ALTEX, accepted manuscript
published May 3, 2021
doi:10.14573/altex.2010221

Research Article

SpheraCosmolife: A New Tool for the Risk Assessment of Cosmetic Products

Gianluca Selvestrel[1], Federica Robino[2], Diego Baderna[1], Serena Manganelli[1,3], David Asturiol[4], Alberto Manganaro[5], Matteo Zanotti Russo[2], Giovanna Lavado[1], Cosimo Toma[1], Alessandra Roncaglioni[1] and Emilio Benfenati[1]

[1]Laboratory of Environmental Chemistry and Toxicology, Environmental Health Department, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milano, Italy; [2]Angel Consulting s.a.s., Milano, Italy; [3]current affiliation: Chemical Food Safety Group, Nestlé Research, Lausanne, Switzerland; [4]European Commission, Joint Research Centre (JRC), Ispra, Italy; [5]Kode s.r.l, Pisa, Italy

Abstract

A new, freely available software for cosmetic products has been designed considering the regulatory framework for cosmetics. This software allows an overall toxicological evaluation of cosmetic ingredients without the need of additional testing and, depending on the product type, it applies defined exposure scenarios to derive risk for consumers. It takes regulatory thresholds into account and uses either experimental values, if available, or predictions. Based on experimental or predicted no observed adverse effect level (NOAEL), the software can define a point of departure (POD), which is useful to calculate the margin of safety (MoS) of the query chemicals. The software also provides other toxicological properties, such as mutagenicity, skin sensitization and the threshold of toxicological concern (TTC) to provide an overall evaluation of the potential chemical hazard. Predictions are calculated using *in silico* models implemented within the VEGA software. The full list of ingredients of a cosmetic product can be processed at the same time, at the effective concentration in the product given by the user. SpheraCosmolife is designed as a support tool for the safety assessors of cosmetic products and can be used to prioritize the cosmetic ingredients or formulations according to their potential risk for the consumers. The major novelty of the tool is that it wraps a series of models (some of them new) into a single user-friendly software system.

1 Introduction

The European Directive on cosmetics represented a paradigm shift in Europe for the safety assessment of cosmetics, which transitioned from the classical toxicological approach based on animal testing, towards a completely novel strategy, where the use of animals for toxicity testing is completely banned (EC, 2009). The European strategy has been followed by an increasing number of countries in the world[1]. However, this regulation does not provide details on which alternative methods are to use. Several ambitious projects have addressed sophisticated alternative testing strategies, such as SEURAT-1[2] European initiative (Berggren et al., 2017) and EU-ToxRisk[3]. The outcomes of a workshop organized by the Scientific Committee on Consumer Safety (SCCS) have now been published, indicating some areas of research to be addressed (Rogiers et al., 2020). The possibility of using *in silico* tools is particularly appealing because they can generate safety data for cosmetic ingredients without testing (Gellatly and Sewell, 2019; Taylor and Rego Alvarez, 2020). There are many tens of thousands of cosmetic ingredients, and this number poses an incredible challenge not only for *in vivo* methods, but for other test methods too. The possibilities that *in silico* models offer to face this task have been demonstrated by calculating a number of properties for about 20,000 botanical ingredients of cosmetics (Raitano et al., 2019).

However, apart from hazard identification, risk assessment of cosmetic ingredients requires both hazard characterization and estimation of exposure. In fact, the assessment of the risk posed by one or more components in the cosmetic product depends on the concentration of the ingredient and the exposure must be examined, starting from skin permeation of the substance of interest. Furthermore, the assessor should use different software systems, to get different values for the several toxicological endpoints and exposure values. Within the EC funded LIFE project VERMEER[4], a strategy was drawn up, offering the possibility to integrate several tools into a single computer system able to support safety assessors of cosmetic products. To comply with the cosmetics regulation, we planned to reproduce as closely as possible the procedure employed for cosmetics, with specific reference to the lists of already regulated ingredients, thresholds to be respected, and the specific endpoints to be evaluated. As a result, we built the software system presented here.

Received October 22, 2020; Accepted April 29, 2021;
Epub May 3, 2021;

ALTEX 38(#), ###-###. doi:10.14573/altex.2010221

Correspondence: Gianluca Selvestrel, PhD
Laboratory of Environmental Chemistry and Toxicology,
Environmental Health Department,
Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milano, Italy
(gianluca.selvestrel@marionegri.it)

[1] https://www.cosmeticsdesign-europe.com/Article/2019/03/05/Global-ban-on-animal-testing-where-are-we-in-2019
[2] http://www.seurat-1.eu/
[3] https://www.eu-toxrisk.eu/
[4] https://www.life-vermeer.eu

2 Methods

The novel software system combines an expert-system approach, which refers to the sequence of steps done by the assessors, with some machine-learning and statistical models, to provide predictions. Thus, at the basis of the novel system, there is a sound theoretical basis, derived from the procedure defined by the regulators, currently done manually in most of the cases. In order to demonstrate the safety of a cosmetic product, prior to placing it on the market, the responsible person has to ensure that the cosmetic product has undergone a safety assessment. ANNEX I of the Regulation (EC) No. 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products defines the aspects to be evaluated within the cosmetic product safety report (EC, 2009). Moreover, as indicated in Article 11 of the Regulation 1223/2009, the responsible person shall keep a Product Information File (PIF) for it.

The procedure for the safety evaluation of a cosmetic product is defined in detail by "The SCCS Notes of Guidance for the testing of cosmetic ingredients and their safety evaluation 10th revision" (SCCS, 2018). Table 1 shows the workflow that a risk assessor goes through and explains how these steps have been translated into SpheraCosmolife.

The risk assessment procedure followed by the regulators is largely replicated by the software. Indeed, the four main steps (hazard identification, exposure assessment, dose-response assessment and risk characterization) are covered. *In silico* structure of the tool represents a forward-looking and future-oriented idea and helps the assessor with a remarkable time saving as well as with a considerable number of data and information that facilitate the decision-making procedure. The assessor has to evaluate the values provided, their uncertainty (also considering the remarks indicated by the software), and the software should not substitute the work of the assessor in an automatic way. We will provide more data and information in the future. New technical and scientific issues, for example new toxicological endpoints and new evaluation approaches (e.g., read-across) will be implemented in future versions in order to better cover the evaluation procedure and offer a more complete and innovative assessment. The following steps describe the software system structure.

Tab. 1: Risk assessment steps defined by regulators and their translation into SpheraCosmolife

Regulatory procedure	Technical guidance/ Regulation references	SpheraCosmolife	Comparison consideration
HAZARD IDENTIFICATION: is carried out to identify the intrinsic toxicological properties of the substance.	SCCS NoG* (relevant toxicological studies on cosmetic ingredients) / Annex I. 8 of Reg. 1223/2009 (toxicological profile of the substances)	Experimental values are retrieved within the internal database. In case experimental data are missing, application of *in silico* models (expert and statistical-based tools) to predict toxicological properties is provided.	Novelty of computational approaches to refine the assessment. Other toxicological endpoints will be added in future versions.
EXPOSURE ASSESSMENT: exposure is calculated based on the declared functions and uses of a substance as cosmetic ingredient, the amount present in the respective cosmetic product categories and their frequency of use.	SCCS NoG* (Exposure assessment) / Annex I. 6 and 7 of Reg. 1223/2009 (exposure to the cosmetic product, exposure to the substances)	Application of the official equations and implementation of predefined exposure scenarios within the software. Introduction of a refined approach, based on new models for skin permeation, to estimate the internal exposure.	Novelty of computational approaches to refine the assessment. Automatic creation of exposure scenarios.
DOSE-RESPONSE ASSESSMENT: calculation of the POD	SCCS NoG*	Automatic calculation of the POD (NOAEL) using a QSAR model. Experimental values are also present within the internal database.	Novelty of computational approaches to refine the assessment. New models for NOAEL and LOAEL will be implemented in future versions. Organ-specific toxicity will be also indicated.
RISK CHARACTERIZATION: the focus is on systemic toxicity. The Margin of Safety is calculated.	SCCS NoG* (General principles for the calculation of the margin of safety and threshold of toxicological concern)	Margin of Safety is automatically calculated using the previous results. A decision tree is implemented to provide a TTC assessment.	SpheraCosmolife offers an automatic calculation of the MoS. A decision tree approach gives the possibility to estimate the TTC class and compare the TTC thresholds with the exposure values, incorporating dermal bioavailability into the use of TTC.

