REPORT FOR THE
INTERNATIONAL COOPERATION ON COSMETICS REGULATION

In Silico Approaches for Safety Assessment of Cosmetic Ingredients

Authors:
Renata Teixeira do Amaral; Jay Ansell; Nora Aptula; Takao Ashikaga; Qasim Chaudhry; Akihiko Hirose; Joanna Jaworska; Hajime Kojima; Mark Lafranconi; Edwin Matthews; Stanley Milstein; Cássia Roesler; Eric Vaillancourt; Rajeshwar Verma; Andrew Worth; Jeffrey Yourick

TABLE OF CONTENTS

1.0 Executive Summary

1. The area of computational toxicology has undergone a lot of scientific developments over the past few decades. As a result, a vast range of models, methods, and tools based on *in silico* approaches are now available that offer a rapid, cost-effective, and ethical alternative to animal testing of chemical substances - including cosmetic ingredients.
2. Various regulatory frameworks in different ICCR countries have a commitment to adhere to 3Rs principles[1], and to promote the use of alternative means for assessing safety of chemical substances instead of testing on animals. In Europe, animal testing of cosmetic ingredients/ products, and marketing of new cosmetic ingredients/ products tested on animals is now banned.
3. The usefulness of structure-activity modelling and other *in silico* (computational) methods as an alternative to testing chemical toxicity in animals has been recognised both by industry and regulators worldwide. This has also been identified as one of the main alternative routes to cosmetic safety assessment in the ICCR-5 report 'Principles of Cosmetic Safety Assessment'[2].
4. In view of the importance of this field, this joint industry-regulator Working Group (WG) carried out an appraisal of the current status of *in silico* approaches in regard to their relevance and use in safety assessment of cosmetic ingredients within the ICCR jurisdictions. This capability review is meant to provide a basis for development of a roadmap for the eventual application of *in silico* methods in the safety assessment of cosmetic ingredients.
5. The main approaches that allow *in silico* assessment of chemical toxicity include computational models based on structure-activity relationship (SAR), or quantitative structure-activity relationship (QSAR), and other *in silico* tools e.g. for mining large chemical toxicity databases and for a 'read-across' from experimental data on structurally and/or functionally similar compounds. Also available are toxicity expert systems that combine different structure-activity based rules, approximations, and/or (Q)SAR models. These versatile *in silico* models, methods, and tools are available both as commercial and free-access software platforms.
6. The appraisal showed that the available *in silico* systems cover a wide range of chemical structure space and majority of the toxicological endpoints that are normally considered in the safety assessment of cosmetic ingredients.
7. The development and use of *in silico* methods requires a thorough consideration of quality and validation of the data and the resulting models and tools, and whether the chemical structure and toxicological space contained within the model's applicability domain adequately covers a given query chemical structure.
8. In this regard, the WG recognises that there is a large number of *in silico* methods referenced in the literature that might be useful for toxicity assessment of cosmetic ingredients. However, this report is limited to an overview of those relevant approaches that are currently under common use. The description of any model or system in this report is, therefore, only meant to provide a general overview and not an endorsement of quality or a recommendation for use as such.
9. Despite a lot of advancements in this field, and the need and the drivers for alternative methods for safety assessment of cosmetics, the current use of *in silico* approaches is

[1] Replacement, Refinement and Reduction of the use animals in laboratory procedures
[2] http://ec.europa.eu/consumers/sectors/cosmetics/files/pdf/iccr5_safety_en.pdf

largely limited to internal decision making both at the industry and the regulatory levels in most ICCR jurisdictions.

10. The lack of full adoption of *in silico* approaches as a mainstream alternative to safety assessment of cosmetic ingredients on animals appears to be for two main reasons:
    a. The current availability of models and tools that have been developed under stringent quality and validation criteria, such as those laid out by the OECD, is limited to only a handful that can be considered of suitable quality to merit use as a routine alternative to *in vivo* testing for safety assessment of cosmetics.
    b. Different models and systems sometime tend to yield conflicting results, which makes it difficult for a user to choose a particular *in silico* system in preference to the others, or to rely on its predicted estimates in a safety assessment. This is because each model and system is generally built on a different dataset and algorithm(s), and therefore tends to decipher and interpret the chemical structure and toxicological information in a different way. Each model/system also reflects a different level of uncertainty and variability associated with the data used in developing it, and the modelling process itself, and may also have a different applicability domain within which the predicted estimates of toxicity are reliable.
11. It practice, however, some of these limitations can be addressed; for example, *in silico* assessments may be drawn from those models and systems that meet the stringent quality and validation standards and have clearly defined applicability domains. Although this may limit the number of workable *in silico* tools, it should provide more confidence in the predicted estimates. It may also be possible to add more confidence to the *in silico* assessments through the use of a combination of appropriate (Q)SAR models, expert systems, and/or read-across approaches, rather than relying on a single model/system. A 'weight of evidence' gathered this way should provide sufficient basis for a reliable *in silico* assessment of chemical toxicity for safety evaluation.
12. It needs to be emphasised that the development, use, and interpretation of the results of *in silico* approaches requires a skilled approach and expert knowledge of toxicology and (bio)chemistry. Thus, despite the ease of use offered by certain advanced *in silico* platforms, the assessment of toxicity by *in silico* approaches must not be reduced to a 'black box' routine.
13. Further work in this area is recommended in relation to development of a uniform and standardised approach that allows the selection and use of appropriate *in silico* system(s), and interpretation of the results from a safety assessment perspective. In addition, a framework also needs to be developed that allows integration of different *in silico* approaches in a consistent scheme to collate sufficient weight of evidence against relevant toxicological endpoints for use in safety assessment of cosmetic ingredients.

2.0 Introduction

The International Cooperation on Cosmetic Regulation (ICCR) held its sixth annual meeting (ICCR-6) July 10-13, 2012 in Rockville, Maryland, U.S.A. to discuss issues related to cosmetics and cosmetic-like drug/quasi-drug products.

As part of the meeting, FDA made a presentation on the current "state of play" of QSAR and "*in silico*" technologies in the safety assessment of cosmetic ingredients. Further it was noted that recent peer-review publications and public presentations including those at the 2012 Personal Care Products Council Regulatory Science Summit (March 29, 2012, Bethesda, MD) and at the

124th Meeting of the Cosmetic Ingredient Review (CIR) Expert Panel (March 5, 2012, Washington, DC) confirm that "*in silico*" technologies relevant to cosmetics safety assessment are being explored in Europe by the EC-JRC, and in the U.S. by FDA-CFSAN and –CDER and EPA as well as by other U.S. federal agencies.

Therefore ICCR proposed that the topic of "*in silico*" computational toxicology tools be introduced to the ICCR Agenda.

2.1 Purpose

The purpose of this WG is to explore suitability of the various available *in silico* approaches for assessing the safety of ingredients intended for use in cosmetics.

2.2 Scope

The main task of this WG is to carry out a review of the existing *in silico* methods that are available and in use for suitability in relation to assess the safety of cosmetic ingredients. Although an account of the different relevant *in silico* approaches will be taken, the work of the WG will focus on general evaluation of the approaches, and not on the assessment of any specific methods, models, or tools. The WG will also aim to propose a tiered approach based on selected methods that can be useful in screening/ prioritization, and potentially in higher tier safety assessment of cosmetic ingredients. For this, the WG will also explore possible integration of the different *in silico* methods, models, and tools so that limitations of individual approaches can be overcome, and sufficient 'weight of evidence' generated to underpin reliability of the *in silico* assessments.

2.3 Approach

To achieve the objectives set out by this WG, the members met via audio-conferences, and discussed and developed a consensus on the *in silico* approaches that can be useful for assessing the safety of cosmetic ingredients.

The WG in their approach aimed to:

1. Identify those *high priority endpoints* from the list of toxicological endpoints routinely assessed for cosmetics for which *in silico* methods are already available and being used[3].
2. Define the scope of the report in terms of the different types of *in silico* approaches that can be useful in safety assessments (e.g. QSAR & expert systems, biokinetic modelling[4], molecular modelling). Summarize the strengths as well as limitations of the approaches[5]. Indicate how their quality (including the uncertainty) is being assessed and reported.
3. Develop an "activity matrix" that summaries, for the priority endpoints, the major ongoing or recently completed projects in the ICCR regions (including the regulatory authorities and industry associations) that are either developing *in silico* tools and/or exploring their applicability relevant to the safety assessment of cosmetics. This will include projects

[3] The development of new methods is outside the scope of this activity.

[4] This includes the modeling of internal exposure, and *in vitro-in vivo* extrapolation, thereby providing a basis for risk assessment based on alternative methods.

[5] This will involve a general evaluation of the approaches, not focussed on specific methods, models or tools

that may be focused specifically on cosmetics, as well as those methods, models and tools that could be indirectly applicable to cosmetic ingredients.
4. Briefly describe each project, with supporting references and web links.
5. Briefly describe each selected method in a systematic way[6], including the tier(s) in which it can be used, and the type of assessment it can support (hazard identification, potency classification, and/or risk assessment)
6. Identify data/knowledge gaps in terms of our ability to assess the suitability of *in silico* approaches for assessing cosmetics.
7. Make recommendations for further work needed by the ICCR (if any).

2.4 Timelines:

- Scoping document to be discussed at ICCR 7 (Japan, July 8-10, 2013)
- Report to be drafted by ICCR-8, 2014.

2.5 Authors:

Renata Teixeira do Amaral	Brazilian Association of Personal Cosmetics, Toiletry and Fragrance, Av. Paulista, 1313. Cj.1080. zipcode:01311-923; São Paulo – SP – Brazil
Jay Ansell:	Personal Care Products Council, 1620 L Street, NW, Washington, DC 20036, USA.
Nora Aptula:	SEAC, Unilever, Colworth Science Park, United Kingdom.
Takao Ashikaga:	Shiseido, Quality Assessment Center, Safety Technology Development Group, Japan.
Qasim Chaudhry (Chair):	The Food and Environment Research Agency, Sand Hutton, York Y041 1LZ, United Kingdom.
Akihiko Hirose:	Division of Risk Assessment, Biological Safety Research Center, National Institute of Health Sciences, Japan.
Joanna Jaworska (Co-Chair):	Procter & Gamble, Modeling & Simulation, Biological Systems, Temselaan 100, 1853 Strombeek-Bever, Belgium.
Hajime Kojima:	Japanese Center for the Validation of Alternative Methods, Biological Safety Research Center, National Institute of Health Sciences, Japan.
Mark Lafranconi:	Tox Horizons, LLC, Maineville, OH 45039, U.S.A.
Edwin Matthews:	Office of Food Additive Safety, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, 5100 Paint Branch Parkway, College Park, MD 20740, USA.
Stanley Milstein:	Office of Cosmetics and Colors, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, 5100 Paint Branch Parkway, College Park, MD 20740, USA.

[6] The QSAR Model Reporting Format, or simplified version of this, could be an option: http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/qsar_tools/QRF

Cássia Roesler: Brazilian Health Surveillance Agency, General Office of Cosmetics, SIA Trecho 5, Área especial 57, Lote 200, CEP: 71205-050; Brasília – DF – Brazil.

Eric Vaillancourt: Health Canada, Risk Assessment Division, Risk Assessment Bureau, Consumer Product Safety Directorate, 269 Laurier Ave. West, 7th floor, 7-101, Ottawa (Ontario) K1A 0K9, Canada.

Rajeshwar Verma: Office of Cosmetics and Colors & Office of Food Additive Safety, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, 5100 Paint Branch Parkway, College Park, MD 20740, USA.

Andrew Worth: European Commission, Joint Research Centre, Institute for Health and Consumer Protection, Via Enrico Fermi 2749 21027 Ispra, Italy.

Jeffrey Yourick: Office of Applied Research and Safety Assessment, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, 8301 Muirkirk Road, Laurel, MD 20708 USA.

3.0 In silico Methods

3.1 (Q)SAR based models

The chemical structure of a compound carries a lot of information – some of which is not very obvious *per se*. This embedded information can be deciphered in the form of numerical 'descriptors' of physicochemical parameters, which ultimately determine the properties, behaviour, and biological activity of a given compound. Such a description of the chemical structure space of a group of compounds in the form of various descriptors (e.g. steric, hydrophobic, electronic, topologic, etc) allows modelling the measured data on a specific biological endpoint to find mathematical relationship(s) between the chemical structure and its biological activity. The resulting *in silico* (computational) models may either be based on structure-activity relationship (SAR), or quantitative structure-activity relationship (QSAR)[7]. Other *in silico* approaches allow read-across from experimental data on structurally and/or functionally similar compounds to derive estimates of toxicity of an untested chemical.

A typical QSAR model describes quantitative relationship(s) between a chemical structure and biological activity, and generally takes the form of a linear equation:

Relative Biological Activity = Constant + (C_1,P_1) + (C_2,P_2) + …. (C_n,P_n)

where the parameters P_1 to P_n are computed for each molecule in the series, and coefficients C_1 to C_n are calculated by fitting variations in the parameters and the physicochemical property or the biological activity. However, a structure-activity relationship could also be non-linear.

An SAR model on the other hand describes qualitative relationship(s) between a chemical structure and biological activity. In its simplest form, an SAR takes the form of a 'structural alert', which represents a distinctive feature in a molecule that bears a relationship with its biological activity.

[7] For introduction to theory and applications of QSARs see Benfenati (2012).

Since early developments in this field in the 1960s, many more reliable chemical property/effect databases, powerful data-mining tools, linear and non-linear algorithms, and soft-computing techniques that can decipher relational patterns in large and complex datasets have become available. These, combined with increasing computational power, have led to the development of more versatile and reliable *in silico* methods and tools for the assessment of chemical toxicity. The (Q)SAR approaches have also progressed from modelling a few physicochemical properties of closely related compounds, and endpoints at the molecular level (e.g. enzyme inhibition, or receptor binding), to more diverse chemical structures and complex toxicological endpoints in whole organisms. Linear algorithms, such as multiple regression, are generally useful for modelling numerical and continuous biological data (e.g. LD50), whereas description of discrete response (e.g. active/inactive, weak/moderate/strong) generally requires decision trees, neural networks, support vectors, or clustering algorithms, etc[8]. The *in silico* modelling approaches have also enabled the assessment of other toxicity modulating factors, such as absorption, distribution, metabolism and elimination (ADME) in live organisms. As a result, *in silico* methods and tools are increasingly seen as an important alternative to testing safety of chemical substances in animals. As such, they offer a rapid, cost-effective, and ethical alternative to animal testing. A number of *in silico* models and systems are currently available, which cover a wide range of chemical space and are therefore valuable for assessment of a variety of chemical substances – including cosmetic ingredients.

3.1.1 Mechanistic vs statistical QSARs

Although multidimensional QSARs developed using molecular structure descriptors and regression analysis techniques have found wide utility and acceptance, it is often difficult to extract a physical interpretation of such models because of the types of descriptors involved and the multidimensional nature of the model. Such models are referred to as statistical. In contrast, mechanistic QSARs are based on a mechanistic hypothesis that drives the choice of descriptors. A good discussion on this subject is provided by Lipnick (1999).

3.1.2 Model validation

Once developed, (Q)SAR models are generally subjected to rigorous testing for robustness and predictivity, and a clear description of the 'applicability domain', which is the biological response and chemical structure space within which a model makes predictions with a given reliability (Netzeva et al., 2005). A fully tested and validated (Q)SAR model would generally yield good predictive assessment of the toxicity of an untested chemical as long as the query compound falls within the domain of the model's prediction space. Thoroughly tested and validated (Q)SAR models therefore a valuable means for predicting the biological activity (including toxicity) of untested chemicals within the bounds of the chemical structure space covered by the model. This also means that each model carries certain limitations and boundaries, and in respect to the chemicals space it covers, and may not perform reliably when used against chemicals that are outside the model's applicability domain. It is, however, possible to address some of these limitations in practice – e.g. through the use of a combination of different (Q)SAR models, expert systems, and read-across approaches to derive sufficient 'weight of evidence' that can provide basis for a reliable *in silico* assessment of chemical toxicity.

