Computational Toxicology 26 (2023) 100272

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.sciencedirect.com/journal/computational-toxicology

The role of a molecular informatics platform to support next generation risk assessment

Chihae Yang [a,b], James F Rathman [b,c], Bruno Bienfait [a], Matthew Burbank [d], Ann Detroyer [d], Steven J. Enoch [e], James W. Firman [e], Steve Gutsell [f], Nicola J. Hewitt [g], Bryan Hobocienski [b], Gerry Kenna [g], Judith C. Madden [e], Tomasz Magdziarz [a], Jörg Marusczyk [a], Aleksandra Mostrag-Szlichtyng [b], Christopher-Tilman Krueger [h], Cathy Lester [i], Catherine Mahoney [j], Abdulkarim Najjar [h], Gladys Ouedraogo [d], Katarzyna R. Przybylak [f], J. Vinicius Ribeiro [b], Mark T.D. Cronin [e,*]

[a] *MN-AM, Nürnberg, Germany*
[b] *MN-AM, Columbus, OH, USA*
[c] *Ohio State University, Columbus, OH, USA*
[d] *L'Oreal, Aulnay-sous-Bois, France*
[e] *Liverpool John Moores University, Liverpool, UK*
[f] *Unilever Research, Sharnbrook, UK*
[g] *Cosmetics Europe, Brussels, Belgium*
[h] *Beiersdorf, Hamburg, Germany*
[i] *Procter and Gamble, Cincinnati, OH, USA*
[j] *Procter and Gamble, Egham, UK*

ARTICLE INFO

Keywords:
Next generation risk assessment
Chemical safety
Read-across

ABSTRACT

Chemoinformatics has been successfully employed in safety assessment through various regulatory programs for which information from databases, as well as predictive methodologies including computational methods, are accepted. One example is the European Union Cosmetics Products Regulations, for which Cosmetics Europe (CE) research activities in non-animal methods have been managed by the Long Range Science Strategy (LRSS)

Abbreviations: 2D, 2-dimensional; 3D, 3-dimensional; AD, applicability domain; ADI, Acceptable Daily Intake; AhR, aryl hydrocarbon receptor; AI, artificial intelligence; AOP, Adverse Outcome Pathway; AQ, Analogue Quality; AR, androgen receptor; AUC, area under the curve, CDER Center for Drug Evaluation and Research; CIR, Cosmetics Ingredients Review; C_{max}, maximum concentration (of substance) in blood/plasma; COC, cohort of concern; CPDB, Carcinogenicity Potency Database; CSRML, Chemical Structure and Reaction Mark-Up Language; CMS, COSMOS; ΔH_f, Heat of Formation; DART, Developmental and Reproductive Toxicity; DPRA, Direct Peptide Reactivity Assay; ECHA, European Chemicals Agency; EC3, Effective Concentration to Induce 3-Fold Increase of Stimulation Index; EDSP, Endocrine Disruption Screening Programme; EFSA, European Food Safety Authority; ER, estrogen receptor; EU, European Union; h-CLAT, Human Cell Line Activation Test; HESS, Hazard Evaluation Support System; HGPRT, hypoxanthine–guanine phosphoribosyltransferase; HOMO, Highest Occupied Molecular Orbital; HTS, High Throughput Screening; IARC, International Agency for Research on Cancer; GRAS, Generally Regarded as Safe; ICCR, International Cooperation on Cosmetics Regulation; ICH, International Council for Harmonisation; ivtCA, *in vitro* chromosome aberration; ivMN, *in vivo* micronucleus; JECFA, The Joint FAO/WHO Expert Committee on Food Additives; LLNA, Local Lymph Node Assay; LO(A)EL, Lowest Observed (Adverse) Effect Level; log Kp, logarithm of the skin permeability coefficient; log P, logarithm of the octanol–water partition coefficient; LUMO, Lowest Unoccupied Molecular Orbital; MIE, molecular initiating event; ML, Machine Learning; MOA, Mode of Action; MOS, Margin of Safety; NDA, New Drug Approvals; NGRA, Next Generation Risk Assessment; NO(A)EL, No Observed (Adverse) Effect Level; NTP, US National Toxicology Program; PBK, Physiologically-Based Kinetic; PK, pharmacokinetic; pNEG, probability of being negative; PPAR, peroxisome-proliferator-activity; pPOS, probability of being positive; PPV, positive predictivity; PR, progesterone receptor; pUNC, probability of being uncertain; qRAX, quantitative read-across; (Q)SAR, (Quantitative) Structure-Activity Relationship; QMRF, QSAR Model Reporting Format; RAX, Read-Across; RDT, Repeat Dose Toxicity; REACH, Registration, Evaluation, Authorisation and Restriction of Chemicals; RfD, Reference Dose; RR, Read-Across Reliability; SAR, structure–activity relationship; SCCS, Scientific Committee of Consumer Safety; SQ, Study Quality; SyGMa, Systematic Generation of potential Metabolites; TF SysTox, Cosmetics Europe Task Force on Systemic Toxicity; TDI, Total Daily Intake; T_{half}, half-life; TK, toxicokinetic; T_{max}, time to reach Cmax in blood/plasma; TPSA, Total Polar Surface Area; TTC, Threshold of Toxicological Concern; US EPA, US Environmental Protection Agency; US FDA, US Food and Drug Administration; UVCB, Unknown or Variable Composition, complex reaction products, or Biological Materials; WOE, Weight of Evidence.

* Corresponding author.
E-mail address: m.t.cronin@ljmu.ac.uk (M.T.D. Cronin).

https://doi.org/10.1016/j.comtox.2023.100272
Received 8 January 2023; Received in revised form 30 March 2023; Accepted 11 April 2023
Available online 17 April 2023

Chemoinformatics
In silico safety assessment

program. The vision is to use mechanistic aspects of existing non-animal methods, as well as New Approach Methodologies (NAMs), to demonstrate that safety assessment of chemicals can be performed using a combination of *in silico* and *in vitro* data. To this end, ChemTunes•ToxGPS® has been adopted as the foundation of the safety assessment system and provides a platform to integrate data and knowledge, and enable toxicity predictions and safety assessments, relevant to cosmetics industries. The ChemTunes•ToxGPS® platform provides chemical, biological, and safety data based both on experiments and predictions, and an interactive/customizable read-across platform. The safety assessment workflow enables users to compile qualified data sources, quantify their reliabilities, and combine them using a weight of evidence approach based on decision theory. The power of this platform was demonstrated through a use case to perform a safety assessment for *Perilla frutescens* through the workflows of threshold of toxicological concern (TTC), *in silico* predictions (QSAR and structural rules) and quantitative read-across (qRAX) assessment for overall safety. The system digitalizes workflows within a knowledge hub, exploiting advanced *in silico* tools in this age of artificial intelligence. The further design of the system for next generation risk assessment (NGRA) is scientifically guided by interactions between the workgroup and international regulatory entities.

1. Introduction

Molecular informatics, which incorporates technologies from both chemistry (chemoinformatics) and biology (bioinformatics), has become an advanced tool and a key component in modern structure-based design of new chemical entities. Chemo- and bio- informatics bring together technologies to capture information from chemical structure, molecular and physicochemical properties, as well as biological activities and effects. Specifically, chemoinformatics encompasses techniques to develop databases of chemical structures with associated tools to mine these data and develop predictive modeling strategies. As such, chemoinformatics has become integral in many industrial sectors to leverage chemistry knowledge to make new discoveries. Many databases of chemical structures are linked to biological data which have been measured in a multitude of assays ranging from omics techniques, *in vitro* tests, *in vivo* responses. The power of informatics approaches is to capture, harness and interrogate the information to learn new knowledge and assist in decision making using that knowledge [1], then incorporating the new knowledge back into the system beyond the initial cycle.

Whilst originally focused on drug discovery with the aim of identifying and optimizing pharmacological activity, many computational tools based on molecular informatics have also been applied to safety assessment of chemicals. These tools utilize the strengths of molecular informatics, which enable efficient use of resources to search complex databases, support making inferences and draw conclusions, identify potential hazards, and support safety assessments. Informatics tools find particular use in product development, but also in regulatory submissions across numerous diverse sectors. Several international regulations, such as the European Union (EU) Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) and Cosmetics Products Regulation, as well as International Council for Harmonisation (ICH) guideline M7(R1) on assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals, encourage or mandate the use of non-animal methods, placing a greater reliance on *in silico* toxicology [2]. As a result, the need to satisfy regulatory requirements has motivated the development of predictive methodologies implemented in public and commercial chemoinformatics systems for chemical safety. Such predictive *in silico* methodologies encompass a variety of approaches including read-across, (quantitative) structure–activity relationships ((Q)SARs) and modeling of pharmacokinetics and exposure [1].

With regard to the promotion of the need for computational methods for safety assessment, one of the most significant pieces of legislation was European Regulation EC N°1223/2009 [3]. The Cosmetics Directive was originally adopted in 1976 and this Directive was replaced in 2009 to enable further harmonization and the EU-wide Cosmetics Products Regulation entered into force in July 2013. A full animal testing and marketing ban on cosmetic ingredients, or combinations thereof, entered into force on 11 March 2013. This prohibited generation of animal data solely for the safety of cosmetic products and their ingredients [4]. To this end, among the array of computational and *in silico* tools, several systems were specifically developed to meet the needs for assessing the safety of cosmetics ingredients, which are often data-poor. A full review of the *in silico* tools available to address the diverse perspectives of cosmetics safety assessment is available [5].

To illustrate the potential for using computational approaches, in combination with targeted *in vitro* data, the International Cooperation on Cosmetics Regulation (ICCR) proposed a set of principles for the risk assessment of cosmetics ingredients [6]. The ICCR principles built upon the SEURAT-1 workflow [7] and have been a key driving force behind Next Generation Risk Assessment (NGRA). Through Cosmetic Europe's (CE's) Long Range Science Strategy (LRSS) initiative, a workflow for NGRA has been assessed using several case studies [6,8-10]. The workflow is intended to handle the complex nature of safety/risk assessment of target substances for which specific data are not available. The main components of the workflow are summarized in Fig. 1.

As the NGRA workflow relies heavily on informatics and computational tools, the CE Task Force on Systemic Toxicity (TF SysTox) identified the need to develop a chemoinformatics platform capable of integration and federation of relevant data, prediction of properties and effects along with quantifiable reliability measures, application of exposure-based approaches such as the threshold of toxicological concern (TTC), and that provides a read-across workflow that draws upon all of the above. These issues are well represented systematically in a recent review by Cronin et al. of the landscape of informatics and computational tools available for safety assessment of cosmetics-related chemicals [5]. To this end, the chemoinformatics workgroup of the TF SysT established in Cosmetics Europe (2017–2022) decided to adopt ChemTunes•ToxGPS® [11] as the basis of their platform to digitalize the complex NGRA process defined in Fig. 1.

As illustrated in Fig. 1, the workflows adopted by domain experts (SEURAT-1 workflow [12]) are mapped to the functional workflows of a platform (digitalized software workflow). Whilst current informatics systems focus predominantly on chemistry, there are opportunities to extend functionality with regard to the needs of the NGRA (and other) safety assessment workflows.

There are also other opportunities for further developing a molecular informatics platform in the current age of sophisticated artificial intelligence (AI). The workflow for NGRA should be equipped with data and functionalities of both chemistry and biology, not only to provide data and knowledge, but also to ultimately assist or facilitate timely decision-making in both domains within a single framework. There should be a seamless conduit to integrate and federate both chemical and biological domains as the core resources of the system that should include the following components:

- Compound safety data, study data (chemical-specific and chemical-agnostic biology), alerts/profilers/models, underlying rules (meta data)

- Safety data: human-based limit values for a variety of product use types involved in regulatory offices and authorities
- Toxicity data – *in vitro*, historical *in vivo* data from animals, toxicogenomics/pharmacology
- Metabolism data – *in vitro* (human/animal), historical *in vivo* data from animals, human data
- Biology data – transcriptomics, pathways
- Prediction models to fill data gaps – e.g., QSARs and structure-based rules for human health-based endpoints, metabolites, etc.
- Prediction models to predict external and internal exposure – e.g., physiologically-based kinetic (PBK) models
- Safety assessment workflows – TTC, read-across

The current paper describes the development of an informatics system to address the needs of safety assessors applying NGRA to cosmetics ingredients. Specifically, it has been designed and built using case studies focused on the application on TTC and conventional read-across (RAX) (Tier 0 in Fig. 1). System requirements were delineated with respect to TTC analysis, conducting a read-across of multiple steps with a variety of considerations, and finally estimating the assessment reliability in order to arrive at a final outcome statement. The aim was to allow these methods to be applied systematically not only to well-characterized structures, but also to complex mixtures observed in botanicals or unknown or variable composition, complex reaction products, or biological materials (UVCB) substances. In this study, we selected compounds often found in botanical extracts and demonstrated how a set of multiple related structures can be handled within the workflow based on the supporting evidence from historical *in vivo* animal, *in vitro*, and *in silico* data. Many of these materials tend to lack compound-specific safety data; whilst some may be considered as Generally Regarded as Safe (GRAS) in the USA [13] or having no safety concerns considering estimated exposure levels, some may be flagged as a potential "constituent of concern". As an illustrative example, the safety assessment of the use of *Perilla frutescens* in a cosmetics formulation is demonstrated in Section 3. The identity and structures of substances for the case study were investigated through the platform resources described in Sections 2.2 – 2.4 to retrieve existing information, apply TTC and allow for the profiling and calculation of compound properties and read-across to fill data gaps.

2. Molecular informatics platform

2.1. ChemTunes•ToxGPS® platform

An NGRA workflow requires a large number of functionalities and diverse resources accessible via complex integration and federation of tools. For this reason, the desired system was built as an extension to the existing ChemTunes•ToxGPS® platform since the existing platform already met many of the pre-requisites. The NGRA process has been captured via contextual interviews with users and this feedback was then transformed into user requirements, which were used to derive functionalities to design the workflows. As illustrated in Fig. 1, the workflows adopted by domain experts are mapped to the functional workflows of a platform system (digitalized software workflow).

The ChemTunes•ToxGPS® platform uses a 3-tier architecture namely chemoinformatics backend, middle layer for communication, and the frontend web application (Fig. 2). The three key application components of this platform are database (ChemTunes), *in silico* workflows for predictions and assessments (ToxGPS), and computational solutions (Express). A high level architectural diagram is illustrated in Fig. 2.

The ChemTunes database is implemented in the PostgreSQL version 13.4 [14] object-relational database system with RDKit-2021.03.5 [15]. The PostgreSQL system combined with the RDKit chemistry cartridge facilitates structure searching capability based on a set of molecular topological fingerprints. In addition to RDKit, ChemTunes•ToxGPS® also employs the internally developed MOSES [16] and CORINA [17] chemistry libraries developed at MN-AM. They are the essential underlying foundation for calculation of physico-chemical properties and molecular descriptors in ToxGPS workflows.

The platform backend is also capable of communicating with external libraries and tools. These individual components are available in the platform function as the engine underlying the knowledge hub, one purpose of which is to capture and exchange data across multiple knowledge sources to serve diverse users from multidisciplinary and cross-cutting sectors. The conceptual framework is designed to expand the domains of the knowledgebase to also include biology, toxicogenomics, transcriptomics, and molecular pathways although only the chemistry-centric portions of the backend are discussed in this section. This versatility to adapt and expand to non-chemistry domains necessary for molecular informatics is also important since, eventually, molecular awareness of the system in the next generation should include both chemo- and bioinformatics knowledge. Ontologies are also interpreted and stored within the meta-database, which is beyond the scope

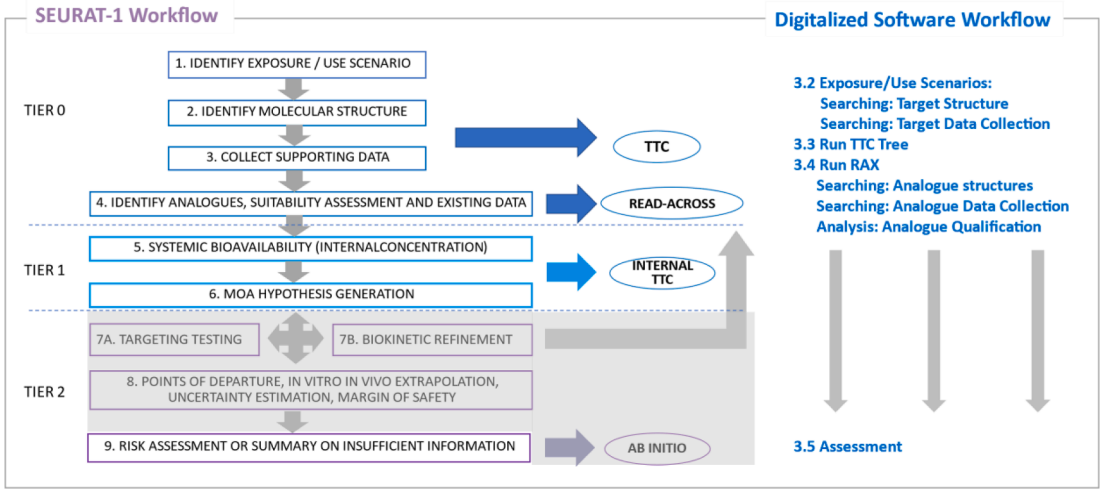


Fig. 1. Expert workflow for Next Generation Risk Assessment and its digitalization procedures [9]. Numeric indices (3.2–3.5) in the digitalized software workflow correspond to the use case sections in this paper.

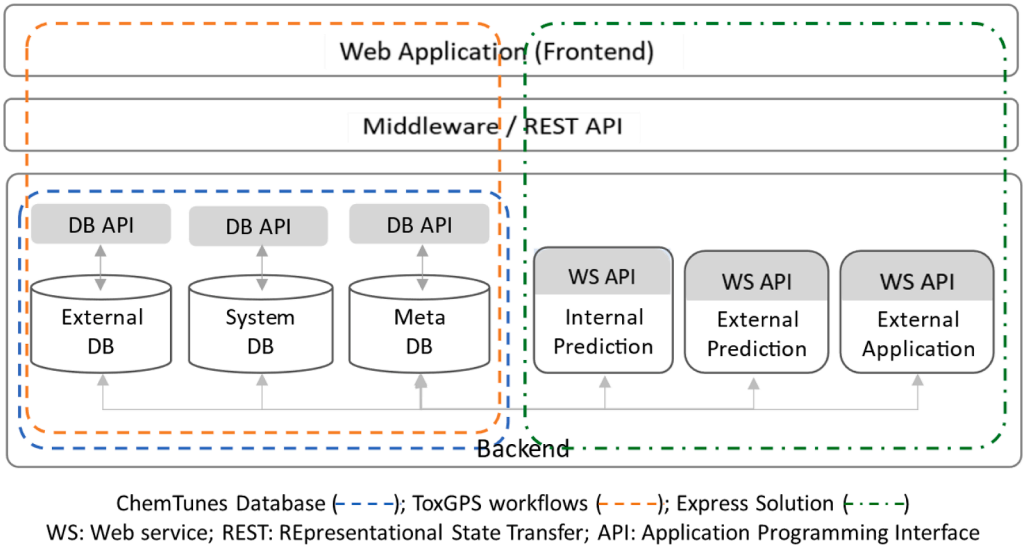


Fig. 2. High level system architecture of the ChemTunes•ToxGPS® platform.

of this manuscript.

2.2. Databases to support NGRA

2.2.1. ChemTunes chemical database

Descriptions of fundamental database architectures have been published previously [18]. Highly curated chemistry content in terms of structures and compound information is available for users of the platform. Each chemical record represents a test substance and is linked to information containing regulatory and biological endpoints including safety assessment, toxicity, bioavailability, and metabolic studies. Sophisticated searching functionalities enable flexible and highly customized searches against the various information types. Since ChemTunes is a chemistry-centered database, a robust compound registry system is required.

Compound Registry: The ChemTunes chemistry database follows the COSMOS (CMS) registry standard [18] and captures rigorous annotations related to structures, compounds, and test articles through a systematic quality control/quality assurance process. The registry system allows both automatic searching and duplicate detection with a systematic quality scoring system to map many sources, as well as a manual expert-mode review of the underlying chemistry database. Chemical structure attributes such as stereochemistry, double bond geometry, material types, and composition types are annotated in detail along with the names and identifiers. Data sources include United States Environmental Protection Agency (US EPA), United States Food and Drug Administration (US FDA), European Chemicals Agency (ECHA) [19,20], European Food Safety Authority (EFSA) OpenFoodTox [21,22], European Commission cosmetic ingredient database [23], PubChem Project, ChemSpider, and CAS Common Chemistry [24]. Over 500,000 curated substance records are currently stored in the registry and serve as the supporting resources for chemistry databases of ChemTunes and COSMOS NG [25]. Nearly 120,000 unique CMS-IDs with associated information are made available in ChemTunes•ToxGPS® v.2022.

2.2.2. Regulatory and safety evaluation information

Risk assessment results leading to human safety are easily retrievable in the ChemTunes database. Specifically, the Safety Evaluation browser displays the critical information in this meta database populated with data from opinions and monographs from the regulatory entities. Along with the safety results, quantitative human health-based limit/guidance values are also included, e.g., Acceptable Daily Intake (ADI) (EFSA, The Joint FAO/WHO Expert Committee on Food Additives; (JECFA)), Total Daily Intake (TDI) (EFSA, Hazard Evaluation Support System (HESS [26])), Reference Dose (RfD) (US EPA) [27], Margin of Safety (MOS) ([28]). Also available are benchmark doses for carcinogenicity studies from Carcinogenicity Potency Database [29], US National Toxicology Program (NTP) [30], and International Agency for Research on Cancer (IARC) [31]. Each safety assessment record provides the associated point of departure value from a critical study based on critical effects which are supported by the underlying toxicity studies in the ChemTunes database.