*ref: SCCS, 2018.

Abbreviations: LOAEL = lowest observable adverse effect level; MoS = Margin of Safety; NOAEL = no observable adverse effect level; NoG = Notes of Guidance; POD = point of departure; QSAR = quantitative structure-activity relationship; SCCS = Scientific Committee on Consumer Safety; TTC = Threshold of Toxicological Concern.

2.1 The identification of the ingredients, products and exposure scenarios

The ingredients of a product are identified through the International Nomenclature of Cosmetic Ingredients (INCI), Simplified Molecular Input Line Entry System (SMILES) format, or the Chemical Abstract Service (CAS) number. The identification of the product types derives from the Tables 2A, 2B and 3 of "The SCCS Notes of Guidance for the testing of cosmetic ingredients and their safety evaluation 10th revision" (SCCS, 2018). The exposure scenarios are those defined in this document, for each product type.

2.2 The database

A key component of the software system is its database. It was populated with data retrieved from the COSMOS Cosmetic Inventory[5] (Worth et al., 2012) and the information in the Annexes to Regulation (EC) No. 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products (EC, 2009). This refers to the following Annexes in particular:

- Annex II: List of substances prohibited in cosmetic products.
- Annex III: List of substances which cosmetic products must not contain except subject to the restrictions laid down.
- Annex IV: List of colorants allowed in cosmetic products.
- Annex V: List of preservatives allowed in cosmetic products.
- Annex VI: List of UV filters allowed in cosmetic products.

Further sources of data refer to these repositories:

- CLP harmonized classification[6]
- Safer Chemical Ingredient List (SCIL)[7]

These data sources refer to official repositories from the EU and the US Environmental Protection Agency (EPA). Overall, in the database of SpheraCosmolife, at the moment there are data related to 5000 substances but the database is regularly updated with new and upgraded information and data. Most of these substances are cosmetic ingredients, and in particular these categories are well represented: skin conditioning, skin protecting, surfactants, emulsifying, and perfuming. On a chemical point of view, quite represented are alcohols, amines, ketones, and substances with aliphatic chain of at least 8 carbons.

The information in the Annexes concerns the maximum concentration, the product type in which the ingredient is allowed, and the wording of conditions of use and warning. For inorganic and organometallic compounds, polymers and data related to mixtures of chemicals, if the substance is present in the database of the system, SpheraCosmolife recognizes it; if not, the software is not able to process this kind of substances. For salts, the original structure is stored in the database too, so the input provided by the user can be a salt. However, the final assessment is done using the neutralized structure, without the cation or anion. Semi-automated data curation and quality checking workflow using the KNIME platform (Gadaleta et al., 2018) and expert-based knowledge were used to retrieve neutralized structures in the database. However, for substances not present in the database, the user should insert the structure of the neutral substance, without ions.

2.3 Risk prediction models

The software evaluates a number of properties, described below. For these properties the source of the data is the VEGA platform[8], and information on the specific property can be found from the description of the *in silico* models of the individual property, available from VEGA. New *in silico* models have been developed too (see below).

The SpheraCosmolife system uses a number of equations and *in silico* models. Some models estimate exposure, other predict hazard. Combining the results of these models, the SpheraCosmolife system can assess the risk associated with the ingredients of a product. If available in the system database, experimental values are used, otherwise the system provides predictions. The software can deal with multiple ingredients in a given cosmetic product, for different categories of products. Although the system treats multiple ingredients, interactions between them are not considered since each cosmetic product is considered as an individual combination of cosmetic substances, as described in the SCCS Notes of Guidance for the testing of cosmetic ingredients and their safety evaluation 9th revision" (SCCS, 2015)

2.3.1 Exposure evaluation

2.3.1.1 The list of product types

The user is asked to indicate the product type, as defined in Tables 2A, 2B and 3 of "The SCCS Notes of Guidance for the testing of cosmetic ingredients and their safety evaluation 10th revision" (SCCS, 2018) to identify an exposure scenario.

2.3.1.2 External exposure

External exposure is obtained as indicated by the SCCS (SCCS, 2018). SpheraCosmolife implements equation 1:

External Dermal Exposure (mg/kg bw/day) = $E_{product}$ * C/100 (Eq. 1)

where $E_{product}$ is the calculated relative daily exposure (mg/kg bw/day). It is tabulated in the tables of the SCCS Notes of Guidance (SCCS, 2018). C is the concentration of the ingredient (%).

2.3.1.3 Systemic exposure dose (SED)

SED is obtained as indicated by the SCCS (SCCS, 2018), from equation 2:

[5] https://cosmosdb.eu/cosmosdb.v2/comptools/
[6] https://echa.europa.eu/it/information-on-chemicals/cl-inventory-database
[7] https://www.epa.gov/saferchoice/safer-ingredients
[8] www.vegahub.eu

$$\mathrm{SED(mg/kg\,bw/day)} = \mathrm{Eproduct} * \frac{C}{100} * \frac{DA}{100} \quad \text{(Eq.2)}$$

This equation takes account of the amount of the finished cosmetic product applied per day ($E_{product}$), the concentration (C) of the substance under study in the product category, expressed as a percentage, and the dermal absorption (DA) expressed as a percentage.

A tiered approach is followed to calculate the systemic exposure dose. The SED value is provided by SpheraCosmolife considering three scenarios, where absorption is taken (i) as 100% (for oral and inhalation exposure), or (ii) 50% (for dermal exposure, a default value defined by the SCCS Notes of guidance (SCCS, 2018)), or (iii) 10, 40 or 80% (for dermal exposure, in a more accurate way, calculated according to the Kroes approach (Kroes et al., 2007), which requires information on the skin permeation, as below).

2.3.1.4. Skin permeation

Two models are used, which implement the models described by Potts and Guy (Potts and Guy, 1992) and ten Berge (ten Berge, 2009; Vecchia and Bunge, 2002a), according to equations 3 and 4, respectively. They provide the constant of permeation, K_p.

$$\log K_p = 0.71 \log(K_{ow}) - 0.0061\ MW - 2.7 \text{ (cm/h)} \quad \text{(Eq. 3)}$$

$$\log Kp = \left[\frac{1}{\frac{1}{Klip+Kpol}+\frac{1}{Kaq}}\right] \quad \text{(Eq. 4)}$$

where:

- $K_{pol} = \frac{b4}{MW^{b5}}$
- $K_{aq} = \frac{b6}{MW^{b7}}$
- $K_{lip} = 10^{[b1+b2*10\log(K_{ow})+b3*MW]}$
- K_{lip} = permeation coefficient lipid medium
- K_{pol} = permeation coefficient corneocytes [proteins]
- K_{aq} = permeation coefficient epidermis [aqueous]
- K_{ow} = octanol/water partition coefficient.
- MW = molecular weight.
- b1, b2, b3, b4, b5, b6, b7 = regression coefficients:

- b_1 = -2,694
- b_2 = 0,9809
- b_3 = -7,868*10^{-3}
- b_4 = 0,05523
- b_5 = 1,383
- b_6 = 1,121*10^{3}
- b_7 = 1,957

Once K_p is obtained, the worst case scenario, considering the most conservative value, is employed to proceed with the workflow and the software calculates the maximum flux of the substance, J_{max}, according to equation 5:

$$J_{max}\ (mg/cm^2/h) = K_p * C_{water,sat} \quad \text{(Eq. 5)}$$

where $C_{water,sat}$ is the saturated water solubility, in the unit of mg/cm³, obtained using the model implemented in the VEGA platform.

Once J_{max} is obtained, the software applies the Kroes approach (Kroes et al., 2007), according to equation 6, which provides the percentages of absorption (%A):

$$\%A = \begin{cases} 10\% \text{ if } J_{max} \leq 0.1\ \mu g/cm^2/h \\ 40\% \text{ if } 0.1 < J_{max} \leq 10\ \mu g/cm^2/h \\ 80\% \text{ if } J_{max} > 10\ \mu g/cm^2/h \end{cases} \quad \text{(Eq. 6)}$$

The SpheraCosmolife software provides the percentages of absorption for these three cases, which represent low, medium and high dermal absorption, as defined by Kroes (Kroes et al., 2007; Shen et al., 2014).