[8] For further reading on the choice of different statistical algorithms for QSAR building see Chaudhry et al. (2007).

3.1.3 Availability and currents use

A variety of (Q)SAR based models and expert systems are available for predicting a range of toxicity endpoints. These include both free access and commercial software platforms. For *in silico* assessments intended for regulatory purposes, the (Q)SAR models generally considered relevant are those that have been developed in accordance with the stringent quality criteria and validation principles, e.g. those laid down by the OECD (2004). The OECD's 'Guidance Document on the Validation of (Q)SAR Models' published in 2007 aims at providing further guidance on how specific QSAR models can be evaluated with respect to the OECD principles.

The use of (Q)SAR models as an alternative approach to testing chemical toxicity on animals has been promoted by regulatory authorities in a number of countries. In Europe, the use of *in silico* approaches for assessment of chemical toxicity has come into focus due to the ban on testing cosmetic ingredients and products on animals. The use of validated (Q)SARs is currently allowed under certain provisions of some European regulations, e.g. the chemicals Regulation REACH. Also in Europe, the Danish EPA has produced a comprehensive set of QSAR models for their 'self-classification system' of industrial chemicals. This database is now incorporated in to the OECD QSAR Toolbox. Other examples of the use of (Q)SAR models within the ICCR jurisdictions is provided in section 4.2. It is however worth highlighting that, despite substantial advancements in this field, and the need for suitable alternatives to animal testing, the use of (Q)SAR models has still not been fully adopted as a mainstream method for safety assessment of cosmetic ingredients in any of the ICCR's jurisdiction. They are, however, being applied by many as screening tools for internal decision making and as important elements for development of a weight of evidence (WoE) for risk assessment. A number of projects have made some inroads into the application of this field in relation to safety assessment of cosmetics. For example considerable work has been done on fragrance allergens, and progress at the EU level can be seen in the specialist Cosmetic Sector reports of ECVAM[5]. The potential application of *in silico* methods for safety assessment of cosmetic ingredients has also been mentioned in the Preamble of the SCCS Memorandum on "Alternative test methods in human health safety assessment of cosmetic ingredients in the EU" (SCCS/1294/10).

3.2 Expert systems

An expert system has been defined as any formalised system that is often, but not necessarily, computer based, and that can be used to make predictions on the basis of prior information (Deaden *et al*., 1997). *In silico* toxicology expert systems may combine different approaches to predict toxicity of a substance from chemical structure. The expert systems (and their implementation in software tools) are based on three main modelling approaches referred to rule-based, statistically-based, or hybrid methods.

Rule-based systems contain "if-then-else" rules that have been derived from toxicological knowledge, and combine them with expert judgment and/or fuzzy logic. Commonly used software tools based on this approach include OncoLogic (Woo *et al*. 2005), Derek Nexus® (Sanderson & Earnshaw 1991; Ridings *et al*., 1996) and HazardExpert (Smithing & Darvas 1992). Derek Nexus® and HazardExpert can be used in conjunction with their related programs Meteor and Metabolexpert to predict the toxic potential of metabolites as well as the parent compounds. In addition to these commercial tools, many of the models included in the freely available Toxtree software and the OECD QSAR Toolbox are rule-based. An example is the Benigni-Bossa rulebase for genotoxicity and carcinogenicity (Benigni *et al*., 2008).

Statistically-based systems use a variety of statistical, rule-induction, artificial intelligence, and pattern recognition techniques to build models from non-congeneric databases. Statistically based systems are included in the commercial tools MultiCASE and TOPKAT, and the publicly

available Lazar and CAESAR models. In addition, many models published in the scientific literature, but not (yet) implemented in software are statistically based.

Hybrid models are based on a combination of knowledge-based rules and statistically-derived models. These are based on the general idea that, within the structural space of a single structural alert (considered to represent a single interaction mechanism), statistically derived models can quantitatively predict the variation in the reactivity of the alert conditioned by the rest of the molecular structure. Examples of the hybrid approach include models implemented in OASIS TIMES (Patlewicz *et al*., 2007).

The advantages and disadvantages of these approaches are summarised in Table 1.

Table 1. Comparison of three main approaches used in expert systems

Approach	Advantages	Disadvantages
Rule-based	– mechanistically connected to the predicted endpoint – provide reasoning for the predictions – in many cases support the prediction with literature references or expert knowledge	– often restricted and/or ill-defined applicability domain – usually cannot explain differences of the activity within a chemical class – usually have lower accuracy of the prediction than statistical models
Statistical	– usually have high accuracy of the predictions – can be used for preliminary research when mechanism of action is unknown	– usually difficult to interpret the model predictions – often do not provide mechanistically reasoning of the predictions – often non-transparent to the end-user
Hybrid	– combines advantages of rule-based and statistical approaches, including mechanistic interpretability (for SA part), and overall accuracy	– likely to have restricted applicability domain

3.3 Read Across

Over the past decades, enormous amount of information relating to physicochemical properties and toxicity of chemicals has been collated in the form of extensive databases. In addition to being useful as data resources, they also provide a means for estimating the toxicity of an untested chemical by drawing parallels from the measured data on other tested chemicals. Such a 'read across' inevitably needs a tight category of already tested chemical analogues that are structurally and/or functionally related to the untested query chemical. The OECD QSAR ToolBox is an example of a versatile suite of programs which can also predict various toxicity endpoints based on read-across using a substantial set of high quality databases. The users may import additional databases into the toolbox to expand its usefulness. The ToolBox allows identification of analogues for a given chemical, retrieve available experimental data for the analogues, and fill data gaps by read-across or trend analysis. The use of read across has become more common in predictive toxicology. In comparison to QSAR programs that use groups of analogues as a source from which to derive mathematical relationships, read across uses analogues directly to predict endpoints for the target. For this reason, it is most important that groups or categories of analogues are carefully assembled taking into account not only

structural similarity, but where possible, also mechanistic similarity. Read across predictions are not possible where close analogues of the target compound cannot be found in the databases.

3.4 Important considerations

3.4.1 Data and Model Quality

The reliability and robustness of *in silico* predictive models is intrinsically dependent on the quality of the data used in building and validating the model. This applies to data on chemical identity, structure, descriptors, biological activity, statistical algorithms used in building the model, and the degree to which the model was tested and validated. Thus whilst good quality data are essential for (Q)SAR model building, it also need to be remembered that each model also reflects the variability and uncertainty within the chemical and toxicity data (due to inter- and intra- species differences, and differences in the experimental protocols used, etc).

The scarcity of good quality toxicity data for a sufficient number of related compounds is often the main limiting factor in the development of robust and reliable *in silico* models. A data quality algorithm has been proposed by Malazizi et al (2006). For *in silico* model development, biological test data from good laboratory practice (GLP) is often sufficiently reliable.

Accurate measurement of toxicity in a living organism is intrinsically complex, as a measured endpoint, such as mortality, may have involved a number of processes and mechanisms. However, uncertainty in the measurements does not preclude the use of the data in modelling. As long as the models do not produce false negative predictions, they are generally acceptable because any false positive predictions will still be on the conservative side in a risk assessment.

Other factors, such as the existence of a chemical in different enantiomeric forms, cannot be resolved easily by *in silico* methods, the vast majority of which is based on descriptors derived from two-dimensional structures. Compared to 2D structures, the use of parameters relating to 3D structures require much a greater effort because it is generally difficult to obtain an optimised configuration of a given molecule that can be regarded as resembling to its form in 'real life'. Generally, 3D modelling approaches use the lowest energy conformation of a chemical structure, and where available, derive the descriptors from its crystal structure. Certain specialist *in silico* systems, such as CoMFA (Comparative Molecular Field Analysis) can consider chirality and hence can deal with different enantiomeric forms of a chemical. Different tautomeric forms also present a problem in terms of reliability of *in silico* assessment as they can co-exist or can readily interchange from one form to another. Again selecting a structure at the lowest energy conformation is generally used in modelling to overcome such issues.

In November 2004, the OECD Member Countries agreed on the principles for validating QSAR models for their use in the regulatory assessment of chemical safety. The internationally agreed principles provide Member Countries with a consistent and scientifically motivated framework for evaluating the regulatory applicability of QSAR models.

The OECD Principles for QSAR Model Validation, which are intended to be read in conjunction with the guidance document, are as follows:

"To facilitate the consideration of a (Q)SAR model for regulatory purposes, it should be associated with the following information:

1. a defined endpoint;
2. an unambiguous algorithm;
3. a defined domain of applicability;
4. appropriate measures of goodness-of-fit, robustness and predictivity;

5. a mechanistic interpretation, if possible.

Further information on the OECD principles is given on the OECD website (http://www.oecd.org/env/ehs/risk-assessment/validationofqsarmodels.htm).

3.4.2 Description of Software/Database Version

The acceptability of a (Q)SAR or read across prediction relies, in part, on a transparent description of the software tool and the process to generate the prediction. It is therefore important that details are provided on the software version/databases used in any *in silico* assessment to keep consistency between assessments. As software platforms may evolve and become available in different versions, it is very important to document the version used, and if possible the training set of compounds used, to make the prediction. The same applies to database versions. Even using the same process, one may reach different conclusions due to differences in data availability because of the use of different databases. Therefor database source and the exact version used should be documented in an *in silico* assessment. This will not only provide transparency, but will also allow other users to reproduce the results if needed.

3.4.3 Documenting the validity of QSAR models

Although numerous QSAR models have been developed and published in the scientific literature, and some models have been used in regulatory assessment of chemicals in some countries for many years, a transparent validation process and objective determination of the reliability of QSAR models are crucial to further enhance their regulatory acceptance.

The **QSAR Model Reporting Format (QMRF)** was developed by the JRC and EU Member State authorities as a harmonised template for summarising and reporting key information on QSAR models, including the results of any validation studies. The information is structured according to the OECD validation principles. Further information on the QMRF is given on the JRC website (http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/qsar_tools/QRF).

3.4.4 Quantification/ characterization of uncertainty

Until now modelling techniques used in applications to toxicology have been predominantly deterministic (e.g. classification trees; regression models). A deterministic technique treats every piece of information at face value and generates a single result for given a set of input parameters.

However, the fact that deterministic models ignore uncertainty in knowledge and variability in data is not sufficient to justify the use of probabilistic models in place of deterministic models. The most important aspect of the model is how it affects decision-making. Deterministic frameworks are generally sufficient when dealing with single endpoint prediction. However, in the WoE approach, the need for probabilistic modelling may arise from the aim of placing knowledge and complex, multifaceted, and mutually dependent data in a systematic, evidence-based, reasoning framework. Therefore, probabilistic WoE frameworks are more realistic than the deterministic ones. In a probabilistic WoE, each piece of evidence is used together with its uncertainty. Deterministic models calculate uncertainty but this information is not conveyed to the user. As such, WoE would greatly benefit from probabilistic approaches that can provide predictions with uncertainty, as this would result in improved weighing different pieces of evidence.

4.0 METHODOLOGY AND APPROACH

4.1 Data requirements for cosmetic ingredients

With the exception of Europe, where animal testing of cosmetic ingredients is banned, safety assessment of cosmetics in other ICCR jurisdictions relies on toxicological data from animal tests, and/or in vitro models. Where available, human clinical and epidemiological data is also used to provide additional information on the toxicological effects of chemicals intended for use in cosmetic products. These conventional methods, however, involve a lot of time, effort, and costs. There are further ethical and other constraints to animal testing, such as questions about predictivity for human risk assessment and the general commitment to the 3Rs principle[9]. This is where the use of *in silico* approaches can be very useful in terms of filling some of the data gaps. In the context of mapping the current and projected uptake of *in silico* methods and tools in regulatory perspective, it is important to consider the current requirements for safety data in the different ICCR jurisdictions. A brief description of the data requirements for safety assessment of cosmetic ingredients is provided below:

4.1.1 CANADA

- All cosmetics sold in Canada must be safe to use and must meet the requirements of the *Food and Drugs Act*, the *Cosmetic Regulations*, and the *Consumer Packaging and Labelling Act* and Regulations. All intentional ingredients must be listed on the label, and the manufacturer must notify Health Canada that it is selling the product and provide information about the product's formulation.
- Health Canada uses the List of Prohibited and Restricted Cosmetic Ingredients (also known as the Cosmetic Ingredient Hotlist) to communicate to manufacturers that certain substances may cause injury to the health of the user, which is in contravention of the general prohibition in the *Food and Drugs Act*. Departmental officials closely follow international scientific and regulatory reports, and regularly review the safety of cosmetic ingredients.
- Health Canada takes a risk-based approach to regulating cosmetics, considering both the properties of substances in products, as well as the amounts to which people are exposed under normal conditions of use, to determine whether there is a risk that needs to be addressed by the Government. If the scientific evidence suggests that a cosmetic ingredient may be unsafe, the Department would conduct a review and, where warranted, take action. This could include banning ingredients or restricting their use, requiring certain labelling communicating the risk to Canadians, or having the cosmetic product removed from stores.
- The Canadian government launched the Chemicals Management Plan (CMP) to strengthen efforts to protect human health and the environment from the risks of harmful chemicals, including those considered to be carcinogenic or those that pose reproductive or developmental hazards. As a result of the CMP, 24 substances have been added to the Cosmetic Ingredient Hotlist and two existing Hotlist items have been amended to be more protective of health. For more information, see: www.chemicalsubstanceschimiques.gc.ca/.

[9] Replacement, Refinement and Reduction of the use animals in laboratory procedures

- Data are not requested at the time cosmetic products are notified to Health Canada. However, when notified product formulations contain ingredients that are subject to a Hotlist restriction or when incidents are reported, the following information *may* be requested:

    - Physicochemical properties (*e.g*. formulation pH, salivary peroxide levels)
    - Skin irritation (*e.g*. human patch tests)
    - Skin sensitization
    - Phototoxicity
    - Eye irritation

- There are currently no policies/guidelines for the validation or acceptance of *in silico* models for cosmetic products, and Health Canada has not yet requested, received or accepted any such model.

4.1.2 EUROPE

- As mentioned before, animal testing of cosmetic ingredients is banned in Europe and testing must be replaced by alternative methods. The Directive 76/768/EEC, and as of 11 July 2013 the Cosmetics Regulation ((EC) No 1223/2009), prohibits the testing of finished cosmetic products and cosmetic ingredients on animals (testing ban), and prohibits the marketing in the European Community, of finished cosmetic products and ingredients included in cosmetic products that were tested on animals (marketing ban). The testing/ marketing bans apply irrespective of the availability of alternative non-animal tests.
- Besides the Cosmetics legislation, Article 7 of the Council Directive 86/609/EEC provides for the protection of animals used for experimental and other scientific purposes 'an animal study shall not be performed if another scientifically satisfactory method of obtaining the result sought, not entailing the use of an animal, is reasonable and practically available'. Directive 86/609/EEC has been repealed as of 1 January 2013 and replaced by Directive 2010/63/EU on the protection of animals used for scientific purposes, which in Article 4 contains the principles of replacement, reduction and refinement (3Rs).
- The toxicological endpoints that need assessing in a safety dossier in Europe include:

    - Acute toxicity (oral; inhalation; dermal)
    - Corrosivity and irritation (Skin corrosivity and skin irritation; Mucous membrane irritation, eye irritation)
    - Skin sensitisation
    - Dermal/ percutaneous absorption
    - Repeated dose toxicity
    - Mutagenicity/genotoxicity
    - Carcinogenicity
    - Reproductive toxicity
    - Toxicokinetic studies
    - Photo-induced toxicity (Phototoxicity (photoirritation) and photosensitisation; Photomutagenicity/ Photoclastogenicity)

4.1.3 JAPAN

- Under the Japanese cosmetics regulation, the safety standards of cosmetic ingredients consist of both positive lists (only for antiseptics, ultraviolet absorbents and tar dyes) and negative lists. In addition, medicinal compounds, non-standardized biological materials, and any prohibited substances under other chemical related laws are forbidden.
- Chemicals in the positive list and medicinal compounds may be allowed with restricted concentrations. Data on other substances is needed.
- Some exemptions of the specific endpoints and some alternatives to in vivo tests due to scientific and/or ethical reasons may be permitted, although any official statement or assessment guidance about those issues has not been issued.
- Regarding a quasi-drug, data on the list will be required as well as positive list, though some endpoints are can be omitted.