2.2.3. ChemTunes toxicity database

Experimental toxicity data are compiled from regulatory sources as well as the open literature. The *in vivo* and *in vitro* experimental toxicity data cover over 80 endpoints including, but not limited to, genetic toxicity, carcinogenicity, skin sensitization, skin/eye irritation, repeated-dose toxicity for target organs and developmental and reproductive toxicity (DART). The toxicity database data model has been published previously [18]. Over 12,000 substances have some level of toxicity data from 117,762 studies. Also available is a curated no observed (adverse) effect level (NO(A)EL) database for >2,400 test substances, which acts as the underlying data for the NO(A)EL Bounds Estimation *in silico* tool [32]. A summary of the endpoints available for toxicity data is given in Supplementary Information Table S1.

2.2.4. ChemTunes ADME database

Both *in vivo* and *in vitro* experimental data are provided in ChemTunes. Available endpoints include dermal and oral absorptions, bioavailability, PBK and metabolic transformations. A summary of the available bioavailability data is given in Supplementary Information Table S2.

2.2.4.1. Metabolism database. The ChemTunes metabolism database currently has two biotransformation domains, i.e., endogenous and exogenous reactions of xenobiotics. Metabolism data for drugs are mostly based on human hepatocytes from the literature [33] and Pharmacological Reviews for New Drug Approvals (NDA) from US FDA Center for Drug Evaluation and Research (CDER) [34]. Data for pesticides are from both *in vitro* and *in vivo* studies (laboratory rodents, primates, and farm animals) found in EFSA opinions. Combining both drugs and pesticides, transformations of >500 substances by nearly 3,000 phase I and II reactions are covered in the database. The endogenous database contains transformations from, prokaryotes, vertebrates, plants and yeasts, based on 14,540 molecules, 3,912 reactions, and 640

pathways. The metabolism database provides information on reactants, products, and reaction pathways from both *in vitro* and *in vivo* studies where pertinent study design information are captured, including species, sex, strain, cell lines, route of exposure, organ/tissue, compartment, activation, microsome, cofactor, medium, duration, dosage, and for data from experimental animals, information on group sizes.

2.2.4.2. Physiologically-Based Kinetics (PBK) database. The PBK database contains both pharmacokinetic (PK) and toxicokinetic (TK) studies. The sources of the database are US FDA CDER [34], and various types of EFSA opinions [35]. Substances include drugs (pharmacological reviews of NDA), industrial chemicals, plant protection products (EFSA PPR) and residues, food contact materials (EFSA CEP) and additives (EFSA FAF), food contaminants (EFSA CONTAM), and industrial chemicals [36]. Over 900 total substances provide assay information including kinetics, absorption/bioavailability, excretion, distribution, and bioaccumulation. The database captures study design parameters (route of administration, study duration, doses, dosing regimen, species, sex, number of animals in test and control groups, animals age and/or weight), analytical methods, and results for each assay, e.g., maximum concentration (of substance) in blood/plasma (C_{max}), time to reach C_{max} in blood/plasma (T_{max}), area under the curve (AUC) and half-life (T_{half}).

2.2.5. *In vitro assays*

In vitro assays covering a wide range of endpoints and including AOP-related assays and the Tox21/CAST high throughput assays database, are available in the ChemTunes database, including, but not limited to, additional assay results from the Cosmetics Europe LRSS activities, e.g., Caco-2 cell permeability and *in vitro* skin sensitization assays.

The *in vitro* ToxCast/Tox21 database is available for 9,298 substances and 1,569 assays developed from ToxCast/Tox21 and their collaborators [37]. Due to the sparse nature of this database (i.e., many missing values), the signals from related assays were aggregated to increase sensitivity. Each assay result was transformed to a binary response (active/inactive) so that the data are represented as a compound-assay array, where each compound is represented as a vector of 0/1 values across all *in vitro* assays, with missing values assigned denoted "NaN". For each assay, the mean and standard deviation (ignoring "NaNs") was calculated treating the 0/1 as floating point values, so the mean for a particular assay is then the proportion of active (1) results.

The next step was to group related assays into assay categories. For this aggregation step, the mode of action terms developed by NTP [38] were used. This process enabled the mapping of 900 assays with 27 NTP High Throughput Screening (HTS) Mode of Action (MOA) category terms spanning 43 mechanistic targets. The rest of the unannotated assays were grouped by their assay targets, resulting in a total of 37 category terms covering 1,479 assays. The assay hierarchy is listed in the Supplementary Information Table S3. For example, assays related to endocrine disruption screening used in the Endocrine Disruption Screening Program (EDSP) (64 in total) and DNA binding/damage/repair (86 in total) were aggregated into categories containing fewer groups of related assays. The responses for individual assays within each group were combined to enhance the signal so that the resulting data array has a much lower proportion of missing values than the original unaggregated compound-assay array.

2.3. *ToxGPS in silico knowledge*

ToxGPS offers *in silico* knowledge in the areas of molecular properties, chemotype profilers, QSAR and rule-base models as well as workflows and tools for read-across.

2.3.1. *Molecular & physicochemical properties*

Three types of molecular properties are calculated in ToxGPS: 1) whole molecule 2-dimensional (2D) or 3-dimensional (3D)); 2) shape (3-dimensional (3D)); 3) quantum mechanics-based (3D) descriptors. Whole molecule properties are further divided into three different groups to address bonding interactions, interfacial/bioavailability, and molecular size. Shape descriptors calculated from 3D structural conformations capture molecular characteristics that play a role in long-range interactions important in macromolecule binding (asphericity eccentricity, diameter, radius of gyration, moments of inertia). Various energy and physicochemical properties obtained by quantum mechanical calculations provide chemical reactivities (Heat of Formation (ΔH_f), energies of the Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO), HOMO/LUMO Gap) [39]. Three-D structures are generated by CORINA [40] implemented within ToxGPS. More details on the molecular descriptors are available in the Supplementary Information and have been previously published [41]. Additional physicochemical properties are also implemented in the Express solution in the ChemTunes•ToxGPS® platform. Supplementary Information Table S4 in summarizes the available properties and their possible use.

2.3.2. *Chemotype profilers*

Chemotype profiling attempts to identify structural (sub-)fragments that are associated with some activity, property or PK effects. Three types of structure profiling methods are available in ToxGPS, ranging from general feature mapping to endpoint-specific rules coded with sensitivity and specificity statistics derived from testing the rules against knowledgebase datasets. These chemotype profilers are coded in Chemical Structure and Reaction Mark-Up Language (CSRML), in which electronic, atomic, and bond properties can be encoded in addition to molecular connectivity [42].

2.3.2.1. Profilers. In ToxGPS, the term "profilers" is used in general where chemotypes are applied to associate structural features with chemical binding or reactivity towards biological target molecules. Profilers are not specifically intended to reflect endpoint associations, nor the likelihood of such events. Examples are DNA and/or protein binders [43-45]. ToxPrint chemotypes are offered only to allow visualization of each chemotype defined for fingerprinting structures. Although ToxPrint chemotypes cover the chemical space of large diverse toxicity databases and are designed with toxicity modes of action in mind, these profilers should not be interpreted as endpoint-related alerts unless this is justified by further analysis against a training set to demonstrate a sufficiently strong statistical association [46].

2.3.2.2. Alerts. A structural alert is traditionally a fragment matching a particular structure feature that is known to trigger a warning for a certain endpoint or is associated with a molecular initiating event [47]. However, it should be noted that an alert does not provide a measure of reliability or likelihood that the possibility of an adverse event may occur. Chemotype profilers in ToxGPS such as liver toxicity [48-53], DART decision tree [54], and mitochondrial toxicity [55] are comprised of alerts.

2.3.2.3. Rules. In ToxGPS, a rule is defined as a chemotype that has been associated quantitatively with a biological activity or toxicity endpoint. Whilst the structural category is visualized in the chemotype profiler frontend in the same way as any profiler or alert, a rule is in fact coded with the positive predictivity (PPV) and an odds ratio as determined from the underlying knowledgebase. These ToxGPS rules include steatosis, skin permeability, and liver and skin metabolic reactivities. Toxicity endpoint-specific rules available in the Chemotype Profilers may overlap with some of the *in silico* prediction methods, but, for the most part, the profilers are more general than the rule-base predictions and can be used in the combination method described in Section 2.4.

Fig. 3 illustrates various chemotype profiler matches for beta-ionone,

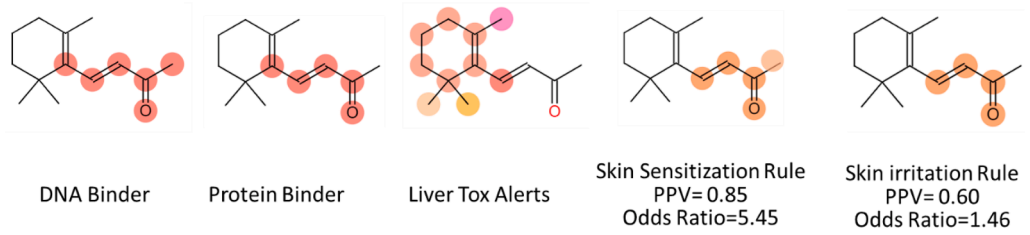


Fig. 3. Profiling of beta-ionone by various chemotypes (profilers, alerts and rules).

one of the constituents of the *Perrilla frutescens* extract (a case study in Section 3), where the fragment(s) matched by the profiler are highlighted.

2.3.2.4. Biotransformation rules. ***Liver Metabolism Rules:*** Liver BioPath rules consist of a total of 144 human-relevant biotransformation rules that represent enzymatic reaction sites similar to the public Systematic Generation of potential Metabolites (SyGMa) rule definitions [56]. The top transformation categories include cleavage, conjugation, hydrolysis, hydroxylation, lactam formation, lactone formation, oxidation/reduction, and rearrangement. The rules are represented in chemotypes embedded with physicochemical properties according to the CSRML syntax [42]. The rules are used as profilers and fingerprints to allow comparison of metabolic reactivities between chemical structures. Within the ChemTunes Metabolism database, these human-based transformations can be explored quickly against other species. Furthermore, in the Liver BioPath metabolizer, the biotransformation rules are coded as reactions and the reaction products are presented as possible metabolites along with the likelihood measures. The list of rules is documented in Supplementary Information Table S5.

Skin Metabolism Rules: A total of 87 structural rules for dermal transformations were derived from the literature [57]. Similar rules are observed as for liver; these include biotransformations from oxidoreductases, transferases (e.g., conjugation reactions), hydrolases, lyases, and isomerases. Again, the rules are represented in CSRML and are available as profilers and fingerprints within ToxGPS. The list of rules is available in Supplementary Information Table S5.

2.3.3. Cramer classification and threshold of toxicological concern (TTC)

The TTC framework is implemented within the informatics platform as a decision tree. This approach requires chemical structure to assign Cramer classes [58] and knowledge of estimated daily intake. ToxGPS provides all Cramer classification methods available in Toxtree v.3.1 including the original, revised, and extended rules [59]. Warnings will be given for known existing issues of Toxtree assignments. There are two types of TTC trees available: the original Kroes non-cancer tree and the modified tree for cosmetics-related chemicals based on SCCS Notes of Guidance [60]. Whilst the original TTC tree applies the thresholds from Munro's approach for the three Cramer classes [61], the new cosmetics tree employs the threshold values analyzed by the COSMOS TTC effort [62]. The TTC values for Cramer Class I and Class III were assigned to thresholds of 46 and 2.3 μg/kg-bw/day respectively, whereas the Cramer Class II still followed the Munro's original value of 9 μg/kg-bw/day.

The TTC tree addresses a series of questions in the framework. The first question in the TTC decision tree excludes a substance if it is a metal or metal-containing compound, or if the substance belongs to the cohort of concern (COC) defined by Kroes et al. [63]. Five "structural groups" are excluded from the TTC consideration if a substance is a known potent carcinogen (nitroso, azoxy, aflatoxin-like) or strongly bioaccumulating (polyhalogenated dibenzodioxins/benzofurans and steroids).

The second step of the TTC framework is to check whether the substance contains structural alerts for genotoxicity. This step is executed by checking DNA reactive mutagenicity and/or clastogenicity either by performing manual SAR or employing *in silico* methods such as structural rules and/or QSAR models when experimental data are not available [64].

2.3.4. In silico predictions

2.3.4.1. Machine learning (ML) hybrid rules. The CSRML methodology allowed development of novel hybrid rules combining expert-guided knowledge with machine learning (ML) approaches. During the expert-driven rule extraction stage, a highly curated training set containing target response data (binary, multinomial, or continuous) about a clearly defined endpoint is constructed. Initial chemotypes are first selected by substructure searches within the training set using the ToxPrint chemotypes. Associations of these initial chemotypes with the target endpoint are explored and the chemotypes are further refined to enhance the SAR (structure–activity relationship) and enrich target association. These enriched chemotypes are then used to subset the datasets into more homogeneous target responses by splitting the set using compound properties based on machine-learning techniques, e.g., recursive partitioning. The full development process and results are described elsewhere [65]. These ML-hybrid rules are used in predictions of various endpoints in genetic toxicity, liver toxicity, DART as well as dermal toxicity (irritation and sensitization) and permeability. Furthermore, when these structural rules are associated with numerical point of departure values, the rulebase gives insights on potency far beyond the simple correlations. For example, the quantitative correlation of these structural rules for potency, e.g., NOAEL values have been previously published for agrochemicals and food-related materials [66].

2.3.4.2. Mode-of-action QSAR approaches for toxicity models. The prediction methods in ToxGPS include QSAR models built from mode-of-action driven training sets as well as rule-base system with predictivity measures to result in a probabilistic outcome. These ToxGPS models have been developed with regulatory perspectives to give rationales and transparency for predictions, which are also documented following the core QSAR Model Reporting Format (QMRF) standard. The general approach combines QSAR results from a global model and multiple MOA models built within the structural neighborhoods of compounds sharing potentially the same or similar molecular initiating event (MIE). Overall, the ToxGPS QSAR models yield a probability estimate for each of the possible outcomes, which are then combined with the rule-base structural alerts, each of which has an associated positive predictivity value. The evidence from both QSAR model and rule-base predictions is combined by Dempster-Shafer Theory [67,68].

The general modeling algorithm for the ToxGPS models is logistic partial least squares regression, a learning method that has the advantage of being more easily interpretable than many other machine learning methods. QSAR models yield a probability estimate for each of the possible outcomes. The prediction outputs include the probability of being positive (endpoint toxic), uncertainty, and the applicability domain (AD). AD is defined based on the model descriptor space of the global model whilst the structural requirements of MIE neighborhood for the MOA models define the AD of such models. All toxicity

predictions in ToxGPS follow the same paradigm although other machine learning or neural networks modeling methods are also employed.

An example of the ToxGPS approach, as applied to skin sensitization of perillaldehyde, is illustrated in Fig. 4. The first step in the *in silico* model for skin sensitization in ToxGPS is the prediction of the skin sensitization hazard (a binary classification: sensitizer/non-sensitizer); these results are shown in the top portion of Fig. 4 (skin sensitization, Local Lymph Node Assay (LLNA) hazard). The colored probability bars depict the results from evidence: the global model, two local MOA models, two alerts, and the overall result obtained by combination of the five individual results. In the hazard model, the red and green bars indicate, for each result, the strength of the evidence for a positive (sensitizer) and negative (non-sensitizer) outcome, respectively. The yellow bar represents the uncertainty in the prediction.

For perillaldehyde, the two reactivity neighbor models (Schiff base formation; Michael acceptor) tend to be positive; for example, the probability of being positive is reported as the range 0.74–0.91 for Michael acceptors. The global model was positive for sensitization with the positive probability range of 0.72–0.85. The overall outcome estimates perillaldehyde to be sensitizing with high probability and low uncertainty. A positive LLNA hazard prediction in ToxGPS triggers application of a second model designed to estimate LLNA potency by predicting whether the query compound is likely to be classified as either GHS 1A (strong sensitizer with an effective concentration to induce 3-fold increase of Stimulation Index (EC3) $\leq$ 2 wt/vol %) or GHS 1B (weak-moderate sensitizer with EC3 $>$ 2 wt/vol %) [69]. As shown in Fig. 4, perillyl aldehyde is predicted to be a weak-moderate sensitizer with probability of being GHS 1A of 0.08, GHS 1B of 0.82, and with uncertainty 0.10. This prediction is consistent with the published LLNA results with EC3 values ranging from 7.9 to 8.7% [70-72]. These publications report none-to-moderate reactivities from various studies including Adverse Outcome Pathway (AOP)-related assays (Direct Peptide Reactivity Assay (DPRA), Human Cell Line Activation Test (h-CLAT and KeratinoSens), human maximization test (positive at 4%) and confirmatory human repeat insult patch test (negative at 0.6%). Supplementary Information Table S6 lists all toxicity prediction endpoints available in ChemTunes•ToxGPS®.

2.3.4.3. ADME predictions. ***Bioavailability Predictions:*** Whilst the ToxGPS approach for toxicity predictions focuses on transparency and rationales required in regulatory-related workflows, the bioavailability models are implemented with machine learning methods targeted for improved predictivity such as artificial neural network, SVM/SVR (support vector machine/support vector regression), and random forest approaches. A wide range of endpoints are available covering physicochemical properties, oral/dermal absorptions and permeabilities, blood brain barrier, and plasma protein binding. The list of these Express models is available in Supplementary Information Table S6.

Liver BioPath Metabolizer: Liver BioPath [73] rules for human *in vitro* hepatocytes are coded as reactions such that metabolism products can be generated stoichiometrically. Both Phase I and II transformation rules are considered for two successive levels of reactions. To prioritize various metabolites from a transformation, PPVs and odds ratios were determined using the underlying ChemTunes metabolism database [67].

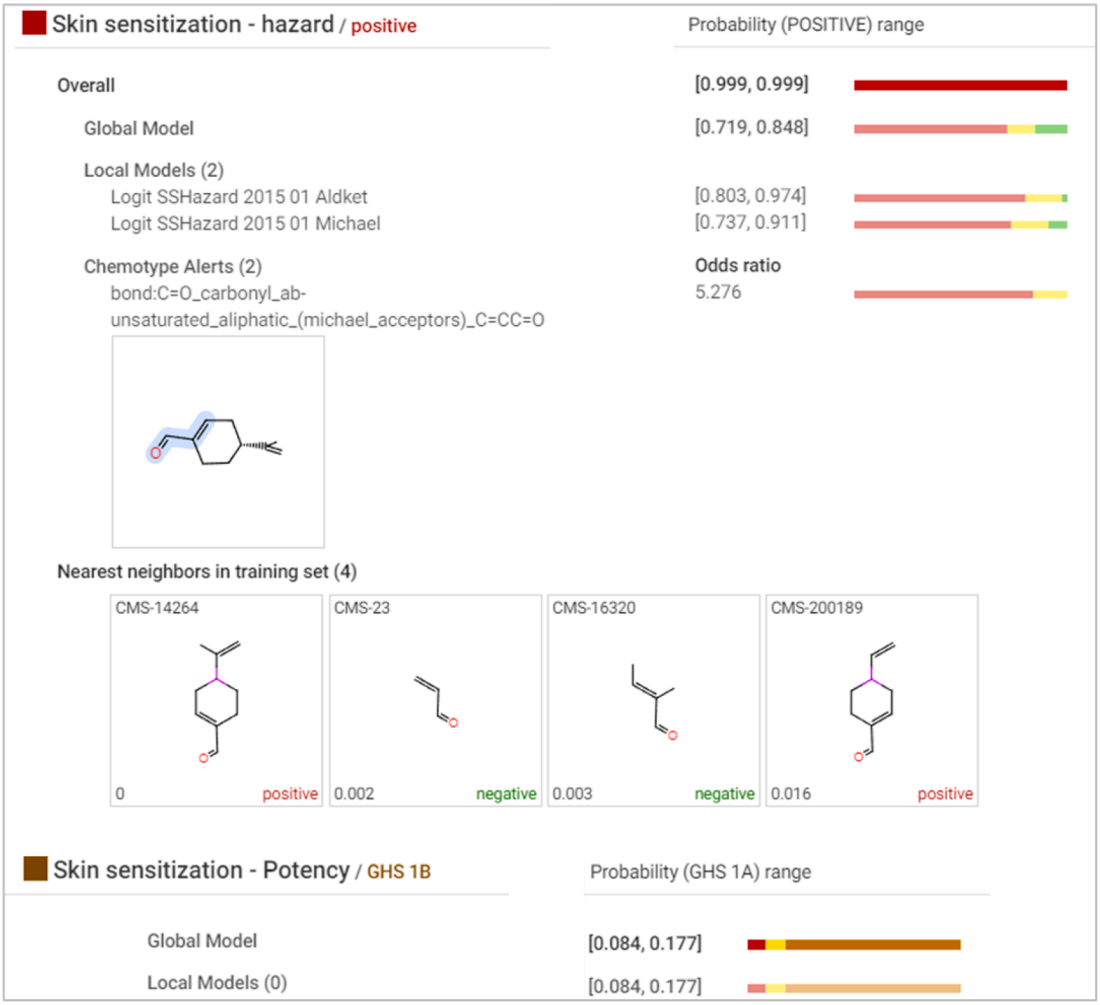


Fig. 4. Prediction browser in ToxGPS for Local Lymph Node Assay (LLNA) models.

