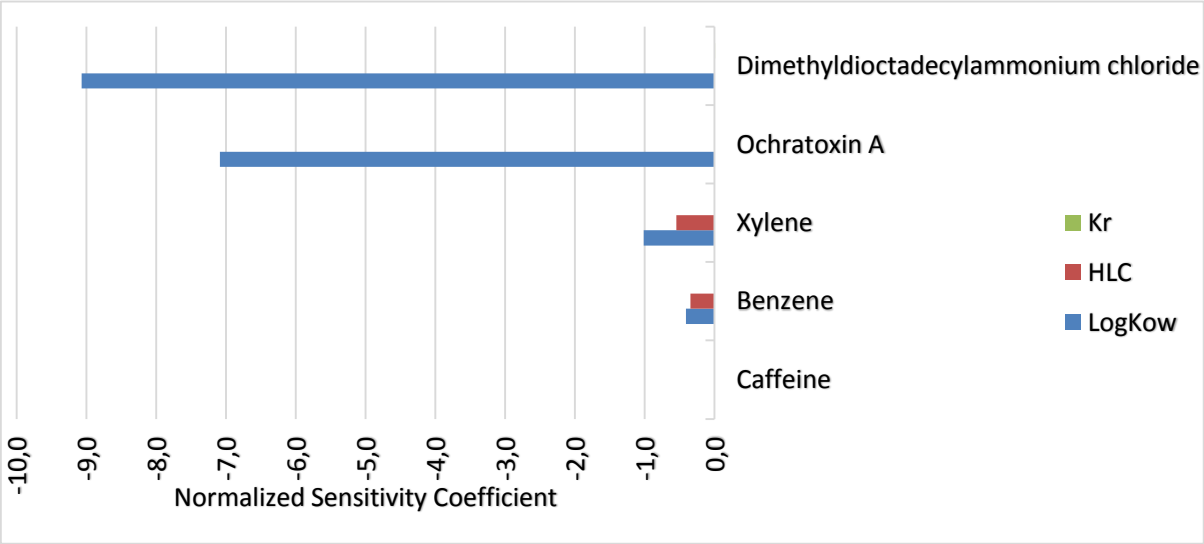

Figure 16- Normalized Sensitivity Coefficient of the dissolved concentration upon 10 % increase in input parameters: Kr, HLC and $logK_{ow}$ for 5 chemicals.

HLC has impact on the output on the dissolved concentration of the compounds herein previously shown to evaporate, benzene and xylene. This supports the claim that under a certain threshold of HLC, in which the chemical does not evaporate significantly, it makes no difference how low the value is.

A variation of 10% in the parameters of SVcomp and MV has no/very little impact on chemicals dissolved concentration, adding more confidence to the extrapolation method used to obtain these parameters from the Molecular Weight. Both these parameters are used in the equations of gas-liquid diffusion

which describes evaporation, hence theoretically these values will only have impact once HLC is high enough. However, this sensitivity analysis indicates that even in the case of compound with high evaporation, these parameters have very low or negligible impact on the output.

Killing rate has a detectable impact on the output but still it is quite low. This shows that although the optimization step is important, obtaining an absolute minimum is not essential as long as the value is closed to it. This is noteworthy as an analysis on the optimization process, which resulted in a 3D graph with the minimum error for several values of Kr and NEC, showed that in general VCBA minimum "lays in a very flat area" meaning that for a relative wide range of NEC and Kr the minimum error does not change significantly making it quite difficult to find the absolute minimum and thus the optimized Kr (graph not shown).

4.6. Assumptions and uncertainties

The VCBA is a model that allows analysis of any chemical as long as the physicochemical characteristics are obtained, additionally allowing incorporation of toxicological data (concentration-response curve). However, to be such a generic model, it was built based on several assumptions and for some groups of chemical compounds the uncertainty on the output may increase substantially.

