EUROPEAN COMMISSION

Community Research

SEVENTH FRAMEWORK PROGRAMME

RISKCYCLE (#226552)

Deliverable 4.1. - Report containing a discussion on the identified criteria and their scores for alternative methods

Deliverable of WP4

D.4.1: Report containing a discussion on the identified criteria and their scores for alternative methods

Contributors: Emilio Benfenati, Diego Baderna, Susanne Heise

Grant Agreement number: 226552

Project Acronym: RISKCYCLE

Project title: Risk-based management of chemicals and products in a circular economy at a global scale

Funding Scheme: 7th Framework Program (FP7)

Project starting date: 01 September 2009

Project duration: 36 months

Name of the scientific representative of the project's coordinator and organisation:

Prof. Dr.-Ing. habil. Dr. h.c. Bernd Bilitewski

Dresden University of Technology
Institute of Waste Management and Contaminated Sites Treatment

Period covered by the report: 0 – 12

Table of Contents

1 Introduction on Alternative Methods

Alternative tests can be included in two categories: in vitro and in silico.

In vitro methods refers to the fact that experiments are done in a tube, generally.

In silico methods refer to the use of the computer to model a certain property of interest.

Below, we will analyse these two categories, and which criteria can be used to choose a suitable methodology.

1.1 The regulatory status of non-animal methods

The term alternatives includes “all procedures which can completely replace the need for animal experiments, reduce the number of animals required, or diminish the amount of distress or pain suffered by animals in meeting the essential needs of man and other animals”[i] . This adds up to the three R’s concept which asks for replacement, reduction and refinement alternatives. An important point, made by the European Centre for the Validation of Alternative Methods (ECVAM) is that these 3 aspects should not be considered as alternatives that could replace each other, but as parts of an integrated system which should lead to progress in the development of non-animal tests and testing strategies.

Non-animal tests therefore comprise[ii] :

1) Maximising the use of existing information, including the reasons for producing a chemical and its uses, as well as knowledge of its toxic hazard potential.

2) The use of data concerning the physicochemical properties of chemicals (for example, stability, solubility, pH, octanol-water partition coefficient, protein binding).

3) Predictions based on structure-activity relationships, including qualitative and quantitative mathematical models, and the use of read-across data from related chemicals.

4) The biokinetic modelling of physiological, pharmacological and toxicological processes.

5) Experiments on lower organisms not classed as ‘protected animals’ (bacteria, fungi, plants, invertebrate animals).

6) Studies on vertebrates at early stages of development (before they become .protected animals.).

7) Studies on in vitro systems of various kinds (including whole perfused organs, tissue slices, cell, tissue and organotypic cultures, and subcellular fractions).
8) Human studies (including estimations of occupational and environmental exposure, epidemiological investigations, post-marketing surveillance for medicines, cosmetics and household and agricultural products, and the ethical and properly controlled use of human volunteers).

ECVAM as part of the JRC fulfils the task to validate alternative methods. Its advisory group ESAC advises ECVAM scientifically and gives expert judgement on the different proposed non-animal tests.
A survey of the regulatory status is given below, indicating those procedures that are accepted by EU, OECD and / or the US (based on http://www.ccac.ca; http://ecvam.jrc.ec.europa.eu/).