2.4. Methods for quantitative read-across

This section describes in detail the approach to apply read-across quantitatively within a NGRA process such that uncertainties involved in the read-across process can be quantitatively estimated. Therefore, we present methods to establish a quantitative read-across (qRAX) based on approaches available from ToxGPS.

2.4.1. Structure similarity and measures

Various similarity searching methods are available in ChemTunes•ToxGPS®. From the main query builder, similar structures can be queried against the whole database, which can be further constrained to include particular endpoints in the database. Within the ToxGPS workflow, this first query result can also be augmented by analogue searching tools that constrain search hits based on a specified similarity threshold and, if desired, substructure scaffolds representing relevant biology. In the Express service, finding analogues via ensemble searching of nearest neighbors from a specific endpoint dataset (NOAEL or EC3 datasets) is also possible.

ToxGPS provides a number of fingerprinting schemes for quantifying structure-based similarity, including both extended connectivity fingerprints (RDKit molecular topological and Morgan [15,74,75]) and the pre-defined features (ToxPrints [42,46] or public 166 MACCS keys [76]). The capability of evaluating similarity with multiple fingerprinting schemes is important because different schemes may capture different perspectives. Once the structures have been fingerprinted, Tanimoto [77] and Dice indices [78] are applied to calculate pairwise similarities. Based on similarity calculations for all structure pairs in the full ChemTunes database using the above four fingerprints, a similarity index of approximately 0.7 is recommended as a reasonable threshold for defining similar structures [32].

The issue over the values of similarity indices such as Tanimoto and Dice is important and must be acknowledged when applying such a method. Mellor et al. [79] undertook an analysis of various calculation methods for molecular similarity. The study demonstrated that similarity indices are dependent on the fingerprint (or other information) that they are computed from. Individual similarity values are comparable only when computed from the same fingerprint and metric (as is undertaken in ToxGPS). Mellor et al. [79] also demonstrated that an inappropriate molecular fingerprint may result in poor quality analogues. In addition, the cut-offs for similarity of analogues must be approached in a pragmatic manner. The value of 0.7 applied in ToxGPS is in line with previous studies and recommendations for cut-offs in molecular similarity [80].

2.4.2. Property and assay similarity

Chemical similarity is estimated by structural and molecular/physicochemical properties whereas biological similarity can be based on *in vitro* or HTS assays. An analogue in ChemTunes•ToxGPS® is defined as a structure sufficiently similar to the target to justify read-across, where similarity takes into account structure connectivity, physicochemical properties, and biological assay information, when available. Additionally, evidence that the target and analogue share the same MOA or metabolic pathways is desirable.

2.4.2.1. Metrics for property or assay similarity. Within a well-defined structural neighborhood, molecular and physicochemical properties are useful to further differentiate structures in ways that cannot be captured by structural features alone. ToxGPS calculates properties described in Section 2.3.1 (Supplementary Information Table S4) and enables evaluation of property-based similarity to further constrain the retrievals from the structure-based similarity searching. When an endpoint is known to be driven by one particular property or assay, this can be easily compared across various structures. When comparisons of profiles of multiple properties across structures are desired, the profiling method by skyline plots were used.

To compare similarity quantitatively, we can use a measure based on Pearson Correlation coefficient to calculate Pearson Similarity [32] when correlation of profiles between structures is important. Pearson similarity is the Pearson correlation coefficient scaled 0 and 1 so that the similarity value itself represents a probabilistic estimate of the relevance of the analogue to the target with respect to physicochemical properties. To address cases where the absolute distance between the property sets is important, a Euclidean Similarity metric based on Euclidean distance is defined as in Eqs. (1) and (2):

$$Euclidean_{distance} \equiv d = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \tag{1}$$

$$Euclidean_{similarity} = \left(\frac{1}{1 + \frac{d}{\sqrt{N}}} \right) \tag{2}$$

where N is the number of selected variables (properties or assays), and x_i and y_i denote standardized values of variable i for molecules x and y, respectively. In our previous publication [32], Euclidean Similarity was calculated from the distance (d) using the common relation of $1/(1 + d)$; however, this is too simplistic for our purpose because similarity would then tend to decrease rather sharply as the number of variables (N) used to calculate d increases. Equation (1) thus represents a robust improvement for property- or assay-based similarity.

It is important to note that when calculating property or assay-based similarity values, the variables must first be scaled by standardization or some other appropriate normalization method since the original variables have different units and may vary widely in magnitude. Before computing the profile similarity for properties, the original values were therefore standardized using means and standard deviations computed from a large random sample of approximately 50,000 structures in the ChemTunes database so that robust and reproducible pairwise similarities can be calculated. For bioavailability properties such as Caco-2 cell permeability, skin permeability, hepatic clearance, and fraction unbound, ChemTunes provides large datasets so that the smaller datasets can be standardized in a systematic and reproducible manner.

2.4.2.2. Profile comparisons of property similarity. The profiles of 12 molecular properties described in Section 2.3.1 are plotted in Fig. 5. We refer to these bar charts as “skyline plots”. Three structures are compared using both Euclidean and Pearson similarities. All three are quite similar in property space according to both metrics (above 80% by Euclidean and 97% by Pearson similarity). Two tertiary alcohols with aliphatic chains ($C > 6$) were compared with the primary alcohol. Farnesol and nerolidol both have a $C > 12$ backbone chain with the same number of double bonds and were recognized as more similar to each other than to linalool. Three properties that differentiate these structures were log P [81], HOMO, and HOMO/LUMO Gap, which relate to the length of the aliphatic chain and number of double bonds. In this property space, the primary vs. tertiary alcohol seemed to have less effect.

2.4.2.3. Profile comparisons of assays. Biological similarity can also be quantified based on data from activity or screening assays. For example, patterns for binding activities related to EDSP assays were selected including the 21 estrogen (ER), 12 androgen (AR), 13 progesterone (PR), 6 aryl hydrocarbon receptor (AhR), and 12 peroxisome-proliferator-activity (PPAR) assay categories [82]. In this analysis, only the assays having data for at least 30% of the structures in the database were included. The Endocrine Disruption Screening Assays and others selected in this analysis contained a large number of data points: AR_BLA Activity (8,306), ER_Era_BLA Activity (8,305), PR_BLA Activity (7,871),

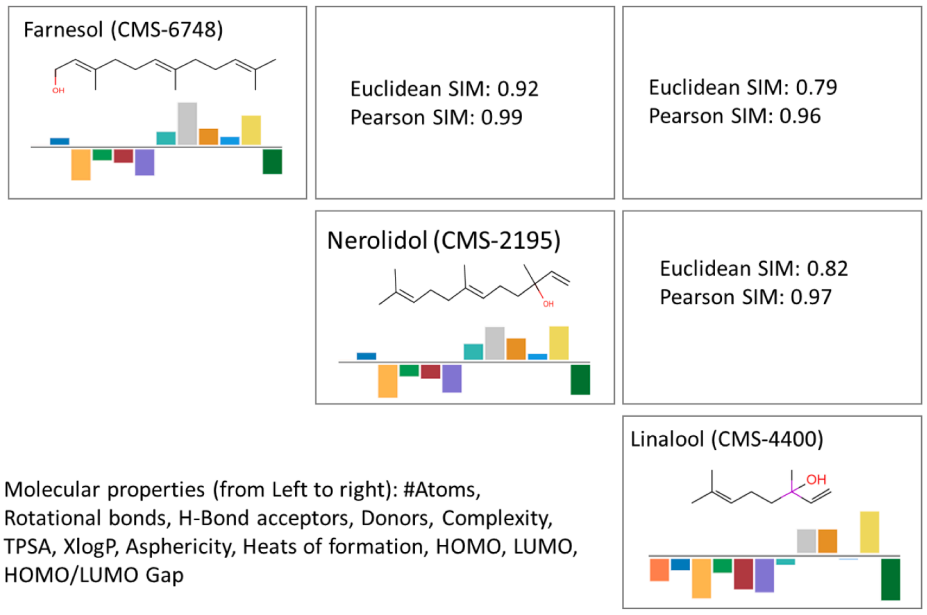


Fig. 5. Comparison of properties based on Euclidean and Pearson similarities ("similarity" abbreviated to SIM in the Figure).

Ahr Activity (8,887), and PPARg Activity (9,179).

For each substance, all available assay results within each group of aggregated assays are averaged and normalized from −0.5 to +0.5. Each substance can then be visualized with a skyline plot where each bar represents a different assay group [65]. Fig. 6 illustrates the 23 aggregated assay categories. Pooling individual assays in this manner increases the signals from the above 64 assays to 33 assay categories (6 AR, 14 ER, 2 AhR, 5 PR, 6 PPAR assays). The values above the horizontal axis depict the assays whose activities were greater than the mean of the available database substances. The EDSP assay categories are grouped by AR, ER, AhR, PR, and PPAR.

The assay-based similarity between any two substances can then be calculated from the respective profiles in the same way as was done for property-based similarity, using either Pearson or Euclidean similarity measures. When the activities are binned into categories, variability among the observations with many missing values is minimized and Euclidean similarity is a better choice because it depends on absolute distance, not simply correlation. The EDSP assay profiles in Fig. 6 seem to be similar between the three aliphatic alcohols: 0.78 for farnesol-nerolidol, 0.74 for farnesol-linalool, and 0.83 for nerolidol and linalool by Euclidean similarity. Whilst the property similarity found farnesol and nerolidol more similar (Fig. 5), the biological metric (i.e., EDSP assays) found nerolidol and linalool (the two tertiary alcohols) more similar. These assays demonstrated the power of inclusion of assay patterns as a way to capture biological similarities.

2.4.3. *Analogue quality*

As described in the previous sections, chemical similarity can be quantified in various ways, considering structure, property, and biological profiles (assays, MOA, AOP, alerts). In ToxGPS, the analogue quality (AQ) is defined as the geometric mean of all similarity measures selected to quantify a given qualified analogue [18,32] relative to a target of interest.

$$AQ = \sqrt[n]{S_1 S_2 \ldots S_n} \tag{3}$$

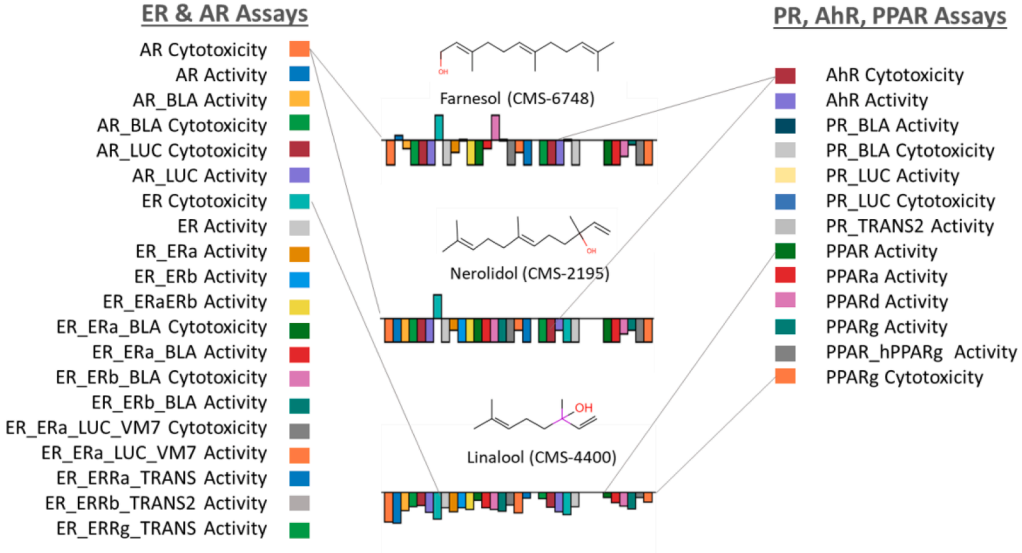


Fig. 6. Profiles of ToxCast EDSP assays.

where each S_i can be a Tanimoto coefficient or Dice index for structural similarity between analogue and target, or a Pearson or Euclidean similarity from property or assay profiles of the analogue and target. In all cases, each S_i, and therefore AQ, is a value between 0 and 1. Note that analogue quality can be assessed for each analogue candidate without consideration of the endpoint(s) of interest.

2.4.4. Study quality (SQ)

The next important step in qualifying analogue candidates for read-across is to evaluate the quality of the evidence provided by the studies. To obtain a read-across outcome with a quantitative uncertainty, it is necessary to apply objective study quality schemes. As expected, this involves some subjective judgement, although certain aspects of the study designs can be evaluated systematically and more objectively. For example, guideline studies conducted in GLP environments by a well-documented source are more highly rated than experiments from non-guideline investigations. Certain assays are also known to generate higher quality data due to the nature of the assay complexity, e.g., results from Ames studies are generally viewed as more reliable than results from experiments for *in vitro* chromosome aberrations. The values for study quality (SQ) were assigned after reviewing the study protocols and call rationales from the sources. A rigorous and systematic process for extracting quantitative study quality measures has been published previously [32].

In a ToxGPS workflow, SQ assignments are made by the user, with options for SQ being HIGH (1.0), MED-HIGH (0.95), MED (0.85), MED-LOW (0.8), and LOW (0.70 or lower). Many data sources provide Klimisch scores [83] and study completeness scores or reliability which can be mapped to this five-category scale. In general, HIGH and MED-HIGH usually correspond to Klimisch scores of 1, and MED and MED-LOW to Klimisch scores of 2. An SQ value lower than 0.5 may provide justification for excluding that particular study; if used in the weight of evidence combination process, a study with a low SQ could substantially increase the uncertainty in the read-across result.

2.4.5. Read-Across (RAX) reliability & outcome

2.4.5.1. RAX reliability. The reliability of a given evidence source for RAX depends primarily on the AQ and SQ. In ToxGPS, the joint probability of these two is defined as the RAX reliability (RR), a measure of confidence in reading across using the selected analogue(s) and their study data to the target. This reliability can be calculated per analogue either individually for each study or by averaging using a decision theory method across multiple studies available for the analogue [68]. High RAX Reliability supports the suitability of the selected analogue evidence for the purposes of read-across to the specified target.

2.4.5.2. RAX outcome. Once selected analogues are qualified and read-across reliabilities are estimated, the read-across outcome will depend on the endpoint of interest for the target. For repeated-dose toxicity, point of departure values such as NOAEL concentrations are typically of interest. For genetic toxicity endpoints, the outcomes will often be binary (positive or negative). For skin sensitization, several outcomes are envisioned, namely, the binary hazard call (sensitizer or non-sensitizer), multinomial ordinal calls (non, mild, moderate, or strong sensitizers), or continuous values (e.g., EC3 values) if the substance was considered a sensitizer.

In Section 3, three use cases are presented which consistently apply the same systematic procedure using our proposed quantitative methods for performing a read-across: 1) calculate AQs for each analogue based on structure, property, and assay information; 2) assign study quality (from 0 to 1) for each contributing study; 3) calculate RR for each analogue over available studies; 4) apply Weight-of Evidence (WOE) method to calculate overall RR for all selected analogues with their contributing studies; 5) calculate the final outcome by combining all sources of evidence, each weighted by its respective RR, using a decision theory approach based on Dempster Shafer Theory.

3. Use cases: demonstrating the support of NGRA of the major constituents of *Perilla frutescens* by molecular informatics

The informatics system, *in silico* knowledge and methodology described in Section 2 can be applied in many chemical safety assessment scenarios, notably for identifying existing data and filling data gaps. One of the most challenging areas for safety assessment is the evaluation of complex chemical mixtures, botanicals and UVCBs. Botanicals are particularly important in cosmetics safety due to their widespread use in these products [84]. Botanicals are complex mixtures, often containing several thousand chemical constituents, the relative concentrations of which may vary depending on site and time of harvesting, growing conditions, and many other factors. These complicating aspects explain the previous lack of informatics solutions for botanicals [85].

This section demonstrates how molecular informatics tools may be applied to fill some of the frequently observed data gaps for a botanical. The goal here is not to cover a full NGRA process, but rather to illustrate how *in silico* approaches contribute to finding solutions for Tier 0 in Fig. 1, capitalizing on the techniques described in Section 2. *Perilla frutescens*, a herb with a variety of uses in cosmetics and other industrial products, was selected, and the following steps which specifically support NGRA are described:

- Characterization of the constituents of a botanical substance (relating, in part, to Steps 2 and 3 in Fig. 1).
- Determining the exposure/use scenario of a botanical substance in a cosmetics product (relating, in part, to Step 1 in Fig. 1).
- Performing TTC for the constituents of a botanical substance (relating, in part, to Steps 2 and 3 in Fig. 1).
- Undertaking quantitative read-across for selected constituents of a botanical substance for genetic toxicity, repeat dose toxicity and skin sensitization (relating, in part, to Step 4 in Fig. 1).

3.1. Problem formulation: safety assessment of the essential oil of Perilla frutescens

This investigation and associated use cases attempt to demonstrate how molecular informatics, implemented through a software platform, can assist elements of Tier 0 in the NGRA of a botanical used in a cosmetic product. Specifically, these use cases considered a hypothetical use case of a leave-on hand cream containing *P. frutescens* at 0.01%. In this instance, the substance was characterized in the initial step (otherwise considered to be Steps 2 or 3 of the NGRA) such that the exposure to individual components could be considered (Section 3.2).

P. frutescens largely consists of naturally occurring monoterpenes and monoterpenoids. One of its constituents is perillyl alcohol, a monoterpene derived from the volatile components of the essential oil of many botanicals including lavender, peppermint, cherries, sage, lemongrass as well as *P. frutescens*. It has been widely used in flavoring and fragrance formulations in toiletries [86], and topical insecticide repellant along with the aldehyde form (perillyl aldehyde) and other monoterpenoids [87], Recently, the potential use of perilla alcohol for treatment of cancer via inhalation has been also reviewed [88].

The composition of several variants of *P. frutescens* (L.) Britton have been reported in the literature. A 2021 study of a variant from Vietnam was used [89] to prepare the case study presented here for a botanical mixture. Although components and compositions vary greatly depending on sources of accessions, e.g., specific locations in China [90], Japan [91], Korea [91,92], Lithuania [93], and US [94], there are constituents common to the reported variants containing volatile components such as alicyclic and aliphatic chains along with their functionalized forms of

alcohols, aldehydes, and acids. In a particular variant from Vietnam, the most abundant constituents were reported to be perillaldehyde (62%), caryophyllene (6.7%), limonene (5%), caryophyllene oxide (2.5%), linalool (1.4%), humulene (1.1%), perillyl alcohol (0.8%), nerodilol (0.65%), and alpha-terpineol (0.5%). Table 1 shows the list of 18 major identified constituents based on the 24 peaks (including 3 unknowns) in the GC MS chromatogram from Dat et al. [89].

Some notable differences in composition among the reported variants include perillaldehyde (0.89–62.1 %), caryophyllene (0.14–16%), caryophyllene oxide (0.45–12%), limonene (0.049–12.6%), phytol (0.18–4.76%), and linalool (0.46–4.6%). In particular, a remarkable difference in the concentration of perillaldehyde, i.e., 0.89% vs 62.1% was observed between "*var. Japonica*" and "Thai Binh Province Vietnam", respectively. In addition to the effect of harvesting location difference, it is also worth noting that the difference in extraction methods, such as aqueous/solvent vs. microwave-assisted distillation, used to prepare GC MS samples might have played a role since perillyl alcohol can be oxidized to perillaldehyde abiotically under certain conditions. Interestingly, perillyl alcohol was found in much lower concentrations (0.8 – 2.2 %) and the variation in this amount between variants was also relatively small. These observations are good examples of some of the complex issues faced when assessing botanical mixtures quantitatively.

Limonene, myrcene, and farnesene can be transformed to monoterpenoids. In human metabolism (summarized in Fig. 7), limonene is oxidized to perillyl alcohol, which can be further oxidized to perillic acid by aliphatic primary alcohol oxidase [95,96]. Similar transformations occur in skin via alcohol and aldehyde dehydrogenases as well as aldo–keto reductase and aldehyde oxidase (Table 2).

In summary, the components of the *P. frutescens* mixture can be roughly grouped into approximate classes within the monoterpene and monoterpenoids, i.e., alicyclic ring, aliphatic chain, functionalized alicycles and alcohols. Conventional cluster analysis can be used to group these constituents, where the clustering algorithm uses a distance metric calculated from either structural features or pertinent physicochemical/molecular properties. Hierarchical clustering was applied using Jaccard distances calculated based on 129 ToxPrint chemotypes, followed by agglomerative nesting using single linkage. The dendrogram in Fig. 8 reveals five clusters at a distance of approximately 0.4 to 0.5, namely four clusters and a singleton that is perillene. The four clusters were alicylic rings (Cluster I), long aliphatic (C > 6) chains (Cluster II); terpenoid alcohols (Cluster III) and terpenoid aldehydes & ketones (IV). This grouping is reasonable from a structure–activity relationship (SAR) perspective, differentiating alicycles and aliphatics and their derivatives with functional groups.

Therefore, the assessments were performed for constituents as members of these clusters. We will employ TTC and Read-Across approaches following the workflow presented in Fig. 1 by the Cosmetics Europe LRSS task force.