Firstly, the QSARs used in VCBA to predict partition coefficients for lipid, serum, and plastic, were based mostly on logKow, not including other physico-chemical properties important for the substance fate, such as if the chemical is a H-donor/acceptor[47]. Besides, as shown below, these QSARs were derived from specific sets of chemicals.

For plastic, Kramer *et al*[26] analyzed 7 polycyclic aromatic hydrocarbons (PAHs) chemicals with $\log K_{ow}$ ranging 3.33 to 6.13 (and HLC from 0.034 to 45 Pa m3 mol^{-1}), measuring the binding constants at 1 % maximum water solubility while Jonker *et al* [48] analyzed 13 PAHs with $\log K_{ow}$ ranging approximately 4.5 to 7. Solely for the protein partition's QSAR, other chemicals than just PAHs were considered, as it was based on a 6 independent studies, including 36 chemicals with $\log K_{ow}$ ranging -1.3 to 5.1. PAHs are a family of neutral non-polar chemicals, hence it is not clear to what extent can we rely on VCBA outputs for chemicals with different characteristics. As an example, we considered protein binding as a non-specific interaction which is not necessarily true for polar, charged and more lipophobic chemicals [29]. For metallorganic and inorganic chemicals this uncertainty further increases. Similarly, to the QSARs included in the VCBA, some of predictive tools/methods for the physical-chemical input parameters, such as the ones in EPIsuite and the Fuller method for calculation of molecular diffusion volume, are based on organic chemicals.

The cell uptake in the VCBA model is based on passive diffusion and active transport is presently not being considered. This cell uptake rate is based on the specific cell surface and the permeability equation[49]:

Log permeability = $-1.1711+0.98\ \log K_{ow} - 0.0011MW$

This equation considers the molecular size and the lipophilicity ($\log K_{ow}$) of the chemical molecule, parameters essential for this hallmark description, such that they constitute two of the five rules of the Lipinski[50]. However other chemicals characteristics such as Van der Walls surface areas also influence the cell uptake but as these calculations require powerful computational methods, for the sake of a more throughput model they were not considered [49].

Moreover, these QSARs are not considering saturation, thus VCBA estimations are representative of the maximum concentration which will partition to other elements than aqueous. To eventually include saturation into these models additional experimental measurements would have to be performed such as determination of the maximum number of binding sites on serum protein or maximum concentration that can be bound to plastic[26].

HLC, as previously indicated, is a parameter measured/predicted at temperatures of 25 °C, but it increases with higher temperatures. Since *in vitro* experiments were obtained at 37 °C, the use of this constant at 25 °C, might cause the VCBA to under-predict this endpoint, as observed by comparing with literature indications of volatile chemicals.

As for metabolism, 3T3BALB/c cells are not metabolically competent, therefore, in the VCBA code the rate of metabolism was set equal to 0.

Finally, although the VCBA is run set to the same conditions of the experimental *in vitro* set up, contamination or other (e.g. cell handling) factors which could influence in a negative way the cell culture, are not taken into account.

5. Conclusions

From the simulations carried out with the VCBA model and the analyses presented in this report the following could be concluded:

Modelling chemical fate in the *in vitro* set up showed how much the dissolved concentration can deviate from the nominal concentration for several chemicals, emphasising the importance of using this kind of kinetic data in the interpretation of *in vitro* studies and in *in vivo-in vitro* extrapolations.

The sensitivity analysis performed showed that HLC and $\log K_{ow}$ are the most critical parameters.

It is very likely that the used HLC is not representative of the most common temperature used in *in vitro* experiments (i.e. 37°C). In the future, a possible HLC "temperature correction factor" could be developed by analysing/studying how HLC of several chemicals change across temperatures.

A higher uncertainty is in the applicability of the QSAR equation describing lipid, protein and plastic binding. By being based on experiments made mostly with neutral organic compounds, such as PAH, it is not clear how will these equations describe chemicals prone to other types of interactions. Besides, saturation which is also not considered in the model, can make the model drift from reality. Hence, in the future, some of the obtained partitions should be tested *in vitro* using compounds spreading across a wide range of $\log K_{ow}$ and HLC.