    - Phys-chem properties
    - Single dose toxicity
    - Repeated dose toxicity
    - Reproductive/ developmental toxicity
    - Skin irritation
    - Skin sensitization
    - Photo toxicity
    - Photo sensitization
    - Eye irritation
    - Genotoxicity
    - Human patch tests
    - ADME(Absorption, Distribution, Metabolism and Excretion)

4.1.4 USA

- Under the U.S. Federal Food, Drug, and Cosmetic Act (FD&C Act), cosmetic products and ingredients are not subject to FDA premarket approval authority, with the exception of color additives (other than those intended for use as coal tar hair dyes). However, they must be safe for consumers under labeled or customary conditions of use.
- In general, except for color additives and those ingredients which are prohibited or restricted from use in cosmetics by regulation, a manufacturer may use any ingredient in the formulation of a cosmetic provided that the ingredient and the finished cosmetic are safe, the product is properly labeled, and the use of the ingredient does not otherwise cause the cosmetic to be adulterated or misbranded under the laws that FDA enforces.
- In addition, regulations prohibit or restrict the use of several ingredients in cosmetic products and require warning statements on the labels of certain types of cosmetics. Neither the law nor FDA regulations require specific tests to demonstrate the safety of individual products or ingredients. Rather, FDA has consistently advised manufacturers to use whatever testing is necessary to ensure the safety of their products and ingredients. Firms may substantiate safety in a number of ways. FDA has stated that "the safety of a product can be adequately substantiated through (a) reliance on already available toxicological test data on individual ingredients and on product formulations that are similar in composition to the particular cosmetic, and (b) performance of any additional toxicological and other tests that are appropriate in light of such existing data and information.

- More information on FDA Authority over Cosmetics is provided at: http://wcms.fda.gov/FDAgov/Cosmetics/GuidanceComplianceRegulatoryInformation/ucm074162.htm. Link to Product Testing: http://wcms.fda.gov/FDAgov/Cosmetics/ProductandIngredientSafety/ProductTesting/default.htm.

4.1.5 BRAZIL

- The Brazilian cosmetics regulation states that cosmetics must be safe for consumers in normal or foreseeable conditions of use. To guarantee the quality of these products, the Brazilian Health Surveillance Agency (Anvisa) works at registration, notification, market inspection and post marketing surveillance. The Agency also creates norms and rules applicable to production processes, techniques, methods and use of those products by consumers.
- Cosmetics products must comply with a set of lists that establishes which ingredients may or may not be used, or even some that may be used if they meet some specific restrictions and conditions (available at http://s.anvisa.gov.br/wps/s/r/lJp). These requirements undertake a systematic review of ingredients and a battery of data of international benchmarks, such as European Union Directive, United States of America laws, and technical criteria recognized by the scientific community of Mercosur States Parties (Resolution GMC N° 51/08 e Resolution GMC N.º133/96)
- The Safety Evaluation Guidelines published by Anvisa (2nd edition published in 2013) recommends a wide variety of tests in the cosmetic ingredients and also in the final product. All the safety evaluation tests must be performed before the cosmetic is placed into the market. In addition, Anvisa's Affairs Law for cosmetics place responsibility on the manufacturer and/or importers of cosmetic products to ensure that their products are safe for the consumer when used as intended (RDC 04/2014).
- In Brazil there are no guidelines for validation of *in silico* models and the respective acceptance criteria for regulatory agencies in the scientific and regulatory dimension. Nevertheless, we believe that knowledge about such premise can design regulatory actions and decision making in the context of *in silico* reports, in addition to promote the development of its predictive ability.

Relevant data required include:

- Acute toxicity
- Skin corrosivity and skin irritation
- Skin sensitisation
- Dermal/ percutaneous absorption
- Repeated dose toxicity
- Mutagenicity/genotoxicity
- Subacute and subchronic toxicity
- Eye irritation
- Mucous membrane irritation
- Photo-induced toxicity (phototoxicity , genotoxicity, photosensitization)
- Carcinogenicity
- Reproductive toxicity
- Toxicokinetic and toxicodynamic studies

The above overview of the regulation of cosmetic related ingredients/products in the different ICCR jurisdictions shows that where there are specified requirements of safety data, they relate to physiochemical data and toxicological endpoints that range from a few simple acute endpoints (e.g. irritation/sensitisation of eye or skin) to more complex chronic endpoints (e.g. carcinogenicity, mutagenicity, developmental toxicity). This information is mapped in section 4.2 and Annex-I to identify whether or not an *in silico* model or expert system is available that can be used to draw estimates on toxicity. It is, however, worth highlighting that it was outside the scope of this WG to develop new, validate, or provide recommendation for the use of a specific *in silico* method, model or system. The description of few selected (Q)SAR models and approaches provided in Annex-I is meant to be a general overview and as such must not be seen as an endorsement of quality or a recommendation for use. It is also recognised that there are other *in silico* models and tools not covered by this report but may be relevant to use in *in silico* assessment of cosmetic ingredients.

4.2 THE CURRENT USE OF IN SILICO METHODS

4.2.1 THE CURRENT USE OF *IN SILICO* METHODS – Europe

4.2.1.1 JRC QSAR Model database

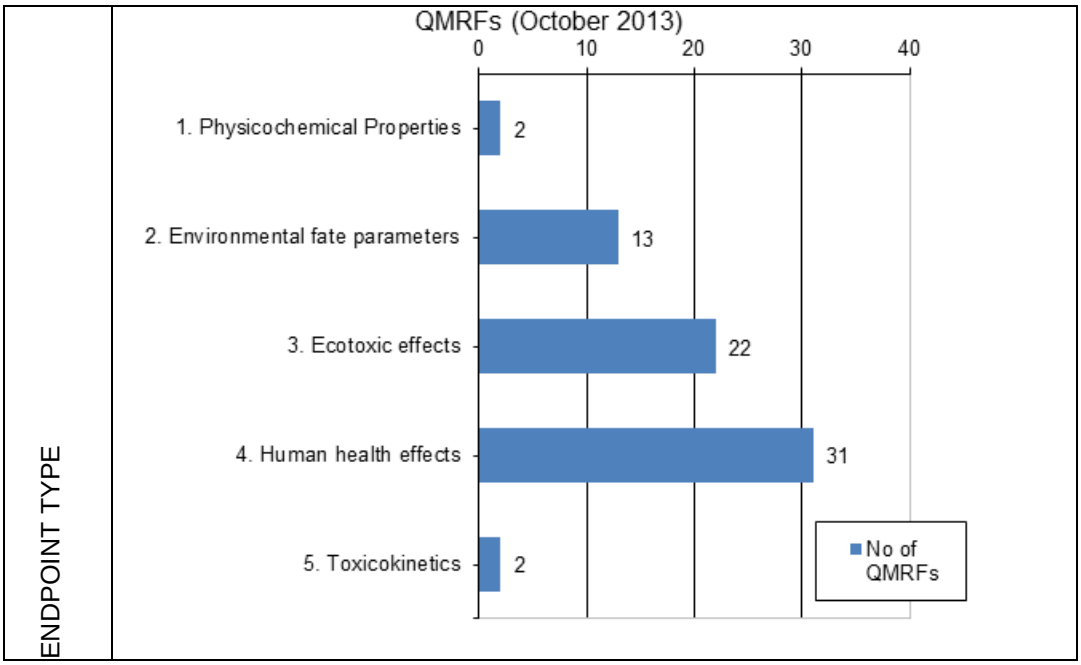
The JRC QSAR Model Database (http://qsardb.jrc.ec.europa.eu/qmrf/) is a freely accessible web application that enables users to submit, publish, and search QMRF[10] reports. Developers and users of QSAR models can submit to the dedicated mailbox information on QSARs by using the QMRF. A downloadable **QMRF editor** (http://sourceforge.net/projects/qmrf/files/QMRF%20Editor/2.0.0/) is used for this purpose. The JRC then performs a quality control (i.e. adequacy and completeness of the documentation) of the QMRF submitted. Properly documented QMRFs are included in the JRC QSAR Model Database. Inclusion of the model does not imply acceptance or endorsement by the JRC or the European Commission, and responsibility for use of the models lies with the end-users.

4.2.1.2 Status of QMRFs in the JRC QSAR Model Database

At the time of writing (October 2013), the JRC QSAR Model Database contains 70 QMRFs (Figure 1). A number of additional QMRFs will also be uploaded in a new version of the database that will be available soon from the same webpage.

[10] (Q)SAR Model Reporting Format (QMRF) is a QSAR model documentation format in line with the OECD principles.

Figure 1. Status of QMRFs in the JRC QSAR Model Database



4.2.2 THE CURRENT USE OF *IN SILICO* METHODS - JAPAN

The following two approaches have been developed in Japan with support from the Japanese government:

- Hazard Evaluation Support System Integrated Platform (HESS) for assessment of repeated-dose toxicity
- Combination Approach for assessment of mutagenicity

Both applications are still at research stage, but the 'Combination Approach' will be used for prioritizing the existing chemicals within the occupational health area.

4.2.2.1 The Hazard Evaluation Support System Integrated Platform – HESS:

The project on 'Development of hazard assessment techniques using structure-activity relationship methods (2007-2011)', sponsored by New Energy and Industrial Technology Development Organization (NEDO) in Japan, led to the development of 'Hazard Evaluation Support System Integrated Platform (HESS)' to provide decision support information to Experts for evaluation of the repeated dose toxicity of chemicals by the category approach. The HESS system has two main databases - the Toxicity Knowledge Information Database which contains repeated dose toxicity test results (>500 reports) and toxicity mechanism information, and Metabolism Knowledge Information Database which contains metabolic maps and kinetic information. HESS also includes support tools for establishing the chemical category for toxicity evaluation as the Category Approach Support Function. HESS is also designed to be compatible with the OECD (Q)SAR Application Toolbox.

The system originally focused on supporting the evaluation of industrial chemical under the 'Chemical Substance Control Law' in Japan. As the system can import additional toxicity test

results of other type of chemicals, the HESS would be applicable to the evaluation of cosmetic ingredients.

4.2.2.2 The Combination Approach:

Different (Q)SAR software systems may generate different evaluation for the same chemical, mainly due to different embedded evaluation rules and algorithms - such as rule-based systems, discriminant-based systems, etc. It was therefore considered that *in silico* evaluation could be optimized by combining the evaluations from multiple systems. A project supported by the Health and Labour Sciences Research Grants Sensitivity, a 'Combination Approach' was developed in Japan for prediction of chemical mutagenicity. The system uses DEREK, MultiCASE, and ADMEWorks, to assesses chemical mutagenicity with increased overall predictively of the three *in silico* systems for a final decision on mutagenicity. This approach has been considered useful in prioritizing the chemicals for testing on the basis of potential mutagenicity among the large numbers of chemical that have no experimental data on mutagenicity.

4.2.3 THE CURRENT USE OF *IN SILICO* METHODS – UNITED STATES

FDA/CFSAN is currently using several commercial software programs to estimate the potential toxicities of organic chemicals, and to construct new QSAR models for toxicological endpoints not covered by the commercial models. These software programs and the commercial QSAR models have been available to FDA/CFSAN under Research Collaboration Agreements (RCAs) with the software vendors. Among these the following seven QSAR and/or rule-based *in silico* programs are routinely used at the FDA:

- CASE-Ultra (Multicase Inc.)
- Leadscope Model Applier (Leadscope, Inc.)
- ADMET-Predictor (Simulations-Plus, Inc.)
- Percepta (Advanced Chemistry Development, ACD, Inc.)
- Symmetry (Prous Institute for Biomedical Research)
- SciQSAR (SciMatics, Inc.)
- Derek Nexus (Lhasa Ltd.)

A brief description of the methods is provided in Annex I.

4.2.4 THE CURRENT USE OF *IN SILICO* METHODS – CANADA

Under the Chemicals Management Plan Substance Grouping Initiative, Health Canada and Environment Canada use the read-across approach to predict endpoint or property information for various substances. Additionally, modelled data may be generated through the use of quantitative structure-activity relationships (QSAR). Such read-across and QSAR data are used both for ecological and human health assessments.

However, in cases specific to cosmetic products, *in silico* approaches have thus far not been used. Health Canada usually requests the highest trade level (e.g. manufacturer or importer) to provide data specifically for the product formulation, and that studies based on key ingredients in the product formulation, or using non-human subjects are not to be considered sufficient evidence of safety. A project in currently underway to use the Threshold of Toxicological Concern (TTC) approach with repeated-dose dermal toxicity data (see Appendix for details).

4.2.5 THE CURRENT USE OF *IN SILICO* METHODS – BRAZIL

As stated in the Brazilian regulations, the safety evaluation of cosmetic products is responsibility of the manufacturer's. That means that Anvisa does not execute any tests by itself, but it is responsible for analyzing all the results provided by a company and to evaluate if they are valid.

There is no specific regulation regarding the use of *in silico* methods for cosmetic products, and this approach has not been fully adopted by most of the companies in Brazil. Recently, the government has implemented some initiatives focused at the development of *in silico* tools (see Annex II), which may alter the way toxicological studies are conducted in Brazilian jurisdiction. The expectation is that these efforts will not only bring scientific benefits, but may also bring advances in the regulatory aspects.

4.2.6 THE CURRENT USE OF *IN SILICO* METHODS – THE COSMETICS INDUSTRY

A preliminary list of *in silico* methods most commonly used by cosmetics industry is provided below. These methods are primarily used by the industry for internal decision-making and screening purposes. Where *in silico* assessments are used for regulatory purposes, they form part of a larger WoE package. The preference is to use open source applications where the assumptions and algorithms are transparent and verifiable. Commercial packages are also used but generally only after extensive verification with historical information.

Method	**Endpoint**								
	Skin Sens.	Skin Irritation	Eye Irritation	Acute Oral	Repeat Dose	Mutagenicity Genotoxicity	Carc.	Dev/ Repro	ADME
Public									
OECD Toolbox	✓	✓	✓	✓	✓	✓	✓	✓	
CAESAR (EU)	✓					✓	✓		
Toxtree (JRC)		✓	✓			✓	✓		
TEST (US EPA)				✓		✓		✓	
Commercial									
DEREK	✓					✓			
METEOR									✓
Topkat	✓	✓	✓	✓	✓	✓	✓	✓	
Multicase							✓		
Academic & Misc									
Kasting Model									✓
TTC						✓	✓		

* Does not include methods such as ACTOR (US EPA) used to gather relevant information from the literature or databases. Also does not include applications such as AIM (US EPA) used to help identify common chemical domains or analogs (AIM).