3.2. Exposure/use scenarios

Step 1 of NGRA, as presented in Fig. 1, is the definition of the exposure/use scenario. As noted above, for the purposes of this investigation, a hypothetical use case of a leave-on hand cream containing *P. frutescens* at 0.01% was considered. The assumed external exposure to leave-on hand creams is 32.70 mg/kg-bw/day in Europe [60]. Thus, for this hand cream scenario, an exposure of 3.27 μg/kg-bw/day collectively to the various constituents from *P. frutescens* may be assumed. Based on the constituent compositions given in Section 3.1, examples of possible daily exposures (μg/kg-bw/day) are estimated for each constituent: perillaldehyde (2.03), caryophyllene (0.22), limonene (0.16), caryophyllene oxide (0.08), linalool (0.05), humulene (0.04), perillyl alcohol (0.03), nerolidol (0.02), and alpha-terpineol (0.02). The possible exposures of all constituents are listed in Table 1.

Since most of the constituents are well-known fragrance and flavoring agents, risk assessments on many of these substances have been published by RIFM and EFSA, including information on general exposure estimations. For example, RIFM reported the total aggregated systemic exposure (oral, dermal, and inhalation) for perilla alcohol to be 0.047 μg/kg/day [86], 6.7 μg/kg/day for alpha terpineol [97], and 38 μg/kg/day for dl-limonene [98]. Considering these general estimations, the above hypothetical exposures from a hand cream containing the perilla extracts are within the normal ranges even if 100% absorption through the skin is assumed for those substances. These estimated daily intakes will be analyzed in terms of the thresholds for toxicological concerns in Section 3.3.

The skin permeability rules and QSAR models in ToxGPS were applied to find that the perillyl alcohol and terpineols are predicted to have logarithmic values of the skin permeability coefficient (log Kp where the units of Kp are cm/h) in the range −3 to −1.5 (medium penetration), whereas large alicyclic rings (e.g., beta-caryophyllene) were predicted to have log Kp values below −3 (low penetration) [99,100]. For example, Lucca et al. reported that beta-caryophyllene extracted from Copaiba oil was penetrating only through the stratum corneum layer at 0.0001% of the amount delivered on top of the skin [101]. When a microemulsion is prepared and applied, there was 0.07% penetration to the epidermis and dermis. Thus, the maximally conservative assumption of 100% absorption is significantly over-protective for some compound classes.

3.3. Thresholds of toxicological concern

For substances whose exposures are sufficiently low, a substance-specific data waiver, such as the use of TTC may be appropriate at Steps 2 and 3 of Tier 0 of NGRA. TTC was performed on the identified constituents of *P. frutescens* according to the method described in Section 2.3.3. A component-based approach was applied by assessing all well-defined components individually for their genotoxic potential [64]. The outcome of the TTC analysis along with the genotoxic outcomes is summarized in Table 1. Two constituents were found to be associated with genotoxic potential according to predictions from Ames mutagenicity, *in vitro* chromosome aberration, and *in vivo* micronucleus QSAR models and rule-base outcomes in ToxGPS as well as experimental data, where available, in the ChemTunes database and literature.

For all non-genotoxic cases, the Cramer classes were determined by the revised Cramer tree and appropriate thresholds were assigned for the non-cancer cosmetics TTC tree (Section 2.3.3). As illustrated in Fig. 9, the estimated daily exposure of all non-genotoxic constituents would pass the TTC threshold comfortably even assuming 100% bioavailability. Bury et al. also analyzed the exposure scenarios of perillyl alcohol and confirmed that even under the worst-case assumption of 100% bioavailability, the daily intake of this substance would be much lower than the TTC threshold of Class I, hence perillyl alcohol is safe to use given its actual low bioavailability from cosmetics formulations [9].

For structures that are deemed genotoxic either by experimental data or predictions, the non-cancer tree is not pursued, as in the case of perillene for which ToxGPS yielded an equivocal result due to potential clastogenicity. It should be noted that an EFSA panel considered that perillene, as one of many constituents of *Zingiber officinale* Roscoe, would be of no concern for genetic toxicity safety, citing predictions from OECD Toolbox [102].

If we assume the exposure of perillene to be 100%, the estimated daily intake would be 0.0036 μg/kg-bw/day, which is above the threshold of 0.0025 μg/kg-bw/day (0.15 μg/person/day) below which a less than one-in-a-million cancer risk may be assumed [63]. If a nominal bioavailability of 50% of oral absorption is assumed, the use of perillene found in the *P. frutescens* would be acceptable if the botanical was formulated at 0.01% in a hand cream, even if perillene were genotoxic. To be more conservative, the bioavailability of perillene was further investigated employing skin metabolic and permeability rule bases as well as QSAR models. According to the skin metabolic rules in the Chemotype profiler, perillene does not undergo substantial metabolic

Table 1
Characterization of constituents in *Perilla frutescens* and the threshold of toxicological concern approach.

Compound Names	CMS ID	Approx. Composition (%)[1]	Cramer Class[2]	TTC(μg/kg-bw/day)[2]	Daily Intake[3](μg/kg-bw/day)	GenotoxicityPrediction (ToxGPS)[4]	GenotoxicityExperiment Summary[5]
Cluster 1: Alicyclic alkenes							
C1. D-cadinene	CMS-42433	0.20	Low (I)	46 (30)	0.0065	**NEGATIVE** • Ames: NEG • *in vitro* chromosome aberration (ivtCA): NEG • *in vivo* micronucleus (ivMN): NEG	• Ames: NEG; EQ aliphatic and alicyclic hydrocarbons — not genotoxic [103]
C2. Beta-Elemene	CMS-51192	0.24	Low (I)	46 (30)	0.0078	**NEGATIVE** • Ames: NEG • ivtCA: NEG • ivMN: NEG	• iv MN NEG [104]
C3. Limonene	CMS-797	4.99	Low (I)	46 (30)	0.16	**NEGATIVE** • Ames: NEG • ivtCA: NEG • ivMN: NEG	• Ames: NEG • ivtCA NEG • ivtMM (MLA): NEG
C4. Beta-caryophyllene	CMS-3991	6.72	Low (I)	46 (30)	0.22	• Ames: NEG • ivtCA: NEG • ivMN: NEG	• Ames: NEG • ivCA NEG; • ivMN: NEG; ivt MN NEG; UDS NEG
C5. Humulene	CMS-42579	1.11	Low (I)	46 (30)	0.036	**NEGATIVE** • Ames: NEG • ivtCA: NEG • ivMN: EQ	aliphatic and alicyclic hydrocarbons — not genotoxic [105-107]
Singleton: Heterocycle alkene							
C6. Perillene	CMS-54767	0.11	High (III)	2 (TTC);0.0025 (TOR)	0.0036	**Predicted: EQUIVOCAL** • Ames: NEG • ivtCA: POS • ivMN: POS	• Ames QSAR: POS [108] • Ames QSAR: NEG [Data from the OECD QSAR Toolbox [109]] • Genotoxic endpoints: NEG [Data from the OECD QSAR Toolbox [109]]
Cluster 2: Aliphatic (>C6) chains & alcohols							
C7. Farnesene	CMS-11498	0.22	Low (I)	46 (30)	0.0072	**NEGATIVE** • Ames: NEG • ivtCA: NEG • ivMN: NEG	• No safety concern [107] • ivtCA: NEG; ivtMN: NEG [110] • Comet Assay: NEG [110]
C8. Nerol	CMS-4979	0.16	Low (I)	46 (30)	0.0052	**NEGATIVE** • Ames: NEG • ivtCA: NEG • ivMN: NEG	• Ames: NEG • ivtCA: NEG • ivt MN NEG
C9. 1-octene-3-ol	CMS-7321	0.11	Int (II)	9 (9)	0.0036	**NEGATIVE** • Ames: NEG • ivtCA: NEG • ivMN: NEG	Ames QSAR: NEG [Data from the OECD QSAR Toolbox [109]Experimental data not found
C10. Phytol	CMS-8161	0.18	Low (I)	46 (30)	0.0059	**NEGATIVE** • Ames: NEG • ivtCA: NEG • ivMN: NEG	• ivtCA: NEG • No data found • ivtMM NEG [111]
C11. Linalool	CMS-4400	1.41	Low (I)	46 (30)	0.046	**NEGATIVE** • Ames: NEG • ivtCA: NEG • ivMN: NEG	• Ames: NEG • ivtCA (CHL) NEG • ivtMN NEG
C12. Nerolidol (E & Z)[6]	CMS-2195	0.78	Low (I)	46 (30)	0.021	**NEGATIVE** • Ames: NEG • ivtCA: NEG • ivMN: NEG	• Ames: NEG (Cytotoxic) • ivtCA: No data found • ivMN: NEG [112]
Cluster 3: Terpenoid Alcohols and Derivatives							
C13. Alpha-Terpineol	CMS-4937	0.51	Low (I)	46 (30)	0.16	**NEGATIVE** • Ames: NEG • ivtCA: EQ • ivMN: NEG	• Ames: NEG • ivtCA: No data found • ivtMN NEG [97]
C14. Beta-Caryophyllene Oxide	CMS-13699	2.45	High (III)	2 (1.5)	0.080	**NEGATIVE** • Ames: POS • ivtCA: POS • ivMN: POS	• Ames: NEG [113,114] • ivtMN NEG [113,114]
C15. Perillaldehyde	CMS-14264	62.1	Int (II)	9 (9)	2.03	**NEGATIVE** • Ames: NEG; • ivtCA: EQ • ivMN: NEG	• Ames: NEG [70]; POS [103] • ivtMM: EQ [103]; NEG [113] • ivMN: NEG [113]; ivtMN NEG; DNA Repair NEG; DNA Damage EQ; [103]

(*continued on next page*)

Table 1 (*continued*)

Compound Names	CMS ID	Approx. Composition (%)[1]	Cramer Class[2]	TTC(μg/kg-bw/day)[2]	Daily Intake[3](μg/kg-bw/day)	GenotoxicityPrediction (ToxGPS)[4]	GenotoxicityExperiment Summary[5]
C16. Perillyl alcohol	CMS-15331	0.80	Low (I)	46 (30)	0.026	**NEGATIVE** • Ames: NEG • ivtCA: EQ • ivMN: NEG	• Ames: NEG [86] • ivtMN: NEG [86]
C17. P-menth-1-en-9-ol	CMS-32553	0.12	Low (I)	46 (30)	0.0039	**NEGATIVE** • Ames: NEG • ivtCA: EQ • ivMN: NEG	Chemical-specific experimental data not found
C18. Cadinol	CMS-62435	0.34	Low (I)	46 (30)	0.011	**NEGATIVE** • Ames: NEG • ivtCA: EQ • ivMN: NEG	Experimental data not found

[1] The composition was taken from the literature [89].
[2] Toxtree v3.1.0 with revised Cramer decision rules was used. TTC values used the SCCS [60] recommendation (values in parenthesis are Munro original).
[3] Expected daily intake was estimated in the section 3.2.
[4] Predicted by ToxGPS models and alerts for Ames mutagenicity, *in vitro* chromosome aberration, and *in vivo* micronucleus.
[5] Unless specified in the table, genetic toxicity data are available within ChemTunes database. The data sources include US FDA, EFSA, SCCS [115] ECHA, and literature.
[6] Both E-and Z-isomers are combined: E- nerolidol (0.65%); Z-isomer (peruviol) (0.13%).

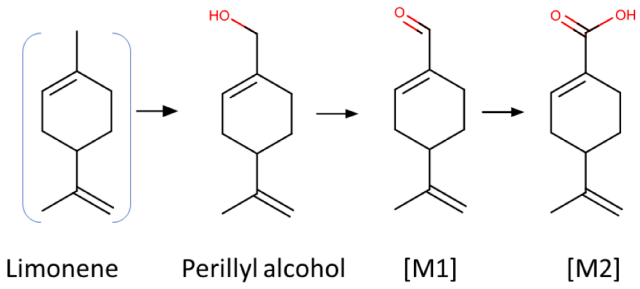


Fig. 7. Partial human metabolic pathways of perillyl alcohol to acid.

transformations in the skin. When applying skin permeability chemotype rules (Section 2.3.4.1), hits are observed for the hybrid rules associated with medium and low Kp classes. When applying the quantitative log Kp model (Section 2.3.4.3 and Supplementary Information Table S8), we obtain log Kp of −1.98 cm/h, which would be classified as medium Kp class. Based on both the skin permeability and the estimated daily exposure from the hand cream, even if we assume that perillene were genotoxic, the use of *P. frutescens* would not have been likely to raise a safety concern.

3.4. Quantitative read-across (qRAX)

Botanicals in general are data-poor substances whose chemical structures and compositions are often ill-defined. The case of *P. frutescens* is fortunate since structure and composition data have been published. As shown in Fig. 8, further characterization of the constituents into several structurally similar clusters facilitated the read-across approach for substance-specific assessment. We selected three clusters for the use-case: 1) alicyclic alkenes; 2) aliphatic chain alcohols; 3) terpenoid alcohol. The structures in three clusters and their data profile are listed in Table 2.

To establish a broad profile of data availability, all constituents of *P. frutescens* were searched for similar structures in the ChemTunes database (over 100,000 structures), selecting structures with pairwise similarity >0.70 (Tanimoto coefficients) against the constituent of interest. From this query, additional structures were identified including p-menthane-3,8-diol (CMS-7392), dihydro alpha-terpineol (CMS-9118), 4-terpineol (CMS-10854), perillic acid (CMS-59861), menthadienol (CMS-14225), and isocyclogeraniol (CMS-2635). When the similarity searching is constrained to the ToxCast *in vitro* assay database of 9,298 substances, most structures in Cluster-2 (aliphatic chain alcohols) had data on endocrine disruption screening, DNA binding/damage/repair, liver/biliary metabolism as well as immune system/inflammation assays. Other clusters also return data although not necessarily for all substances in the cluster. As described in Section 2.4.2.3, these assays were used as a metric for biological similarity whenever possible. Structural similarity was calculated based on both RDKIT molecular fingerprints and ToxPrint chemotypes, while property-based similarity was calculated using the same 12 properties depicted in Fig. 5.

For the purpose of this use-case, and in order to present a stepwise demonstration of qRAX, each cluster was assigned to a major target endpoint, based on the data availability in the areas of genetic toxicity, repeated-dose, DART and skin sensitization. Toxicity and safety data were compiled from the ChemTunes database and augmented with additional public data from ECHA, EFSA, HESS, SCCS, US Cosmetics Ingredients Review (CIR) [116], US EPA, and open literature. The toxicity data availability is summarized in Table 2 and more details are described in the Supplementary Information Table S7.

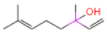
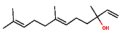
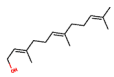
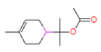
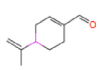
The three case studies for this qRAX were undertaken using the informatics tools/methods described previously. The same approach was applied consistently for read-across over the three case studies:

- STEP 1: calculate AQ for each analogue based on structure, property, and assay information;
- STEP 2: assign study quality (from 0 to 1) for each contributing study;
- STEP 3: calculate RR for each analogue over available studies;
- STEP 4: apply WOE method to calculate overall RR for all selected analogues with their contributing studies; and
- STEP 5: calculate the final outcome by combining all sources of evidence, each weighted by its respective RAX Reliability, using a decision theory approach based on Dempster Shafer Theory

3.4.1. Use case 1: read-across to fill a data gap for genetic toxicity potential of beta-elemene

Constituents in the Cluster-1 including beta-elemene (T-1) are monoterpenes that are found in many herbs and botanicals. They are mostly data-poor with one exception, i.e., various (d-, l-, racemic) forms of limonene. Conducting a similarity searching against ChemTunes database, trivinylcyclohexane (A-1.2) was identified as an analogue

Table 2
Compiled evidence for target-analogue pairs for each cluster used in the case studies.

Substance [1]	Name & IDs	Similarity RDKit[2,3]	Similarity ToxPrint[3]	Liver metabolic rules[4]	Skin metabolic rules[4]	Tox data availability[5]	Property Pearson similarity	Tox21/Cast EDSP Euclidean Similarity
Cluster 1: Alicyclic alkenes								
C2. T-1 	beta-Elemene CMS-51192 CAS: 33880–83-0	1.0	1.0	9 rules: Aliphatic hydroxylations; carboxylations; vinyl oxidation…	No rules	Genetox Skin Sens.	1.0 	No data
C3. A-1.1 	D-limonene CMS-797 CAS: 5989–27-5	0.58	0.25	9 rules: Aliphatic hydroxylations; carboxylations	No rules	Genetox Repeat Dose Toxicity (RDT) DART Skin Sens.	0.91 	
A-1.2 	Trivinylcyclohexane CMS-15106 2855–27-8	0.67	0.8	9 rules: Aliphatic hydroxylations; vinyl oxidations	No rules	Genetox Skin Sens.	0.99 	No data
Cluster 2: Aliphatic (>C6) alcohols								
C9. T-2 	octene-3-ol amylvinyl carbinol CMS-7321 CAS: 3391–86-4	1.0	1.0	9 rules: Aliphatic hydroxylations; *sec*-OH & vinyl oxidations	2 rules: Aldo-keto reductase; Alcohol dehydrogenase	Genetox (read-across) RDT insuff. DART insuff.	1 	
C11. A-2.1 	Linalool CMS-4400 CAS: 78–70-6	0.41	0.53	11 rules: Vinyl oxidation	No rules	Genetox RDT DART Skin Sens.	0.91 	0.88 
C12. A-2.2 	CMS-2195 Nerolidol CAS: 7212–44-4	0.37	0.47	16 rules: Aliphatic hydroxylation; vinyl oxidation	No rules	Ames assay RDT DART Skin Sens.	0.90 	0.80 
A-2.3 	Farnesol CMS-6748 CAS: 4602–84-0	0.27	0.44	18 rules: Aliphatic hydroxylations; dehydrogenase; p-OH oxidation	3 rules: Alcohol dehydrogenase	read-across from nerolidol	0.91 	0.74 
Cluster 3: Alicyclic Alcohols								
C13. T-3 	alpha-Terpineol CMS-4938 CAS: 98–55-5	1.0	1.0	11 rules: Aliphatic hydroxylation; Carboxylation	No rules	Genetox RDT DART Skin Sens.	1.0 	1.0
A-3.1 	Isocyclogeraniol CMS-26351 CAS: 68527–77-5	0.74	0.75	15 rules: Alipharic hydroxylations; carboxylations	1 rule: Alcohol dehydrogenase	Ames assay RDT insuff. DART insuff. Skin Sens.	0.93 	No Data
A-3.2 	Terpinyl acetate CMS-4888 CAS: 80–26-2	0.66	0.36	12 rules: Alipharic hydroxylations; carboxylations	2 rules: Carboxyesterase	Genetox RDT DART insuff. Skin Sens.	0.93 	
A-3.3 	4-Terpineol CMS-10854 CAS: 562–74-3	0.77	0.78	10 rules: Alipharic hydroxylations; carboxylations	No rules	Genetox RDT DART insuff. Skin Sens.	 0.98	
C16. A-3.4 	CMS-15331 Perillyl alcohol CAS: 536–59-4	0.54	0.60	11 rules: Aliphatic hydroxylation; p-OH oxidation; carboxylation	1 rule: Alcohol dehydrogenase	Ames assay RDT insuff. DART insuff. Non-Skin Sens. Human	0.96 	No data
C15 M-A-3.4 	Perillaldehyde CMS-14264 CAS: 2111–75-3	0.42	0.33	10 rules: Aliphatic hydroxylations; Aldehyde oxidation & reduction	3 rules: Aldo-keto reductase; Aldehyde DH; Aldehyde oxidase	Ames assay RDT insuff. DART insuff. Skin Sens.	0.77 	No data

[1] Structures codes: "C" for constituents of the *P. frutescens*; "T" for target'; "A" for analogues; "M" for metabolites. Chemical Records for structurable constituents of *P. frutescens* are provided in Supplementary Information Table S9.
[2] RDKit topological fingerprints were used. Other fingerprints from RDKit included Morgan and MACCS keys.
[3] Structure similarity is quantified by pairwise Tanimoto coefficients. The numbers in parenthesis are Dice indices.

[4] List of metabolic rules are provided in the Supplementary Information Table S5.
[5] Available toxicity data are listed in the Supplementary Information Table S7.