2.3.2 Hazard evaluation

The hazard assessment of chemicals is performed with several *in silico* models and profilers available on the VEGA platform. A brief description of each model is presented below.

2.3.2.1 Mutagenicity, bacterial reverse mutation test – Ames test

The consensus model available in VEGA is used for Ames mutagenicity prediction. It combines the results of four models, taking account of the reliability of each prediction for the target substance (Manganelli et al., 2018). It integrates models using both expert- and statistical-based tools, following the recommendation of the International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) (ICH, 2017).

2.3.2.2 Chromosomal aberration

The integrated model available in VEGA is used for chromosomal aberration. It was built up with the CORAL software using SMILES based attributes. The classification model is based on a dataset of 477 organic compounds (223 active and 254 inactive in chromosomal aberrations test). The data were collected from the Genotoxicity OASIS Database and from the Toxicity Japan MHLW that include experimental data for chromosomal aberrations determined by *in vitro* test using Chinese hamster lung (CHL) and ovary (CHO) cells, with and without S9 metabolic activation (Toropov et al., 2019).

2.3.2.3 In vitro micronucleus genotoxicity test

The model developed with the SARpy software, based on structural alerts, is used for the *in vitro* micronucleus genotoxicity test. The model provides a qualitative prediction of genotoxicity as induction of micronucleus in mammalian cells *in vitro* (MNvit). The model was built on a dataset containing 380 organic chemicals with genotoxicant and non-genotoxicant MNvit experimental data (153 inactive and 227 active chemicals). The experimental data were collected, according to the OECD 487 guideline, from eChemPortal inventory, peer-reviewed literature, SCCS and European Food Safety Authority (EFSA) opinions, European Centre for Validation of Alternative Methods (ECVAM) guidelines and review. The fragment- based model uses 138 structural alerts, including 82 active and 56 inactive fragments (Baderna et al., 2020).

2.3.2.4 Skin sensitization

Two models are used, available within VEGA. The first is the CAESAR model (Chaudhry et al., 2010). The second is a new model, described here: it is a decision tree (DT) based on a data set of 332 chemicals with data on local lymph node assay (LLNA) (226 sensitizers and 106 non-sensitizers). Data were collected from the CAESAR model database (209 substances) and 269 substances from Asturiol et al. (2016). The dataset was split into a training (80%) and test (20%) set. To partition the chemicals into training set and test set, assuring high diversity and keeping the ration sensitizer/non-sensitizer, the used procedure was the following. The chemicals were initially separated into sensitizer and non-sensitizer. The subsequent steps were carried out separately for sensitizer and non-sensitizer. Each group was clustered based on the chemical similarity defined by their fingerprints (RDKit atomic pairs) into as many clusters as the number of chemical divided by 10. Subsequently, 80% of clustered chemicals were assigned randomly to the training set using the assigned cluster as stratification variable. The remaining 20% of chemicals were assigned to the test set. The chemicals were structurally diverse and the distribution between sensitizers and non-sensitizers was preserved. Structural diversity refers to diversity in terms of chemical classes. Details have been described by Asturiol et al. (2016). The RDKit atom pairs fingerprints[9] were used to cluster similar chemicals with a kNN algorithm. The chemicals were randomly selected from each cluster in the same proportions of sensitisers/non-sensitisers as in the dataset. The DT model was built using the Recursive PARTitioning (rpart) module included in R software[10] (Therneau and Atkinson, 2015) to develop CART (Classification and Regression Trees) models. The model is based on 2D descriptors calculated using DRAGON (Dragon v. 7.0.8, Kode srl[11]) and a stepwise variable selection using linear discriminant analysis (LDA) and a bootstrap technique (based on balanced resampling) was used for validation to select the best variables, with an in-house code implemented in R (Manganelli et al., 2019).
Molecular descriptors used in the DT are:

- nDB, number of double bonds
- IC1, Information Content index
- SIC2, Structural Information Content index (neighborhood symmetry of 2-order)
- GATS8s, Geary autocorrelation of lag 8 weighted by I-state
- nRCHO, number of aldehydes (aliphatic)
- CATS2D_01_NL, CATS2D Negative-Lipophilic at lag 01
- F04[C-O], Frequency of C - O at topological distance 4
- MLOGP, Moriguchi octanol-water partition coefficient (logP).

2.3.2.5 No observed adverse effect level (NOAEL)

The NOAEL model available in VEGA was used, built up with the CORAL software using SMILES attributes. Repeated dose 90-day oral toxicity study in rodents was considered to build the model. Studies for duration of 28 days of treatment were also considered, and they were divided by a factor of 3, in order to approximate the 90-day, as specified in the SCCS Notes of Guidance (SCCS, 2018). The dataset consists of 140 organic compounds with experimental values collected from the US EPA's Integrated risk information system (IRIS) database, the Hazard evaluation support system (HESS) and Munro databases. The regression model is based on optimal descriptors calculated by the Monte Carlo method with SMILES attributes and the graph of atomic orbitals (Toropov et al., 2015).

2.3.3. Risk characterization

As last step, SpheraCosmolife software carries out the risk characterization of chemicals or formulation considering the process defined within the SCCS Notes of Guidance (SCCS, 2018). The margin of safety (MoS) and the threshold of toxicological concern (TTC) are provided to the users.

2.3.3.1. Calculation of the margin of safety

The MoS is the ratio of the point of departure (POD) and systemic exposure dose (SED). It is commonly used in human health risk assessment and in particular in the cosmetics risk assessment and it is a key value within the Product Information File

[9] Landrum G. (2016). RDKit: open-source cheminformatics. https://www.rdkit.org/
[10] R Core Team (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/
[11] Kode srl. (2017). Dragon (Software for Molecular Descriptor Calculation) Version 7.0.8. https://chm.kode-solutions.net

(PIF), which is essential to be able to put the product on the European market. The software uses as a POD the No Observed Adverse Effect Level (NOAEL). The MoS is calculated according to equation 7:

$$\mathrm{MoS} = \frac{\mathrm{POD(PointOfDeparture)}}{\mathrm{SED(SystemicExposureDose)}} \qquad \text{(Eq. 7)}$$

Within SpheraCosmolife the POD can be an experimental, if available, or a predicted NOAEL. To consider a substance not risky, the MoS has to be higher than 100, as indicated in the SCCS Notes of Guidance (SCCS, 2018).

2.3.3.2. Threshold of toxicological concern (TTC)

The software includes an assessment considering the TTC approach, that refers to the establishment of a level of exposure below which there would be no appreciable risk to human health (Kroes et al., 2004). It uses the Cramer decision tree, implemented in VEGA. The tool for TTC is the same algorithm implemented in Toxtree v. 3.1.0 (Patlewicz et al., 2008), using the original Cramer classes (Cramer et al., 1978; Munro et al., 1996; Patlewicz et al., 2008). The assessment for the TTC refers to the classes listed in Table S1[12].

3 Results

We developed SpheraCosmolife, a unified software system intended to assist risk assessors of cosmetic products, and companies to boost the safety of their products, avoiding ingredients which at a given concentration may be of concern. For this purpose, some new models have been developed and implemented in the VEGA website[8]. The overall scheme follows the steps done by the assessor. The evaluation is easy to be performed: firstly, the user is asked to provide the information regarding the ingredient inserted using the SMILES, the INCI, or the CAS format, its concentration and the product type. Then, the software checks whether the ingredient is present in the Annexes of the Regulation (EC) No. 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products (EC, 2009). Next, the software searches for values in its database useful for the properties for risk characterization, as described in the Methods section. If there are no experimental values, the software predicts the properties of interest. The focus is on systemic toxicity and the MoS is calculated. The user is assisted, providing a series of outcomes relative to the product of interest. We describe below the components of the new software and its outcomes.

3.1 Performance of new skin sensitization model

The new decision tree for the prediction of skin sensitization potential as expressed by the LLNA gave good results (Tab. 2): the results on the training set are quite balanced, while on the test set this model gave more false negatives than false positives. Conversely, the CAESAR model is over-conservative, so it is appropriate to look at the results of both models. To facilitate the use of the results, a system for integrating the two models has been developed, following the approach for the integration of the results of the models for mutagenicity described above.

