Food and Chemical Toxicology 132 (2019) 110718

Contents lists available at ScienceDirect

Food and Chemical Toxicology

journal homepage: www.elsevier.com/locate/foodchemtox

Profiling 58 compounds including cosmetic-relevant chemicals using ToxRefDB and ToxCast

Ly L. Pham[a,b], Lisa Truong[a,b,c], Gladys Ouedraogo[d], Sophie Loisel-Joubert[e], Matthew T. Martin[a,f], Katie Paul Friedman[a,*]

[a] National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, 27711, USA
[b] ORISE Postdoctoral Research Participant, USA
[c] Currently at Oregon State University, Department of Environmental Toxicology, Corvallis, OR, 97333, USA
[d] L'Oréal Safety Research & Innovation, 1 Avenue E. Schueller, 93600, Aulnay-sous-Bois, France
[e] L'Oréal Safety Research & Innovation, 9 Rue Pierre Dreyfus, 92110, Clichy, France
[f] Currently at Global Investigative Toxicology, Drug Safety Research and Development, Pfizer Inc, 445 Eastern Point Road, Groton, CT, 06340, USA

ARTICLE INFO

Keywords:
High-throughput screening
ToxCast
Cosmetics-relevant chemicals
ToxRefDB

ABSTRACT

Safety assessment for cosmetic-relevant chemicals (CRCs) in the European Union has been reshaped by restrictions on animal testing, and new approach methodologies (NAMs) for predicting toxicity are critical to ensure new cosmetic product safety. To demonstrate NAMs for safety assessment, we surveyed *in vitro* bioactivity and *in vivo* systemic toxicity data in the US Environmental Protection Agency's (EPA's) Toxicity Forecaster (ToxCast) and Toxicity Reference databases (ToxRefDB), respectively, for 58 chemicals identified as CRCs, including cosmetic ingredients as well as trace contaminants. CRCs were diverse in use types as suggested by broad chemical use categories. In terms of both target organ effects and study type, the median of the lowest effect level (LEL) doses in ToxRefDB for CRCs tended to be slightly higher than the median for the remaining 928 chemicals with study data in ToxRefDB, though the ranges of LELs were similar. For 17 of the 58 CRCs, high-throughput toxicokinetic data were used to calculate administered equivalent doses (AEDs) in mg/kg/day units for the *in vitro* bioactivity observed, and these AEDs served as conservative estimators of the systemic LELs observed *in vivo*. This work suggests that NAMs for bioactivity may inform a conservative point-of-departure estimate for diverse CRCs.

1. Introduction

In the last 15 years, there has been a major shift in the European Union (EU) regulatory toxicology community regarding cosmetic-relevant chemicals (CRCs), from requiring *in vivo* toxicity studies to a complete ban on *in vivo* studies for defining hazard. In 2009, a ban on animal testing of finished cosmetic products was extended to ingredients or combinations of ingredients, and was accompanied by a marketing ban for cosmetic products tested on animals (EU Regulation and Union, 2009). The deadline for phasing out all *in vivo* testing was set for March 2013, irrespective of the maturity of new approach methodologies (NAMs) for use (ECHA, 2016). A systematic review of available NAMs for toxicokinetic, repeat dose toxicity, carcinogenicity, skin sensitization, and reproductive toxicity demonstrated great advancement of *in vitro* models and computational tools to assess the hazard; however, they also identified substantial gaps that prevent full replacement of *in vivo* studies (Adler et al., 2011; Hartung et al., 2011). Thus, continuing efforts focus on developing and evaluating NAMs, e.g. *in vitro* and *in silico* predictive models. The main objectives of the work herein are: one, to provide a reproducible workflow for using NAMs, including *in vitro* screening assays and *in silico* predictions, to derive an estimate of a dose that might correspond to bioactivity *in vivo* for a set of CRCs; and, two, to compare these NAM-based approaches for screening level assessment to effect levels from traditional *in vivo* studies, with the intention of building confidence and examining potential limitations.