5.0 Combined use of in silico approaches

The availability of numerous models poses a question which model to use for a given query compound given that often the predicted values between models vary significantly. This is because different models use different statistical methodologies and molecular descriptors, and may have different predictive power towards the same toxicological endpoint. To circumvent arbitrary selection of model, one can combine predictions from several (Q)SAR models and or read-across approaches. Consensus, (or ensemble), modelling as it is sometimes called, takes multiple predictions (from different models) and yields a single prediction value. The theory behind such an approach is that the weaknesses (poor predictions) observed in one model are generally compensated for in another, resulting in a more robust and predictive consensus model. Therefore, combining multiple models is only beneficial when these models are sufficiently different from one another and adequately cover the chemical/biological space. Although reasonable predictive performance is still a prerequisite, models used in this manner should ideally complement each other, i.e. each showing excellent predictivity in unique areas of its applicability domain. Methods include bagging (Breiman, 1996) and boosting (Freund, 1996). In bagging, each model receives equal weighting, whereas in boosting, weighting are used to give more influence to the best performing models. One can also use hierarchical Bayesian models (Harol et al 2009). Numerous other studies have shown model combinations to be more successful than the use of a single model (Abshear et al., 2006; Khedkar et al., 2014). In addition to increased performance the combination of multiple predictions in this manner, forms a weight-of-evidence which can greatly improve the confidence in a correct prediction.

Examples of the current use of combined (Q)SAR systems in ICCR jurisdictions include the use of The Combination Approach in Japan (section 4.2.2). A case study in this regard relates to the combined use of (Q)SAR models by Shiseido (Japan), who have developed a systemic way for in-house assessment of repeated-dose toxicity. This system is designed on the assumption that chemicals which have a chance of exposure to human are in scope of target. This scheme is described as follows:

1. A target chemical is first checked for existing toxicological data using published data set and in-house dataset.
2. The extent of exposure is considered in view of the product type (e.g., leave-on or rinse-off) and percutaneous absorption of the chemical.
3. NOEL is calculated by using the *in silico* model. Shiseido has adopted an artificial neural network (ANN) analysis for calculating NOEL from chemical structure. Over 400 data were used for constructing the *in silico* model. Learning was implemented using QwikNet software ver. 2.23 (Craig Jensen, Sammamish, WA).
4. At the same time, analogous chemicals are identified in the HESS dataset (section 4.2.2) and others. This is based on structural similarity, toxicological alert pattern from DEREK Nexus, intestinal absorption, etc. NOEL values of the analogous chemicals are obtained from HESS.
5. NOEL of the target chemical is determined by referring not only to the calculated NOEL by the ANN model, but also to the NOELs of the analogous chemicals.
6. Finally, MoS of the chemicals is calculated using the estimated NOEL and percutaneous absorption to determine overall safety.

On the (draft[11]) Guideline 'assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk' under the 'International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use' (ICH), an analysis of computational toxicology assessment was introduced in order to predict the mutagenicity of impurities in the new drug products, on which no experimental data was available.

The guideline stated that two types of (Q)SAR prediction models should be applied. One model should be expert rule-based, and the second should be statistical-based. The used (Q)SAR models should conform to the validation principles published by the Organisation for Economic Co-operation and Development (OECD 2007). The absence of structural alerts from the two complementary (Q)SAR models is considered to be sufficient for confirming no concern of mutagenicity of impurities. In order to provide additional supportive evidence on relevance of any predictions and to elucidate underlying reasons in case of conflicting results, any results of computer-based analysis should be reviewed.

Amongst the several published examples of the use of combined (Q)SAR systems is in the Integrated Testing Strategies (ITSs) developed to provide evidence from different information sources such as QSAR models, in vitro and in vivo test results in a WoE approach in the EU FP6 project ORISIS (Rorije et al., 2013). This approach was applied to skin sensitisation endpoint using Bayesian statistics to calculate the probability of the predictability to estimate the reliability of a conclusion on the skin sensitisation potential of a chemical. Rorije et al. (2013) concluded that the use of two or more positive predictions from QSAR models actually gave a higher probability. The calculated Bayesian probability that a substance will test positive in the LLNA test was 85.9%, when three models (DEREKfW, TIMES-SS and the SMARTs rules implemented in the OECD Toolbox) agreed with each other – i.e. either all three gave a positive result or a negative result (280 out of 522 compounds). For the remaining substances, the QSARs yielded conflicting predictions, preventing the WoE conclusion to reach the required 80% or 90% probability, therefore needing additional information to make the WoE sufficient for the REACH regulatory framework.

6.0 Summary and Conclusions

It is evident from the above overview that *in silico* methods and tools for assessment of chemical toxicity have gone through enormous progress over the past few decades. As a result, extensive structure-activity databases, data-mining algorithms, (Q)SAR models, expert systems, read-across methods, and other versatile *in silico* tools are now available in the form of commercial and free-access software platforms. As such, they offer a rapid, cost-effective, and ethical alternative for generating chemical toxicity estimates without the need to test on animals. The *in silico* models and systems currently available for assessment of chemical toxicity cover a wide range of structure-activity space and are equally relevant and valuable for generating toxicity estimates for cosmetic ingredients.

There are a number of drivers for moving away from testing toxicity of cosmetic ingredients in animals – such as the improved predictivity, cost, time, and ethical implications – and for deriving data and estimates by alternative in vitro and *in silico* means. Availability of such alternative methods is more crucial in Europe where animal testing of cosmetic

[11] The draft document is expected to be finalized at the next ICH meeting in the beginning of June, and could be adopted by three authorities (EU, JP, US) as a final document. (The harmonized guideline moves to immediately to the regulatory implementation step.) Therefore, the word of (draft) may be removed from the document in July.

ingredients/products, and marketing of new cosmetic ingredients/products tested on animals, is now banned. However, despite the need and the drivers, the current use of *in silico* approaches is largely limited to internal decision making both at the industry and at the regulatory levels in most ICCR jurisdictions, and has still not been fully adopted as one of the mainstream methods for safety assessment of cosmetic ingredients.

One likely reason is that only a handful of the currently available models and tools have been developed in line with the stringent quality and validation standards to merit their consideration for adoption as a routine alternative to *in vivo* testing for safety assessment. Also, different models and systems sometime tend to yield conflicting results, which makes it difficult for a user to choose a particular *in silico* system in preference to the others, and to rely on its predicted estimates in a safety assessment. This is because each model and system is generally built on a different dataset and algorithm(s), and therefore tends to decipher and interpret the chemical structure and toxicological information in a different way. Each model/system also reflects a different level of uncertainty and variability associated with the data used in developing it, and the modelling process itself, and may also have a different applicability domain within which the predicted estimates of toxicity are reliable. It practice, however, some of these limitations can be addressed through appropriate selection and use of the *in silico* systems. For example, *in silico* assessments may only be derived from those models and systems that meet the stringent quality and validation standards, and have clearly defined applicability domains. Although this may limit the number of workable *in silico* tools, it should provide more confidence in the predicted estimates. It may also be possible to add more confidence to *in silico* assessments through the use of a combination of appropriate (Q)SAR models, expert systems, and/or read-across approaches, rather than relying on a single model/system. A 'weight of evidence' gathered this way should provide sufficient basis for a reliable *in silico* assessment of chemical toxicity for safety evaluation.

In this regard, it also needs to be emphasised that the development and use of *in silico* methods require a thorough consideration of the quality of the data and the algorithms behind a system, the amount of testing/validation carried out, and the adequacy of the chemical structure and biological activity space covered by the model's applicability domain. This inevitably requires expert knowledge and a degree of skilled approach based on a good understanding of toxicology and (bio)chemistry. Thus, despite the ease of use offered by certain advanced *in silico* platforms, the assessment of toxicity by *in silico* approaches must not be reduced to a 'black box' routine.

Further work in this area is recommended in relation to development of a uniform and standardised approach that allows the selection and use of appropriate *in silico* system(s), and interpretation of the results from a cosmetics safety assessment perspective. In addition, an appropriate framework also needs to be developed that allows integration of different *in silico* approaches in a consistent scheme that allows sufficient weight of evidence to be gathered against relevant toxicological endpoints for use in safety assessment of cosmetic ingredients.

7.0 References

Abshear T., Banik G.M., D'Souza M.L., Nedwed K. and Peng C. (2006) A model validation and consensus building environment. SAR QSAR Environ. Res. 17, 311-321.

Benfenati, E. (2012) E-Book: Theory, guidance and applications on QSAR and REACH, ISBN 978-88-902405-4-6; Orchestra Project website: http://www.orchestra-qsar.eu/documents/333.

Benigni R, Bossa C, Jeliazkova N, Netzeva T & Worth A (2008). The Benigni / Bossa rulebase for mutagenicity and carcinogenicity – a module of Toxtree. EUR 23241 EN. http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/publications/

Breiman, L. (1996) Bagging Predictors. Machine Learning, 123—140.

Chaudhry, Q., Chrétien, J., Craciun, M., Guo, G., Lemke, F., Müller, J-A, Neagu, N. Piclin, N., Pintore, M., Trundle, P. (2007) Chapter 4: Algorithms For (Q)SAR Model Building; in Quantitative Structure-Activity Relationship (QSAR) for Pesticide Regulatory Purposes, Benfenati, E. (Ed), pp 111, Elsevier. ISBN 13: 978-0444-52710-3.

Cotterill, J.V., Chaudhry, M.Q., Matthews, W., Watkins, R. (2008) *In silico* assessment of toxicity of heat-generated food contaminants, Food and Chemical Toxicology 46: 1905 –1918.

Dearden JC, Barratt MD, Benigni R, Bristol DW, Combes RD, Cronin MTD, Judson PN, Payne MP, Richard AM, Tichy M, Worth AP & Yourick JJ (1997). The Development and Validation of Expert Systems for Predicting Toxicity. Alternatives to Laboratory Animals 25, 223-252.

Freund, Y. and Schapire, R.E. (1996) Experiments with a New Boosting Algorithm, Machine Learning: Proceedings of the Thirteenth International Conference, 1996, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.7636&rep=rep1&type=pdf

Harol A., J. Jaworska[1], M. Hewitt[2], M.Cronin Bayesian Integration of QSAR Models to Predict Aqueous Solubility. SETAC 2009 poster

Hayashi M, Kamata E, Hirose A, Takahashi M, Morita T, Ema M. (2005) *In silico* assessment of chemical mutagenesis in comparison with results of Salmonella microsome assay on 909 chemicals. Mutat Res. 588: 129-135.

Hayashi, M and Sakuratani, Y. (2011): Development of an evaluation support system for estimating repeated-dose toxicity of chemicals based on chemical structure. In New Horizons in Predictive Toxicology: Current Status and Application (Res Drug Discovery) (Wilson, A.G.E., Ed.) Chap. 3, Royal Society of Chemistry, Cambridge.

Hayashi, M and Sakuratani, Y. (2011): Development of an evaluation support system for estimating repeated-dose toxicity of chemicals based on chemical structure. In New Horizons in Predictive Toxicology: Current Status and Application (Res Drug Discovery) (Wilson, A.G.E., Ed.) Chap. 3, Royal Society of Chemistry, Cambridge.

ICH guideline (M7), Assessment and Control of DNA Reactive (Mutagenic) Impurities in Pharmaceuticals to Limit Potential Carcinogenic Risk (draft consensus guideline. Step 2 Version. Feb. 2013). Available from: www.ich.org/products/guidelines/multidisciplinary/article/multidisciplinary-guidelines.html.

Khedkar VM, Joseph J, Pissurlenkar R, Saran A, Coutinho EC; (2014) How good are ensembles in improving QSAR models? The case with eCoRIA. J Biomol Struct Dyn. 2014 Apr 22:1-21. [Epub ahead of print].

Lipnick RL (1999) Correlative and mechanistic QSAR models in toxicology. SAR QSAR Environ Res. 10(2-3): 239-248.

Malazizi, L., Neagu, D., Chaudhry, Q. (2006) A Data Quality Assessment Algorithm with Application in Predictive Toxicology, TASK Quarterly Journal (Scientific Bulletin of Academic Computer Center in Gdansk, Poland), ISSN 1428-6394, volume 11, number 1-2/2007 Advances in Artificial Intelligence, www.task.gda.pl/files/quart/TQ2007/01-02/tq111p-g.pdf.

Netzeva TI, Worth AP, Aldenberg T, Benigni R, Cronin MTD, Gramatica P, Jaworska JS, Kahn S, Klopman G, Marchant CA, Myatt G, Nikolova-Jeliazkova N, Patlewicz GY, Perkins R, Roberts DW, Schultz TW, Stanton DT, van de Sandt JJM, Tong W, Veith G, Yang C 2005. Current status of methods for defining the applicability domain of (Quantitative) Structure-Activity Relationships. ATLA 33: 155-173.

OECD (2004) OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models, available at www.oecd.org/env/ehs/risk-assessment/37849783.pdf.

OECD, Environment Health and Safety Publications Series on Testing and Assessment No. 58: Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models in the Assessment of New and Existing Chemicals. ENV/JM/MONO(2006)25, OECD, Paris, France 2007. Available from: http://www.oecd.org/chemicalsafety/risk-assessment/38131728.pdf/

Patlewicz G, Dimitrov SD, Low LK, Kern PS, Dimitrova GD, Comber MIH, Aptula AO, Phillips RD, Niemela J, Madsen C, Wedebye EB, Roberts DW, Bailey PT & Mekenyan OG (2007). TIMES-SS – A promising tool for the assessment of skin sensitization hazard. A characterisation with respect to the OECD Validation Principles for (Q)SARs. Regulatory Toxicology and Pharmacology 48, 225–239.

Ridings JE, Barratt MD, Cary R, Earnshaw CG, Eggington CE, Ellis MK, Judson PN, Langowski JJ, Marchant CA, Payne MP, Watson WP & Yih TD (1996). Computer prediction of possible toxic action from chemical structure: an update on the DEREK system. Toxicology 106(1-3), 267-279.

Rogiers V. and Beken S. (Editors and authors) (2000) Alternative Methods to Animal experiments - Actual status, development and approach in Belgium, VUB Press, Brussels. ISBN 90-5487-264-0.

Rorije, E., Aldenberg, T., Buist, H., Kroese, D., Schüürmann, G. (2013) The OSIRIS Weight of Evidence approach: ITS for skin sensitisation, Regulatory Toxicology and Pharmacology 67: 146–156.

Sanderson DM & Earnshaw CG (1991). Computer prediction of possible toxic action from chemical structure; the DEREK system. Human and Experimental Toxicology 10, 261-273.

Smithing MP & Darvas F (1992). Hazardexpert:an expert system for predicting chemical toxicity. In: Food Safety Assessment (JW Finlay, SF Robinson and DJ Armstrong). Washington, American Chemical Society: 191-200.

Woo Y-T & Lai DY (2005). OncoLogic: A mechanism-based expert system for predicting the carcinogenic potential of chemicals. In Predictive Toxicology (C Helma, ed), 385-413.

Yamada T, Tanaka Y, Zhang H, Hasegawa R, Sakuratani Y, Mekenyan O, Yamazoe Y, Yamada J, Hayashi M. (2012): A category approach to predicting the hemolytic effects of ethylene glycol alkyl ethers in repeated-dose toxicity. J Toxicol Sci. 37: 503-515.