Tab. 2: The performance of the decision tree model for skin sensitization

	Training set	Test set
n	264	68
TP	145	30
TN	66	18
FP	18	4
FN	35	16
Accuracy	0.80	0.71
Specificity	0.79	0.82
Sensitivity	0.81	0.65
MCC	0.57	0.44

n = number of compounds; TP = true positives; TN = true negatives; FP = false positives; FN = false negatives; MCC = Matthews correlation coefficient.

The results of the *in silico* model for skin sensitization were analyzed considering the true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). We calculated these statistical parameters: sensitivity, specificity, accuracy, and Matthews correlation coefficient (MCC):

$$\mathrm{Sensitivity} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})}$$

$$\mathrm{Specificity} = \frac{\mathrm{TN}}{(\mathrm{TN} + \mathrm{FP})}$$

[12] doi:10.14573/altex.2010221s

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{\text{Total}}$$

$$\text{MCC} = \frac{(\text{TP} * \text{TN}) - (\text{FP} * \text{FN})}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}$$

Statistics reported refers to the performance on the training set. However, Cross-Validation (CV) is performed during the model building using the Recursive PARTitioning (rpart) module included in R software[10] (Therneau and Atkinson, 2015).

3.2 The results of the other *in silico* models

SpheraCosmolife is a novel software, which includes a new integrated software system for skin sensitization and a new model for this endpoint, as described previously. Other *in silico* models have been recently developed (e.g. *in vitro* micronucleus genotoxicity test model) and they are individually available in VEGA. They are organized into a unified workflow here. The detailed information on each VEGA model is available within the VEGA platform[8]. Tables 3 and 4 summarize the performance of each model, in classification and regression, respectively.

Tab. 3: Summary of the statistical parameters of the classification models implemented into SpheraCosmolife

Model		n	Accuracy	Specificity	Sensitivity
Skin sensitization model CAESAR	Training set	167	0.91	0.74	0.95
	Test set	42	0.93	0.75	0.97
**Mutagenicity, bacterial reverse mutation test – Ames test*	Training set	NA	NA	NA	NA
	Validation set 1 (Cassano et al., 2014)	532	0.80	0.61	0.84
	Validation set 2 (Benfenati et al., 2018)	~2000	0.70	0.59	0.72
	Validation set 3 (Carnesecchi et al., 2020)	673	0.72	0.72	0.72
Chromosomal aberration	Training set	407	0.75	0.80	0.70
	Calibration set	35	0.94	0.95	0.94
	Validation set	35	0.94	1	0.87
In vitro micronucleus genotoxicity test	Training set	293	0.88	0.73	0.97
	Test set	87	0.83	0.63	0.94

*This is an integrated model which combines the predictions of four models (Cassano et al., 2014). The statistics of each individual model are reported in VEGA. This integrated model has been used to predict substances not present in the training set within three studies (as in the references) and the values are reported here.

Tab. 4: Summary of the statistical parameters of the regression models implemented into SpheraCosmolife

Model		n	R^2	RMSE
NOAEL	Training set	97	0.53	0.61
	Test set	16	0.73	0.49
	Validation test	27	0.60	0.43
Skin permeation logKp model (Potts and Guy)	Training set	271	0.51	0.79
Skin permeation logKp model (ten Berge)	Training set	271	0.58	0.72

3.3 User inputs for SpheraCosmolife

In the first form the user has to insert the product type, the list of ingredients of the product and their concentrations. Figure 1 shows where to insert the information. The product type needs to be selected from a list reporting the product types (e.g. body lotion, face cream, hand cream, shampoo, shower gel, deodorant, etc.) according to the product types in Tables 2A, 2B and 3 of "The SCCS Notes of Guidance for the testing of cosmetic ingredients and their safety evaluation 10th revision" (SCCS, 2018), to define an exposure scenario. The ingredients can be inserted using the SMILES, the INCI, or the CAS format. If INCI or CAS are provided as the only input, it will be possible to process the molecule only if there is a match in the database, otherwise the SMILES of the ingredient is needed. The user can insert a single ingredient or a set of ingredients corresponding to a specific formulation. Another required input is the concentration of each ingredient. Then, the software searches the ingredients in the internal database.

3.4 The SpheraCosmolife software summary results

At the beginning the software provides a summary of the results for all the ingredients in the product in html format. The results depend on the product type and on the concentrations of the ingredients indicated by the user. Figure 2 gives an example of the output table with the summary of the hazard and exposure features of the ingredients indicated in the input file. This summary table reports the ingredients and their concentrations as indicated by the user. It shows whether an ingredient is present in an Annex of Regulation (EC, 2009), if it is mutagenic (Ames test), if it is a sensitizer, the dermal absorption according to the Kroes approach, the MoS, and TTC. No assessment is provided for substances for which no SMILES is retrieved, e.g. for inorganic compounds and mixtures (unless already present in the database). All the assessments refer to the neutralized structure of the compounds as they are neutralized in the process.

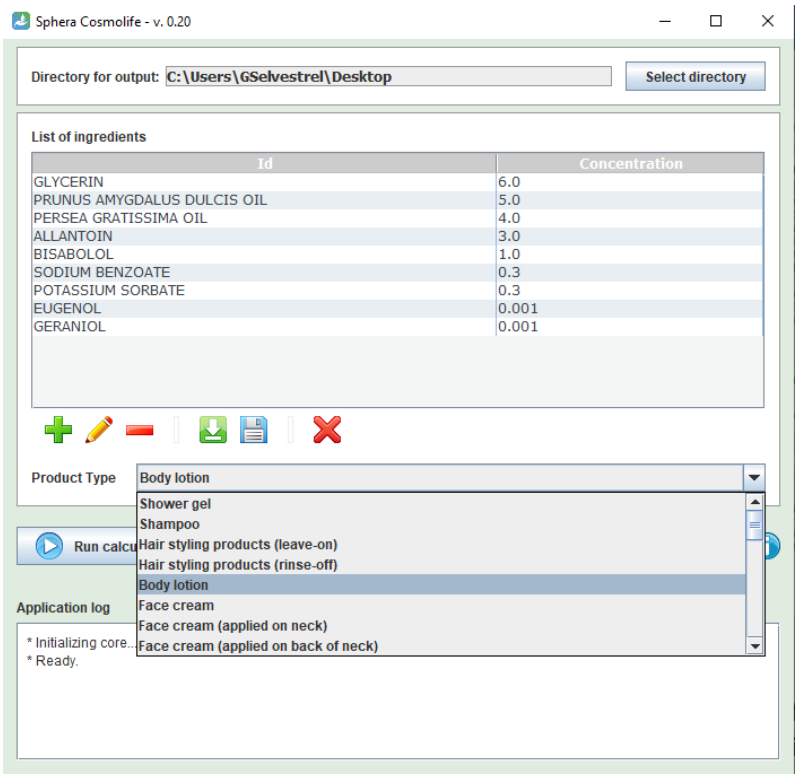

Fig. 1: An example of input for a cosmetic product to be assessed, with hypothetical ingredients

Processed product

Product type: **Body lotion**

Ingredients:

	Ingredient Id	CAS	INCI	Conc. %	Annex	Mutagenicity	Skin Sensitization	Dermal abs.	MoS	TTC
Details	GLYCERIN	56-81-5	GLYCERIN	8.0	-	NON-Mutagen (EXPERIMENTAL value)	NON-Sensitizer (EXPERIMENTAL value)	80%	32.73	0.03 mg/kg bw/day
Details	PRUNUS AMYGDALUS DULCIS OIL		PRUNUS AMYGDALUS DULCIS OIL	5.0	-	-	-	-	-	
Details	PERSEA GRATISSIMA OIL		PERSEA GRATISSIMA OIL	4.0	-	-	-	-	-	
Details	ALLANTOIN	97-59-6	ALLANTOIN	3.0	-	NON-Mutagen (moderate reliability)	Sensitizer (low reliability)	40%	268.64	0.0015 mg/kg bw/day
Details	BISABOLOL	23089-26-1	BISABOLOL	1.0	-	NON-Mutagen (good reliability)	Sensitizer (moderate reliability)	40%	133.5	0.0015 mg/kg bw/day
Details	SODIUM BENZOATE	532-32-1	SODIUM BENZOATE	0.3	V	NON-Mutagen (EXPERIMENTAL value)	NON-Sensitizer (EXPERIMENTAL value)	80%	93.55	0.03 mg/kg bw/day
Details	POTASSIUM SORBATE	24634-61-5	POTASSIUM SORBATE	0.3	V	NON-Mutagen (EXPERIMENTAL value)	Sensitizer (good reliability)	40%	3382.03	0.03 mg/kg bw/day
Details	EUGENOL	97-53-0	EUGENOL	0.001	III	NON-Mutagen (EXPERIMENTAL value)	Sensitizer (EXPERIMENTAL value)	80%	233360.39	0.03 mg/kg bw/day
Details	GERANIOL	106-24-1	GERANIOL	0.001	III	NON-Mutagen (EXPERIMENTAL value)	Sensitizer (EXPERIMENTAL value)	40%	158441.56	0.03 mg/kg bw/day