Indeed, there are several ongoing efforts to generate and evaluate NAM data for safety assessment. European initiatives such as the SEURAT1 project (2016), EU-ToxRisk (Daneshian et al., 2016), and the Cosmetics Europe Long Range Science Strategy program (Desprez et al., 2018) are a few examples of such efforts. However, the regulatory needs for CRCs in Europe are also supported by other ongoing high-

* Corresponding author. 109 T.W. Alexander Drive, Mail Drop D143-02, Research Triangle Park, NC, 27711, USA.
E-mail address: paul-friedman.katie@epa.gov (K. Paul Friedman).

https://doi.org/10.1016/j.fct.2019.110718
Received 8 March 2019; Received in revised form 17 July 2019; Accepted 26 July 2019
Available online 26 July 2019
0278-6915/ Published by Elsevier Ltd.

throughput screening (HTS) initiatives in the United States, with similar goals to rapidly evaluate many chemicals for bioactivity of interest, such as the Toxicity Testing in the 21st century (Tox21) (Thomas et al., 2018) and the U.S. Environmental Protection Agency (EPA) Toxicity Forecaster (ToxCast) programs (Kavlock and Dix, 2010). In 2007, the US National Research Council issued a challenge to transition away from animal models of toxicity and to develop and use high-throughput approaches to inform prediction of human-relevant toxicity (Council, 2007; National Academies of Sciences and Medicine, 2017). In response, the Tox21 (https://ncats.nih.gov/tox21/) partnership between the National Center for Advancing Translational Sciences, the National Toxicology Program at the National Institute of Environmental Health Sciences, the EPA, and the Food and Drug Administration was established to develop high-throughput technologies and computational models for predicting adverse human health outcomes (Attene-Ramos et al., 2013; Thomas et al., 2018; Tice et al., 2013). The US EPA ToxCast program utilizes a diverse set of *in vitro* HTS assays with the goal to develop approaches to prioritize chemicals with little to no hazard safety information (Judson et al., 2016; Kavlock and Dix, 2010). In Version 2, the ToxCast database termed "invitrodb" structures data for over 1000 assay endpoints, including assay data derived from the Tox21 partnership. Though the number of chemicals screened in each assay endpoint is variable, the total ToxCast chemical library contains 1860 unique chemicals structures, and the Tox21 chemical library approaches 10,000 chemicals (Richard, 2016). Thus, the ToxCast and Tox21 projects comprise a large data resource for advancing the use of HTS for assessment of chemical bioactivity, including cosmetics.

A parallel resource for evaluating the performance of ToxCast/Tox21 is legacy *in vivo* data from animal toxicity studies, stored in the Toxicity Reference Database, or ToxRefDB. ToxRefDB is a large public resource of highly curated, repeat-dose animal toxicity studies and is currently available as ToxRefDB v1 (Martin et al., 2009a, 2009c). The database structures information for 1144 chemicals and 5890 studies, including details on study design features and the study outcome such as chemical identity, treatment group parameters, standardized effect vocabulary, and treatment-related effect levels, i.e., the dose at which an effect was significantly different from control. ToxRefDBv1 provides a foundational resource to examine *in vivo* toxicity observed in multiple study types, including repeat dose study types such as subacute, multigenerational, and chronic designs, and potentially in multiple species (e.g., rat, mouse, and/or dog).

To evaluate predictive tools for assessment of cosmetic-relevant chemicals (CRCs), we focused this work on chemicals used as cosmetic ingredients as well as chemicals that are unintentional trace contaminants. We surveyed information available in the EPA ToxCast (invitrodb version 2) and ToxRefDB (version 1) databases to evaluate the potential *in vitro* and *in vivo* activity of 58 chemicals that were identified as either ingredients or contaminates in personal care products, fragrances, and/or cosmetics based on information from the Personal Care Products Council (PCPC) (Boyer et al., 2017). These 58 chemicals are diverse in their structure and use types. Systemic effect information (effects on target organs, clinical chemistry, or in-life observations such as body weight in adult animals) from ToxRefDBv1 for these 58 chemicals was evaluated to identify the most prevalent target organs. High hit rates within an organ system may indicate a priority for developing *in vitro* tests to screen for early indicators of toxicity. To identify potentially selective bioactivity, i.e., bioactivity independent of cytotoxicity, the range of 50% bioactivity concentrations, or AC50 values, and the active hit rates were collected from the HTS assays in ToxCast and compared to the ToxCast-based estimates of cytotoxicity (Judson et al., 2016) for the CRCs. The premise of this *in vitro* to *in vivo* comparison is that the concentration range in which any *in vitro* bioactivity is observed may be a reasonable starting point for initial estimation of a range of doses that represent a threshold for *in vivo* effects and a bioactivity:exposure ratio (Thomas et al., 2013; Wambaugh et al., 2018; Wetmore, 2015). The intention of performing this retrospective survey for CRCs is to demonstrate the *in vitro* and *in vivo* activities of these chemicals using publicly-available tools and to further demonstrate how NAM-based information from ToxCast may inform conservative estimates of systemic points-of-departure for these chemicals.