8.0 ACRONYMS

3Rs	Refinement, Reduction, Replacement of the use of animals in laboratory procedures
Alternative methods	All those procedures which can completely replace the need for animal experiments, which can reduce the number of animals required, or which can reduce the amount of pain and stress to which the animal is subjected in order to meet the essential needs of humans and other animals [Rogiers and Beken, 2000]
ECVAM	European Centre for the Validation of Alternative Methods
EFSA	European Food Safety Authority
In silico methods	Computational approaches that use (quantitative) structure-activity relationship modelling, and read-across between substances on the basis of structural or functional similarities.
In vitro test methods	Biological method that uses organs, tissue sections and tissue cultures, isolated cells and their cultures, cell lines and subcellular fractions, or non-biological method that uses chemical interaction studies, receptor binding studies, etc [Rogiers and Beken 2000]
LLNA	Local Lymph Node Assay
OECD	Organisation for Economic Co-operation and Development
QSAR	Quantitative Structure-Activity Relationship
Read-across	A method to derive biological activity (e.g. toxicity) values for a target chemical structure from experimental values on structurally- and/or functionally similar chemical analogues.
REACH	Registration, Evaluation, Authorisation and restriction of Chemicals
SAR	Structure-Activity Relationship
SCCS	Scientific Committee on Consumer Safety
WoE	Weight of Evidence

ANNEX–I: RELEVANT IN SILICO METHODS AND TOOLS

OECD QSAR TOOLBOX

Methodology: Categorisation, Read-Across, (Q)SAR

Area of Applicability: A wide range of chemicals

Brief Description: The OECD QSAR Toolbox is a stand-alone software system intended to be used by governments, industry and other stakeholders to fill gaps in toxicity data needed for assessing the hazards of organic chemicals for environmental and human health endpoints. The Toolbox incorporates information and "tools" from various sources into a logical workflow. The Toolbox has multiple functionalities allowing the user to perform a number of operations (e.g., categorize large inventories of chemicals according to mechanisms of action, using or building QSAR models, read-across from structurally related analogous compounds that have already been tested).

The primary function of the Toolbox is to place discrete organic substances into chemical categories for outcomes of regulatory interest and use data from tested category members to aid in filling data gaps for untested category members. It is designed to use 2D structures and chemical properties to group chemicals, including the query substance, into endpoint-specific, meaningful categories, which include tested compounds that are all "similar" in a transparent manner. Thus, a category is constructed on chemical and biological mechanistic information but data gaps are filled by using in vivo test data within the category. The Toolbox contains: 1) Databases with results from experimental studies; 2) Accumulated knowledge for structural characteristics (alerts) that can indicate the presence of hazards and other properties, and 3) Apparatus to estimate missing experimental values.

It is important to remember that the 3 main features of Toolbox are: 1) Identification of relevant structural characteristics and potential mechanism or mode of action of a target chemical, 2) Identification of other chemicals that have the same structural characteristics and/or mechanism or mode of action, and 3) Use of existing experimental data to fill data gap(s).

The Toolbox workflow is a sequence of modules; 1) input, 2) profiling, 3) endpoints, 4) category definition, 5) data gap filling, and 6) reporting. The aim of the module "Filling Data Gap" is to give access to three different data gap filling tools; 1) Read-across, 2) Trend analysis, and 3) QSAR models. Read-across and trend analysis both use the available results in the category data matrix to fill a data gap. QSAR models gives access to a library of QSAR models which have been integrated into the Toolbox. With the module "Report", the user can generate reports on any of the predictions performed with the Toolbox. The Toolbox contains a number of predefined report templates as well as a template editor with which a user-defined template can be formatted.

Version 3.2 of the Toolbox, released at the end phase 2 developments in November 2012, includes, in addition to a number of general improvements to previous versions, the ability for direct data exchange with IUCLID 5.5. The Toolbox provides a module for the applications of adverse outcome pathways (AOPs) to chemical category formation. This pathway approach is based on the concept that toxicity results from a chemical first reaching and then interacting with an initial key target (e.g., membrane, receptor) in the organism; this is defined as the primary molecular initiating event. Further from this primary interaction is a series of other key events; all such key events can be profiled in an AOP workflow within the Toolbox and linked to databases within the Toolbox which have data from test methods that quantify the AOP-specific key events.

Reference: www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm

CRAMER DECISION TREE

Methodology: Expert System (decision tree) based on structural alerts and expert knowledge

Area of Applicability: Food, drug, cosmetics, pesticides, biocides

Brief Description: The Cramer decision tree is probably the most commonly used approach for classifying and ranking chemicals on the basis of their oral toxicity. It consists of 33 questions. Each 'yes' or 'no' answer leads to a further question, or to the final classification into one of the three classes:

Class I: Substances with simple chemical structures and for which efficient modes of metabolism exist, suggesting a low order of oral toxicity.

Class II: Substances which possess structures that are less innocuous than class I substances, but do not contain structural features suggestive of toxicity like those substances in class III.

Class III: Substances with chemical structures that permit no strong initial presumption of safety or may even suggest significant toxicity or have reactive functional groups.

The logic of the sequential questions was based on the then available knowledge on toxicity and on how chemical structures are metabolised in mammalian metabolic pathways. The questions relate mostly to chemical structure, but natural occurrence in the body and in food are also taken into consideration. The decision tree is intended for use with all ingested, structurally-defined organic substances.

Cramer Decision Tree was first described by Cramer et al (1978). It is now implemented in Toxtree and OECD QSAR Toolbox. The approach has been evaluated by JRC (Lapenna & Worth, 2011). It's use is recommended by EFSA in its opinion on TTC, with the recommendation that Cramer class II compounds should be treated as class III (EFSA 2012). It is also recommended by the European Commission's non-food scientific committees in their opinion on TTC, again with the caveat that Cramer class II compounds should be treated as class III (EC, 2012). It is widely used for the assessment and prioritisation of low level contaminants that are not subject to regulatory reporting requirement (e.g. contaminants in food and drinking water, metabolites and degradates of pesticide actives, etc).

Based on a survey of Toxtree/ Cramer users by the JRC (Lapenna et al., 2011), recommendations were made to address the following limitations:

Many of the original Cramer rules are written in a confusing and interdependent way, which leads to difficulties in rationalisation of the predictions. They should be rewritten in a clearer way, possibly with modification and re-ordering. Two of the rules are not based on chemical features, but simply make reference to look-up lists of chemicals (Q1, normal body constituents; Q22, common food components). These two questions could either be omitted, or the lists of chemicals extended with e.g. recently authorised food additives after peer-reviewing. The assessment of whether specified substances can be "generally regarded as safe" should be carried out at a different step in the overall TTC assessment scheme. Some rules make ambiguous references to chemical features (e.g. steric hindrance) which need to be clarified and revised.

References:

Cramer GM, Ford RA & Hall RL (1978). Estimation of toxic hazard--a decision tree approach. Food and Cosmetics Toxicology 16, 255-276.

EC (2012). Opinion on the Use of the Threshold of Toxicological Concern (TTC) Approach for Human Safety Assessment of Chemical Substances with focus on Cosmetics and Consumer Products. SCHER/SCCS/SCENIHR Available from: http://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_092.pdf

EFSA (2012). Scientific Opinion on exploring options for providing advice about possible human health risks based on the concept of Threshold of Toxicological Concern (TTC) *EFSA Journal* 2012;10(7):2750. European Food Safety Authority. Available from: http://www.efsa.europa.eu/en/efsajournal/doc/2750.pdf

Lapenna S & Worth A (2011). Analysis of the Cramer classification scheme for oral systemic toxicity - implications for its implementation in Toxtree. JRC report EUR 24898 EN. Publications Office of the European Union, Luxembourg. Available at: http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/publications/

BENIGNI-BOSSA RULE BASE

Methodology: Expert System based on structural alerts and QSAR for genotoxicity and carcinogenicity

Area of Applicability: Food, Drug, Cosmetics, Industrial Chemicals

Brief Description: The Benigni-Bossa rule base makes generic predictions of both the genotoxic and non-genotoxic carcinogenicity potential of chemicals, based on the rules published in the EC report (Benigni et al., 2008: The Benigni/Bossa rulebase for mutagenicity and carcinogenicity – a module of Toxtree). The present list of structural alerts (over 30) refers mainly to genotoxic carcinogenicity, and also includes a number of structural alerts for potential non-genotoxic carcinogens. In addition to the alerts, the rule base includes optional QSAR models for:

i) mutagenic activity in Salmonella typhimurium TA100 strain (Ames test) of aromatic amines (QSAR6), and alpha,beta-unsaturated aldehydes (QSAR13), and
ii) carcinogenicity in rodents of aromatic amines (QSAR8).

The underlying mechanism(s) for triggering of an alert for genotoxic carcinogenicity are not clearly defined, as these may include different possible mechanisms of genotoxicity (e.g. Ames mutagenicity, chromosomal aberration, chromosomal instability, etc.) which may be linked to carcinogenicity.

The Benigni-Bossa rule base was first described in JRC report (Benigni et al., 2008). QMRF for the rule base are available in the JRC QSAR database (http://qsardb.jrc.ec.europa.eu/qmrf/). It is implemented in Toxtree and OECD QSAR Toolbox, and used in the Kroes et al TTC decision tree which is implemented in Toxtree. The rule base is evaluated by JRC in a study for EFSA (Worth et al, 2010a, 2010b). It has also been evaluated in scientific literature by Hillebrecht et al. (2011); and Bakhtyari et al. (2013).

References:

Bakhtyari NG, Raitano G, Benfenati E, Martin T, Young D (2013).Comparison of *in silico* models for prediction of mutagenicity. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev. 31(1): 45-66.

Benigni R, Bossa C, Jeliazkova N, Netzeva T & Worth A (2008). The Benigni / Bossa rulebase for mutagenicity and carcinogenicity – a module of Toxtree. EUR 23241 EN. http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/publications/

Hillebrecht A, Muster W, Brigo A, Kansy M, Weiser T, Singer T (2011). Comparative evaluation of *in silico* systems for ames test mutagenicity prediction: scope and limitations. Chemical Research in Toxicology 24(6):843-854.

Worth A, Lapenna S, Lo Piparo E, Mostrag-Szlichtyng A & Serafimova R (2010a). The applicability of software tools for genotoxicity and carcinogenicity prediction: case studies relevant to the assessment of pesticides. JRC Technical report EUR 24640 EN. Available at: http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/publications/

Worth AP, Fuart-Gatnik M, Lapenna S, Lo Piparo E, Mostrag-Szlichtyng A & Serafimova R (2010b). Applicability of QSAR analysis to the evaluation of the toxicological relevance of metabolites and degradates of pesticide active substances for dietary risk assessment. Final report of a project carried out by the JRC for the European Food Safety Authority (EFSA). Available at: http://www.efsa.europa.eu/en/scdocs/scdoc/50e.htm

CASE-ULTRA (VERSION 1.4.7.0)

Methodology: Molecular fragment-based, QSAR expert system that generates predictive models automatically by using machine learning techniques from training data sets of non-congeneric compounds associated with biological/toxicological data

Area of Applicability: food, drug, cosmetics, pesticides, biocides

Brief Description: *CASE-Ultra* (version 1.4.7.0) is the latest generation Windows based program from MultiCASE Inc. (www.multicase.com), which is mainly influenced by *MCASE* (Multiple Computer Automated Structure Evaluation) methodology (previously known as *CASE*, *MC4PC* and *CaseTox*). The program can be used to estimate the toxicities of organic chemicals, and to construct new QSAR models (modules). It is a molecular fragment-based, QSAR expert system that generates predictive models automatically by using machine learning techniques from training data sets of non-congeneric compounds associated with biological/toxicological data. - *CASE-Ultra* first examines SMILES codes of active chemicals in the training set and breaks them into all possible linear or branched (non-hydrogen) atom molecular fragments -, then identifies a set of structural alerts that are statistically related to the activity. These fragments are labelled as positive alerts or biophores. The positive alerts are the main building blocks of the QSAR model and are responsible for identifying active chemicals during prediction. The significance of each fragment towards activity is determined using a combination of two-objective criteria comprised of the Shannon's entropy as a fitness measure and the number of the active training set molecules containing this fragment. A set of top fragments (based on the aforesaid two-objective criteria) is selected in order to cover all the active chemicals of the training set. In addition, deactivating alerts are also developed and selected using a similar process but by scanning inactive chemicals and finding fragments that occur mainly in inactive chemicals. *CASE-Ultra* then calculates various descriptors for the training set chemicals that include log*P*, water solubility, molar refractivity, surface area descriptors, vapour pressure, Gasteiger charges, E-State descriptors, fragment descriptors and others. Once the final sets of the positive and deactivating alerts are identified as well as QSAR descriptors are calculated, *CASE-Ultra* attempts to build separate local QSARs for each positive alert using a stepwise regression method in order to explain the variation in activity within the training set chemicals covered by that alert. It has a fully automated procedure for calculating the predictive performance of the QSAR models that computes LMO and LOO cross-validation statistics (Matthews and Contrera, 1998; Chakravarti et al., 2012; Saiakhov et al., 2013).

During *CASE-Ultra* prediction from a particular model, a test chemical is scanned through the list of the model's positive and deactivating alerts and if no positive alerts could be identified, the test chemical is considered to have no evidence of activity and is presumed to be inactive. The presence of a deactivating alert along with a positive alert also predicts the test chemical as inactive. In general, a test chemical is predicted active if it contains either one or more positive alerts. However, this active prediction can be changed if the local QSAR of the positive alert(s) modifies the prediction. On the contrary, if a test chemical contains a positive alert that is present in just one or two active training set chemicals (low statistical confidence and also no local QSAR); the prediction is considered as inconclusive. The program also identifies unusual features/fragments in test chemicals that do not match the training data (unknown structural fragments) and thus explains the domain of applicability (DoA) for overall model. A test chemical is considered as an out of the DoA if it has three or more unknown structural fragments. The prediction coverage of the test set is defined as the percentage of the test chemicals as predicted either an active or inactive (Chakravarti et al. 2012). CASE Ultra is used to predict a single test chemical or a very large test set chemicals (in batch mode) and the results can be exported in the form of a report with varying levels of details. This program provides the following information/options to the users: (i) details about the model, (ii) alerts – statistical significance, average activity, molecules containing particular alert and active molecules containing particular alert, (iii) QSAR equation(s), (iv) validation

statistics, (v) training set chemicals – except proprietary chemicals, (vi) option to add more chemicals to the training set, reprocess, and revalidate the model, and so on. The current version of the *CASE-Ultra* has more than 450 modules that have been undergone internal validation. The models cover various areas of toxicology and pharmacology that includes the followings: Acute toxicity, ADME, Adverse effects in humans, allergies, antibacterial, anti-HIV, carcinogenicity, cytotoxicity, developmental toxicity, eco-toxicity, enzyme inhibition, genetic toxicity, teratogenicity, *etc*.

CASE-Ultra has been evaluated in scientific literature (Saiakhov, et al. 2013; Chakravarti, et al. 2012; Matthews, et al. 2008; Matthews and Contrera, 1998).

References:

Chakravarti, S. K.; Saiakhov, R. D.; Klopman, G. Optimizing Predictive Performance of CASE Ultra Expert System Models Using the Applicability Domains of Individual Toxicity Alerts. J. Chem. Inf. Model. 52: 2609-2618, 2012.

Matthews, E. J.; Contrera, J. F. A new highly specific method for predicting the carcinogenic potential of pharmaceuticals in rodents using enhanced MCASE QSAR-ES software. Regul. Toxicol. Pharmacol. 28:242–264, 1998.

Saiakhov, R.; Chakravarti, S.; Klopman, G. Effectiveness of CASE Ultra Expert System in Evaluating Adverse Effects of Drugs. Mol. Inf. 32: 87-97, 2013.

Matthews, E.J.; Kruhlak, N. L.; Daniel Benz, R.; Contrera, J. F.; Marchant, C. A.; Yang, C. (2008) Combined Use of MC4PC, MDL-QSAR, BioEpisteme, Leadscope PDM, and Derek for Windows Software to Achieve High-Performance, High-Confidence, Mode of Action–Based Predictions of Chemical Carcinogenesis in Rodents. Toxicol. Mech. Methods, 18:189–206.