Fig. 2: An example of the output table with the summary of the hazard and exposure features of hypothetical ingredients of a cosmetic product

3.5 The SpheraCosmolife software detailed results

Clicking on the "Details" box for each ingredient in the first column of the summary table, one gets the detailed results for each ingredient, as an html page. First, the user has information on the substance, with the SMILES, the CAS and INCI, the information on the product type selected, and the concentration of the ingredient, as a percentage (%) and mg/g. Then SpheraCosmolife reports the values from the SCCS Notes of Guidance (SCCS, 2018) related to the product type selected by the user. These values are specific for each product type and serve to calculate the exposure, as described in the next section. These are the parameters provided: the relative daily exposure, the surface area involved, the type of exposure, and the exposure time.

Regarding the presence in an Annex, the presence of a substance in one of these lists does not preclude the possibility of continuing the evaluation. However, the substances in Annex II cannot be used in cosmetic products. This for instance is the case of formaldehyde, and the software provides a warning with a red cell. In other cases, there is a threshold in the Annexes above which the substances cannot be used. For instance, Phenoxyethanol is in Annex V with a maximum concentration of 1.0 %. In this case, the user must use a lower concentration to get a product safe and in compliance with the law.

Moreover, in this section the software checks whether the ingredient is classified according to the Classification, Labelling and Packaging (CLP) Regulation (EC, 2008) and is contained in the Safer Chemical Ingredients List (SCIL); this is a list of chemicals, arranged by functional-use class, that the Safer Choice Program has determined are safer than traditional chemical ingredients[13]. Figure S1[12] gives an example of the information described.

3.5.1. Information on exposure

SpheraCosmolife provides information regarding the exposure to each ingredient, based on the inputs of the user who indicated the product type and the concentration. Figure S2[12] gives an example of the exposure information provided by the software.

[13] https://www.epa.gov/saferchoice/safer-ingredients

The exposure is calculated for different scenarios. External exposure is obtained using the parameters defined by the SCCS for the product type, as described in the Methods section. The SED is obtained for three scenarios: (1) absorption of 100 %, oral or inhalation exposure; (2) absorption of 50 %, a default value for dermal exposure, as indicated by SCCS (SCCS, 2018); (3) or the realistic scenario for dermal absorption based on the models for skin permeation. SpheraCosmolife provides the output of two models for skin permeation, and then chooses the worst case, using the most conservative of these two values. As we explained, for these models and all the other ones, whenever there is an experimental value for the ingredient of interest, the software shows and uses it. Water solubility and the J_{max} are also shown.

Since all these models are implemented in VEGA, the output of the predictions takes into account the reliability of the prediction, measured using the applicability domain index (ADI) calculated by VEGA. ADI is a value that evaluates the reliability of the prediction. For this, it takes into account several parameters, such as the predictions done on similar compounds, the agreement between the predicted value for the target compound and the experimental values of the most similar compounds, the presence of unusual fragments, how similar are the related substances, etc. Full details for each model are given at the VEGA website[8], including the guidance for the users[14]. Thus, for each prediction SpheraCosmolife reports the level of reliability as low, medium or good. This indication should be used as a warning, and the assessor should evaluate if the value can be used or not; in case the assessor needs more information to take the decision, (s)he should use the original VEGA model of interest, which provides full details, while in the SpheraCosmolife report there is only the summary.

3.5.2. Information on hazard values and TTC

The software reports values (experimental if available or predicted) related to: mutagenicity (Ames integrated model), genotoxicity (*in vitro* micronucleus and chromosomal aberration), skin sensitization (CAESAR, DT model and the integrated model), and NOAEL. Figure S3[12] gives an example of the information about hazard assessment. All the values provided are highlighted with a color to assist the user. The cell with the prediction is red if the compound is predicted as mutagenic/genotoxic/sensitizer, or green if it is safe, i.e. if the prediction has good reliability or if an experimental value is provided directly. These colors are orange and yellow respectively if the reliability of the predictions is only moderate or low.

The software carries out a risk characterization, considering the systemic toxicity and the MoS.

The software also includes an assessment considering the TTC approach. In this case, the previous external and internal exposure values are compared with the TTC threshold for the specific ingredient, in order to incorporate dermal bioavailability into the use of TTC for cosmetics. A decision tree is applied (Williams et al., 2016). The output shows a red or green cell, depending on whether external and internal exposure values are above or below the TTC threshold, using the Cramer decision tree, implemented in VEGA. Figure 3 gives an example of the output for the TTC, for the three different SED scenarios. The interest on the use of the TTC value is for substances which are impurities. In general, the evaluation of the other endpoints, if reliable, should be considered more relevant.

Fig. 3: The output of the SpheraCosmolife software for the TTC for a given ingredient (Eugenol in a body lotion scenario)

7. TTC

Cramer class (ToxTree/Vega model prediction	Low (Class I)
TTC according to Cramer classification	0.03 mg/kg bw/day

TTC vs External / Internal exposure	
TTC	0.03 mg/kg bw/day
External exp.	0.0001 mg/kg bw/d
Internal exp. (SED 100%)	0.0001 mg/kg bw/d
Internal exp. (SED 50%)	0.000062 mg/kg bw/d
Internal exp. (SED 10%)	0.000012 mg/kg bw/d

Fig. 4: List of a hypothetical cosmetic product

List of ingredients

Id	Concentration
GLYCERIN	6.0
PRUNUS AMYGDALUS DULCIS OIL	5.0
PERSEA GRATISSIMA OIL	4.0
ALLANTOIN	3.0
BISABOLOL	1.0
SODIUM BENZOATE	0.3
POTASSIUM SORBATE	0.3
EUGENOL	0.001
GERANIOL	0.001

[14] https://www.vegahub.eu/download/vega-interpretation/

Tab. 5: Differences between manual and automated evaluation

Steps	Manual procedure	SpheraCosmolife	Comparison consideration
Input information	Definition of: -product type -ingredients -concentration	User inserts this information. The product type is selected from a picklist. The software automatically retrieves the ingredients within it database. If not present in the database, input structure is based on SMILES strings.	Same conceptual scheme. Same equations and rules to define exposure scenarios. Process is facilitated and much faster. Database checked automatically and not manually.
Output assessment	The assessment is based on *in vitro* tests and data retrieved from the literature and usually no *in silico* inputs are used, for lack of experience. Assessment is time consuming. Reliability and robustness of the data difficult to be considered. Regulatory information needs to be manually retrieved. Often, this situation leads to an incomplete safety report, for lack of data.	The output is obtained in few seconds. Valid and consolidated *in silico* data are available for all endpoints. Reliability of the results is presented using a specific algorithm. In the same table the user can visualize data for exposure, hazard, MoS and TTC. Complete and exhaustive regulatory data (presence in the annexes, maximum threshold, etc.) are automatically retrieved within the internal database.	SpheraCosmolife facilitates the safety assessment process. Additional information is available from *in silico* models, with related uncertainty.
Regulatory information	The assessor has to browse through the regulation to retrieved information.	SpheraCosmolife automatically provides all the detailed information retrieved on the cosmetic Regulation thanks to its internal database.	SpheraCosmolife offers a complete overview of the regulatory conditions of the different ingredients. Moreover, information from US EPA is provided.
Hazard and exposure assessment	Data retrieving is a very time consuming. Data retrieved from the literature need to be curated and correctly interpreted.	The process is much faster. Experimental data have been curated and the predictions report reliability, to facilitate the interpretation of the results.	This tool integrates within the same platform models for hazard and exposure. Risk assessment is done immediately, introducing pieces of information commonly not addressed (internal exposure, Kroes approach, etc.) which improves the safety assessment process. Some experience on the interpretation of the VEGA results is needed.