Endpoint	Non-animal testing method	Validation / Regulation
Acute aquatic toxicity	Fish acute toxicity – UTC-step down approach	EU
	Fish Embryo Toxicity (FET) Test	OECD – Assessing evaluation
Acute oral toxicity	Up-and-Down Procedure for acute oral toxicity testing	EU
	Acute Toxic Class Method for acute oral toxicity testing	EU
	Fixed Dose Procedure for acute oral toxicity testing Fixed Dose Procedure for acute oral toxicity testing	EU
	Neutral Red Uptake (NRU) test with human cells	US
	Neutral Red Uptake (NRU) test with rodent cells	US
Eye irritation	Isolated Chicken eye test	EU US
	The Bovine Corneal Opacity and Permeability (BCOP) and the Isolated Chicken Eye (ICE) test methods for eye irritation	EU US
	Hen's Egg Test - Chorioallantoic Membrane (HET-CAM) Test Method	EU/US: not sufficiently validated
	Isolated Rabbit Eye (IRE) Test Metho	EU/US: not sufficiently validated
	Slug Mucosal Irritation (SMI) Assay	EU: is being evaluated
Chromosomal aberration	Micronucleus Test as an alternative to the *In Vitro* Chromosome Abberation Assay for genotoxicity testing	EU; OECD
Genotoxicity	Bacterial Reverse Mutation Test (Ames test)	OECD, US: Approved
	Saccharomyces cerevisiae Gene Mutation Assay	OECD
	Saacharomyces cerevisiae Mitotic Recombination Assay	OECD
	In Vitro Mammalian Chromosome Aberration Test	OECD, US
	In Vitro Mammalian Cell Gene Mutation Test	OECD, US
	In Vitro Sister Chromatid Exchange (SCE) Test	OECD
	In Vitro Unscheduled DNA Synthesis (UDS) in Mammalian Cells Test	OECD
Monoclonal Antibody production	*In vitro* production of monoclonal antibodies	EU

Endpoint	Non-animal testing method	Validation / Regulation
Haematotoxicity	The Colony Forming Unit-Granulocyte/Macrophage (CFU-GM) Assay for predicting acute neutropenia in humans	EU
Phototoxicity	3T3 Neutral Red Uptake (NRU) phototoxicity	EU, OECD
Pyrogenicity	Five *In Vitro* Pyrogen tests	EU, US
Reproductive and developmental toxicity	Embryonic Stem Cell Test (EST) for embryotoxicity	EU
	Micromass (MM) embryotoxicity assay	
	Whole Rat Embryo embryotoxicity assay	
	Extended One-Generation Reproduction Toxicity Study	OECD: new test guideline
	Combined Repeated Dose Toxicity Study with the Reproduction/ Developmental Toxicity Screening Test	OECD
	Micromass Embryotoxicity Assay	EU
	Whole Rat Embryo Embryotoxicity Assay	EU
	Frog Embryo Teratogenesis Assay- *Xenopus* (FETAX)	Not sufficiently validated
Skin corrosion	EST-1000 method for skin corrosivity testing	EU
	SkinEthic™ Human Skin Model for skin corrosivity testing	EU
	CORROSITEX assay for skin corrosivity	EU, US
	EpiDerm™ skin corrosivity test	
	EPISKIN™ skin corrosivity test	EU, US
	Rat Transcutaneous Electrical Resistance (TER) skin corrosivity test	EU, US
Skin irritation	Two *in vitro* skin irritation tests: EpiDerm SIT and SkinEthic^{TM} RHE assay	EU
	Artificial skin models (EpiSkin$^{®}$, EpiDerm$^{®}$) for skin irritation testing	EU
Skin sensitization	Reduced Local Lymph Node Assay (rLLNA) for skin sensitisation	EU, US
	Local Lymph Node Assay for skin sensitisation (LLNA)	EU, US
Vaccines	The relevance of the target-animal safety test for batch safety testing of vaccines for veterinary use	EU
	ELISA test for batch potency testing of erysipelas vaccines	EU
	ELISA test for batch potency testing of tetanus vaccines for human use	EU
	ELISA test for batch potency testing of tetanus vaccines for human use	EU

2 *In vitro*

The requirements arising from the regulation and the need to better characterise the toxicological, ecotoxicological and environmental properties of an increasing number of chemicals have the consequence of an increased number of animal experiments, to provide answer to these data needs.