LEADSCOPE MODEL APPLIER (LSMA; VERSION 1.7.4-1)

Methodology: Molecular fragment based QSAR system

Area of Applicability: Food, drug, cosmetics, pesticides, biocides

Brief Description: *Leadscope Model Applier* (LSMA; version 1.7.4-1) is a latest version Windows based and commercially available program of Leadscope, Inc. (www.leadscope.com). It is a molecular fragment based QSAR system that uses chemo-informatics approach and 2D features to evaluate chemical compounds and their likelihood to associate with a modeled toxicity endpoint. It must be noted that this program is used only for the toxicity prediction from the pre-developed QSAR models already in the system and cannot be used to construct a QSAR model. However, QSAR models can be developed by using other program called the *Leadscope Enterprise* (LSE; version 3.1.1). LSE classifies chemical structures using a library of more than 27,000 predefined sub-structures (finger prints) that represent functional groups, heterocycles and pharmacophores organized in a chemical hierarchy that serves as a knowledge-base to facilitate compound grouping. LSE has a QSAR model builder that uses partial logistic regression (PLR) algorithm, 2D structural features and eight calculated molecular descriptors in order to develop classification models. It has a fully automated procedure for calculating the predictive performance of QSAR models in random LMO and LOO cross-validation experiments. The DoA of a QSAR model is defined by analyzing the similarity of structural features between the test chemical and the training data set chemicals (Cross et al., 2003; Yang et al., 2004; Matthews et al., 2008).

The current version of the LSMA program provides more than 90 modules in various areas of toxicology that covered under following toxicity suites: developmental toxicity, genetic toxicity, human adverse cardiological effects, human adverse hepatobiliary effects, human adverse urinary effects, neurotoxicity, reproductive toxicity, rodent carcinogenicity, etc. During LSMA prediction from a particular model, program compares the structural features of the training data set with the features of test chemicals, and then identifies and reports the prediction for test chemicals including the number of test chemicals that cannot be predicted and are not within the DoA of the model. The prediction coverage of the test set is defined as the percentage of the test chemicals as predicted either an active or inactive. The LSMA program is used to predict a single test chemical or a very large test set chemicals (in batch mode) and the results can be exported in the form of a report with varying levels of details including prediction explanation. This program provides the following information of a particular model and about the test chemical prediction results to the users: (i) model descriptions including validation results, (ii) highlights the number of chemicals with a positive predicted probability of >= 0.5, (iii) 0/1 distribution of the training set i.e. percentage of training set chemicals having the feature that are zero and those that are one, (iv) probability distribution of model results for those training set chemicals having particular feature, (v) prediction calls for test chemical(s) including positive, negative, not-in-domain or missing descriptors, (vi) explanation of the prediction result(s), and so on.

The programme has been evaluated in scientific literature (Matthews, et al. 2008; Frid and Matthews, 2010).

References:

Cross, K. P.; Myatt, G.; Yang, C.; Fligner, M. A.; Verducci, J. S.; Blower, P. E. Finding discriminating structural features by reassembling common building blocks. J. Med. Chem. 46:4770–4775, 2003.

Matthews, E.J.; Kruhlak, N. L.; Daniel Benz, R.; Contrera, J. F.; Marchant, C. A.; Yang, C. (2008) Combined Use of MC4PC, MDL-QSAR, BioEpisteme, Leadscope PDM, and Derek for Windows Software to Achieve High-Performance, High-Confidence, Mode of Action–Based Predictions of Chemical Carcinogenesis in Rodents. Toxicol. Mech. Methods, 18:189–206.

Yang, C.; Cross, K.; Myatt, G. J.; Blower, P. E.; Rathman, J. F. (2004) Building predictive models for protein tyrosine phosphatase 1B inhibitors based on discriminating structural features by reassembling medicinal chemistry building blocks. J. Med. Chem. 47:5984–5994.

Frid, A. A.; Matthews, E. J. Prediction of drug-related cardiac adverse effects in humans-B: Use of QSAR programs for early detection of drug-induced cardiac toxicities. Regul. Toxicol. Pharmacol. 56:276–289, 2010.

ADMET PREDICTOR™ 6 (VERSION 6.5.0013)

Methodology: *In silico* predictor for estimation of absorption, distribution, metabolism, elimination, and toxicity (ADMET) properties of molecules based on molecular structures and structural descriptors.

Area of Applicability: Food, drug, cosmetics, pesticides, biocides

Brief Description: *ADMET Predictor™* 6 (version 6.5.0013) is the latest version of Windows program available commercially from Simulations Plus, Inc. (www.simulations-plus.com). *ADMET Predictor* is designed to estimate absorption, distribution, metabolism, elimination, and toxicity (ADMET) properties of molecules based on their molecular structures, employing individual models for specific ADMET endpoints. It is a molecular descriptor-based QSAR system that calculates 2D (and 3D, if 3D structures are input) molecular and atomic-level descriptors of test molecules and provides prediction of various ADMET endpoints that include (i) Physico-Chemical and Biopharmaceutical Module: ionization constants (pKa), human effective permeability in jejunum (Peff), MDCK apparent permeability (Papp), corneal permeability, human skin permeability, solubility, logP, logD, molal volume, blood-brain barrier permeation, human plasma protein binding, human volume of distribution, blood-to-plasma concentration ratio, inhibition of HIV integrase, etc., (ii) Metabolism Module: likely sites of metabolic attack by five CYP P450 enzymes, classification of a molecule whether it will be a substrate of one of the five CYP P450 enzymes, Michaelis-Menten kinetic (K_m and V_{max}) constants for hydroxylation reaction catalyzed by five CYP P450 enzymes, intrinsic clearance (CL_{int}) resulting from metabolic activity of five CYP P450 enzymes, classification of a molecule whether it will be glucuronidated by one of the nine isoforms of the Uridine 5'-Diphosphate-Glucuronosyl-transferase (UGT), etc., (iii) Toxicity Module: estrogen and androgen receptor toxicities (qualitative and quantitative), maximum recommended therapeutic dose (MRTD), acute toxicity, carcinogenicity, mutagenic chromosomal aberrations, phospholipidosis, reproductive/developmental toxicity, skin sensitization, hERG-encoded K+ channel affinity, human liver adverse effects of drugs, etc., (iv) Simulation Module: fraction absorbed in human (by simulation at 1 mg, 10 mg, 100 mg, and 1000 mg dose levels) and optimal dose in human (in mg) matching desired efficacious concentration in blood plasma, (v) Customizable ADMET Risk Filters: risk of low absorption from an oral dose, risk of mutagenicity, risk of overall toxicity, risk of metabolic liability, and global ADMET risk summarizing all available ADMET endpoints prediction in one score (ADMET Risk). The ADMET Predictor automatically determines whether a test compound is within the Domain of Applicability (DoA), or outside the DoA, of the model by incorporating minimum and maximum values of each of their descriptor in the training set plus a 10% tolerance. If a test molecule is outside the DoA of the molecular descriptors then the ADMET Predictor evaluates it as outside the DoA and color codes (magenta color) in the prediction. This result is also exported separately in terms of binary values (1s and 0s, where 1 represents outside the DoA and 0 represents within the DoA).

ADMET-Predictor also has a sub-program called the *ADMET Modeler™,* which provides model-building functionality in order to create both classification and regression models using a collection of methodologies such as artificial neural network ensemble (ANNE) and support vector machine (SVM) are used for both the classification and regression analysis, while kernel partial least square (KPLS) and multiple linear regression (MLR) methods are used only for the regression analysis (Agoram et al., 2001; Bolger et al., 2009; Fraczkiewicz et al., 2009; Choi et al., 2013).

The programme has been evaluated in scientific literature by Choi et al. (2013).

References:

Agoram, B.; Woltosz, W.S.; Bolger, M.B. Predicting the impact of physiological and biochemical processes on oral drug bioavailability. Adv. Drug Deliv. Rev. 50(Suppl 1): S41-67, 2001.

Bolger, M.B.; Lukacova, V.; Woltosz, W.S. Simulations of the nonlinear dose dependence for substrates of influx and efflux transporters in the human intestine. AAPS J.: 11(2):353-63, 2009.

Fraczkiewicz, R.; Zhuang, D.; Zhang, J.; Miller, D.; Woltosz, W.S.; Bolger, M.B. Busting the Black Box Myth: Designing Out Unwanted ADMET Properties with Machine Learning Approaches. CICSJ Bulletin 27(4):96-102, 2009.

Choi, S.S.; Kim, J.S.; Valerio(Jr.), L.G.; Sadrieh, N. *In silico* modeling to predict drug-induced phospholipidosis. Toxicol. Appl. Pharmacol. 269, 195–204, 2013.

PERCEPTA

Methodology: Molecular-descriptor and molecular-fragment based QSAR system that generates predictive models automatically by using Pharma Algorithms from training data sets of non-congeneric compounds associated with physical/ biological/toxicological data.

Area of Applicability: Food, drug, cosmetics, pesticides, biocides

Brief Description: *Percepta* is a Windows based computational program that is available commercially by Advanced Chemistry Development Labs, Inc. (ACD/Lab; www.acdlabs.com). It is a molecular-descriptor and molecular-fragment based QSAR system that is designed to estimate certain physicochemical and ADME properties, and toxicity endpoints using a collection of (Q)SAR modules. These modules have been built using the *Algorithm Builder*™ software developed by Pharma Algorithms (now merged with ACD Labs). Prediction modules of *Percepta* are actually a bundle of four suites such as ACD/Impurities Package for Toxicity, ACD/PhysChem Suite, ACD/ADME Suite, and ACD/Tox Suite. The program calculates fragment, molecular and atomic descriptors of test molecules using a variety of input parameters (e.g. name, 2D structure, or SMILES) and provides prediction results that include:

- PhysChem Modules: Absolv (hydrogen bonding acidity, hydrogen bonding basicity, polarity/polarizability, partitioning coefficient between gas phase and hexadecane, McGowan volume and excessive molar fraction), aqueous solubility, boiling point, logP, logD, pKa, sigma, and other PhysChem descriptors (e.g. H-bond donors and acceptors, freely rotatable bonds, density, molar refractivity, molar volume, parachor, rule-of-5, surface tension, etc.);
- ADME Modules: Blood-brain barrier permeation, cytochrome P450 inhibitors, cytochrome P450 substrates, distribution, maximum recommended daily dose, oral bioavailability, passive absorption, P-gp specificity, regioselectivity of metabolism; and
- Toxicity Modules: Acute toxicity, aquatic toxicity, endocrine system disruption, genotoxicity, health effects, hERG inhibition, eye and skin irritation.

Users can evaluate predictions either in a single structure or spreadsheet view. Spreadsheet view offers the additional capability to view predictions from all available modules in one screen, and a number of graphing, sorting, and filtering tools to rank compounds and aid evaluation. In this program, each module provides both a probability of the test chemical being active or inactive, as well as a reliability index (RI) for the prediction, which evaluates respective confidence in the prediction. The RI includes an assessment of the DoA as well as the relative confidence in the probability of the prediction (Lanevskij et al., 2009; Lanevskij et al., 2011; Reynolds et al., 2009).

The programme has been evaluated in scientific literature (Lanevskij, et al. 2009; Reynolds, et al. 2009; Lanevskij, et al. 2011).

References:

Lanevskij, K.; Japertas, P.; Didziapetris, R.; Petrauskas, A. DRUG DISCOVERY INTERFACE: Ionization-Specific Prediction of Blood–Brain Permeability. J. Pharm. Sci. 98(1), 122-134, 2009.

Lanevskij, K.; Dapkunas, J.; Juska, L.; Japertas, P.; Didziapetris, R. QSAR Analysis of Blood–Brain Distribution: The Influence of Plasma and Brain Tissue Binding. J. Pharm. Sci. 100(6), 2147-2160, 2011.

Reynolds, D.P.; Lanevskij, K.; Japertas, P.; Didziapetris, R.; Petrauskas, A. Ionization-Specific Analysis of Human Intestinal Absorption. J. Pharm. Sci. 98(11), 4039-4054, 2009.

SYMMETRY®

Methodology: Descriptors-based, QSAR system that generates predictive models automatically from training data sets of non-congeneric compounds associated with biological/toxicological data by using three different algorithms, such as, logistic regression, similarity regression and combined regression.

Area of Applicability: Food, drug, cosmetics, pesticides, biocides

Description: Symmetry® is the next-generation Windows-based QSAR software that replaces BioEpisteme® and is available from the Prous Institute for Biomedical Research, Barcelona, Spain (www.prousresearch.com).This program employs a bi-functional system that includes a module for test set predictions and a module for QSAR model development. For descriptor calculations, the program utilizes the FDA Mold2 descriptor package (http://www.fda.gov/ScienceResearch/BioinformaticsTools/Mold2/ucm144528.htm) that calculates a diverse set of 777 two-dimensional molecular descriptors from chemical structure information. Symmetry employs three different algorithms for the development of classification models, such as, logistic regression, similarity regression and combined regression. The combined algorithm works comparable well, which involves both the logistic regression and molecular descriptor similarity. The program utilizes three different methods for the selection of descriptors, such as, correlation-based feature selection (CFS), genetic algorithm for CFS, and wrapper descriptor selection. Duplicates are removed automatically from the training set, while outliers are eliminated at >3 inter-percentile range from lower and upper percentiles. Symmetry employs a synthetic minority over-sampling technique (SMOTE) that uses the molecular structures trained with the minority class to generate synthetic structures labeled with that class until the number of positives and negatives is equal. SMOTE method is useful when there are imbalances in the classes in the training data set (*e.g.*, more negatives than positives) and therefore QSAR model development may either be very difficult or not possible. The program automatically performs 10-fold cross-validation and provides validation measures of the performance from the cross-validation, once the model is completed (Choi et al., 2013).

Internal validation report provides details about the QSAR model(s), molecular descriptors used in constructing the model(s), ratio of positive and negative compounds with their individual percentage, graphical representation of the predicted molecules (true positive, true negative, false positive and false negative), model statistics (specificity, sensitivity, concordance, Matthews correlation coefficient, and so on), and comparison between prediction result and actual value for each of the training set compound. Once a model is built and found to have acceptable internal validation performance, the external validation experiment can be performed using a new or different data set (not available in the training set) to see the predictive ability of the model. Symmetry has two types of toxicological suites: (a) preclinical safety suite (62 models) includes rodent carcinogenicity, reproductive toxicity, fetal developmental toxicity, fetal survival toxicity, behavioral toxicity, genetic toxicity and (b) the human adverse effects/clinical suite (24 models) includes adverse cardiological effects, hepatobiliary effects and urinary system effects in humans. The other important models are the molecular mechanism of action (MOA) models that predict the interaction of molecular entities with molecular targets in the areas of type 2 diabetes, cancer, nervous system diseases and behaviour and mental disorders. Symmetry also has pharmacokinetics and disease-based models.

The programme has been evaluated in scientific literature by Choi et al. (2013).

References:

Choi, S.S.; Kim, J.S.; Valerio(Jr.), L.G.; Sadrieh, N. *In silico* modeling to predict drug-induced phospholipidosis. Toxicol. Appl. Pharmacol. 269, 195–204, 2013.

SCIQSAR

Methodology: Descriptors-based, QSAR system that generates predictive models automatically from training data sets of non-congeneric compounds associated with biological/toxicological data by using six different algorithms.