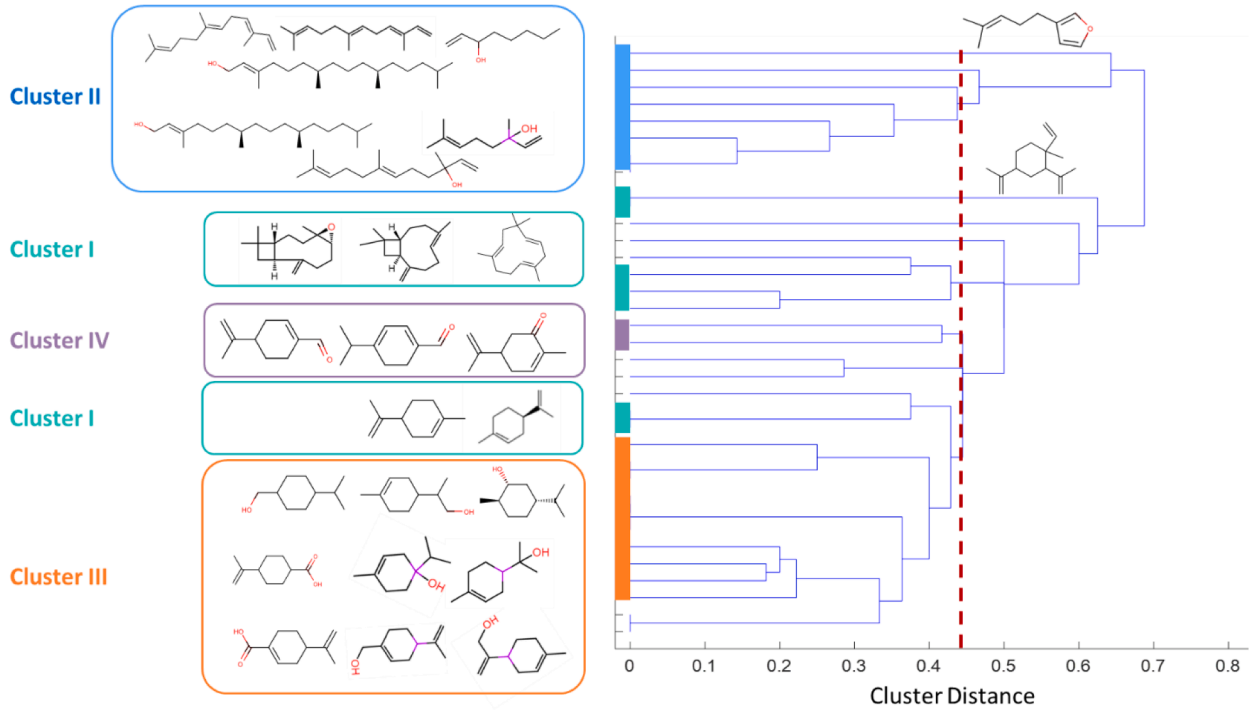


Fig. 8. Hierarchical grouping of *Perilla frutescens* constituents based on Toxprint fingerprints and single linkage.

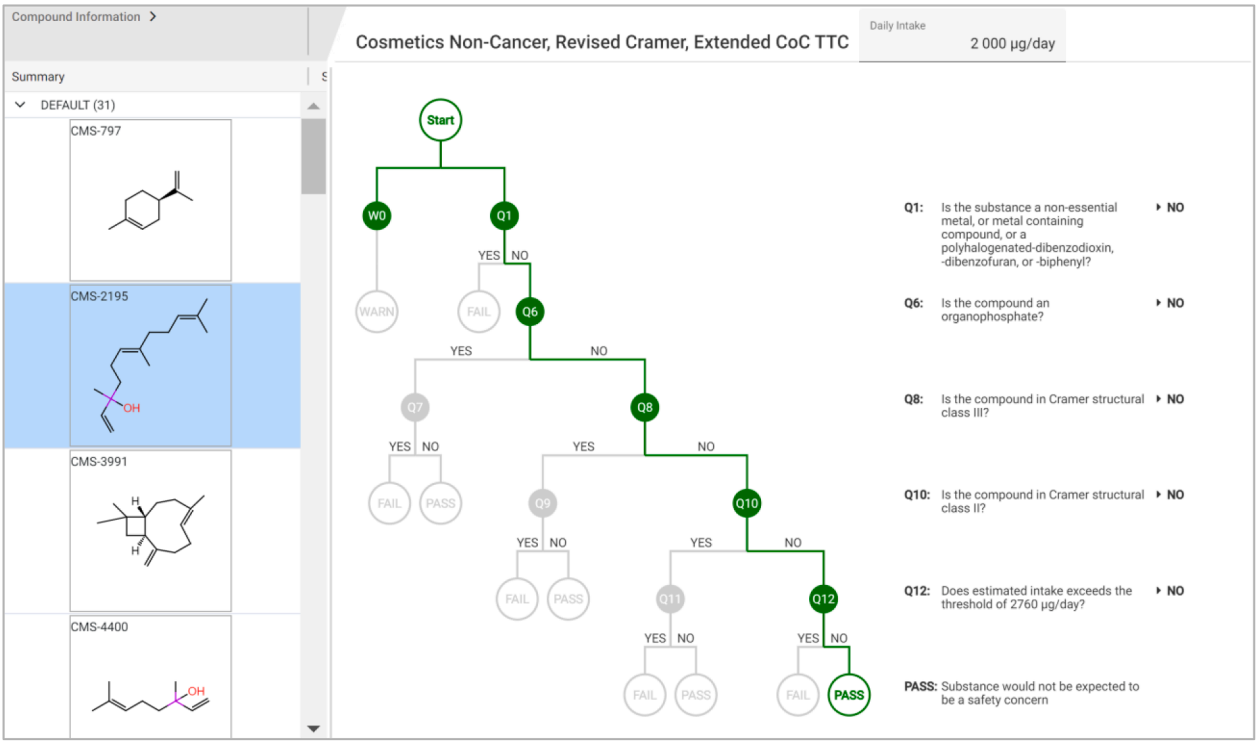


Fig. 9. Cosmetics TTC with revised Cramer Decision Tree in ToxGPS Workflows.

candidate having genetic toxicity data. The T-1 and A-1.2 structures share a common cyclohexyl alkene scaffold, resulting in a structure similarity of 0.80 (ToxPrint fingerprints) and Pearson similarity of 0.93 in property profiles. The structural similarity of d-limonene (A-1.1) to the target was surprisingly low (0.58 by RDKit) although the property-based similarity was higher (0.91 by Pearson similarity). The same set

of properties are calculated as in the Section 2.4.2.2 and their patterns are compared in Table 2 as skyline plots. D-limonene (A-1.1) is included as an analogue for read-across for beta-elemene since it provides rich data on Ames mutagenicity, *in vitro* chromosome aberration, and *in vitro* mammalian mutagenicity. Both target and analogue structures do not contain any DNA binding structural features, which suggests negligible DNA reactivities.

Next, we compiled genetic toxicity data, both mutagenicity and clastogenicity endpoints commonly required in regulatory process, e.g., bacterial reverse mutagenesis (Ames test), *in vitro* mammalian mutagenesis (mouse lymphoma or hypoxanthine–guanine phosphoribosyltransferase (HGPRT)), *in vitro* chromosome aberration, and *in vivo* micronucleus. In addition to the call criteria, study qualities depend on study design parameters including test systems (strains, cells), metabolic activations, dose/concentration ranges, control types, and cytotoxicity. The quality rating approach described in Section 2.4.4 was applied and the data used in the evaluation are summarized in Table 3.

The AQ and SQ were combined to give RR that can be used to judge whether read-across is feasible from the compiled evidence. The qRAX approach further quantifies the confidence in the outcome by estimating the associated uncertainty. In ToxGPS, the read-across is performed first for each analogue per individual endpoint, followed by an assessment that considers multiple analogues. This process can be combined with the outcomes from other related endpoints to arrive at a read-across result for a single overall aggregated outcome. In this case the read-across outcomes for mutagenicity and clastogenicity were further combined to give an aggregated result for the more general genetic toxicity endpoint. This weight-of-evidence analysis followed the Dempster Shafer Theory approach [68]. We present four common examples how the evidence in Table 3 may be applied for assessing the RAX reliability.

Case Study 1: RAX from one analogue for one study. The Ames mutagenicity of the target can be assessed by one analogue A-1.1. Since only this single piece of evidence is considered, the RR is 69% (AQ*SQ = 0.73*0.95).

Case Study 2: RAX from multiple analogues for one study type. Another common approach is to combine results across multiple analogues for one study type, e.g., two Ames studies from each analogue. Simple averaging of the individual RR values would give 75%. However, a basic idea from decision theory allows combination of multiple independent sources of reliable evidence to yield a WOE result with a higher reliability than any individual source alone. In this case, the reliability of the Ames read-across based on both analogues will be 94%. Fig. 10A depicts this process graphically using the WOE calculator within ToxGPS.

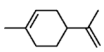
Case Study 3: RAX from multiple studies per one analogue. The third approach is to combine evidence for one analogue across multiple studies or study types. The three genetic toxicity studies for each analogue can be combined for the genotoxic potential of the target compound. The genotoxicity RRs were 94 and 99% per A-1.1 (d-limonene) and A-1.2 (trivinyl cyclohexane), respectively. The combination process is illustrated in Fig. 10B for A-1.1.

Case Study 4: RAX from multiple studies per multiple analogues. Evidence from all analogues across all relevant studies are combined in this case study. When both analogues along with all studies are considered in Table 3, we achieve a higher RR. The decision theory approach results in greater reliability when combining multiple independent studies whose outcomes are in agreement, thereby giving greater confidence than a single study by itself. When a WOE approach is employed to both analogues and including all four studies, the reliability increases to 99%. This result justifies the combination of independent multiple sources of information as shown in Table 3.

The RAX reliability gives the confidence that the genetic toxicity of beta-elemene can be estimated by the selected analogues and their studies. The final step is to estimate the outcome itself. If the purpose of the read-across is to determine the outcome of an Ames assay, it would be difficult to conclude that the target is Ames negative due to the low RAX reliability. However, if the goal of the read-across was to predict the genotoxicity of beta-elemene based on d-limonene and trivinylcyclohexane, then we have very high confidence (99%) that the read-across result will be reliable.

The Assessment Table exported from the ToxGPS RAX workflow is displayed in Fig. 11, detailing the input and output of this process, namely the AQ, SQ and individual study results (negative, positive, equivocal, not used). The individual study results are depicted by colored probability bars with green, red, or yellow denoting, respectively, the strength of the evidence for a negative, positive, or uncertain

Table 3
Weight-of-Evidence assessment of the target (CMS-51192) for RAX Reliability (RR).

Analogues	AQ	SQ-1 (Ames)	SQ-2 (ivtCA)	SQ-3 (ivtMM)	RR (WOE) per analogue (all studies)
A-1.1 (d-limonene) CMS-797	0.73 ($\sqrt{0.58 \times 0.91}$)	SQ-1: 0.95 (Med-High) RR A1_S1: 0.69 4 TA strains ± S9 non-mutagenic	SQ-2: 0.80 (Med -Low) RR A1_S2: 0.58 CHO cells ± S9 non-clastogenic	SQ-2: 0.75 (Low) RR A1_S3: 0.55 Mouse Lymphoma negative	**0.94**
 A-1.2 (trivinyl cyclohexane) CMS-15106	0.81 ($\sqrt{0.67 \times 0.99}$)	SQ-1: 1.0 (High) RR A2_S1: 0.81 4 TA, 2 WP2 strains ± S9 non-mutagenic	SQ-2: 1.0 (High) RR A2_S2: 0.81 CHL V79 ± S9 non-clastogenic	SQ-3: 0.80 (Med-Low) RR A2_S3: 0.65 CHO HPRT negative	**0.99**
					
RR (WOE) for each assay		**0.94**	**0.92**	**0.84**	
Overall RR	**0.99**				

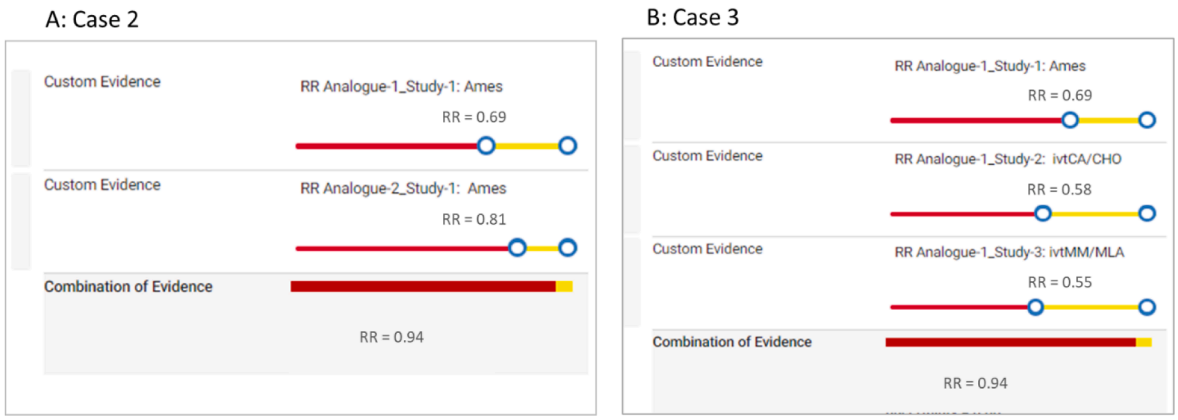


Fig. 10. Weight of Evidence process for read-across reliability estimation for various RAX approaches.

result. The uncertainty is derived from the study quality. The individual probability bars for all selected pieces of evidence are combined to give the final weight of evidence outcome with the associated uncertainty.

As shown in Fig. 11, the outcome of this read-across for the genetic toxicity of beta-elemene is negative based on mutagenicity and clastogenicity evidence and supported by the RAX Reliability >99.9%. The final genetic toxicity outcome is depicted by a probability bar indicating an expected probability of >99.9% for a negative outcome (green) with very low uncertainty (yellow). A positive (genotoxic) study outcome would result in a red bar with uncertainty in yellow. In summary, the genetic toxicity potential of beta-elemene by read-across is negative with very high confidence judging from the RAX Reliability combining the non-genotoxic evidence from d-limonene and trivinylcyclohexane.

Although conventional read-across analyses generally do not consider *in silico* methods as sources of evidence, in the context of NGRA for genetic toxicity, QSAR models and rule-base data can also be included. For example, a prediction for Ames mutagenicity from a reliable QSAR model can be included to strengthen the evidence, improving the read-across for Ames mutagenicity of the target. In cases where Ames mutagenicity is the goal of the read-across using only A-1.2 (CMS-15106), the RAX reliability would be acceptable (77%) but can be improved by combining the QSAR evidence. The decision theory approach explicitly takes into account the strength and the reliability of the *in silico* model itself, as determined from conventional model validation statistics [68]. As shown in Fig. 12, the uncertainty of the predicted outcome is reduced greatly (14%) while confirming the negative Ames mutagenicity outcome.

3.4.2. *Use case 2: read-across to fill a data gap for repeated-dose toxicity of amylvinyl carbinol*

Cluster-2 of the *P. frutescens* extract consists of structures having long aliphatic chains (>= 6 carbons) with one or more double bonds. As shown in Table 2, this class consists of hydrocarbon chains and alcohols that are well known fragrance and flavor additives. Both linalool and nerolidol are data-rich substances, which have been used as read-across materials for other aliphatic alcohols such as amylvinyl carbinol [117-119]. Hence, nerolildol, linalool, and farnesol were considered here as analogues for the target, amylvinyl carbinol, although their reactivities may need to be explored further due to the tertiary- (linalool and nerolidol), secondary- (amylvinyl carbinol), and primary- (farnesol) nature of alcohols. Primary aliphatic alcohols, e.g., farnesol, were labelled for protein binding potential, whereas the secondary alcohol, e.g., amylvinyl carbinol would have different metabolic potential than the tertiary alcohols. ToxPrints distinguished these alcohols for reactivity as well as the degree of branching and number of double bonds in the aliphatic chains. Several quantum mechanical properties, heats of formation and HOMO, differentiated target from the analogues. For biological similarity, the aggregated EDSP assays defined in Fig. 6 were used to quantify assay-based similarity, using the Euclidean similarity measure. When the profile of the aggregated assay vectors was compared, the pattern of the target turned out to be more closely related to nerolidol and linalool rather than to farnesol (cluster-2 in Table 2). The AQ values were calculated as the geometric mean of all three measures of similarities (structure-based, property-based, assay-based). Although not described in detail in this paper, properties related to bioavailability (absorption, permeability, hepatic clearance) and biokinetics parameters can also be included in the AQ determination. For metabolites, the platform is also equipped with metabolite generation capability (Section 2.3.2.4).

We then compiled the toxicity data from short-term, subchronic/chronic, and systematic effects from DART studies. Studies following guidelines were preferred, which resulted in clear NO(A)EL values with systemic/target organ effects. For NOAEL estimation, the original NOAEL values were standardized to chronic NOAEL by applying adjustment factors 3 and 6 for subchronic and short-term studies, respectively. When NOAEL was not established due to the Lowest Observed (Adverse) Effect Level (LO(A)EL) being the lowest tested dose, an adjustment factor of 3 was applied for NO(A)EL extrapolation. The study qualities (SQ) were again estimated in five categories from high to low and mapped to the numeric scale of 0–1 to facilitate the probabilistic assessment as described previously. Table 4 summarizes the steps to estimate RAX reliabilities in consideration of both the analogue qualities (AQ) and the study qualities (SQ) of the data from the analogue candidates. At this point, farnesol was eliminated as an analogue candidate due to insufficient toxicity data as well as its low AQ. Nerolidol provided NOAEL of 105 mg/kg-bw/day from a 59-day repeated-dose/repeated/reproductive study resulting in a chronic NOAEL of 35 mg/kg-bw/day with associated critical effects in liver. The SQ was rated at 0.9 due to a penalty from being shorter than 90-day duration. Two studies were used for linalool, one short-term and the second for a subchronic study. The second 95-day chronic NOAEL was rated at SQ of 0.9 due to a penalty from a free-standing NOAEL (where a LO(A)EL was not established).

The RR for the first analogue (A-2.1 nerolidol) depends only on one study, hence it simply is calculated by AQ*SQ to be 0.63. The RR for the second analogue (A-2.2 linalool) is estimated from the two contribution, namely, AQ*SQ (rat 28-day) and AQ* SQ (90-day). These two terms are then combined using Dempster Shafer Decision Theory to result in RR of 91%. The WOE approach implemented in ToxGPS workflow is analogous to the process depicted in Fig. 10. Again, we observe the benefit of combining multiple sources of evidence so that expected reliability of the WOE result is higher than the individual sources alone.

Unlike the genetic toxicity read-across example in Use Case 1, where the outcome was binary (positive or negative), the outcome of repeated-dose toxicity requires the evaluation of endpoint findings and an associated point of departure value, e.g., NOAELs. Reported findings included clinical chemistry and possible kidney and liver toxicities. The

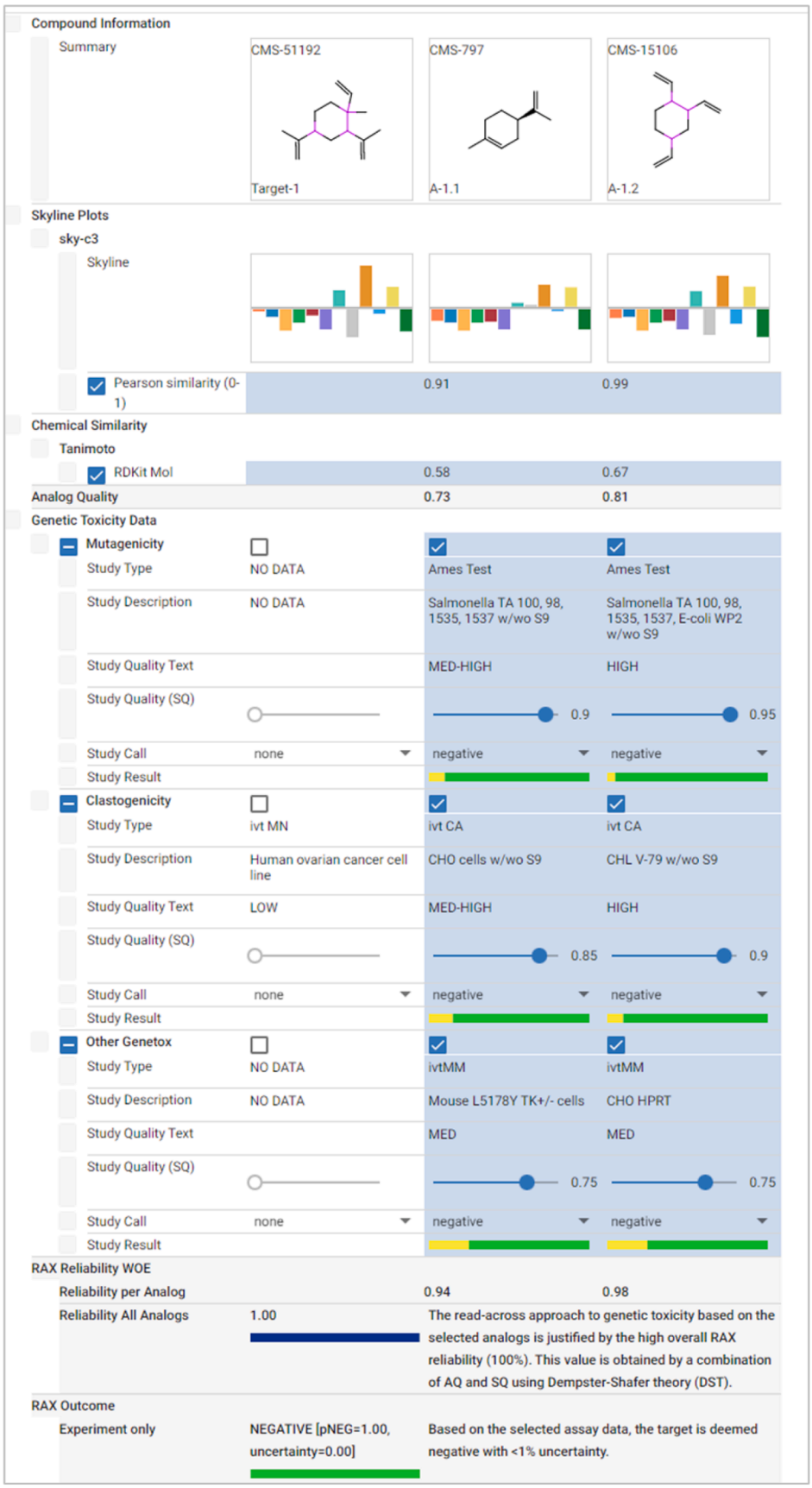


Fig. 11. Read-across assessment table for genetic toxicity (binary endpoints).