However, this request of more animal testing faces several issues. There is an ethical concern on the millions of animals used every year for experiments. These tests are also those more expensive, and thus this poses questions about the costs for these experiments and the resources to cover them. Many of these tests, especially those chronic, require long times, years in some cases. The number of available laboratories in Europe to cover this potential request is insufficient.

For all these reasons, some European regulations foresee the use of methods alternative to animal tests, such as the REACH legislation, and actually the cosmetics directive foresees the complete ban of animal tests for cosmetics by 2013.

In spite of the great effort and advances made on in vitro testing, we are still far to have alternative methods robust enough to cover developmental, neurotoxic, reproductive or carcinogenic potential for the substances evaluated. However the use of some distinct approaches may cover a great part of the potential toxic effects of some environmental pollutants.

2.1 *Advantages and Disadvantages*

In vitro tests give qualitative information on mode of action of a chemical on specific metabolic processes, enzyme processes or electron transport (in mitochondria or chloroplasts) and specific binding reactions (like the binding to a particular receptor). In same case, they are also an useful tool to evaluate the overall effects of a particular substances on an entire cells taking accounting of mechanisms of sorption, transport and metabolism.

Advantages of the such methods are that they are generally short term tests giving results in few hours or days, they require only small amount of chemical and space, they are generally cheap to run compared to in vivo experiments.

On the other hand, It is often very difficult to relate a response to a specific concentration because it is difficult to keep the concentration constant in such systems e.g.

due to use of plastic materials and media with high content of other organic compounds. Moreover toxicokinetics aspects such as distribution in different compartments are also lacking information in these models.

Moreover, disadvantages are that they only reflect part of what's going on the specific tissue or enzymatic process and thus don't take degradation in the whole organism into consideration or other processes that may influence the same process in an intact organism.

3 *In silico*

3.1 Definitions

In silico methods can be also called non-testing methods, because they do not use laboratory experiments. They include a number of approaches. Some *in silico* methods are aimed to model the interactions between the biological macromolecule and the substance which links to it. These methods are called virtual screening and evaluate the docking potential.

Other methods are called quantitative structure-activity-relationships (QSAR).

In case the relationship is only qualitative, the expression SAR is used. In this case the model investigates the existence of a relationship between a certain chemical property, such as a fragment, and the effect, such as carcinogenic effect, without assigning a numerical continuous value to the toxicity.

Other non-testing methods are called read-across and grouping. Read-across establishes A similitude between a few chemicals, and the property of the unknown substance is taken from the property of the similar compounds. In case a number of chemicals from the same chemical group is available, another possibility is to gain information from the behaviour of this family of similar compounds (grouping).

Here we will address in particular QSAR, because are more used. However, the idea is to extract the common criteria for all in silico methods.

3.2 Components of the QSAR models

At the basis of the QSAR models there are three components: the toxicity (or environmental) property of the chemicals, the chemical information associated to these

chemicals, and the mathematical function which links these two components: the property and the chemical information.

Model development involves the use of chemical compounds with known toxicity levels, which are then used as the training set. This is a very important point since models can only be developed based on knowledge – and the bigger the database the better the model. The model is subsequently developed using chemical parameters and a suitable algorithm.

QSAR are rapidly evolving, for a series of advancements in the scientific field and expectations as alternative methods. Indeed, new information technology techniques have been introduced, and new ways to describe the chemical information, offering new perspective, on one hand, and on the other regulations, such as REACH, call for the availability of robust models.

A few decades ago the range of chemical descriptors used was very limited. Let us take the example of Corwin Hansch's studies, in which he described the relationship between ecotoxicity and a series of parameters, including log P (P = partition coefficient between octanol and water). On the basis of this model, toxicity could then be understood by quantifying uptake of the compound into the fish's body.

Over time other descriptors have been investigated in an attempt to better explain certain factors, such as chemical reactivity and molecular size. Nowadays thousands of chemical descriptors can be calculated and thousands of fragments can be obtained using other programs.