Area of Applicability: Food, drug, cosmetics, pesticides, biocides

Description: SciQSAR (formerly MDL-QSAR) is Windows-based QSAR software that is currently available by SciMatics Inc. (www.scimatics.com). The program provides over 240 physicochemical, electro-topological E-state, connectivity and other descriptors; a selection of 6 different algorithms for structure similarity searching; a genetic algorithm for descriptor selection; and a variety of statistical tools including the capability to perform parametric and non-parametric discriminant analysis. The best SciQSAR model can be obtained by searching for the best subset of descriptors and optimization of the discriminant analysis parameters. The selection of the type and optimal parameters for a model with highest accuracy can be achieved by trying various combinations of model building parameters. The performance of a model is evaluated using an automated validation process within the program that calculates the prediction error rate in the training set (i.e., probabilities of misclassification) (Matthews et al., 2008; Contrera, 2013).

The programme has been evaluated in scientific literature (Mathews, et al. 2008; Contrera, 2013).

References:

Contrera, J. F. Validation of Toxtree and SciQSAR *in silico* predictive software using a publicly available benchmark mutagenicity database and their applicability for the qualification of impurities in pharmaceuticals. Regul. Toxicol. Pharmacol. 67:285–293, 2013.

Matthews, E.J.; Kruhlak, N. L.; Daniel Benz, R.; Contrera, J. F.; Marchant, C. A.; Yang, C. Combined Use of MC4PC, MDL-QSAR, BioEpisteme, Leadscope PDM, and Derek for Windows Software to Achieve High-Performance, High-Confidence, Mode of Action–Based Predictions of Chemical Carcinogenesis in Rodents. Toxicol. Mech. Methods, 18:189–206, 2008.

DEREK NEXUS

Methodology: SAR; Expert System

Area of Applicability: Food/ Drug/ Cosmetics/ Industrial Chemicals

Brief Description: Derek Nexus is used to assess the toxicity of a wide range of chemical classes, including food, drug, cosmetic, and industrial chemicals, and the system provides predictions for over 50 toxicological endpoints, including mutagenicity, chromosome damage, carcinogenicity, skin sensitisation and reproductive toxicity.

The broad applicability of Derek Nexus is supported by the wide range of organisations that currently use the software, including agrochemical, cosmetic, food and nutrition, petrochemical, pharmaceutical and tobacco companies.

The performance of Derek has been evaluated, and in many cases these evaluations have been published in peer-reviewed journals. Examples include heat generated food contaminants [Cotterill *et al*], flavour chemicals [Ono *et al*], chemicals released from plastic food packaging [Rothenbacher and Schwack], pharmaceutical impurities [Sutter *et al*, Dobo *et al*], pharmaceuticals [Snyder, Mathews *et al* 2009], cosmetic ingredients [Goebel *et al*, Suarez-Rodriquez], pesticides [Crettaz and Benigni] and industrial chemicals [Hayashi *et al*, Rybacka *et al*].

Derek Nexus is a rule-based expert system for the prediction of toxicity. Its knowledge base is composed of alerts, examples and reasoning rules which may each contribute to the predictions made by the system. Each alert in Derek describes a chemical substructure believed to be responsible for inducing a specific toxicological outcome (often referred to as a toxicophore). Alerts are derived by experts, using toxicological data and information regarding the biological mechanism of action. Where relevant, metabolism data may be incorporated into an alert, enabling the prediction of compounds which are not directly toxicity but are metabolised to an active species. The derivation of each alert is described in the alert comments along with supporting references and example compounds where possible. By reporting this information to the user, Derek provides highly transparent predictions.

There are currently over 800 alerts in Derek Nexus covering over 50 toxicological endpoints (including mutagenicity, chromosome damage, carcinogenicity, reproductive toxicity and skin sensitisation). The system provides species-specific predictions for humans, other mammals and bacteria.

When a query compound is submitted to Derek, its structure is analysed for any toxicophores which match alerts within the knowledge base. In addition, certain physicochemical properties (*e.g*. logKp) are calculated for the query structure and these values are checked against rules within the knowledgebase. The reasoning engine within Derek then assesses the matched alerts and reasoning rules to generate the likelihood of toxicity for a specified species [Judson *et al* 2003]. Likelihood in Derek is expressed using one of nine confidence levels [Judson *et al* 2013]. These are: certain, probable, plausible, equivocal, doubted, improbable, impossible, open and contradicted. If a query compound matches no alerts or reasoning rules in the knowledge base, the program displays a message of “nothing to report”.

The use of Derek Nexus for the assessment of a range of toxicological endpoints is described in numerous publications, including those cited in this report. In addition, QSAR Model Reporting Format (QMRF) reports are available describing Derek’s methodology in specific relation to the endpoints of mutagenicity, chromosome damage, carcinogenicity and skin sensitisation. These reports are available in the Joint Research Centre (JRC) QSAR model database [JRC].

Limitations: As with other statistical models and expert systems, Derek Nexus predictions are restricted to chemicals that are structurally similar to those used to develop the model. In practice, if a query chemical activates an alert in Derek it can be considered to be within the ‘applicability domain’ of the model. With this in mind, in order to ensure that Derek is applicable to a wide range of chemicals, an effort is made to use all available public data when developing alerts. In addition,

whenever possible, proprietary data provided by Derek users are also used to develop alerts: this enables the system to describe the activity of chemical classes for which public data are not available.

While Derek Nexus contains alerts for over 50 toxicological endpoints, these endpoints differ with respect to their level of development. The well-developed endpoints in Derek are mutagenicity, chromosome damage, carcinogenicity and skin sensitisation, each of which has more than 50 associated alerts in the knowledge base.

If a compound does not activate an alert or reasoning rule in Derek, a result of 'nothing to report' is presented to the user. This can be interpreted as a negative prediction or that the query compound is outside the domain of the model. Which of these is more appropriate may depend on the endpoint of interest. For the endpoints of mutagenicity and skin sensitisation, which feature multiple alerts believed to cover most of the mechanisms and chemical classes responsible for activity, 'nothing to report' may be extrapolated to a negative prediction. For other endpoints, it is necessary for the user to use expert judgement on the potential inactivity of the query compound.

Recognized in regulation: Lhasa Limited works closely with regulators to ensure that Derek Nexus continues to meet regulatory requirements for a variety of use-cases. Lhasa Limited is currently involved in collaborative research with the US Food and Drug Association (under a Research Collaboration Agreement) and the National Institute of Health Sciences (NIHS) in Japan.

Derek Nexus predictions are accepted by major regulatory agencies, such as the FDA and the European Medicines Agency (EMA), for the assessment of genotoxic impurities. It is of note that under the draft ICH M7 guidelines for the assessment of pharmaceutical impurities, the results from two *in silico* systems may be used to negate the need for any further *in vitro* or *in vivo* testing [ICH M7 Draft Consensus Guideline].

Derek predictions for multiple toxicological endpoints may be used in support of regulatory submissions for the High Production Volume Challenge (HPV) Program and the Registration, Authorization and Evaluation of Chemicals (REACH) regulatory framework. The suitability of several *in silico* systems, including Derek Nexus, to support REACH submissions has recently been evaluated [Rybacka *et al*].

Validated in research publications: The performance of Derek has been evaluated by several regulatory agencies, including the FDA [Mathews *et al*, 2008, Mathews *et al* 2009] and the NIHS [Ono *et al*, Hayashi *et a*l], along with various governmental organisations, commercial organisations and academic institutions.

These validations have been performed for several toxicity endpoints using both public and proprietary datasets. Some recent studies are listed below, categorised according to the endpoint assessed.

- ***Mutagenicity***: Hansen *et al*, Hillebrecht *et al*, Snyder, Dobo *et al*, Valerio, Sutter *et al*, Ono *et al,* Judson *et al* 2013, Rybacka *et al*
- **Chromosome damage**: Judson et al 2013
- ***Carcinogenicity***: Judson *et al 2013*, Mathews *et al* 2008, Cotterill et al, Crettaz and Benigni, Ryback *et al*
- ***Skin sensitisation***: Judson *et al* 2013, Teubner *et al,* Rorije *et al*, Goebel *et al*, Langton *et al*, Patlewicz et al
- **Hepatotoxicity**: Greene et al, Mathews et al 2009
- Reproductive toxicity: Rybacka et al

Formal validation of individual alerts in Derek Nexus: For the endpoints of mutagenicity, chromosome damage, carcinogenicity and skin sensitisation, the positive predictivity (number of compounds correctly predicted as positive / total number of compounds predicted as positive) of individual alerts is analysed against appropriate datasets of experimental results and reported in the alert validation comments. This information allows the user to assess the robustness of an individual alert.

References:

Basketter et al, Regulatory Toxicology and Pharmacology, **2013**, 67, 531-535.

Cotterill et al, Food and Chemical Toxicology, **2008**, 46, 1905-1918.

Crettaz and Benigni, Journal of Chemical Information and Modelling, **2005**, 45, 1864-1873.

Dobo et al, Regulatory Toxicology and Pharmacology, **2012**, 62, 449-455.

Goebel et al, Regulatory Toxicology and Pharmacology, **2012**, 63, 40-52.

Greene et al, Chemical Research in Toxicology, **2010**, 23, 1215-1222.

Hansen et al, Journal of Chemical Information and Modelling, **2009**, 49, 2077-2081.

Hayashi *et al*, *Mutation Research*, **2005**, *588*, 129-135.

Hillebrecht et al, Chemical Research in Toxicology, **2011**, 24, 843-854.

ICH M7 Draft Consensus Guideline: Assessment and Control of DNA Reactive (Mutagenic) Impurities in Pharmaceuticals to Limit Carcinogenic Risk. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M7/M7_Step_2.pdf

Joint Research Centre, QSAR reporting formats and JRC QSAR Model Database, http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/qsar_tools/QRF

Judson et al, Journal of Chemical Information and Chemical Sciences, **2003**, 43, 1364-1370.

Judson *et al*, *Toxicology Research*, **2013**, *2*, 70-79.

Langton *et al*, *Contact Dermatitis*, **2006**, *55*, 342-347.

Patlewicz et al, SAR and QSAR in Environmental Research, **2007**, 18, 515-541.

Mathews et al, Regulatory Toxicology and Pharmacology, **2009**, 54, 43-65.

Mathews et al, Toxicology Mechanisms and Methods, **2008**, 18, 189-206.

Ono et al, Food and Chemical Toxicology, **2012**, 50, 1538-1546.

Pfuhler et al, Regulatory Toxicology and Pharmacology, **2010**, 57, 315-324.

Rorije et al, Regulatory Toxicology and Pharmacology, **2013**, 67, 146-156.

RothenBacher and Schwack, *Journal of AOAC International*, **2009**, *92*, 941-950.

Rybacka et al, Basic and Clinical Pharmacology and Toxicology, **2014**, epub ahead of print.

Suarez-Rodriquez, 2013, Poster presented at the 2013 EMGS Meeting: Reliability of SAR Predictions for TTC Risk Assessment of New Ingredients.

Snyder, Environmental and Molecular Mutagenesis, **2009**, 50, 435-450.

Sutter et al, Regulatory Toxicology and Pharmacology, **2013**, 67, 39-52.

Teubner et al, Regulatory Toxicology and Pharmacology, **2013**, 67, 468-485.

Valerio, Toxicology and Applied Pharmacology, **2009**, 241, 356-370.

TOPKAT®

Methodology: TOPKAT is a data driven system based on experimental data from open literature and statistically robust and validated QSAR relationships.

Area of Applicability: General chemicals

Brief Description: A statistically based program, TOPKAT® (Toxicity Prediction by Komputer Assisted Technology) is used in the assessment of acute and chronic systemic toxicity. The program is structure-based property-sensitive 'similarity' measure utilizing a set of descriptors includes Kier and Hall electrotopological states (e-states), shape, symmetry, molecular weight, and the octanol-water partition coefficient. Thus, the TOPKAT ® 'similarity' reflects similarity of descriptors between two molecules with respect to a specific property or endpoint. (Demchuk 2011)

Available toxicological endpoints: (Accelrys web site)

- Rodent Carcinogenicity (as a sex and species specific endpoint based on either the NTP dataset or the FDA dataset)
- Weight of Evidence Carcinogenicity
- Carcinogenic Potency TD50
- Ames Mutagenicity
- Developmental Toxicity Potential (DTP)
- Rat Oral LD50
- Rat Maximum Tolerated Dose
- Rat Inhalational LC50
- Rat Chronic Oral LOAEL
- Skin Irritancy
- Skin Sensitization (GPMT)
- Ocular Irritancy
- Aerobic Biodegradability
- Fathead Minnow LC50
- Daphnia magna EC50
- VlogP

TOPKAT modules provide qualitative (yes/ no) output for rodent carcinogenicity, Ames mutagenicity, developmental toxicity, skin sensitization, skin irritancy, ocular irritation, and aerobic biodegradability. The quantitative models provide point estimates for the lowest observed adverse effect level, oral rat lethal dose, lethal concentration, maximum tolerated dose, and octanol-water partition coefficient (V log P) along with 95% confidence limits for each.

A number of studies have shown that TOPKAT gives reasonable predictions for a range of chemicals including pesticides and industrial chemicals (Lapenna, 2010).

In 1996, the FDA's Office of Cosmetics and Colors (OCAC) conducted a pilot project to evaluate toxicity predictions for a set of cosmetic ingredients, color additives, and food additive ingredients in six of the TOPKAT toxicity modules: carcinogenicity, mutagenicity, developmental toxicity potential, skin irritation, eye irritation, and oral LD50. For this, structurally-defined single ingredients were encoded via SMILES. Literature toxicity study data quality was assessed according to FDA "Redbook II" criteria for direct food additives and color additives, and only Red Book II "grade A" studies (i.e., those met at least 80% of the proposed guidelines for a particular toxicity test), were selected from the literature for inclusion in the evaluation test sets with TOPKAT prediction modules. It is worth mentioning that no Red Book II guidelines existed for skin or eye irritation.

The experimental data were compared for concordance with the TOPKAT predictions, and determination of TOPKAT prediction module performance was assessed in terms of specificity and sensitivity. Confidence in the TOPKAT predictions was evaluated via chemical structure similarity search and degree of coverage in "Optimum Prediction Space (OPS)".

FDA found that:

- TOPKAT 1.5 and 3.0 software offers several toxicity prediction modules highly relevant to cosmetic ingredient safety assessment. However, some chemical structure assessments (long-chain aliphatics, polymers and complex ring structures) were not well covered by certain TOPKAT modules.
- Model-dependent TOPKAT predictions may be outside the OPS for one endpoint, but within the OPS for another.
- Many of the chemicals selected from literature sources were found to be already present in the TOPKAT model databases, and TOPKAT predictions were often found to be outside the OPS.
- FDA's experience with TOPKAT 3.0 and 1.5 toxicity endpoint prediction modules were summarized in an internal OCAC "Beta Test Final Evaluation Report" (1998), and a 1997 publication (see, ATLA 25, 223.252, 1997).

From this evaluation and other studies, the use of TOPKAT, as with other computational methods, should primarily be for screening and internal decision-making, and any information generated by these methods should form part of a more comprehensive package in a 'weight of evidence' approach.

References:

- Silvia Lapenna, Mojca Fuart-Gatnik and Andrew Worth, "JRC Scientific and Technical Reports, "Review of QSAR Models and Software Tools for predicting Acute and Chronic Systemic Toxicity", 2010
- Accelrys Inc. found at http://accelrys.com/mini/toxicology/predictive-functionality.html)
- Eugene Demchuk, Patricia Ruiz, Selene Chou, Bruce A. Fowler, "SAR/QSAR methods in public health practice", Toxicology and Applied Pharmacology 254 (2011) 192–197

TOXICITY ESTIMATION SOFTWARE TOOL (T.E.S.T.) V4.0

Methodology: T.E.S.T is an Expert system that estimates toxicity values using an ensemble of QSARs. The user inputs the desired structure and the toxicity is estimated using one of seven available methodologies. The program requires no further external input as the required descriptors are calculated within the program.