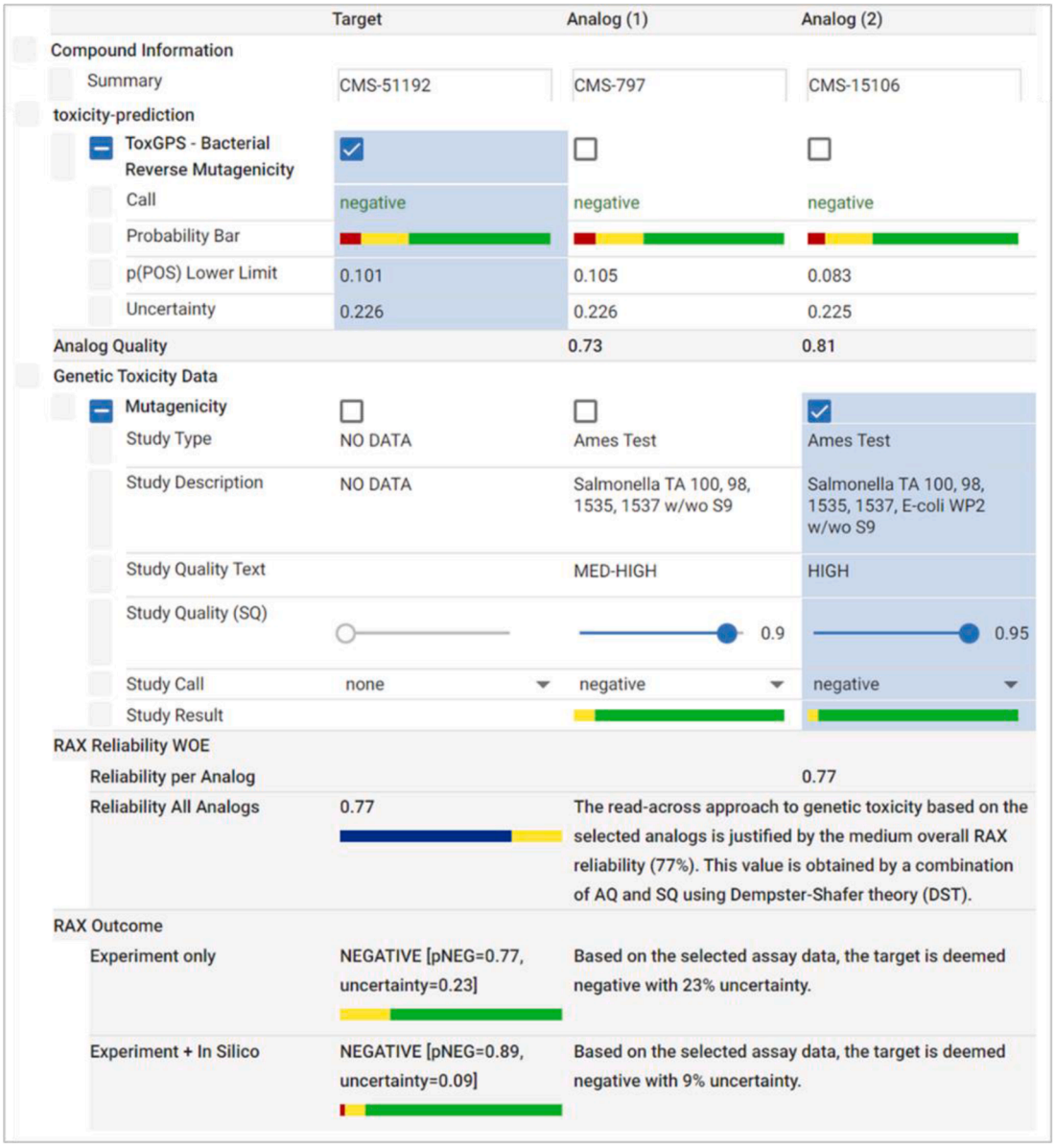


Fig. 12. The effect of including QSAR predictions on read-across reliability and outcome confidence.

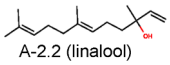
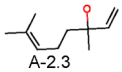
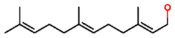
assessment table is customized for the specified endpoint category of interest, so for this read-across on repeated-dose toxicity, the ToxGPS workflow also provides a rough initial estimate of the POD based solely on the NOAEL values provided by users. It's important to stress that the platform does not predict NOAEL values. A 95% confidence interval is obtained by weighting each NOAEL value by its respective individual RR. The target NOAEL is reported based on the data presented in Table 4. The estimated 95% CI (weighted by RR) was 3–765 mg/kg-bw/day with the mean of 48 mg/kg-bw/day. Whilst the RAX reliability quantitatively assesses the robustness of the read-across process, the uncertainty involved in the read-across outcome is, in this case, reflected in the width of the confidence interval. A tight range indicates the various sources complement each other and are in close agreement, whereas a wide confidence interval is the result of significant variability among NOAEL values of the analogues. In this example, there were only three observations with relatively low RR.

This chemoinformatics method has been published previously [32], which describes two complimentary approaches for estimating NOAEL confidence bounds. In the work here, we employ the nearest neighbor ensemble searching method by querying the target amylvinyl carbinol at a confidence level at 95% and the nearest neighbor threshold of 0.6 against the full chronic NOAEL dataset. The similar structures were also constrained to contain the scaffold 3-butene-2-ol, thus limiting search candidates to tertiary or secondary alcohols. The NOAEL bounds estimation service identified additional nearest neighbors, e.g., 2-methyl-e-buten-2-ol (CMS-11757) and linalyl isobutyrate (CMS-11771). Chronic NOAEL bounds were based on 5 observations from repeated-dose toxicity studies. This additional investigation narrowed the 95% confidence interval for the target's NOAEL to 5.1–353 mg/kg-bw/day, a tighter interval than the initial range of 3–765 obtained in the qRAX process. We can say with 95% confidence that this interval contains the true mean chronic NOAEL value for the target. This estimation method allows the refinement of the NOAEL bounds by combining the data from ChemTunes NOAEL dataset with the user provided data, which is the ultimate result of the read-across for repeated-dose toxicity.

3.4.3. Use case 3: read-across to fill a data gap for skin sensitization potential of alpha-terpineol

The risk assessment of alpha-terpineol published by RIFM [97] indicated that the terpineols "do not present a concern" for skin

Table 4
Weight of Evidence assessment of target (CMS-7321) for Repeat Dose Toxicity (RDT) RAX Reliability.

Analogs	AQ	SQ-1 (short-term)	SQ-2 (sub/chronic)	RR per analogue (all studies)
A-2.1 (nerolidol) CMS-2195 	0.70 ($\sqrt[3]{0.47 \times 0.90 \times 0.80}$)	No Data	0.9 (Med-High)RR A1_S-2 (0.63) Rat 59-day oral-diet; liver; chronic NOAEL = 35 mkd	**0.63**
A-2.2 (linalool) CMS-4400 	0.75 ($\sqrt[3]{0.53 \times 0.91 \times 0.88}$)	0.95 (Med-High)RR A2_S-1 (0.71) Rat 28-day oral-gavage; liver, kidney; chronic NOAEL = 20 mkd	0.9 (Med)RR A2_S-2 (0.68) Rat 95-day oral-diet; no adverse effect; chronic NOAEL = 167 mkd	**0.91**
A-2.3 (farnesol) CMS-6748 	0.67 ($\sqrt[3]{0.44 \times 0.91 \times 0.74}$)	Insufficient data	Insufficient data	
RR per study type		0.68	0.88	
Overall RR	**0.96**			

sensitization potential. However, the authors also indicated that "cyclic terpenes could be reasonably anticipated to undergo autoxidation resulting in potentially sensitizing degradation products" [97]. Alpha-terpineol belongs to cluster-3, which consists of terpenyl derivatives. Many substances in this class have been assessed for skin sensitization potential previously and deemed safe for the use of cosmetics ingredients although many of the weight of evidence decisions were qualitative. This study revisited this aspect with a more quantitative treatment. We will here assess alpha-terpineol (CMS-4938) as a neat ingredient.

To identify suitable analogues, dermal profiles of various substances were compared in Table 5 for dermal properties including skin permeability, skin metabolic reactivity, and irritation in addition to protein binding ability. The target and the first two analogues (A-3.1, A-3.2) did not hit any protein binding chemotypes while their skin permeability and irritation profiles were also similar. The analogue A-3.2 (CMS-4888) is alpha terpinyl acetate which is expected to be hydrolyzed to the alcohol form, namely the target, under physiological conditions.

Chemical similarity was again addressed by structural features and molecular properties as in the other clusters. When these candidates include metabolites and/or reacted species, chemical similarity based solely on structural features becomes less indicative or reliable than molecular properties indicating chemical reactivity. In fact, a case is reported in a REACH dossier where repeated dose toxicities of alpha-terpinol were addressed by alpha-terpinyl acetate [112]. Therefore, terpineol acetate was considered in this study as one of the viable analogues although it does not yield high similarity from the structure point of view. The other two candidates were 4-terpineol (CMS-10854) and isocyclogeraniol (CMS-26351). Perilla alcohol and perillaldehyde were excluded due to their protein binding reactivity.

As in the other cases, the same set of molecular properties were employed to address aspects related to size and bonding (H-bond acceptors and donors, number of heavy atoms), interfacial properties for penetrating skin (logarithm of the octanol–water partition coefficient (log P), Total Polar Surface Area (TPSA)), shape of the molecules (e.g., asphericity) important for interacting with proteins, and quantum mechanical properties indicating reactivity, e.g, ΔH_f, HOMO. The skyline plots in Table 2 illustrated patterns of property profiles for analogues and the target. Asphericity and ΔH_f seem to vary within this group. In addition, judging from the rules and QSAR model, their permeabilities were all in the medium log Kp bin ($-3 \leq \log Kp < -1.5$ cm/h). It should be pointed out that the information from skin irritation models or the data from Material Safety Data Sheet is for neat chemicals and were not for a constituent of a botanical extract that is present in a low amount as

Table 5
Dermal Toxicity Relevant Profiles.

Profile Description	CMS-4938(C13) alpha-terpineol†	CMS-26351 isocyclo-geraniol	CMS-4888 terpinyl acetate	CMS-10854 4-Terpineol	CMS-15331 (C16) perillyl alcohol	CMS-14264 (C15) perillaldehyde
Role	**T (target)**	**A-3.1**	**A-3.2**	**A-3.3**	**A-3.4**	**A-3.5**
Protein Binding	0	0	0	0	1 hit	1 hit
Skin Metabolic Reactivity‡	0	1 hit	2 hits	0	1 hit	3 hits
Skin Permeability Prediction†	4 rules; MED log Kp Class	0 rules; MED log Kp Class	0 rules; MED log Kp Class	4 rules; MED log Kp Class	0 rules; MED log Kp Class	0 rules; MED log Kp Class
Skin Irritation†	Positive/Irritating**	Positive/ Irritating**	Positive/Irritating*	Positive/Irritating*	Positive/ Irritating**	Positive/ Irritating**
Tox21/CAST assay (assay vector skyline plots)	Cell surface, growth factor, Immune system, Inflammation, Histone modification 	No Data	Cell surface, growth factor, Immune system, Inflammation, Histone modification Euclid sim. = 0.92	Cell surface, growth factor, Immune system, Inflammation, Histone modification  Euclid sim. = 1.0	No Data	No Data

*Data from ECHA Dossier ** Data taken from Material Safety Data Sheets.
‡ Table S5 in Supplementary Information for metabolic rules.
† Table S8 in Supplementary Information for more details on the models.

in our hand cream scenario.

To bring biological similarity into the analysis, available evidence from Tox21/CAST HTS assays was also included. Although there are no assays directly related to hapten formations or dendritic cell interactions, several assay categories were related to cell surface, growth factors, immune responses, inflammation (e.g., interlukin-2), and histone modifications. These assay data were available for target T-3 and the two analogues (A-3.1 and A-3.2). Pairwise biological similarity was estimated by Euclidean similarity over the assay vectors when data were available. Table 2 shows that assay activity patterns of the target analogues were similar and tend to be lower than the average of the whole *in vitro* dataset.

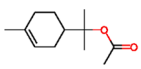
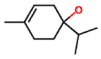
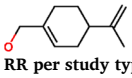
Decisions on the analogues also depended on the availability of reliable local lymph node assay (OECD Test Guideline (TG) 429) data although the AOP-related assays such as DPRA, KeratinoSens (or LuSens) and/or hCLAT assays were also considered. Older *in vivo* studies were also cited when no other data were available, but the study quality was lower. The target was associated with a negative popliteal lymph node assay study in rat as well as negative guinea pig and human studies; no OECD TG 429 or AOP-related assays were found in public literature. Two analogues of isocyclogeraniol (A-3.1) and alpha-terpinyl acetate (A-3.2) have well-characterized LLNA data; A-3.1 was considered a mild sensitizer (EC3 $>$ 25 wt/vol%) whereas A-3.2 was labelled a non-sensitizer (EC3 $>$ 100 wt/vol%). The AOP-related assays alone for A-3.1 indicated possible negative results in conflict with positive *in vivo* studies in LLNA and Buehler Test [120]. The analogue 4-terpeniol (A-3.3) had positive outcomes in DPRA and LuSens assays. The analogue perillyl alcohol (A-3.4) had only human maximization data [121], where no reactions indicative of sensitization were observed with the test material exposure of 2760 μg/cm^2. The toxicity data and SQ are summarized in Table 6 along with the analogue qualities.

The RR values were estimated to be high due to multiple analogues and studies. The next step is to assess the skin sensitization potential of the target. Due to conflicting study results shown in Table 6, the overall RAX outcome based on these studies has high uncertainty. For each given analogue, there are conflicts across the studies (LLNA, AOP, other); for each study, there are conflicts across the analogues. In fact, the main source of the uncertainty originates from the conflicts in LLNA study results for the two analogues (isocyclogeraniol and alpha-terpinyl acetate), hence combining other studies or analogues will not greatly reduce the uncertainty of the outcome. If only one analogue with one LLNA study is selected, the RR decreases, but will artificially decrease the outcome uncertainty by ignoring other qualified information. If only A-3.1 is considered with one LLNA study, the RR will decrease to 0.70 (from 1.00) with the expected outcome of sensitizer with an uncertainty of 0.30. Taking A-3.2 with LLNA study will yield RR of 0.77 and the skin sensitization potential would be non-sensitizer with uncertainty of 0.23. Since the AQ of the two analogues and the SQ of both studies were all similar, the RR values are also comparable. The only difference is that the outcome of LLNA studies were in conflict where the results were deemed both acceptable. The uncertainty involved in this outcome estimation was quantitatively estimated to be nearly 0.5 (48%) and hence the skin sensitization hazard outcome would be considered equivocal. Even when the LLNA model result for the target is added in the combination, the uncertainty does not decrease due to the fact that the prediction was also a weak negative (probability of being positive (pPOS) = 0.37; probability of being negative (pNEG) = 0.60; probability of being uncertain (pUNC) = 0.03).

The target CMS-4938 (alpha-terpineol), however, had a few older pre-LLNA studies, which were not included in the above assessment due to uncertainties associated with the test methods, including guinea pig [122], human and rat (popliteal lymph node assay) [112] studies. Since the analogue data give conflicting information, we decided to include one of these studies for the target in the combination of evidence. Fig. 13 illustrates the impact of inclusion of popliteal local lymph node assay as part of the target information with a very low study quality of 0.5. In Fig. 13-A, the outcome of the two pieces of evidence, i.e., RAX results for 3 analogues in Table 6 and LLNA hazard prediction. The outcome is still equivocal. When we add the study data for target (3rd custom evidence), the combined outcome turned to negative (pNEG = 0.60) with lower but still relatively large uncertainty (21%) as exhibited in Fig. 13-B.

In a sense, this is not a surprising conclusion based on the publications on many of these terpineol analogues [70,86,97,98,123]. Whilst the risk assessment determined there were no safety concerns for skin sensitization under a realistic daily exposure condition, the conclusions also mentioned the possible reactivity of terpenoids. This discussion is

Table 6
Weight of Evidence assessment of CMS-4938 for skin sensitization RAX Reliability (RR).

	AQ	SQ-1 (LLNA)	SQ-2 (AOP assays)	SQ-3 (other)	RR per analogue (all studies)
A-3.1 CMS-26251 (isocyclogeraniol) 	0.86 ($\sqrt{0.74 \times 0.93}$)	1 (High) LLNA **weak sensitizer** EC3 = 25 wt/vol%	0.85 (Med) **negative**	0.75 (Med-Low) Buehler test **positive**	0.98
A-3.2 CMS-4888 (terpineol acetate) 	0.83 ($\sqrt[3]{0.66 \times 0.93 \times 0.92}$)	1 (High) LLNA **non-sensitizer** EC3 > 100 wt/vol%	NA	NA	0.83
A-3.3 (4-terpineol) 	0.90 ($\sqrt[3]{0.77 \times 0.95 \times 1.0}$)	NA	0.85 (Med) **positive**	NA	0.77
A-3.4 (perilla. Alc.) 	0.71 ($\sqrt{0.54 \times 0.96}$)	NA	NA	0.75 Human study **negative**	0.53
RR per study type		0.97	0.93	0.82	
Overall RR	**0.999**				

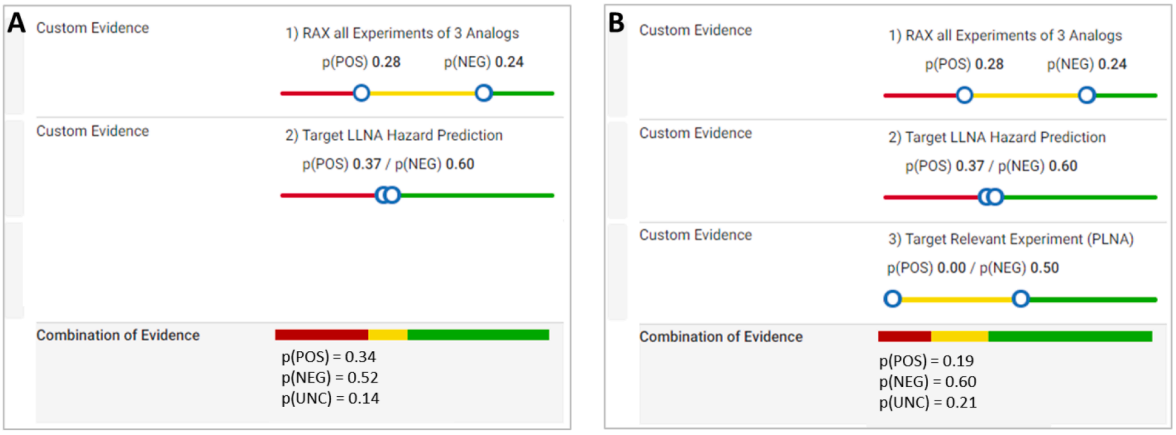


Fig. 13. Effect of a relevant study for the target on the skin sensitization outcome. A: no experimental data for target; B: experimental data for target.

very much in agreement with extensive assessment reports from RIFM on the fragrance materials based on terpenoids.

3.5. Safety assessment of Perilla frutescens

Although the safety assessment of the botanical itself was not our goal, an overall picture can now be described after individual evaluations. For each major cluster, endpoint assessment was feasible by applying the qRAX approach. Whenever possible, our findings were compared with those of the assessment for *P. frutescens*.

Genetic Toxicity. Genetic toxicity data of constituents of *P. frutescens* are presented in Table 3. Most substances did not show mutagenicity or clastogenicity except perillaldehyde. When no genetic toxicity data were available, a substance-specific assessment was carried out. ToxGPS prediction results based on QSAR and chemotype rules along with qRAX were applied using the same set of criteria for analogue quality, i.e., structural features, properties, and DNA binding/damage/repair HTS assay profiles. Beta-elemene, amylvinyl carbinol and terpinyl acetate were the target structures for which the *in silico* safety assessments were carried out in ChemTunes. Their outcome was non-genotoxic with uncertainties in the outcomes range of 0.003–0.01.

From the point of view of the extract as a whole, there is only one recent publication on genetic toxicity evaluation of *P. frutescens* (L.) Britt [124]. On the basis of *in vitro* micronucleus assay in CHL V79 Cell Line, the authors concluded no genotoxic effect of the extract neither with nor without metabolic activation. These data are consistent with our overall weight of evidence outcomes.

Repeated-Dose Toxicity Information on repeated-dose, as well as DART, studies is summarized in the Supplementary Information Table S7. In addition to the analysis of cluster-2 (read-across of RDT for amylvinyl carbinol) in Section 3.4.2, both cluster-1 and cluster-3 structures were also analyzed using the qRAX approach. Given d-limonene and its data from cluster-2, repeated-dose toxicity of members of the cluster-1 whose AQ values are above 0.7, were considered to be related to systemic effects including body weight changes and possible liver effects from short, subchronic, and chronic studies with chronic NOAEL values ranging 167–250 (as compared to the experimentally determined NOAEL values ranging from 250 to 500 mg/kg-bw/day). Studies showing nephrotoxicity in male rats were not selected since these kidney effects identified as alpha 2u-globulin nephropathy are not considered to be relevant for humans [125]. Similarly, evaluation of repeated-dose toxicity of perillyl alcohol was feasible by using the cluster-3 members as analogues, resulting in RAX Reliability of 97%. Again, for the members of cluster-3, applying the qRAX approach for liver effects may be a consideration, yielding an estimated NOAEL range of 58–357 mg/kg-bw/day based on the data employed. Based on data from all three clusters, the NOAEL 95% confidence bounds was 5–357 mg/kg-bw/day.