The growth in the number of chemical descriptors and fragments is also the result of the availability of more powerful modelling algorithms. The older QSAR models used linear equations with a very limited number of parameters, in general one or two. Multilinear regressions have now been developed, which offer the possibility of screening a high number of parameters. Non-linear models and the automatic generation of mathematical solutions have now been made possible by the emergence of other tools such as artificial neural network, fuzzy logic, and data mining algorithms.

3.3 Purposes in predicting models and related criteria

Regulatory models require not only powerful methods, but also certain characteristics related to the intended use, such as use of suitable values as input of the model, great consideration of the output of the model, which should fit into the for-

mat identified by the law, with given thresholds and uncertainty, depending on the use, and other features, such as transparency and reproducibility of the model.

The requirements for regulation may be different from those for other purposes. Academic applications of QSAR models are the commonest. Here no strict restrictions and needs exist, beyond the interests of the scientific community.

Models used within industry are also different. In this case, in most situations, confidential data are used, and the interest is to avoid false positive, i.e. to industry wants to avoid to make studies (and spend money) for chemicals which then result to be non active (false positives).

Regulatory QSAR models are more demanding because of their relationship with the law, which introduces requirements, some internal to the QSAR model process, others external. Internally the model needs a high level of quality control. Externally, the model has to comply to, and be suited for, the regulatory use.

3.4 Criteria for evaluation of QSAR models for regulatory purposes

Within RISKCYCLE we will give major attention to criteria which are suitable for regulatory purposes, due to the basic interest of RISKCYCLE, to cope with a reduction of the risk complying To the EU regulations.

For this, the REACH legislation provides a good guidance on the requirements, since QSAR models are explicitly mentioned within the law, in Annex XI.

According to REACH, a (Q)SAR is valid if:

- the model is recognized scientifically valid;
- the substance is included in the applicability domain of the model;
- results are adequate for classification and labelling and for risk assessment;
- adequate documentation of the methods provided.

Let’s discuss the first requirement, a criterion for us. We notice that it is not requested that the model is validated. Validation is a formal process, which takes many years. The formal validation process of a QSAR models would end after REACH probably.

Thus, according to REACH the validity has to be assessed through scientific criteria, considering the performance of the model in its results in prediction. Of particular interest is the check of the predictive performances of the model.

For regulatory purposes greater attention should be given to models which avoids false negatives. Thus, in the evaluation, preference should be given when the model has lower false negatives.

The fact that the model works, that is predictive, should be quite obvious. It is our opinion that this criterion should be applied to all in silico methods, not only QSAR, and actually to all alternative methods. The same applies to all criteria listed by REACH.

The second criterion is quite interesting. It requires showing that the model, which fulfils the requirements for the model, is appropriate for the chemical it has been applied to. Thus, it is not enough to demonstrate that the model works. It is assumed that it does not work in all cases. Thus, a specific evaluation has to be done.

There are some chemometric (chemometrics is a statistical area which combines statistics and chemistry) tools which use the chemical descriptors and/or fragments of the chemicals used to build up the model, and compare if the chemical descriptors and/or fragments of the target chemical are similar. An example of this approach is given by the freely available software AMBIT. A major disadvantage of this approach is that it is based only on the chemical information.

Another recent tool has been developed within the ORCHESTRA project. The tool keeps into account both the chemometric information and the toxicity predictions done by the model, and in particular what kind of errors have been done by the model. It applies to the CAESAR QSAR models. Furthermore, this tool is based not only on the a priori data and information, as the other approaches, but also on the a posteriori result of the model. The user knows if the model can or cannot be used for a certain compound. In some cases a warning is given, recommending expert opinion. In all cases the reasons for the reliability is given, and it can be evaluated in a transparent way.