The low (none) impact of the 10 % variation MV and SVcomp parameters in the dissolved concentration output, supports the use of extrapolations of the MW, hence increasing the number of chemicals that can be used. Moreover, even if this QSAR is not totally applicable for inorganic and metallo-organic chemicals, it is unlikely that the difference between the predicted and the real value would have a high impact on the VCBA ouput.

Using the VCBA through the development of $\log K_{ow}$-based alert classes and in the future also on HLC-based classes, is promising in refinement of experimental designs and possibly will allow a deeper interpretation of possible problems in the *in vitro* set up, such as reproducibility. It can give chemical-specific indications on whether different percentages of supplementation, different time endpoint or microplates wells geometry will have a strong impact on the *in vitro* bioavailable concentrations and hence if some *in vitro* experiments are directly comparable or not.

Although *in vitro* toxicokinetics plays a role in the discrepancy between *in vivo* data and *in vitro* predictions (e.g. acute oral toxicity prediction), so do the *in vivo* kinetics, and both should be considered. Therefore, the discrepancies found in acute oral toxicity classification between *in vivo* and *in vitro* experiments could be reduced and /or solved by applying PBK models.

In order to promote the use of the VCBA model by the scientific community and its future application in regulatory context, it would be necessary to build confidence on the simulations made by checking them, to some extent,

experimentally. Based on the results presented in this report, a set of 10 chemicals is proposed to start up such verification.

In this work the dissolved concentration after 24-48 hours is assumed to be the "toxic dose". However, other dosimetrics such Cmax (maximum concentration) and the area under the curve (AUC) of chemical concentration have been used in dose-response studies[4] especially in repeated long-term exposure mode, which can be modelled by VCBA. Which dosimetric would be the most appropriate for *in vivo* models toxicity comparison, is an issue that should be further evaluated in the future.

6. References

1. European Commission. White paper: Strategy for a future Chemicals Policy. *COM* **88,** http://eur-europa.eu/LexUriServ/LexUriServ.do? (2001).
2. Ekwall, B. Screening of toxic compounds in mammalian cell cultures. *Ann. N. Y. Acad. Sci.* **407,** 64–77 (1983).
3. Schirmer, K. Proposal to improve vertebrate cell cultures to establish them as substitutes for the regulatory testing of chemicals and effluents using fish. *Toxicology* **224,** 163–183 (2006).
4. Groothuis, F. A. *et al.* Dose metric considerations in in vitro assays to improve quantitative in vitro-in vivo dose extrapolations. *Toxicology* **332,** 30–40 (2013).
5. Louisse, J., Beekmann, K. & Rietjens, I. M. C. M. Use of physiologically based kinetic modeling-based reverse dosimetry to predict in vivo toxicity from in vitro data. *Chem. Res. Toxicol.* acs.chemrestox.6b00302 (2016). doi:10.1021/acs.chemrestox.6b00302
6. Coecke, S. *et al.* Toxicokinetics as a key to the integrated toxicity risk assessment based primarily on non-animal approaches. *Toxicol. Vitr.* **27,** 1570–1577 (2013).
7. Andersen, M. E. & Krishnan, K. Physiologically based pharmacokinetics and cancer risk assessment. *Environ. Health Perspect.* **102,** 103–108 (1994).
8. Lu, J. *et al.* Developing a Physiologically-Based Pharmacokinetic Model Knowledgebase in Support of Provisional Model Construction. *PLoS Comput. Biol.* **12,** 1–22 (2016).
9. Gajewska, M. *et al.* In vitro-to-in vivo correlation of the skin penetration, liver clearance and hepatotoxicity of caffeine. *Food Chem. Toxicol.* **75,** 39–49 (2015).
10. DeJongh, J. *et al.* An Integrated Approach to the Prediction of Systemic Toxicity using Computer-based Biokinetic Models and Biological In vitro Test Methods: Overview of a Prevalidation Study Based on the ECITTS Project. *Toxicol. Vitr.* **13,** 549–554 (1999).
11. Louisse, J. *et al.* The use of in vitro toxicity data and physiologically based kinetic modeling to predict dose-response curves for in vivo developmental toxicity of glycol ethers in rat and man. *Toxicol. Sci.* **118,** 470–484 (2010).
12. Gubbels-van Hal, W. M. L. G. *et al.* An alternative approach for the safety evaluation of new and existing chemicals, an exercise in integrated testing. *Regul. Toxicol. Pharmacol.* **42,** 284–295 (2005).
13. Astashkina, A., Mann, B. & Grainger, D. W. A critical evaluation of in vitro cell culture models for high-throughput drug screening and toxicity. *Pharmacol. Ther.* **134,** 82–106 (2012).
14. Soldatow, V., LeCluyse, E., Griffith, L. & Rusyn, I. In vitro models for liver toxicity testing. *Toxicol. Res. (Camb).* **2,** 23–39 (2013).
15. Blaauboer, B. J. *et al.* The use of biomarkers of toxicity for integrating in vitro hazard estimates into risk assessment for humans. *ALTEX* **29,** 411–425 (2012).
16. deBruyn, A. M. H. & Gobas, F. A. P. C. the Sorptive Capacity of Animal Protein. *Environ. Toxicol. Chem.* **26,** 1803 (2007).
17. Seibert, H., Mörchel, S. & Gülden, M. Factors influencing nominal effective concentrations of chemical compounds in vitro: Medium protein concentration. *Toxicol. Vitr.* **16,** 289–297 (2002).
18. Kramer, N. I., Krismartina, M., Rico-Rico, Á., Blaauboer, B. J. & Hermens, J. L. M. Quantifying processes determining the free concentration of phenanthrene in basal cytotoxicity assays. *Chem. Res. Toxicol.* **25,** 436–445 (2012).