The pharmacological and toxicological profiles of *P. frutescens* have been reviewed from the vantage point of the extract considered as a whole [126]. However, studies leading to understand long-term safety, e.g., chronic or subchronic studies in animals, have not been reported [124,127]. Atypical interstitial pneumonia for human and acute pulmonary emphysema in bovines were reported for *P. frutescens*. Recently a 13-week study was reported for a herb medicine (Samson) containing *P. frutescens* (L.) Britton along with *Panax ginseng* C. A. Mey. and *Peucedanum praeruptorum* Dunn [128] with a NOEL of 1000 mg/kg-bw/day for increased serum creatinine level and overall NOAEL of 4000 mg/kg-bw/day. Overall, very little information is known for the *P. frutescens* extract on studies designed to address long-term safety.

Skin Sensitization Most of the constituents in cluster-1 and 2 had positive skin sensitization potential whereas the members in cluster-3 exhibited varying degrees of skin reactivity and sensitization potential to *in vitro* AOP assays, *in vivo* toxicities of LLNA or guinea pig assays as well as human studies. *P. frutescens* has been reported for occupational dermatitis in Asia [129]. However, the exposure from the extract in the cosmetic formulation is several orders of magnitude lower than the exposure associated with occupational hazard. Our substance-specific evaluations were more relevant to the occupational hazard than the use of this substance in cosmetics formulations. As discussed in Section 3.2, the anticipated exposure of *P. frutescens* would be 3.27 μg/kg-bw/day assuming the use case of its concentration at 0.01% in hand cream lotion with hand surface area of 860 cm^2 [130]. This would be equivalent to an exposure (mass per area of skin) of 0.228 $\mu g/cm^2$, which is 4000-fold lower than the non-reactive Dermal Sensitization Threshold of 900 $\mu g/cm^2$ [131].

4. Conclusions

This investigation demonstrated the pivotal role of molecular informatics designed to facilitate and support safety and risk assessment process for chemical substances. The presented ChemTunes•ToxGPS® platform for NGRA implemented improvements such as new chemical, biological, and safety data based both on experiments and predictions, and the ability to construct read-across workflows with quantitative methods for evaluation. Thus, the safety assessment workflow enables users to compile qualified data sources, quantify their reliabilities, and combine them using a weight of evidence approach based on decision theory. The conceptual framework is also designed to expand the domains of the knowledgebase to new approach methodologies including toxicogenomics, transcriptomics, and molecular pathways.

Further, the power of these approaches to support the assessment of the use of a botanical in a cosmetic product, for which no data currently exist, was demonstrated. Common workflows for chemical safety

assessment of botanical extracts were described following a component-based approach. Constituents of a botanical with known composition were analyzed first for threshold of toxicological concerns based on the exposure estimation from the product formula for non-genotoxic constituents. Substance-specific assessments were then conducted for safety outcomes for genetic toxicity, repeated-dose toxicity, and skin sensitization. Nearly 30 constituents of *P. frutescens* were clustered into five structural groups, from which three clusters were selected to process through the qRAX process. Chemical and biological similarities (AQ) and SQ were combined to give RAX reliability, a measure of the confidence that read-across using the selected analogues and their respective study data will generate reliable outcomes for the target of interest. Each read-across was then completed by exploring combinations of the available evidence to estimate outcomes, with examples presented for genotoxic calls, systemic/target organ findings with estimated NOAEL bounds, and skin sensitization potential with estimated EC3 bounds. Implementation of these workflows is enabled by computational methods available in the ChemTunes•ToxGPS® platform. This paper will be supported by the findings of a further study with use cases in which bioavailability and metabolite generation must be considered. In summary, this study has confirmed the possibility of integration informatics capabilities to support NGRA into a single software platform.

CRediT authorship contribution statement

Chihae Yang: Conceptualization, Writing – original draft. **James F Rathman:** Conceptualization, Writing – review & editing. **Bruno Bienfait:** Writing – review & editing. **Matthew Burbank:** Writing – review & editing. **Ann Detroyer:** Writing – review & editing. **Steven J. Enoch:** Writing – review & editing. **James W. Firman:** Writing – review & editing. **Steve Gutsell:** Writing – review & editing. **Nicola J. Hewitt:** Writing – review & editing. **Bryan Hobocienski:** Writing – review & editing. **Gerry Kenna:** Writing – review & editing. **Judith C. Madden:** Writing – review & editing. **Tomasz Magdziarz:** Writing – review & editing. **Jörg Marusczyk:** Writing – review & editing. **Aleksandra Mostrag-Szlichtyng:** Writing – review & editing. **Christopher-Tilman Krueger:** Writing – review & editing. **Cathy Lester:** Writing – review & editing. **Catherine Mahoney:** Writing – review & editing. **Abdulkarim Najjar:** . **Gladys Ouedraogo:** Writing – review & editing. **Katarzyna R. Przybylak:** Writing – review & editing. **J. Vinicius Ribeiro:** Writing – review & editing. **Mark T.D. Cronin:** Conceptualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: ChemTunes•ToxGPS® is developed and marketed by MN-AM Company. Authors from MN-AM (Yang; Rathman; Bienfait; Hobocienski; Tomasz Magdziarz; Marusczyk; Mostrag-Szlichtyng; Ribeiro) are involved in the design and implementation of this platform. Mark Cronin (LJMU) is a senior science advisor to MN-AM.

Data availability

The data that has been used is confidential.

Acknowledgement

MN-AM wishes to thank Professor Johann Gasteiger and Dr. Oliver Sacher for their earlier contributions to BioPath database and Dr. Johannes Schwöbel for the initial dermal biotransformation rules. This study has been funded by Cosmetics Europe's LRSS program (2017-2022).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.comtox.2023.100272.

References

[1] J.C. Madden, S.J. Enoch, A. Paini, M.T.D. Cronin, A review of *in silico* tools as alternatives to animal testing: Principles, resources and applications, ATLA Altern. Lab. Anim. 48 (2020) 146–172.
[2] K. Taylor, L. Rego Alvarez, 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 (2020), 100112.
[3] E. Commission, Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products, Off. J. Eur. Union L342 (2009) 59–209.
[4] E. Arnesdotter, V. Rogiers, T. Vanhaecke, M. Vinken, An overview of current practices for regulatory risk assessment with lessons learnt from cosmetics in the European Union, Crit. Rev. Toxicol. 51 (2021) 395–417.
[5] M.T.D. Cronin, S.J. Enoch, J.C. Madden, J.F. Rathman, A.-N. Richarz, C. Yang, A review of *in silico* toxicology approaches to support the safety assessment of cosmetics-related materials, Comput. Toxicol. 21 (2022), 100213.
[6] M. Dent, R. Teixeira Amaral, P. Amores Da Silva, J. Ansell, F. Boisleve, M. Hatao, A. Hirose, Y. Kasai, P. Kern, R. Kreiling, S. Milstein, B. Montemayor, J. Oliveira, A. Richarz, R. Taalman, E. Vaillancourt, R. Verma, V.O.C.N. Posada, C. Weiss, H. Kojima, Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients, Comput. Toxicol. 7 (2018) 20–26.
[7] E. Berggren, A. White, G. Ouedraogo, A. Paini, A.N. Richarz, F.Y. Bois, T. Exner, S. Leite, L.A. van Grunsven, A. Worth, C. Mahony, *Ab initio* chemical safety assessment: A workflow based on exposure considerations and non-animal methods, Comput. Toxicol. 4 (2017) 31–44.
[8] C. Alexander-White, D. Bury, M. Cronin, M. Dent, E. Hack, N.J. Hewitt, G. Kenna, J. Naciff, G. Ouedraogo, A. Schepky, C. Mahony, A 10-step framework for use of read-across (RAX) in next generation risk assessment (NGRA) for cosmetics safety assessment, Regul. Toxicol. Pharm. 129 (2022), 105094.
[9] D. Bury, C. Alexander-White, H.J. Clewell III, C. Cronin, B. Desprez, A. Detroyer, A. Efremenko, J. Firman, E. Hack, N.J. Hewitt, G. Kenna, M. Klaric, C. Lester, C. Mahony, G. Ouedraogo, A. Paini, A. Schepky, New framework for a non-animal approach adequately assures the safety of cosmetic ingredients – A case study on caffeine, Regul. Toxicol. Pharm. 123 (2021), 104931.
[10] G. Ouedraogo, C. Alexander-White, D. Bury, H.J. Clewell, M. Cronine, T. Cull, M. Dent, B. Desprez, A. Detroyer, C. Ellison, S. Giammanco, E. Hack, N.J. Hewitt, G. Kenna, M. Klaric, R. Kreiling, C. Lestern, C. Mahony, E. Mombelli, J. Naciff, J. O'Brien, A. Schepky, S. Tozer, B. van der Burg, B. van Vugt-Lussenburg, S. Stuard, C. Europe, Read-across and new approach methodologies applied in a 10-step framework for cosmetics safety assessment – A case study with parabens, Regul. Toxicol. Pharm. 132 (2022), 105161.
[11] ChemTunes, 2023. ChemTunes•ToxGPS®. MN-AM. https://www.mn-am.com/products/chemtunestoxgps (last (accessed 2 January 2023).
[12] T. Gocht, E. Berggren, H.J. Ahr, I. Cotgreave, M.T. Cronin, G. Daston, B. Hardy, E. Heinzle, J. Hescheler, D.J. Knight, C. Mahony, M. Peschanski, M. Schwarz, R. S. Thomas, C. Verfaillie, A. White, M. Whelan, The SEURAT-1 approach towards animal free human safety assessment, ALTEX 32 (2015) 9–24.
[13] GRAS. US FDA Generally Regarded as Safe (GRAS) Notification (GRN). https://www.fda.gov/food/generally-recognized-safe-gras/gras-notice-inventory (accessed 2 January 2023).
[14] PostgreSQL, 2022. https://www.postgresql.org/ (accessed 2 January 2023).
[15] RDKit 2023. RDKit at https://www.rdkit.org/ (accessed 2 January 2023).
[16] K. Boda, T. Seidel, J. Gasteiger, Structure and reaction based evaluation of synthetic accessibility, J. Comput. Aided Mol. Des. 21 (2007) 311–325.
[17] J. Sadowski, J. Gasteiger, G. Klebe, Comparison of automatic three-dimensional model builders using 639 X-ray structures, J. Chem. Inf. Comput. Sci. 34 (1994) 1000–1008.
[18] C. Yang, M.T.D. Cronin, K.B. Arvidson, B. Bienfait, S.J. Enoch, B. Heldreth, B. Hobocienski, K. Muldoon-Jacobs, Y. Lan, J.C. Madden, T. Magdziarz, J. Marusczyk, A. Mostrag, M. Nelms, D. Neagu, K. Przybylak, J.F. Rathman, J. Park, A.-N. Richarz, A.M. Richard, J.V. Ribeiro, O. Sacher, C.H. Schwab, V. Vitcheva, P. Volarath, A.P. Worth, COSMOS Next Generation – A public knowledge base leveraging chemical and biological data to support the regulatory assessment of chemicals, Comput. Toxicol. 19 (2021), 100175.
[19] ECHA, European Chemical Agency) Registered Substances Database, accessed 2 January 2023, https://echa.europa.eu/information-on-chemicals/registered-substances, 2023.
[20] IUCLID, IUCLID 6 & Public API, accessed 2 January 2023, https://iuclid6.echa.europa.eu/public-api, 2023.
[21] OpenFoodTox 2023. EFSA Chemical Hazard Database. https://www.efsa.europa.eu/en/data/chemical-hazards-data (accessed 2 January 2023).
[22] J.L.C.M. Dorne, J. Richardson, A. Livaniou, E. Carnesecchi, L. Ceriani, R. Baldin, S. Kovarich, M. Pavan, E. Saouter, F. Biganzoli, L. Pasinato, M. Zare Jeddi, T. P. Robinson, G.E.N. Kass, A.K.D. Liem, A.A. Toropov, A.P. Toropova, C. Yang, A. Tarkhov, N. Georgiadis, M.R. Di Nicola, A. Mostrag, H. Verhagen, A. Roncaglioni, E. Benfenati, A. Bassan, EFSA's OpenFoodTox: An open source toxicological database on chemicals in food and feed and its future developments, Environ. Int. 146 (2021), 106293.

[23] CosIng. European Commission Cosmetic Ingredient Database. https://ec.europa.eu/growth/sectors/cosmetics/Cosing_en (accessed 2 January 2023).
[24] CAS 2023. https://commonchemistry.cas.org/ (accessed 2 January 2023).
[25] COSMOS NG (2023). http://www.ng.cosmosdb.eu/ (accessed 2 January 2023).
[26] NITE, Japanese National Institute of Technology and Evaluation (NITE) Hazard Evaluation Support System (HESS) Integrated Platform, accessed 2 January 2023, https://www.nite.go.jp/, 2023.
[27] US EPA IRIS, https://www.epa.gov/iris/reference-dose-rfd-description-and-use-health-risk-assessments (accessed 13 March 2023).
[28] SCCS, Scientific Committee of Consumer Safety (SCCS) Opinions, accessed 2 January 2023, https://ec.europa.eu/health/scientific_committees/consumer_safety/opinions_en, 2023.
[29] CPDB. Cancer Potency Database. https://healthdata.gov/dataset/Carcinogenic-Potency-Database-CPDB-/sqjy-rr5s/data (accessed 13 March 2023).
[30] NTP. US National Toxicology Program (NTP) databases. https://ntp.niehs.nih.gov/publications/ (accessed 2 January 2023).
[31] IARC. World Health Organisation International Agency for Research on Cancer (WHO IARC). https://monographs.iarc.fr/ (accessed 6 Jan 2023).
[32] C. Yang, J.F. Rathman, T. Magdziarz, A. Mostrag, S. Kulkarni, T.S. Barton-Maclaren, Do similar structures have similar No Observed Adverse Effect Level (NOAEL) values? Exploring chemoinformatics approaches for estimating NOAEL bounds and uncertainties, Chem. Res. Toxicol. 34 (2021) 616–633.
[33] P. Lee, H. Aizawa, L. Gan, C. Prakash, D. Zhong (Eds.), Handbook of Metabolic Pathways of Xenobiotics, 1st Edition, Wiley, London, 2014.
[34] Drugs@FDA. https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm (last accessed 30 December 2022).
[35] EFSA publications. https://www.efsa.europa.eu/en/publications (last accessed Jan 2023).
[36] J. Lu, M.R. Goldsmith, C.M. Grulke, D.T. Chang, R.D. Brooks, J.A. Leonard, M. B. Phillips, E.D. Hypes, M.J. Fair, R. Tornero-Velez, J. Johnson, C.C. Dary, Y. M. Tan, Developing a physiologically-based pharmacokinetic model knowledgebase in support of provisional model construction, PLoS Comput. Biol. 12 (2016) el004495.
[37] ToxCast *In Vitro* Database. https://www.epa.gov/chemical-research/exploring-toxcast-data#Download (last accessed Jan 2023).
[38] NTP MOA Download. https://ice.ntp.niehs.nih.gov/downloads/MOA/cHTSMOA.xlsx (last accessed Jan 2023).
[39] Hennemann M, Clark T, (2014) EMPIRE: A highly parallel semiempirical molecular orbital program: 1: Self-Consistent Field Calculations. Journal of Molecular Modeling 20: 2331.
[40] CORINA Classic (2023). MN-AM. https://mn-am.com/products/corina/ (accessed 2 January 2023).
[41] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, 2000.
[42] C. Yang, A. Tarkhov, J. Marusczyk, B. Bienfait, J. Gasteiger, T. Kleinoeder, T. Magdziarz, O. Sacher, C. Schwab, J. Schwoebel, L. Terfloth, K. Arvidson, A. Richard, A. Worth, J. Rathman, A new publicly available chemical query language, CSRML, to support chemotype representations for application to data-mining and modeling, J. Chem. Inf. Model. 55 (2015) 510–528.
[43] S.J. Enoch, M.T.D. Cronin, A review of the electrophilic reaction chemistry involved in covalent DNA binding, Crit. Rev. Toxicol. 40 (2010) 728–748.
[44] S.J. Enoch, M.T.D. Cronin, Development of new structural alerts suitable for chemical category formation for assigning covalent and non-covalent mechanisms relevant to DNA binding, Mutation Research/Genetic Toxicology and Environmental Mutagenesis 743 (2012) 10–19.
[45] S.J. Enoch, C.M. Ellison, T.W. Schultz, M.T.D. Cronin, A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity, Crit. Rev. Toxicol. 41 (2011) 783–802.
[46] Toxprint 2023. ToxPrint Chemotypes. https://toxprint.org/ (accessed 2 January 2023).
[47] M.T.D. Cronin, F.J. Bauer, M. Bonnell, B. Campos, D.J. Ebbrell, J.W. Firman, S. Gutsell, G. Hodges, G. Patlewicz, M. Sapounidou, N. Spînu, P.C. Thomas, A. P. Worth, A scheme to evaluate structural alerts to predict toxicity – Assessing confidence by characterising uncertainties, Regul. Toxicol. Pharm. 135 (2022), 105249.
[48] N. Greene, L. Fisk, R.T. Naven, R.R. Note, M.L. Patel, D.J. Pelletier, Developing structure-activity relationships for the prediction of hepatotoxicity, Chem. Res. Toxicol. 23 (2010) 1215–1222.
[49] K.R. Przybylak, M.T.D. Cronin, *In silico* studies of the relationship between chemical structure and drug induced phospholipidosis, Mol. Inf. 30 (2011) 415–429.
[50] M. Hewitt, S.J. Enoch, J.C. Madden, K.R. Przybylak, M.T.D. Cronin, Hepatotoxicity: A scheme for generating chemical categories for read-across, structural alerts and insights into mechanism(s) of action, Crit. Rev. Toxicol. 43 (2013) 537–558.
[51] K.R. Przybylak, A.R. Alzahrani, M.T.D. Cronin, How does the quality of phospholipidosis data influence the predictivity of structural alerts? J. Chem. Inf. Model. 54 (2014) 2224–2232.
[52] C.L. Mellor, F.P. Steinmetz, M.T.D. Cronin, The identification of nuclear receptors associated with hepatic steatosis to develop and extend adverse outcome pathways, Crit. Rev. Toxicol. 46 (2016) 138–152.
[53] C.L. Mellor, F.P. Steinmetz, M.T.D. Cronin, Using molecular initiating events to develop a structural alert based screening workflow for nuclear receptor ligands associated with hepatic steatosis, Chem. Res. Toxicol. 29 (2016) 203–212.
[54] S. Wu, J. Fisher, J. Naciff, M. Laufersweiler, C. Lester, G. Daston, K. Blackburn, Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants, Chem. Res. Toxicol. 26 (2013) 1840–1861.
[55] M.D. Nelms, M.T.D. Cronin, T.W. Schultz, S.J. Enoch, Experimental verification, and domain definition, of structural alerts for protein binding: epoxides, lactones, nitroso, nitros, aldehydes and ketones, SAR QSAR Environ. Res. 24 (2013) 695–709.
[56] L. Ridder, M. Wagener, SyGMa: combining expert knowledge and empirical scoring in the prediction of metabolites, ChemMedChem 3 (2008) 821–832.
[57] S.D. Dimitrov, L.K. Low, G.Y. Patlewicz, P.S. Kern, G.D. Dimitrova, M.H. Comber, R.D. Phillips, J. Niemela, P.T. Bailey, O.G. Mekenyan, Skin sensitisation: modeling based on skin metabolism simulation and formation of protein conjugates, Int. J. Toxicol. 24 (2005) 189–204.
[58] G.M. Cramer, R.A. Ford, R.L. Hall, Estimation of toxic hazard decision tree approach, Food Cosmet. Toxicol. 16 (1978) 255–276.
[59] Toxtree. http://toxtree.sourceforge.net/download.html#Toxtree_3.1.0 (accessed 2 January 2023).
[60] SCCS, 2021. EC Scientific Committee of Consumer Safety (SCCS), The SCCS Notes Of Guidance For The Testing Of Cosmetic Ingredients And Their Safety Evaluation 11Th Revision, 30-31 March 2021, SCCS/1628/21.
[61] I.C. Munro, R.A. Ford, E. Kennepohl, J.G. Sprenger, Correlation of structural class with no-observed-effect levels: a proposal for establishing a threshold of concern, Food Chem. Toxicol. 34 (1996) 829–867.
[62] C. Yang, S.M. Barlow, K.L. Muldoon Jacobs, V. Vitcheva, R. Boobis, S.P. Felter, K. B. Arvidson, D. Keller, M.T.D. Cronin, S. Enoch, A.P. Worth, H.M. Hollnagel, Thresholds of Toxicological Concern for cosmetics-related substances: New database, thresholds, and enrichment of chemical space, Food Chem. Toxicol. 109 (2017) 170–193.
[63] Kroes R, Renwick AG, Cheeseman M, Kleiner J, Mangelsdorf I, Piersma A, Schilter B, Schlatter J, van Schothorst F, Vos JG, Würtzen G, European branch of the International Life Sciences Institute (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: 65-83.
[64] EFSA 2019. More S, Bampidis V, Benford D, Boesten J, Bragard C, Halldorsson T, Hernandez-Jerez A, Hougaard-Bennekou S, Koutsoumanis K, Naegeli H, Nielsen SS, Schrenk D, Silano V, Turck D, Younes M, Aquilina G, Crebelli R, Gurtler R, Hirsch-Ernst KI, Mosesso P, Nielsen E, Solecki R, Carf M, Martino C, Maurici D, Parra Morte J, Schlatter J. 2019. Statement on the genotoxicity assessment of chemical mixtures. EFSA Journal 2019; 17(1):5519, 11 pp. https://doi.org/10.2903/j.efsa.2019.5519.
[65] Yang C, Rathman JF, Ribeiro JV, Mostrag A, Hobocienski B, Magdriarz T, Kulkarni S, Barton-Maclaren TS. (2023) High throughput read-across for screening a large inventory of related structures by balancing machine learning and human knowledge. Submitted to Chemical Research in Toxicology, Feb 2023.
[66] C. Yang, M.A. Cheeseman, J.F. Rathman, A. Mostrag, V. Vitcheva, N. Skoulis, S. Goldberg, A new paradigm in threshold of toxicological concern based on chemoinformatics analysis of a highly curated database enriched with antimicrobials, Food Chem. Toxicol. 143 (2020), 111561.
[67] J. Rathman, C. Yang, V.J. Ribeiro, A. Mostrag, S. Thakkar, W. Tong, B. Hobocienski, O. Sacher, T. Magdziarz, B. Bienfait, Development of a battery of *in silico* prediction tools for drug-induced liver injury from the vantage point of translational safety assessment, Chem. Res. Toxicol. 34 (2021) 601–615.
[68] J.F. Rathman, C. Yang, H. Zhou, Dempster-Shafer theory for combining *in silico* evidence and estimating uncertainty in chemical risk assessment, Comput. Toxicol. 6 (2018) 16–31.
[69] UN GHS. United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS) https://unece.org/about-ghs.
[70] A.M. Api, F. Belmonte, D. Belsito, S. Biserta, D. Botelho, M. Bruze, G.A. Burton Jr., J. Buschmann, M.A. Cancellieri, M.L. Dagli, M. Date, W. Dekant, C. Deodhar, A. D. Fryer, S. Gadhia, L. Jones, K. Joshi, A. Lapczynski, M. Lavelle, D.C. Liebler, M. Na, D. O'Brien, A. Patel, T.M. Penning, G. Ritacco, R.J. Rodriguez-Ropero, N. Sadekar, D. Salvito, T.W. Schultz, I.G. Sipes, G. Sullivan, Y. Thakkar, Y. Tokuram, S. Tsang, RIFM fragrance ingredient safety assessment, p-mentha-1,8-dien-7-al, CAS Registry Number 2111–75-3, Food Chem. Toxicol. 134 (2019), 110711.
[71] G.F. Gerberick, C.A. Ryan, P.S. Kern, H. Schlatter, R.J. Dearman, I. Kimber, G. Patlewicz, D.A. Basketter, Compilation of historical Local Lymph Node data for evaluation of skin sensitization alternative methods, Dermatitis 16 (2005) 157–202.
[72] D. Urbisch, A. Mehling, K. Guth, T. Ramirez, N. Honarvar, S. Kolle, R. Landsiedel, J. Jaworska, P.S. Kern, F. Gerberick, A. Natsch, R. Emter, T. Ashikaga, M. Miyazawa, H. Sakaguchi, Assessing skin sensitisation hazard in mice and men using non-animal test methods, Regul. Toxicol. Pharm. 71 (2015) 337–351.
[73] M. Reitz, O. Sacher, A. Tarkhov, D. Trümbach, J. Gasteiger, Enabling the exploration of biochemical pathways, Org. Biomol. Chem. 22 (2004) 3226–3237.
[74] RDKit fp 2023. RDKit fingerprints documentation at https://www.rdkit.org/docs/ (accessed 2 January 2023).
[75] H.L. Morgan, The generation of a unique machine description for chemical structures - a technique developed at Chemical Abstracts Service, J. Chem. Doc. 5 (1965) 107–113.
[76] J.L. Durant, B.A. Leland, D.R. Henry, J.G. Nourse, Reoptimization of MDL Keys for use in drug discovery, J. Chem. Inf. Comput. Sci. 42 (2002) 1273–1280.
[77] Tanimoto TT (1958) An Elementary Mathematical theory of Classification and Prediction. International Business Machines Corporation, New York NY, November 17, 1958.
[78] L.R. Dice, Measures of the amount of ecologic association between species, Ecology 26 (1945) 297–302.