3.5.3. Examples: comparison of the procedures and application of the software on a real case

As shown in the previous sections, SpheraCosmolife represents an innovative and robust system able to guide the assessor during her/his evaluation analysis. With this tool it is possible to evaluate a single ingredient, but it has been designed to work with a number of ingredients typically used in a cosmetic product. It is important to notice that the tool offers the possibility to process a great number of substances in few seconds, saving time and money, considerably increasing the results of the evaluation.

A first example is the evaluation of a hypothetical cosmetic product (body lotion) with a list of ingredients as described in the Figure 2. Table 5 shows how the formulation is evaluated with or without the tool. Figure 4 reproduces the list of ingredients shown in Figure 2.

Another use of the tool is the evaluation of impurities. The Article 17 of the EU Regulation 1223/2009 on cosmetic products clearly explains that the non-intended presence of a small quantity of a prohibited substance (impurity) which is technically unavoidable in good manufacturing practice, shall be permitted providing that such presence is in conformity with safety requirements. As an indicative real case example, we consider the impurities detected during analytical checks of volatile organic compound (VOC) in a nail product. Figure 5 shows the analytical results for a nail polish. The VOC analysis detected methanol (CAS 67-56-1), styrene (CAS 100-42-5), toluene (CAS 108-88-3) and dichloromethane (CAS 75-09-2).

Looking at figure 5, the regulatory context is immediately clear. Methanol and toluene are substances included in Annex III (list of restricted ingredients). Therefore, even if they are not intentionally added, a concern related to their safety is limited, since all the substances included in this list have been already evaluated by SCCS. In this specific case, it is possible to read that toluene can be used in nail polish up to 25% (Figure 6), while methanol is allowed as impurity of ethanol or isopropyl alcohol up to 5% (Figure 7).

Styrene and dichloromethane are substances prohibited as cosmetic ingredients (included in Annex II), officially known to be substances of concern. However, it is possible to check the specific exposure to the users and define their risk in use. In this case, concentrations are low enough (0.002 % for styrene and 0.0003 % for dichloromethane), to consider them as ‘traces’ and thus an evaluation using the TTC approach is possible.

Processed product

Product type: **Nail polish**

Ingredients:

	Ingredient Id	CAS	INCI	Conc. %	Annex	Mutagenicity	Skin Sensitization	Dermal abs.	MoS	TTC
Details	methanol	67-56-1	methanol	0.2	III	NON-Mutagen (EXPERIMENTAL value)	NON-Sensitizer (good reliability)	80%	62500	0.03 mg/kg bw/day
Details	styrene	100-42-5	styrene	0.002	II	NON-Mutagen (EXPERIMENTAL value)	Sensitizer (moderate reliability)	80%	69500	0.03 mg/kg bw/day
Details	toluene	108-88-3	toluene	6.0E-4	III	NON-Mutagen (EXPERIMENTAL value)	Sensitizer (moderate reliability)	80%	290416.67	0.03 mg/kg bw/day
Details	75-09-2	75-09-2	DICHLOROMETHANE	3.0E-4	II	Mutagen (EXPERIMENTAL value)	NON-Sensitizer (low reliability)	80%	4681666.67	0.0015 mg/kg bw/day

Fig. 5: Analytical results (VOC) for a nail polish

Molecules found in the annex lists: 1

Found molecule no. 1	
CAS	108-88-3
INCI	TOLUENE
Found in annex	ANNEX III: LIST OF SUBSTANCES WHICH COSMETIC PRODUCTS MUST NOT CONTAIN EXCEPT SUBJECT TO THE RESTRINCTIONS LAID DOWN
Annex details	ANNEX III: Product type, body part: Nail products ANNEX III: MAX Concentration in ready for use preparation: 25% ANNEX III: Wording of conditions of use and warnings: Keep out of reach of children To be used by adults only ANNEX III: NOTES: Reprotoxic 2 >1029/06 - Opinion on Toluene (its use as a solvent in nail cosmetics) >1170/08 - Opinion on Toluene (its use as a solvent in nail cosmetics)
Found Safer Chemicals classification	No
Found in CLP list	Yes
CLP details	CLP HARMONIZED CLASSIFICATION (Link): https://echa.europa.eu/it/information-on-chemicals/cl-inventory-database/-/discli/details/30426 HAZARD CLASS AND CATEGORY CODE (CLASSIFICATION): Flam. Liq. 2 Repr. 2 Asp. Tox. 1 STOT SE 3 STOT RE 2 * Skin Irrit. 2 HAZARD STATEMENT CODE (CLASSIFICATION): H225 H361d *** H304 H336 H373 ** H315 PICTOGRAM, SIGNAL WORD CODE (LABELLING): GHS02 GHS08 GHS07 Dgr HAZARD STATEMENT CODE (LABELLING): H225 H361d *** H304 H373 ** H315 H336

Fig. 6: Regulatory information provided by SpheraCosmolife for toluene

Molecules found in the annex lists: 1

Found molecule no. 1	
CAS	67-56-1
INCI	METHYL ALCOHOL
Found in annex	ANNEX III: LIST OF SUBSTANCES WHICH COSMETIC PRODUCTS MUST NOT CONTAIN EXCEPT SUBJECT TO THE RESTRINCTIONS LAID DOWN
Annex details	ANNEX III: Product type, body part: Denaturant for ethanol and isopropyl alcohol ANNEX III: MAX Concentration in ready for use preparation: 5 % (as a % ethanol and isopropyl alcohol)
Found Safer Chemicals classification	No
Found in CLP list	Yes
CLP details	CLP HARMONIZED CLASSIFICATION (Link): https://echa.europa.eu/it/information-on-chemicals/cl-inventory-database/-/discli/details/37212 HAZARD CLASS AND CATEGORY CODE (CLASSIFICATION): Flam. Liq. 2 Acute Tox. 3 * Acute Tox. 3 * Acute Tox. 3 * STOT SE 1 HAZARD STATEMENT CODE (CLASSIFICATION): H225 H331 H311 H301 H370 ** PICTOGRAM, SIGNAL WORD CODE (LABELLING): GHS02 GHS06 GHS08 Dgr HAZARD STATEMENT CODE (LABELLING): H225 H331 H311 H301 H370 **

Fig. 7: Regulatory information provided by SpheraCosmolife for methanol

Figure 5 shows that according to the Cramer decision tree styrene is in class I (low level of concern; TTC = 0.03 mg/kg bw/day), while dichloromethane is in class III (high level of concern; TTC = 0.0015 mg/kg bw/day). Comparing these TTC thresholds with the exposure calculated by SpheraCosmolife (see Figure 8 and 9), an indication of the risk associated with the presence of an unexpected substance is shown, helping the safety assessor during the evaluation process.

A last example of the use of the tool is the evaluation of a botanic derivative ingredient. Botanical extracts are largely used in cosmetic industry because they are relatively easy to obtain, they are generally considered safe (as they are of natural origin) and they are used commercially for marketing descriptions of cosmetic products. SpheraCosmolife allows a rapid evaluation of the botanic ingredient, as shown in the example described in Figure 10. The result is a safety profile of the botanic derivative ingredient that can be taken into consideration by the safety assessor to define the safety of the cosmetic product where the botanic ingredient is used.