19. Gülden, M., Mörchel, S. & Seibert, H. Factors influencing nominal effective concentrations of chemical compounds in vitro: cell concentration. *Toxicol. Vitr.* **15,** 233–243 (2001).

20. Fischer, F. C. *et al.* Modeling Exposure in the Tox21 in Vitro Bioassays. *Chem. Res. Toxicol.* (2017). doi:10.1021/acs.chemrestox.7b00023

21. Comenges, J. M. Z., Joossens, E., Benito, J. V. S., Worth, A. & Paini, A. Theoretical and mathematical foundation of the Virtual Cell Based Assay – A review. *Toxicol. Vitr.* (2016). doi:10.1016/j.tiv.2016.07.013

22. Truisi, G. L. *et al.* Understanding the biokinetics of ibuprofen after single and repeated treatments in rat and human in vitro liver cell systems. *Toxicol. Lett.* **233,** 172–186 (2015).

23. Stadnicka-michalak, J., Schirmer, K. & Ashauer, R. Toxicology across scales : Cell population growth in vitro predicts reduced fish growth. *Sceince Adv.* 1–8 (2015). doi:10.1126/sciadv.1500302

24. Armitage, J. M., Wania, F. & Arnot, J. a. Application of mass balance models and the chemical activity concept to facilitate the use of in vitro toxicity data for risk assessment. *Environ. Sci. Technol.* **48,** 9770–9 (2014).

25. Broeders, J. J. W. *et al.* Biokinetics of chlorpromazine in primary rat and human hepatocytes and human HepaRG cells after repeated exposure. *Toxicol. Vitr.* **30,** 52–61 (2015).

26. Kramer, N. I. Measuring , Modeling , and Increasing the Free Concentration of Test Chemicals in Cell Assays. (Utrecht University, 2010).

27. Teeguarden, J. G. & Barton, H. A. Computational modeling of serum-binding proteins and clearance in extrapolations across life stages and species for endocrine active compounds. *Risk Anal.* **24,** 751–770 (2004).