Area of Applicability: General chemicals

Brief Description: The ensemble of QSARs within T.E.S.T are:

- Hierarchical method: The toxicity for a given query compound is estimated using the weighted average of the predictions from several different models. The different models are obtained by using Ward's method to divide the training set into a series of structurally similar clusters. A genetic algorithm based technique is used to generate models for each cluster. The models are generated prior to runtime.
- FDA method: The prediction for each test chemical is made using a new model that is fit to the chemicals that are most similar to the test compound. Each model is generated at runtime.
- Single model method: Predictions are made using a multilinear regression model that is fit to the training set (using molecular descriptors as independent variables) using a genetic algorithm based approach. The regression model is generated prior to runtime.
- Group contribution method: Predictions are made using a multilinear regression model that is fit to the training set (using molecular fragment counts as independent variables). The regression model is generated prior to runtime.
- Nearest neighbor method: The predicted toxicity is estimated by taking an average of the 3 chemicals in the training set that are most similar to the test chemical.
- Consensus method: The predicted toxicity is estimated by taking an average of the predicted toxicities from the above QSAR methods (provided the predictions are within the respective applicability domains).
- Random forest method: The predicted toxicity is estimated using a decision tree which bins a chemical into a certain toxicity score (i.e. positive or negative developmental toxicity) using a set of molecular descriptors as decision variables. The random forest method is currently only available for the developmental toxicity endpoint.

T.E.S.T allows to estimate the value for several toxicity end points:

1. 96 hour fathead minnow LC_{50}
2. 48 hour Daphnia magna LC_{50}
3. 48 hour Tetrahymena pyriformis IGC_{50}
4. Oral rat LD_{50}
5. Bioaccumulation factor
6. Developmental toxicity
7. Ames mutagenicity

T.E.S.T. also allows to estimate several physical properties, such as Normal boiling point; Density; Flash point; Thermal conductivity; Viscosity; Surface tension; Water solubility; Vapor pressure; Melting point.

While each method has its own advantages and disadvantages, providing multiple prediction methodologies can provide higher confidence in estimates where predicted toxicities are fairly similar from different methods or alternatively allow the user to select estimates from methods the research may have more confidence based on personal experience.

Validation results are provided for each end point against the seven methods utilized. (EPA 2102):

- The Rat LD_{50} the prediction was the weakest endpoint with no single model or group fitting the entire training set of 7413 chemicals. The authors suggest this may not be surprising since this endpoint has a higher degree of experimental uncertainty and has been shown to be more difficult to model than other endpoints.
- In general, the prediction statistics for the physical properties were best and considered by the authors to be 'excellent'.

References:

U.S. Environmental Protection Agency, User's Guide for T.E.S.T. (version 4.1), 2012 http://www.epa.gov/nrmrl/std/qsar/TEST-user-guide-v41.pdf

VEGA

Methodology: A platform to provide (Q)SARs and other in silico tools for safety assessment of chemical substances

Area of Applicability: General chemicals

Brief Description:

The VEGA platform has been developed by the Istituto di Ricerche Farmacologiche Mario Negri in Milan with a number of collaborating organisations and through a series of EU-funded projects. The models used in VEGA for carcinogenicity and mutagenicity originated in the EU project CAESAR, (www.caesar-project.eu/), with subsequent improvements and additions from contributing organisations. The models were developed in line with the OECD principles using high quality datasets with the aim to use them for regulatory purposes. VEGA platform also incorporates some of the models in the US-EPA Toxicity Estimation Software Tool; T.E.S.T (www.epa.gov/nrmrl/std/qsar/qsar.html).

In addition to mutagenicity and carcinogenicity, VEGA also provides (Q)SAR models for developmental toxicity, skin sensitisation, as well as ecotox endpoints (e.g. Daphnia magna LC50 and Fathead minnow LC50), environmental endpoints (bioconcentration factor, ready biodegradability), and models for calculation of physicochemical parameters (logP).

VEGA also allows user to build own (Q)SAR models using SARpy for classification models, and CORAL (CORrelation And Logic) for regression based models. In addition, it provides chemoinformatics tools for SMARTS matching, similarity calculation, etc.

VEGA models generate transparent, reproducible, and verifiable results. The system comprises a series of tools that have been optimised so that the results obtained for a target compound can also be related to those for other structurally related compounds. VEGA also has a comprehensive 5-point validation system that allows the user to assess the reliability of predictions.

Reference:

The Vega Platform www.vega-qsar.eu.

.

THRESHOLD OF TOXICOLOGICAL CONCERN (TTC)

Methodology: The Threshold of Toxicological Concern (TTC) approach is a structure-activity relationship (SAR) tool

Area of Applicability: Food, drugs, cosmetics and industrial chemical.

Brief Description: The Threshold of Toxicological Concern (TTC) approach is a structure-activity relationship (SAR) tool which originated as a "Threshold of Regulation" for the safety assessment of components of food-contact materials that pose a negligible risk and for which toxicological data are unavailable, whereby a substance of known chemical structure and for which valid intake (*i.e*. chronic exposure) estimates are available is compared to a group of structurally-related compounds.

The compounds used for comparison are grouped in order of increasing toxicity potency in "Cramer classes" labelled I, II and III, and their classification is based on a detailed, 33-question "decision tree". Substances included in Class I have relatively simple chemical structures and possess well-known metabolic pathways and innocuous end-products, thereby suggesting a low likelihood of toxicity. Substances in Class II have intermediate structural complexity, may contain reactive moieties, but lack the structural attributes generally associated with toxicity. Substances included in Class III have relatively complex chemical structures that permit no strong initial presumption of safety, or may even suggest significant toxicity (Cramer et al., 1978; Munro et al., 1996). Heavy metals (e.g. arsenic, cadmium, lead and mercury), proteins, aflatoxin-like compounds, N-nitroso compounds, azoxy compounds, polyhalogenated-dibenzo-p-dioxins, -dibenzofurans and -biphenyls are explicitly excluded from the three classes used for TTC evaluations and require substance-specific data (Kroes et al., 2004).

Using the aforementioned classification method and the toxicological data for 613 substances, a TTC was defined for each Cramer structural class by plotting the cumulative distribution of the no-observed-effect levels (NOELs) for the substances and by fitting a log-normal distribution (Munro et al., 1996). The 5th percentile value for each NOEL distribution was chosen, because it provides a 95% confidence that the NOEL of substances of unknown toxicity within the same Cramer class would be greater or equal to that of the 5th percentile and, therefore, of a lower toxicity potency. The 5th percentile values were divided by an uncertainty factor of 100 and then adjusted to a 60-kg body weight to determine TTC values for each Cramer class.

TTC values are defined as the maximum oral doses for which there is no appreciable risk to human health following daily exposure over a lifetime. Although the TTC concept was originally developed for the safety assessment of dietary chemical contaminants, it may also be suitable to assess or prioritize the requirement for toxicity testing and assessment for chemicals that are present in cosmetic ingredients or final formulations, including contaminants or degradation products (Kroes et al., 2007). In fact, a preliminary analysis of the applicability of the TTC approach to cosmetic and personal care products has already been undertaken as part of the COSMOS project. It was concluded that the current TTC approach is broadly applicable to cosmetics, but that improvements (e.g. quality control of core datasets, moderate alterations/additions of Cramer classification scheme, and development of clear guidance for its application) could be made (Worth et al., 2012).

One possible *in silico* approach applicable to cosmetic products involves the use of ToxTree, an open-source software commissioned by the European Commission Joint Research Centre (JRC) which estimates toxic hazard of chemical substances, to rapidly estimate the Cramer class of the chemical structure of interest and to determine the associated TTC value. The chronic systemic dose for this chemical can then be modelled (e.g. using ConsExpo), and the outcome of the chosen scenario can be compared to the TTC value. If the chronic systemic dose is lower than the TTC value, toxicity can be excluded solely based on this assessment. When the modelled chronic systemic dose is greater than the TTC value, toxicity cannot be excluded, and a higher tier approach is needed to refine the model. Higher tier approaches are presented in other sections of this document.

The robustness of the Threshold of Toxicological Concern (TTC) approach has been confirmed for

rodent carcinogens (Cheeseman et al., 1999). The robustness of TTC values for various specific non-cancer endpoints has been reviewed in details by the European Food Safety Authority (EFSA, 2012). As for any modelling approach, both the selection of a suitable TTC value and the reliability of the structural alert scheme depend on the robustness and comprehensiveness of the underlying database (MST, 2011). Theoretically, the TTC approach could be used to assess the safety of any chemical, if the chemical structure is known and human exposure can be accurately and robustly estimated.

The TTC approach, initially introduced as a "Threshold of Regulation", has gained broad international recognition (e.g. US EPA; Danish Ministry of the Environment (MST); European Commission) and has been reviewed and improved through peer-reviewed scientific publications (e.g. Cheeseman et al., 1999; Cramer et al., 1978; Kroes et al., 2004, 2007; Munro et al., 1996).

References:

Cheeseman, M.A., Machuga, E.J. and Bailey, A.B. 1999. A tiered approach to threshold of regulation. Food and Chemical Toxicology 37(4):387-412. http://dx.doi.org/10.1016/S0278-6915(99)00024-1

ConsExpo, The Netherlands National Institute for Public Health and the Environment (Dutch: Rijksinstituut voor Volksgezondheid en Milieu; RIVM), http://www.rivm.nl/consexpo (accessed 25 February 2014).

Cramer, G.M., Ford, R.A. and Hall, R.L. 1978. Estimation of toxic hazard – A decision tree approach. Food and Cosmetics Toxicology 16(3):255-276. http://dx.doi.org/10.1016/S0015-6264(76)80522-6

Danish Ministry of the Environment, Environmental Protection Agency (Dutch: Miljøministeriet; MST). 2011. The Threshold of Toxicological Concern (TTC) concept: Development and regulatory applications. http://www2.mst.dk/udgiv/publications/2011/03/978-87-92708-86-1.pdf (accessed 25 February 2014).

European Commission Joint Research Center (JRC), Institute for Health and Consumer Protection, http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/qsar_tools/qsar-tools (accessed 25 February 2014).

European Food Safety Authority (EFSA). 2012, Scientific opinion on exploring options for providing advice about possible human health risks based on the concept of Threshold of Toxicological Concern (TTC). ESFA Journal 10(7):2750. http://dx.doi.org/10.2903/j.efsa.2012.2750.

Kroes, R., Renwick, A.G., Cheeseman, M., Kleiner, J., Mangelsdorf, I., Piersma, A., Schilter, B., Schlatter, J., van Schothorst, F., Vos, J.G., Würtzen, G. 2004. Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food and Chemical Toxicology 42(1):65-83. http://dx.doi.org/10.1016/j.fct.2003.08.006

Kroes, R., Renwick, A.G., Feron, V., Galli, C.L., Gibney, M., Greim, H., Guy, R.H., Lhuguenot, J.C., van de Sandt, J.J.M. 2007. Application of the threshold of toxicological concern (TTC) to the safety evaluation of cosmetic ingredients. Food and Chemical Toxicology 45(12):2533-2562. http://dx.doi.org/10.1016/j.fct.2007.06.021

Munro, I.C., Ford, R.A., Kennepohl, E. and Sprenger, J.G. 1996. Correlation of structural class with no-observed-effect levels: A proposal for establishing a threshold of concern. Food and Chemical Toxicology 34(9):829-867. http://dx.doi.org/10.1016/S0278-6915(96)00049-X

ToxTree, http://toxtree.sourceforge.net/ (accessed 25 February 2014).

Worth, A., Cronin, M., Enoch, S., Fioravanzo, E., Fuart-Gatnik, M., Pavan, M., Yang, C. 2012. Applicability of the Threshold of Toxicological Concern (TTC) approach to cosmetics – preliminary analysis. COSMOS project. http://dx.doi.org/10.2788/5059

Annex–II: Brief Description of Relevant Projects in ICCR Jurisdictions

EU1 - COSMOS

Project description:

The COSMOS project (http://www.cosmostox.eu/), jointly funded by the European Commission and Cosmetics Europe, is developing publicly available computational workflows based on the integrated use of open-access and open-source models for the prediction of repeated dose toxicity (Anzali et al, 2012). This includes: a) the establishment of an inventory of cosmetic substances (including identifiers and chemical structures) and a repeat dose toxicity database (including oral and dermal data); b) the development of novel ways of establishing thresholds of toxicological concern (TTC), based on innovative chemistry based prediction approaches and biokinetic modelling.

Key reference:

Anzali S, Berthold MR, Fioravanzo E, Neagu D, Péry A, Worth AP, Yang C, Cronin MTD & Richarz A-N (2012) Development of computational models for the risk assessment of cosmetic ingredients. IFSCC Magazine 15 (4): 249-255. Available at: http://www.cosmostox.eu/publications/printed/

Timeframe: January 2011-December 2015

Contact: Prof. Mark Cronin, Liverpool John Moores University, UK

EU2 - ANTARES

Project description:

ANTARES (http://www.antares-life.eu/) was an EC (DG Environment) funded project which has assessed the validation characteristics of a range of (Q)SAR models for the ecotoxicological, toxicological and environmental endpoints that are relevant to REACH. Human health endpoints considered were acute oral toxicity (rodent LD50), mutagenicity (Ames test) and carcinogenicity (rodent bioassay). A wide range of software tools were evaluated, including both commercial and non-commercial tools, and taking applicability domain considerations into account wherever possible.

Timeframe: January 2010-December 2012

Contact: Prof. Emilio Benfenati, Mario Negri Institute, Milan, Italy

Canada – Repeated-dose dermal toxicity

Project description:

The objective of this project was to determine whether the Cramer classification scheme can be applied to systemic toxicity resulting from dermal exposure. For this purpose, a reference database containing NO(A)EL values for systemic toxicity subsequent to dermal administration in various mammalian models (*i.e.* rats, mice and rabbits) was assembled by Health Canada, whereby 52 of the 140 substances (37%) are known to be used in cosmetic products available on the Canadian market. The project demonstrated that, although repeated-dose dermal toxicity data are not readily available, the robust relationship between oral and dermal NOAELs in the Cramer classification scheme supports the use of the oral Threshold of Toxicological Concern (TTC) for cosmetic products intended for dermal application.

Key reference:

Faith. M. Williams, A. Chiodini, G. Barrett, M.T.D, Cronin, R.H. Guy, N. Montiero-Riviere, J. Plautz, C. Roper, H. Rothe, D. Rua, J. Westerhout , C. Yang (*in preparation*). Application of a decision tree with systemic exposure prediction to oral TTC for the safety evaluation of cosmetic chemicals (*tentative title*).

Timeframe: Manuscript in preparation.
Contact: Gordon Barrett, Health Canada, Canada.

BRAZIL – RENAMA

Project description:

The National Network on Alternative Methods (RENAMA – Rede Nacional de Métodos Alternativos), created in 2012 by a Ministry of Science, Technology and Innovation act, promotes initiatives to develop the use of alternative methods in Brazil and efforts to validate then in the regulatory scenario. Recently, RENAMA worked with the Brazilian Biosciences National Laboratory (LNBio – Laboratório Nacional de Biociências) selecting proposals of *in silico* tests focused on the prediction of toxicological and pharmacokinetic properties of molecules to be used as drugs or cosmetics. These projects are still to be initiated, but represent a strong effort to promote the consolidation of *in silico* approaches in Brazil.

Key Reference (Portuguese only):

http://renama.org.br/

http://lnbio.cnpem.br/wp-content/uploads/2014/01/EDITAL-RENAMA_LNBio.pdf