4 Discussion

4.1 The novelty of the tool

For the assessment of cosmetic products *in silico* models offer a unique opportunity, as they can process a very large number of ingredients quickly, without the need of new, additional tests, saving money and time (Raitano et al., 2019; Gellatly and Sewell, 2019; Taylor and Rego Alvarez, 2020). The assessment of cosmetic ingredients can benefit from the hundreds of *in silico* models available predicting the properties of interest for many endpoints. Some of these models are freely available, others are commercial, or require a fee. For instance, the models to predict Ames test are the most numerous, and they have been reviewed in some papers (Cassano et al., 2014; Honma et al., 2019). In spite of the number of models available, their application can be complicated as the user should run separate models in different platforms, which may require different input

Fig. 8: TTC evaluation for styrene

7. TTC

Cramer class (ToxTree/Vega model prediction	Low (Class I)
TTC according to Cramer classification	0.03 mg/kg bw/day

TTC vs External / Internal exposure	
TTC	0.03 mg/kg bw/day
External exp.	0.0001 mg/kg bw/d
Internal exp. (SED 100%)	0.0001 mg/kg bw/d
Internal exp. (SED 50%)	0.00005 mg/kg bw/d
Internal exp. (SED 80%)	0.00008 mg/kg bw/d

Fig. 9: TTC evaluation for dichloromethane

7. TTC

Cramer class (ToxTree/Vega model prediction	High (Class III)
TTC according to Cramer classification	0.0015 mg/kg bw/day

TTC vs External / Internal exposure	
TTC	0.0015 mg/kg bw/day
External exp.	0.000015 mg/kg bw/d
Internal exp. (SED 100%)	0.000015 mg/kg bw/d
Internal exp. (SED 50%)	0.000007 mg/kg bw/d
Internal exp. (SED 80%)	0.000012 mg/kg bw/d

Fig. 10: Saffron (CROCUS SATIVUS STIGMA EXTRACT) components

Processed product

Product type: **Face cream**

Ingredients:

	Ingredient Id	CAS	INCI		Annex	Mutagenicity	Skin Sensitization	Dermal abs.
Details	Ingredient no. 1	27876-94-4	CI 75100	*crocetin*	IV	Mutagen (moderate reliability)	Sensitizer (low reliability)	40%
Details	Ingredient no. 2			*crocin*	-	NON-Mutagen (EXPERIMENTAL value)	Sensitizer (low reliability)	10%
Details	Ingredient no. 3			*picrocrocin*	-	Mutagen (moderate reliability)	NON-Sensitizer (low reliability)	10%
Details	Ingredient no. 4	116-26-7	2,3-DIHYDRO-2,2,6-TRIMETHYLBENZALDEHYDE	*safranal*	-	NON-Mutagen (good reliability)	Sensitizer (EXPERIMENTAL value)	40%

formats, specific instructions, or provide outputs that may be difficult to integrate and compare. Furthermore, for risk assessment independent models for hazard and exposure assessment need to be run. All these difficulties represent a barrier to the use of *in silico* models for cosmetics.

Within the EC project VERMEER, we wanted to integrate models for hazard and exposure. The SpheraCosmolife software system presented here is one of the tools we are developing to increase the use of *in silico* tools for risk assessment, improving at the same time the robustness and the reliability of the results.

A major innovative aspect introduced by SpheraCosmolife is that the system does not provide the output for one single endpoint, as in the usual situation, but a battery of models specific for the application are wrapped into a single system. The user is not required to learn multiple programs, because the input is the same for all the models which run automatically in the background.

A second innovative aspect is that SpheraCosmolife integrates models for hazard and exposure within the same platform. Historically, models for exposure have been developed separately, typically within commercial platforms, while the platforms for hazard models do not contain tools for exposure. SpheraCosmolife offers a novel perspective, because it integrates both.

A third innovative aspect is that this system is dedicated for the specific sector of cosmetics. Thus, it contains legislative thresholds and specific references to the European directive, proceeding in the direction of a practical application. Usually, the existing platforms are generic, and intended to be used for multiple purposes.

A fourth innovative aspect of this system is that it copes with products, not only ingredients, and thus it addresses simultaneously a series of substances which are present in the cosmetic product. This also helps the user, interested in the practical case where multiple substances are used.

The safety evaluation is performed easily: user provides the structure of the chemicals of interested, the type of product in which the chemical will be used as ingredients and the concentrations in the final formulations. Then, SpheraCosmolife executes automatically several analyses and predictions providing an overall evaluation of the formulation in a structured output report.

SpheraCosmolife is opening up fresh avenues to research on *in silico* models, moving towards the real case application, facing the practical problems of a sector urgently demanding solutions to assess products, and introducing novel topics to take account of the peculiarities of a focused sector. The effort to apply models on practical cases will provide solutions lowering the barriers to the use of *in silico* models, which tend to be theoretical tools. *In silico* models have to "learn" the user's requirements, and too often the developers oblige the user to learn programs which only partially assist them. A dialogue has to be established between the users and the developers, and this has been accomplished within the VERMEER project. Development of the overall architecture should start with the very initial phases, understanding the users' needs; we codified the different endpoints used by the assessor of cosmetic products, and analyzed how the assessment is done without *in silico* model. Then, we tried to replicate these steps using different software modules, some already existing, or developed on purpose. This process is close to the development of an expert system software, and we inserted a number of tools that are statistical-based, not only expert-based rules.

4.2 The limitations of the tool

The novel system should be considered as a help for the assessor, and not as a replacement of the human evaluation. In the case that experimental values are reported, the evaluation is easier. The major novelty is in the organized series of *in silico* models. However, particularly in this case, the user should consider the reliability of the values, as indicated by the system. The fact that there are elements of uncertainty clearly indicates that the system is not perfect. The assessor should use the uncertainty in the final assessment. It boosts the objectivity of the assessment, showing systematically the different levels of evidence, experimental and calculated. The assessor should use expert knowledge particularly when the uncertainty is high. Since SpheraCosmolife is supported by the VEGA platform, the ADI tool to evaluate the reliability of the prediction can be used, and further elements can be obtained using the VEGA models individually, downloaded from the VEGA website. In this case, the individual models show as further elements the similar compounds, which can be used for read across, and more details on the ADI. This is a quantitative value, based on three fundamental components at the basis of any *in silico* models: the chemical information, the toxicological/property information and the algorithm. Each of these components is used within the ADI. VEGA evaluates similar chemicals and the uncommon features from a chemical point of view. The ADI also investigates the toxicological profiles and features of the related substances (thus not only based on the chemical similarity) using, for instance, toxicological alerts. Finally, the uncertainty of the algorithm is addressed too. Combining all these analyses regarding the applicability domain, VEGA provides an assessment of the reliability of the predictions, which is reported by SpheraCosmolife.

The user should be aware that for some endpoints the uncertainty is higher, and Tables 2 and 3 provide a first indication. Examples of areas with larger uncertainty are the predictive tool for NOAEL, which is a difficult endpoint, because it is affected by the natural variability, and by the choice of the doses used within the experimental test. The current version of the model is based on a limited training set. We are working to reduce uncertainty and in the future new models for NOAEL will be added. Skin sensitization is another endpoint with uncertainty. Both models for this endpoint implemented in the system are quite conservative, and as a consequence there are some false positives. Also in this case, we are developing new models, which will be implemented in a new version. As we said, in case of uncertainty the user can run the individual models within VEGA and look at the similar compounds, taking advantage of the read across approach. In the future, read across will be also implemented into SpheraCosmolife, and further endpoints added.

5 Conclusions

We introduced for the first time a single software system, SpheraCosmolife, to facilitate and harmonize the safety assessment of cosmetic products. The tool is specific for this kind of product, and refers to the thresholds and requirements relative to cosmetics. The software system provides the MoS, based on the systemic exposure dose, and includes a number of models for both exposure and hazard prediction. Different scenarios are provided based on the use or not of the results of the skin permeation models. The software aims to be user-friendly, requiring a very limited number of inputs from the user, while it refers to the internal database and models, to provide an evaluation even when experimental values are missing. Development of the system started since its early stages phase on the user's requirements, rather than of offering a set of prebuilt tools which could be useful. The level of uncertainty of the results is indicated, and the assessor has supporting information for the final evaluation. SpheraCosmolife has been designed with flexible capabilities for future extensions in mind and more features will be added in future versions. New functionalities will be added to accommodate the requests of the end-users. Frequency of the updates depends on the progress of the work necessary to create new models and evaluate the consistency of new toxicological endpoints, but a version 2.0 is planned for the end of 2021. Nevertheless, we feel that it can already help in the assessment of cosmetic products.

References

Asturiol, D., Casati, S. and Worth, A. (2016). Consensus of classification trees for skin sensitisation hazard prediction. *Toxicology in Vitro 36*, 197–209. doi:10.1016/j.tiv.2016.07.014.

Baderna, D., Gadaleta, D., Lostaglio, E. et al. (2020). New in silico models to predict *in vitro* micronucleus induction as marker of genotoxicity. *J Hazard Mater 385*, 121638. doi:10.1016/j.jhazmat.2019.121638.