28. Minne B. Heringa, *,† *et al.* Toward More Useful In Vitro Toxicity Data with Measured Free Concentrations. **38,** 6263–6270 (2004).

29. Gülden, M. & Seibert, H. Impact of bioavailability on the correlation between in vitro cytotoxic and in vivo acute fish toxic concentrations of chemicals. *Aquat. Toxicol.* **72,** 327–337 (2005).

30. Zaldivar, J., MENNECOZZI, M., Robim, M. R. & Mounir, B. *A biology-based dynamic approach for the modelling of toxicity in cell-based assays. Part I: Fate modelling*. (2010). doi:10.2788/94002

31. Zaldívar, J. M. *et al. A Biology-Based Dynamic Approach for the Modelling of Toxicity in Cell Assays: Part II: Models for Cell Population Growth and Toxicity*. (2011). doi:10.2788/61603

32. Prieto, P. *et al.* The value of selected in vitro and in silico methods to predict acute oral toxicity in a regulatory context: Results from the European Project ACuteTox. *Toxicol. Vitr.* **27,** 1357–1376 (2013).

33. Halder, M., Kienzler, A., Whelan, M. & Worth, A. *EURL ECVAM Strategy to replace, reduce and refine the use of fish in aquatic toxicity and bioaccumulation testing*. (2014). doi:10.2788/084219

34. Fuller, E. N., Schetler, P. D. & Giddings, J. C. A New Method for Prediction of Binary Gas - Phase Diffusion. *Ind. Eng. Chem.* **58,** 18–27 (1966).

35. Fuller, E. N., Ensley, K. & Giddings, J. C. Diffusion of halogenated hydrocarbons in helium. The Effect of structure on collision cross sections. *J. Phys. Chem.* **73,** (1969).

36. Tang, M. J., Shiraiwa, M., Pöschl, U., Cox, R. A. & Kalberer, M. Compilation and evaluation of gas phase diffusion coefficients of reactive trace gases in the

atmosphere: Volume 2. Diffusivities of organic compounds, pressure-normalised mean free paths, and average Knudsen numbers for gas uptake calculations. *Atmos. Chem. Phys.* **15,** 5585–5598 (2015).

37. Tang, M. J., Cox, R. A. & Kalberer, M. Compilation and evaluation of gas phase diffusion coefficients of reactive trace gases in the atmosphere: Volume 1. Inorganic compounds. *Atmos. Chem. Phys.* **14,** 9233–9247 (2014).

38. Mansouri, K., Grulke, C. M., Richard, A. M., Judson, R. S. & Williams, A. J. An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modelling. (2016). doi:10.1080/1062936X.2016.1253611

39. OECD, D. E. *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)Sar] Models*. *OECD Environment Health and Safety Publications: Series on Testing and Assessment* **69,** (2007).

40. Rietjens, I. M. C. M., Louisse, J. & Punt, A. Tutorial on physiologically based kinetic modeling in molecular nutrition and food research. *Mol. Nutr. Food Res.* **55,** 941–956 (2011).

41. Balaz, S. Modeling kinetics of subcellular disposition of hemicals. *Chem. Rev.* **109,** 1793–1899 (2009).

42. Hulscher, T. E. M. Ten, Van Der Velde, L. E. & Bruggeman, W. A. Temperature Dependence of Henry's Law Constants for Selected Chlorobenzenes, Polychlorinated Biphenyls and Polycyclic Aromatic Hydrocarbons. *Environ. Toxicol. Chem.* **11,** 1595–1603 (1992).

43. Stadnicka-Michalak, J., Tanneberger, K., Schirmer, K. & Ashauer, R. Measured and modeled toxicokinetics in cultured fish cells and application to in vitro - In vivo toxicity extrapolation. *PLoS One* **9,** (2014).

44. International Organization for Standardization, G. ISO 14442 Water quality -- Guidelines for algal growth inhibition tests with poorly soluble materials, volatile compounds, metals and waste water. (2006).