2. Methods

2.1. Selection of the chemical list

The 58 CRCs were manually identifed through expert review of: ToxRefDBv1 (Martin et al., 2009b; Martin et al., 2009c) to identify cosmetically relevant chemicals or trace contaminates; chemical status designated by the Personal Care Products Council (PCPC) (Boyer et al., 2017); and, overall availability of information in ToxCast and ToxRefDBv1. Data from ToxRefDBv1(USEPA, 2014) were considered for studies with oral administration to adult animals that were considered of acceptable quality for data extraction purposes during ToxRefDB development. All data from ToxCast (invitrodb_v2; https://www.epa.gov/chemical-research/exploring-toxcast-data-downloadable-data) were considered.

2.2. Use categories

CRCs were mapped to one or more broad use categories, based on previous efforts to consolidate use information (Wambaugh et al., 2014) available on the EPA's Aggregated Computational Toxicology Online Resource (ACToR) (Judson et al., 2008). Chemicals can have a wide variety of associated uses and therefore can be associated with multiple categories (Fig. 1). This analysis was intended to illustrate functional use diversity for the CRCs in this work.

2.3. In vivo data source

The Toxicity Reference Database (ToxRefDBv1; ftp://newftp.epa.gov/comptox/High_Throughput_Screening_Data/Animal_Tox_Data/) (Martin et al., 2009a) was used as a source of legacy *in vivo* data. ToxRefDBv1 contains data largely from studies conducted in accordance with or by specifications similar to the EPA Office of Chemical Safety and Pollution Prevention (OCSPP) 870 series Health Effects Test Guidelines, with over 5000 studies (guideline, guideline-like, and other) associated with 1144 chemicals. The single largest source of these study data were extracted from reviews of registrant-submitted toxicity studies, known as data evaluation records (DERs), from the US EPA's Office of Pesticide Programs (OPP) within OCSPP, with additional study data obtained from studies available from the National Toxicology Program, the pharmaceutical industry, and the publicly-available literature (Martin et al., 2009b; USEPA, 2014). The database includes information regarding the study design, chemical identity, treatment group parameters, and treatment-related effect levels, i.e. the dose at which an effect was significantly different from control (Martin et al., 2009b; Martin et al., 2009a). To prepare the data for use in this analysis, "data usability", "data entry level", and "data entry status" were required to be "acceptable", "all effects", and "complete," respectively, which were simply cautions on the data extraction process and not the content of the studies themselves. These data quality filters are commonly applied to data from ToxRefDBv1 for use in modeling applications. Only studies that examined systemic effects in adult animals (in-life observations including body weight and clinical signs, clinical chemistry, organ weight, and/or gross or microscopic changes in tissues from adult or parental animals) following oral routes of administration were used. Animal species included in this analysis were dogs, mice, rabbits, and rats. All lowest effect levels (LELs) from ToxRefDBv1, or the lowest doses at which significant treatment-related effects on organs or in-life observations were observed *in vivo*, were $\log_{10}$ transformed. Additionally, organ level LELs were defined in this work by the lowest dose at which a treatment-related effect was