The third criterion is that the model should target an endpoint relevant for REACH. Only models which address the endpoints of interest for REACH are appropriate within this purpose. We notice that REACH mentions different purposes for the QSAR models: classification and labelling, is one possible target of the model, and risk assessment in another. In the first case models are classifiers; in the second

case a regression more is more suitable. Indeed, in the first case the output of the model is a class, while in the second one it is a continuous value. A continuous value is necessary to get the ratio between the effect dose and the exposure level.

The fourth criterion asks to transparency. This is reasonable, since all documentation at the basis of the assessment of the properties of a chemical should be clearly available and checkable. One of the driving forces of REACH was to have the correct knowledge on the properties of the chemical substances on the market. If some of the information is hidden this clearly goes against the spirit of REACH.

Besides the criteria, we may identify other criteria, which are related to the presence of desirable characteristics of the QSAR model.

QSAR models with a regulatory purpose should mimic the in vivo (and occasionally, in vitro) data, which are typically used in the context identified by the law. As a consequence it should be very much preferable that also the data at the basis of the QSAR models are experimental data suitable for the regulation. In any case, their quality should be very high, and a check should be done on it.

Furthermore, knowledge on the variability and uncertainty associated to each component of the model should be addressed, and described. Within RISKCYCLE, but also for any risk assessment process, the uncertainty of the component is fundamental.

Another criterion is the model reproducibility. This refers in a certain extent to the uncertainty, which was mentioned before, relatively to the knowledge on the input parameters. Here we address the reproducibility of the final result.

Related to this is the easiness of the model. If we imagine a model which is complicated, and has several parameters to be chosen, we may easily get different results. Thus, ideally this is further criterion.

The clarity of the result should be another criterion. It may happen that the output of the model is of difficult interpretation.

The access to the model is another criterion. Some models are free, others very expensive.

The time necessary to get the results (speed of the model) is another desirable criterion.

Another useful feature is the possibility to run predictions in batch, in order to save time.

Within RISKCYCLE we will give particular attention to a series of chemicals present in the waste. The integration of the criteria above mentioned will guide us. In addition, since the list of chemicals is well-defined within RISKCYCLE, this will be a further element, which implicitly is present in what above said. Indeed, referring to the applicability domain of the model, we will evaluate if a certain model specifically address a certain chemical class, which is important for the chemical under assessment.

Other principles for the validation of QSAR models for regulatory purposes were edited by OECD in 2007[iii]. A good model must have:

- A defined endpoint
- An unambiguous algorithm
- A defined domain of applicability
- Appropriate measures of goodness-of-fit, robustness and predictivity
- A mechanistic interpretation.

The definition if the endpoint is essential to understand what kind of experimental systems is being modelled by the in silico method.

Second principle ensure transparency of the model, describing the algorithms used to generate predictions. This information is critical to evaluate uje performance of a the model. In the case of commercial models, the used algorithms is not always made publicly available but its reproducibility must be explained in the guidance material.

The description of the applicability domain is needed to express the limitations in terms of the types of chemicals, properties of mechanism can be generated by the model with an acceptable reliability.

The information about internal and external validation for the model is used to evaluate the performance of the in silico tool.

Regarding the last but not the least principle, it is recognized that it is not always possible to provide a mechanistic interpretation of a given QSAR model but the absence of this information does not preclude the use of the given model in the regulatory context.

OECD also provide a check list for the application of its principles in the context of QSAR validation. This checklist can be useful to help scientists and regulators during the selection of a QSAR model and to evaluate its robustness/validity.

[i] Smyth, D.H. (1978). *Alternatives to Animal Experiments*, 218pp. London, UK: Scolar Press.
[ii] Worth, A. Pl and Balls, M. (Eds) (2002): Alternative (Non-animal) Methods for Chemicals Testing: Current Status and Future Prospects. A Report prepared by ECVAM and the ECVAM working Group on Chemicals. ATLA 30 Supplement 1. pp.125.
[iii] OECD, 2007. GUIDANCE DOCUMENT ON THE VALIDATION OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIPS [(Q)SAR] MODELS.