45. OECD. Guidance document on aquatic toxicity testing of difficult substances and mixtures. *Environ. Heal. Saf. Publ. Ser. Test. Assess. - N°23* 53 (2000). doi:ENV/JM/MONO(2007)10

46. Pirovano, A., Huijbregts, M. A. J., Ragas, A. M. J. & Hendriks, A. J. Compound lipophilicity as a descriptor to predict binding affinity (1/ K m) in mammals. *Environ. Sci. Technol.* **46,** 5168–5174 (2012).

47. Goss, K. & Schwarzenbach, Ä. P. Linear Free Energy Relationships Used To Evaluate Equilibrium Partitioning of Organic Compounds. *Environ. Sci. Technol.* **35,** 1–9 (2001).

48. Jonker, M. T. O. & Van Der Heijden, S. A. Bioconcentration factor hydrophobicity cutoff: An artificial phenomenon reconstructed. *Environ. Sci. Technol.* **41,** 7363–7369 (2007).

49. Yazdanian, M., Glynn, S. L., Wright, J. L. & Hawi, A. Correlating Partitioning ad Cacao-2 cell permeability of structurally diverse smal molecular weight compounds.pdf. *Pharm. Res.* **15,** 1490–1494 (1998).

50. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Develop ment Settings. *Adv. Drug Deliv. Rev.* **23,** 3–25 (1997).

7. List of abbreviations and definitions

ADMET - absorption, distribution, metabolism, excretion and toxicity

VCBA - Virtual Cell Based Assay

NEC –No-effect Concentration

Kr – killing rate

AUC - area under the curve

SVcomp-Molecular Diffusion Volume

HLC- Henry Law Constant

MV – Molar Volume

MW – Molecular Weight

logK_{ow} – Logarithm of the partition octanol/water

PAH – Policyclic aromatic hydrocarbons

TP- True Positive

TN - True Negative

FP - False Positive

FN - False Negative

LD50 – Half maximal lethal dose

IC50 – Half maximal inhibitory concentration

NRU – Neutral Red Uptake

EU CLP - Classification, Labelling and Packaging of Substances and Mixtures

QSARs – Quantitative Structure-Activity Relationships

8. List of figures

Figure 15 - Comparison of the in vivo LD_{50} (mg/kg) with the in vitro nominal and the predicted dissolved IC_{50} (mg/kg).

Figure 16 - Normalized Sensitivity Coefficient of the dissolved concentration upon 10 % increase in input parameters: Kr, HLC and $logK_{ow}$ for 5 chemicals.

9. List of tables

+

10. Annexes

Table 11 – Sensitivity Coefficient for parameters $logK_{ow}$, HLC, Svcomp, MV and Kr in Caffeine, Benzene, Xylene, Ochratoxin A, Dimethyldioctadecylammonium chloride.

Chemicals	Parameters					Variation
	$logK_{ow}$	HLC	Svcomp	MV	Kr	
Caffeine	2.1E-04	-2.0E-04	0	-2.0E-04	-2.3E-04	-10%
	-1.8E-04	2.0E-04	0	2.0E-04	1.7E-04	+10%
Benzene	-0.304	-0.370	0	0	-2.1E-04	-10%
	-0.410	-0.345	0	0	-1.9E-04	+10%
Xylene	-0.608	-0.607	1.6E-05	0	-3.6E-04	-10%
	-1.015	-0.548	1.0E-05	0	-2.4E-04	+10%
Ochratoxin A	-19.375	0	0	0	-0.007	-10%
	-7.087	0	0	0	-0.007	+10%
Dimethyldioctadecylammonium chloride	-99.831	0	0	0	0.003	-10%
	-9.068	0	0	0	0.006	+10%

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: http://europea.eu/contact

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: http://europa.eu/contact

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: http://europa.eu

EU publications
You can download or order free and priced EU publications from EU Bookshop at: http://bookshop.europa.eu. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see http://europa.eu/contact).

KJ-NA-28694-EN-N

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi: 10.2760/475757

ISBN 978-92-79-70867-1