[79] C.L. Mellor, R.L. Marchese Robinson, R. Benigni, D. Ebbrell, S. Enoch, J. W. Firman, J.C. Madden, G. Pawar, C. Yang, M.T.D. Cronin, Molecular fingerprint-derived similarity measures for toxicological read-across: Recommendations for optimal use, Regul. Toxicol. Pharm. 101 (2019) 121–134.

[80] S.J. Enoch, M.T.D. Cronin, J.C. Madden, M. Hewitt, Formation of structural categories to allow for read-across for teratogenicity, QSAR Comb. Sci. 28 (2009) 696–708.

[81] R. Wang, Y. Gao, L. Lai, Calculating partition coefficient by atom-additive method, Perspect. Drug Discovery Des. 19 (2000) 47–66.

[82] EDSP. US Environmental Protection Agency Endocrine Screening Program. https://www.epa.gov/endocrine-disruption/endocrine-disruptor-screening-program-edsp-overview (accessed 2 January 2023).

[83] H.J. Klimisch, M. Andreae, U. Tillmann, A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data, Regul. Toxicol. Pharm. 25 (1997) 1–5.

[84] C.L. Galli, N.J. Walker, N.H. Oberlies, A.L. Roe, J. Edwards, S. Fitzpatrick, J. C. Griffiths, A.W. Hayes, C. Mahony, D.S. Marsmand, L. Keeffe, Development of a consensus approach for botanical safety evaluation – A roundtable report, Toxicol. Lett. 314 (2019) 10–17.

[85] J.J. Kellogg, M.F. Paine, J.S. McCune, N.H. Oberlies, N.B. Cech, Selection and characterization of botanical natural products for research studies: A NaPDI Center recommended approach, Nat. Prod. Rep. 36 (2019) 1196–1221.

[86] A.M. Api, D. Belsito, D. Botelhoa, M. Bruze, G.A. Burton Jr, J. Buschmanne, M. L. Dagli, M. Date, W. Dekant, C. Deodhar, M. Francis, A.D. Fryer, L. Jones, K. Joshi, S. La Cava, A. Lapczynski, D.C. Liebler, D. O'Brien, A. Patel, T. M. Penning, G. Ritacco, J. Romine, S. Sadekara, D. Salvitoa, T.W. Schultzk, I. G. Sipes, G. Sullivan, Y. Thakkar, Y. Tokura, S. Tsang, RIFM fragrance ingredient safety assessment, p-mentha-1,8-dien-7-ol, CAS Registry Number 536–59-4, Food Chem. Toxicol. 122 (2018) S342–S347.

[87] S. Ebadollahi, M. Ziaee, F. Palla, Essential oils extracted from different species of the *Lamiaceae* plant family as prospective bioagents against several detrimental pests, Molecules 25 (2020) 1556.

[88] A.O. Durco, L.S.R. Conceiçao, D.S. de Souza, C.A. Lima, J. de Souza, S. Quintans, M.R. Viana dos Santos, Perillyl alcohol as a treatment for cancer: A systematic review, Phytomedicine Plus 1 (2021), 100090.

[89] U.T. Dat, T.K.N. Tran, T.H. Tran, Q.L. Pham, T.T.T. Dinh, Q.T. Tran, Volatile composition of *Perilla frutescens* (L.) essential oil in Thai Binh Province, Vietnam extracted by microwave-assisted distillation method Sentences, IOP Conf. Series, Mater. Sci. Eng. 1092 (2021), 012093.

[90] C. You, Y. Wang, W. Zhang, K. Yang, Y. Wu, C.H. Geng, H. Jiang, S. Du, Z. Deng, Z. Liu, Chemical constituents and biological activities of the Purple Perilla essential oil against *Lasioderma serricorne*, Ind. Crop. Prod. 61 (2014) 331–337.

[91] B.K. Ghimire, C.Y. Yu, I.M. Chung, Assessment of the phenolic profile, antimicrobial activity and oxidative stability of transgenic *Perilla frutescens* L. overexpressing tocopherol methyltransferase (γ-tmt) gene, Plant Physiol. Biochem. 118 (2017) 77–87.

[92] W.H. Seo, H.H. Baek, Characteristic aroma-active compounds of Korean perilla (*Perilla frutescens* Britton) leaf, J. Agric. Food Chem. 57 (2009) 11537–11542.

[93] L. Bumblauskienė, V. Jakstas, V. Janulis, R. Mazdzierienė, O. Ragazinskienė, Preliminary analysis on essential oil composition of *Perilla* L. cultivated in Lithuania, Acta Poloniae Pharmaceutica - Drug Research 66 (2009) 409–413.

[94] N. Tabanca, B. Demirci, A. Ali, Z. Ali, E.K. Blythe, I.A. Khan, Essential oils of green and red *Perilla frutescens* as potential sources of compounds for mosquito management, Ind. Crop. Prod. 65 (2015) 36–44.

[95] T.B. Adams, J.B. Hallagan, J.M. Putnam, T.L. Gierke, J. Doull, I.C. Munro, P. Newberne, P.S. Portoghese, R.L. Smith, B.M. Wagner, C.S. Well, L.A. Woods, R. A. Rord, The FEMA GRAS assessment of alicyclic substances used as flavour ingredients, Food Chem. Toxicol. 34 (1996) 763–828.

[96] Z. O'Brien, Pharmacokinetics, *in vitro* absorption and metabolism of perillyl alcohol a chemopreventive and chemotherapeutic agent, The Ohio State University USA, 2005. PhD Dissertation.

[97] Api AM, Belsito D,Bhatia S, Botelhoa D, Browne D, M. Bruze M, Burton Jr A, Buschmanne J, Calow P, Dagli ML, Date M, Dekant W, Deodhar C, Fryer AD, Joshi K, L. Kromidas L, La Cava S, Lalko JF, Lapczynski A, Liebler DC, Miyachi Y, O'Brien D, Parakhia R, Patel A, Penning TM, Politano VT, Ritacco G, Romine J, Salvito D, Schultz TW, Shen J, Sipes JG, Thakkar Y, Tokura Y, Tsang S, Wahler J, Wall B, Wilcox DJ (2017) RIFM fragrance ingredient safety assessment, terpineol, CAS Registry Number 8000-41-7. Food and Chemical Toxicology 110: S392-S402.

[98] A.M. Api, D. Belsito, B.M. Botelho, G.A. Burton Jr., J. Buschmann, M. A. Cancellieri, M.L. Dagli, M. Date, W. Dekant, C. Deodhar, A.D. Fryer, L. Jones, K. Joshi, M. Kumar, A. Lapczynski, M. Lavelle, I. Lee, D.C. Liebler, H. Moustakas, M. Na, T.M. Penning, G. Ritacco, J. Romine, N. Sadekar, T.W. Schultz, D. Selechnik, F. Siddiqi, I.G. Sipes, G. Sullivan, Y. Thakkar, Y. Tokura, RIFM fragrance ingredient safety assessment, dl-limonene (racemic), CAS Registry Number 138–86-3, Food Chem. Toxicol. 161 (2022), 112764.

[99] G. Ates, F.P. Steinmetz, T.Y. Doktorova, J.C. Madden, V. Rogiers, Linking existing in vitro dermal absorption data to physicochemical properties: Contribution to the design of a weight-of-evidence approach for the safety evaluation of cosmetic ingredients with low dermal bioavailability, Regul. Toxicol. Pharm. 76 (2016) 74–78.

[100] J. Shen, L. Kromidas, T. Schultz, S. Bhatia, An *in silico* skin absorption model for fragrance materials, Food Chem. Toxicol. 74 (2014) 164–176.

[101] L.G. Lucca, S.P. de Matos, B.T. Borille, D. de O. Dias, H.F. Teixeira, V.F. Veiga Jr., R.P. Limberger, L.S. Koester, Determination of β-caryophyllene skin permeation/retention from crude copaiba oil (*Copaifera multijuga* Hayne) and respective oil-based nanoemulsion using a novel HS-GC/MS method, J. Pharm. Biomed. Anal. 104 (2015) 144–148.

[102] EFSA 2020. Safety and efficacy of essential oil, oleoresin and tincture from *Zingiber officinale* Roscoe when used as sensory additives in feed for all animal species. EFSA Journal 2020;18(6):6147.

[103] EFSA 2017. Scientific Opinion on Flavouring Group Evaluation 208 Revision 2 (FGE.208Rev2): Consideration of genotoxicity data on alicyclic aldehydes with α,β-unsaturation in ring/side-chain and precursors from chemical subgroup 2.2 of FGE.19. EFSA Journal 2017;15(5):4766.

[104] B. Zou, Q. Li, J. Zhao, J.M. Li, C.F. Cuff, E. Reed, β-Elemene and taxanes synergistically induce cytotoxicity and inhibit proliferation in ovarian cancer and other tumor cells, Anticancer Res. 33 (2013) 929–940.

[105] EFSA 2015. Scientific Opinion on Flavouring Group Evaluation 78, Revision 2 (FGE.78Rev2): Consideration of aliphatic and alicyclic and aromatic hydrocarbons evaluated by JECFA (63rd meeting) structurally related to aliphatic hydrocarbons evaluated by EFSA in FGE.25Rev3. EFSA Journal 2015;13(4):4067.

[106] JECFA 2003. Joint FAO/WHO Expert Committee on Food Additives (JECFA). Speijers GJA, Renwick A. Alicyclic Primary Alcohols, Aldehydes, Acids and Related Esters. WHO Food Additives Series: 50.

[107] JECFA 2005. Joint FAO/WHO Expert Committee on Food Additives (2005) Compendium of food additive specifications. Addendum 12. Joint FAO/WHO Expert Committee of Food Additives 63rd session. Rome, 8-17 June 2004. FAO Food and Nutrition paper 52 Add. 12.

[108] S. Balaji, B. Chempakam, Toxicity prediction of compounds from turmeric (*Curcuma longa* L), Food Chem. Toxicol. 48 (2010) 2951–2959.

[109] OECD. OECD QSAR Toolbox ver 4.5. https://qsartoolbox.org/ (accessed 2 January 2023).

[110] T.K. Celik, B. Togar, H. Turkez, N. Taspinar, *In vitro* cytotoxic, genotoxic, and oxidative effects of acyclic sesquiterpene farnesene, Turk. J. Biol. 38 (2014) 253–259.

[111] ECHA 2017. Dossier 19523/7/7/2. https://echa.europa.eu/registration-dossier/-/registered-dossier/19523/7/7/2 (accessed 2 January 2023).

[112] ECHA 2021. Dossier 10490/7/6/1. https://echa.europa.eu/registration-dossier/-/registered-dossier/10490/7/5/2 (accessed 2 January 2023).

[113] A.M. Api, F. Belmonte, D. Belsito, S. Biserta, D. Botelho, M. Bruze, G.A. Burton Jr., J. Buschmann, M.A. Cancellieri, M.L. Dagli, M. Date, W. Dekant, C. Deodhar, A. D. Fryer, S. Gadhia, L. Jones, K. Joshi, A. Lapczynski, M. Lavelle, D.C. Liebler, M. Na, D. O'Brien, A. Patel, T.M. Penning, G. Ritacco, R.J. Rodriguez-Ropero, N. Sadekar, D. Salvito, T.W. Schultz, I.G. Sipes, G. Sullivan, Y. Thakkar, Y. Tokuram, S. Tsang, RIFM fragrance ingredient safety assessment, caryophyllene oxide, CAS Registry Number 1139–30-6, Food Chem. Toxicol. 138 (Suppl 1) (2020), 111102.

[114] A. Di Sotto, F. Maffei, P. Hrelia, F. Castelli, M.G. Sarpietro, G. Mazzanti, Genotoxicity assessment of β-caryophyllene oxide, Regul. Toxicol. Pharm. 66 (2013) 264–268.

[115] SCCS 2011. SCCS Opinion on Fragrance Allergens in Cosmetic Products. 13-14 Dec 2011.

[116] US CIR. United States Cosmetic Ingredient Review (CIR). https://www.cir-safety.org/ingredients (accessed 2 January 2023).

[117] ECHA 2019a. Dossier 1931/7/6/2. https://echa.europa.eu/registration-dossier/-/registered-dossier/1931/7/6/2.

[118] A.M. Api, D. Belsito, D. Botelhoa, M. Bruze, P. Calow, M.L. Dagli, W. Dekant, A. D. Fryer, K.L.S. La Cava, J.F. Lalko, A. Lapczynski, D.C. Liebler, Y. Miyachi, V. T. Politano, G. Ritacco, D. Salvito, J. Shen, T.W. Schultz, I.G. Sipes, B. Wall, D. K. Wilcox, RIFM fragrance ingredient safety assessment, Linalool, CAS registry number 78–70-6, Food Chem. Toxicol. 82 (2015) S29–S38.

[119] A.M. Api, D. Belsito, S. Bhatia, M. Bruze, P. Calow, M.L. Dagli, W. Dekant, A. D. Fryer, L. Kromidas, S. La Cava, J.F. Lalko, A. Lapczynski, D.C. Liebler, Y. Miyachi, V.T. Politano, G. Ritacco, D. Salvito, T.W. Schultz, J. Shen, I.G. Sipes, B. Wall, D.K. Wilcox, RIFM fragrance ingredient safety assessment, 3,7-dimethyl-1,6-nonadien-3-ol, CAS Registry Number 10339–55-6, Food Chem. Toxicol. 97 (2016) S168–S179.

[120] ECHA 2019. Dossier 1931/7/7/3. https://echa.europa.eu/registration-dossier/-/registered-dossier/1931/7/7/3 (accessed 2 January 2023).

[121] RIFM 1977. Report on Human Maximization Studies. Report to RIFM. RIFM report number 1702. RIFM, Woodcliff Lake, NJ, USA.

[122] Klecak G (1979) The open epicutaneous test (OET), a predictive test procedure in the guinea pig for estimation of allergenic properties of simple chemical compounds, their mixtures and of finished cosmetic preparations. International Federation Society Cosmetics Chemists 9/18/79.

[123] A.M. Api, D. Belsito, S. Biserta, D. Botelho, M. Bruze, G.A. Burton Jr., J. Buschmann, M.A. Cancellieri, M.L. Dagli, M. Date, W. Dekant, C. Deodhar, A. D. Fryer, S. Gadhia, L. Jones, K. Joshi, A. Lapczynski, M. Lavelle, D.C. Liebler, M. Na, D. O'Brien, A. Patel, T.M. Penning, G. Ritacco, F. Rodriguez-Ropero, J. Romine, N. Sadekar, D. Salvito, T.W. Schultz, F. Siddiqi, I.G. Sipes, G. Sullivan, Y. Thakkar, Y. Tokura, S. Tsang, RIFM fragrance ingredient safety assessment, 2,4,6-trimethyl-3-cyclohexene-1-methanol, CAS Registry Number 68527–77-5, Food Chem. Toxicol. 134 (2019), 110951.

[124] P. Puchadapirom, M. Suttajit, S. Thongpraditchote, W. Kitphati, A. Tammasakchai, Genotoxicity evaluation of ethanolic leaf extract of Thai Perilla (*Perilla frutescens* (L.) Britt.) using Micronucleus Assay in V79 cell line, Thai J. Pharmacol. 42 (2020) 5–16.

[125] C. Klaassen, Casarett and Doull's Toxicology: The Basic Science of Poisons, seventh ed., McGraw-Hill, New York NY, 2008.

[126] H. Yu, J.-F. Qiu, L.-J. Ma, Y.-J. Hu, P. Li, J.-B. Wan, Phytochemical and phytopharmacological review of *Perilla frutescens* L. (Labiatae), a traditional edible-medicinal herb in China, Food Chem. Toxicol. 108 (2017) 375–391.
[127] H.M. Ahmed, Ethnomedicinal, phytochemical and pharmacological investigations of *Perilla frutescens* (L.), Britt. Molecules 24 (2018) 102.
[128] J.-H. Nho, H.-J. Lee, S.-Y. Lee, J.-H. Jang, B.-D. Yang, J.-H. Jeong, G.-Y. Lee, H.-W. Cho, J.-C. Kim, H.-K. Jung, A 13-week repeated oral dose toxicity evaluation and a 4-week recovery evaluation of the Sam So Eum (SSE) in male and female rats, J. Ethnopharmacol. 260 (2020), 112988.
[129] T. Kanzaki, S. Kimura, Occupational allergic contact dermatitis from *Perilla frutescens* (shiso), Contact Dermatitis 26 (1992) 55–56.
[130] SCCS, 2018. The SCCS Notes of Guidance for the Testing of Cosmetic Ingredients and Their Safety Evaluation. 10th revision. 2018.
[131] R.J. Safford, A.M. Api, D.W. Roberts, J.F. Lalko, Extension of the dermal sensitization threshold (DST) approach to incorporate chemicals classified as reactive, Regul. Toxicol. Pharm. 72 (2015) 694–701.