Benfenati, E., Golbamaki, A., Raitano, G. et al. (2018). A large comparison of integrated SAR/QSAR models of the Ames test for mutagenicity. *SAR QSAR Environ Res 29, 591-611*. doi:10.1080/1062936X.2018.1497702.

Berggren, E., White, A., Ouedraogo, G. et al. (2017). Ab initio chemical safety assessment: A workflow based on exposure considerations and non-animal methods. *Comput Toxicol 4*, 31–44. doi:10.1016/j.comtox.2017.10.001.

Carnesecchi, E., Raitano, G., Gamba, A. et al. (2020). Evaluation of non-commercial models for genotoxicity and carcinogenicity in the assessment of EFSA's databases. *SAR QSAR Environ Res 31, 33-48.* doi:10.1080/1062936X.2019.1690045.

Cassano, A., Raitano, G., Mombelli, E. et al. (2014). Evaluation of QSAR models for the prediction of Ames genotoxicity: a retrospective exercise on the chemical substances registered under the EU REACH regulation. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 32*, 273–298. doi:10.1080/10590501.2014.938955.

Chaudhry, Q., Piclin, N., Cotterill, J. et al. (2010). Global QSAR models of skin sensitisers for regulatory purposes. *Chem Cent J 4 Suppl 1*, S5. doi:10.1186/1752-153X-4-S1-S5.

Cramer, G. M., Ford, R. A. and Hall, R. L. (1978). Estimation of toxic hazard--a decision tree approach. *Food Cosmet Toxicol 16*, 255–276. doi:10.1016/s0015-6264(76)80522-6.

EC - European Commission (2009). Regulation (EC) No.1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. Official Journal of European Union. L, 342, 59-209. http://data.europa.eu/eli/reg/2009/1223/2020-05-01

EC - European Commission (2008). Regulation (EC) No. 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006. Official Journal L 353, 31/12/2008, p. 1-1355. http://data.europa.eu/eli/reg/2009/1223/2020-05-01

Gadaleta, D., Lombardo, A., Toma, C. et al. (2018). A new semi-automated workflow for chemical data retrieval and quality checking for modeling applications. *J Cheminformatics 10*, 60. doi:10.1186/s13321-018-0315-6.

Gellatly, N. and Sewell, F. (2019). Regulatory acceptance of in silico approaches for the safety assessment of cosmetic-related substances. *Comput Toxicol 11*, 82–89. doi:10.1016/j.comtox.2019.03.003.

Honma, M., Kitazawa, A., Cayley, A. et al. (2019). Improvement of quantitative structure-activity relationship (QSAR) tools for predicting Ames mutagenicity: outcomes of the Ames/QSAR International Challenge Project. *Mutagenesis 34*, 3–16. doi:10.1093/mutage/gey031.

ICH guideline M7 (R1) on assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk. EMA/CHMP/ICH/83812/2013. (2017) https://www.ema.europa.eu/en/documents/scientific-guideline/ich-guideline-m7r1-assessment-control-dna-reactive-mutagenic-impurities-pharmaceuticals-limit_en.pdf

Kroes, R., Renwick, A. G., Cheeseman, M. et al. (2004). Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. *Food Chem Toxicol 42*, 65–83. doi:10.1016/j.fct.2003.08.006.

Kroes, R., Renwick, A. G., Feron, V. et al. (2007). Application of the threshold of toxicological concern (TTC) to the safety evaluation of cosmetic ingredients. *Food Chem Toxicol 45*, 2533–2562. doi:10.1016/j.fct.2007.06.021.

Manganelli, S., Schilter, B., Benfenati, E. et al. (2018). Integrated strategy for mutagenicity prediction applied to food contact chemicals. *ALTEX 35*, 169–178. doi:10.14573/altex.1707171.

Manganelli, S., Roncaglioni, A., Mansouri, K. et al. (2019). Development, validation and integration of in silico models to identify androgen active chemicals. *Chemosphere 220*, 204–215. doi:10.1016/j.chemosphere.2018.12.131.

Munro, I. C., Ford, R. A., Kennepohl, E. et al. (1996). Correlation of structural class with no-observed-effect levels: a proposal for establishing a threshold of concern. *Food Chem Toxicol 34*, 829–867. doi:10.1016/s0278-6915(96)00049-x.

Patlewicz, G., Jeliazkova, N., Safford, R. J. et al. (2008). An evaluation of the implementation of the Cramer classification scheme in the Toxtree software. *SAR QSAR Environ Res 19*, 495–524. doi:10.1080/10629360802083871.

Potts, R. O. and Guy, R. H. (1992). Predicting skin permeability. *Pharm Res 9*, 663–669. doi:10.1023/a:1015810312465.

Raitano, G., Roncaglioni, A., Manganaro, A. et al. (2019). Integrating in silico models for the prediction of mutagenicity (Ames test) of botanical ingredients of cosmetics. *Comput Toxicol 12*, 100108. doi:10.1016/j.comtox.2019.100108.

Rogiers, V., Benfenati, E., Bernauer, U. et al. (2020). The way forward for assessing the human health safety of cosmetics in the EU - Workshop proceedings. *Toxicology 436*, 152421. doi:10.1016/j.tox.2020.152421.

SCCS - Scientific Committee on Consumer Safety (2018). SCCS Notes of Guidance for the Testing of Cosmetic Ingredients and their Safety Evaluation 10th revision, 24-25 October 2018, SCCS/1602/18. Available at: https://ec.europa.eu/health/sites/health/files/scientific_committees/consumer_safety/docs/sccs_o_224.pdf

SCCS - Scientific Committee on Consumer Safety (2015). SCCS Notes of Guidance for the Testing of Cosmetic Ingredients and their Safety Evaluation 9th revision, Revised version of 25 April 2016, SCCS/1564/15. Available at: https://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_190.pdf

Shen, J., Kromidas, L., Schultz, T. et al. (2014). An in silico skin absorption model for fragrance materials. *Food Chem Toxicol 74*, 164–176. doi:10.1016/j.fct.2014.09.015.

Taylor, K. and Rego Alvarez, L. (2020). Regulatory drivers in the last 20 years towards the use of in silico techniques as replacements to animal testing for cosmetic-related substances. *Comput Toxicol 13*, 100112. doi:10.1016/j.comtox.2019.100112.

Ten Berge, W. (2009). A simple dermal absorption model: derivation and application. *Chemosphere 75*, 1440–1445. doi:10.1016/j.chemosphere.2009.02.043.

Therneau, T. M., Atkinson, E. J. (2015). An Introduction to Recursive Partitioning Using the RPART Routines. Mayo Foundation, Rochester. https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf

Toropov, A. A., Toropova, A. P., Raitano, G. et al. (2019). CORAL: Building up QSAR models for the chromosome aberration test. *Saudi J Biol Sci 26*, 1101–1106. doi:10.1016/j.sjbs.2018.05.013.

Toropov, A. A., Toropova, A. P., Pizzo, F. et al. (2015). CORAL: model for no observed adverse effect level (NOAEL). *Mol Divers 19*, 563–575. doi:10.1007/s11030-015-9587-1.

Vecchia, B.E. and Bunge A.L. (2002a). Skin absorption databases and predictive equations. Chapter 3 in Transdermal Drug Delivery, edited by Guy RH and Hadgraft J, Publisher Marcel Dekker. https://www.researchgate.net/publication/272149756_Skin_Absorption_Databases_and_Predictive_Equations

Williams, F. M., Rothe, H., Barrett, G. et al. (2016). Assessing the safety of cosmetic chemicals: Consideration of a flux decision tree to predict dermally delivered systemic dose for comparison with oral TTC (Threshold of Toxicological Concern). *Regul Toxicol Pharmacol 76*, 174–186. doi:10.1016/j.yrtph.2016.01.005.

Worth, A., Cronin, M., Enoch, S., et al. (2012). Applicability of the Threshold of Toxicological Concern (TTC) approach to cosmetics – preliminary analysis. European Union. doi:10.2788/5059

Conflicts of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

This work was developed under the framework of the LIFE+ Projects VERMEER [LIFE16/ENV/IT/000167] and CONCERT REACH [LIFE17/GIE/IT/000461].