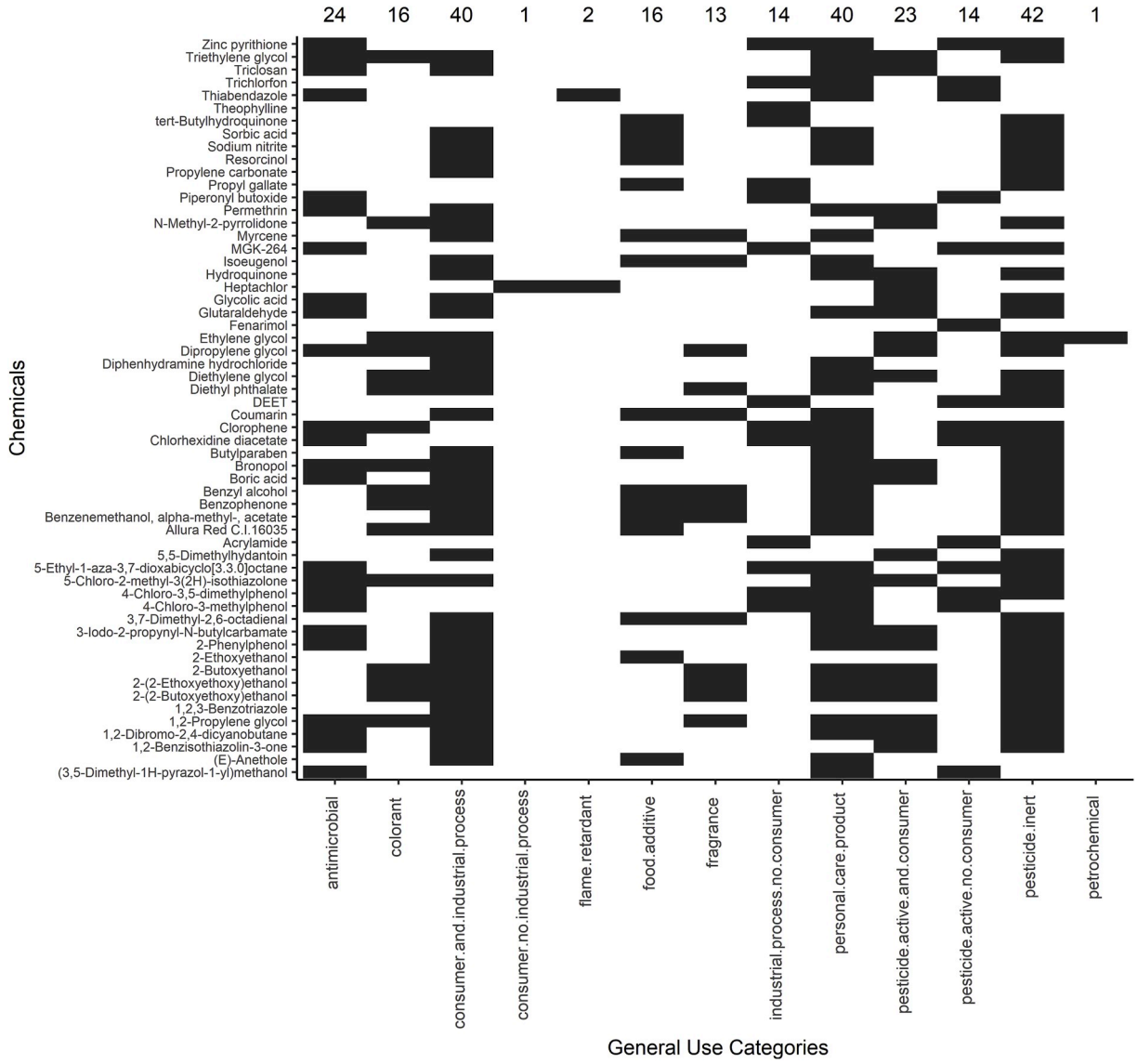

Fig. 1. Binary Use Information for CRCs with ACToR data. The 58 CRCs were evaluated for affiliated use terms in ACToR, revealing that each chemical can be in multiple general use categories. Number line at the top of the figure indicates how many chemicals are in each category.

observed for that organ (including changes in weight or gross or microscopic pathology). In total, there are 928 chemicals and 4157 studies with data that passed these filters for ToxRefDBv1, of which 58 were CRCs associated with 295 studies. All of the ToxRefDBv1 data used in this analysis are included as an R data file (.RData) (Supplemental File 3).

2.4. Comparison of the in vivo effects for CRCs versus the ToxRefDBv1 library

Three analyses were performed to provide a perspective on the potential *in vivo* hazard of CRCs with available information in ToxRefDBv1: (1) a qualitative review of the incidence of systemic, organ-level endpoint effects observed in adult animals for CRCs; (2) a quantitative comparison of the distributions of the CRC set versus the remaining ToxRefDBv1 chemical library LEL values for target organs that affected by at least 10 CRC chemicals; and, (3) a quantitative comparison of the distributions of LELs from the CRC set and the ToxRefDBv1 chemical library by study type. LEL (mg/kg/day) values were all adjusted to a human eqivalent dose (HED) using allometric scaling, per the following set of equations by species (Nair and Jacob, 2016):

$$HED_m = Dose_{mouse} * 0.081 \quad (1a)$$

$$HED_r = Dose_{rat} * 0.162 \quad (1b)$$

$$HED_{rb} = Dose_{rabbit} * 0.324 \quad (1c)$$

$$HED_d = Dose_{dog} * 0.541 \quad (1d)$$

For the first analysis, the percentage of the 58 cosmetic chemicals that affected each target organ, including positive findings on organ weight, gross pathology, and/or microscopic pathology, was compared to the percentage of the remaining ToxRefDBv1 chemical inventory that affected that organ level endpoint target (870 chemicals). Not all 97 organs with available data in ToxRefDBv1 were affected by CRCs (Fig. 2). For the second analysis, distributions of the LEL values for 14 key organ-level endpoint targets (liver, kidney, stomach, spleen, testes, uterus, thymus, adrenal gland, lung, brain, heart, bone marrow, lymph node, and thyroid gland) for CRCs were compared to the distributions of LEL values for the remaining ToxRefDBv1 inventory. These 14 organs were selected because they demonstrated the highest incidence rates of

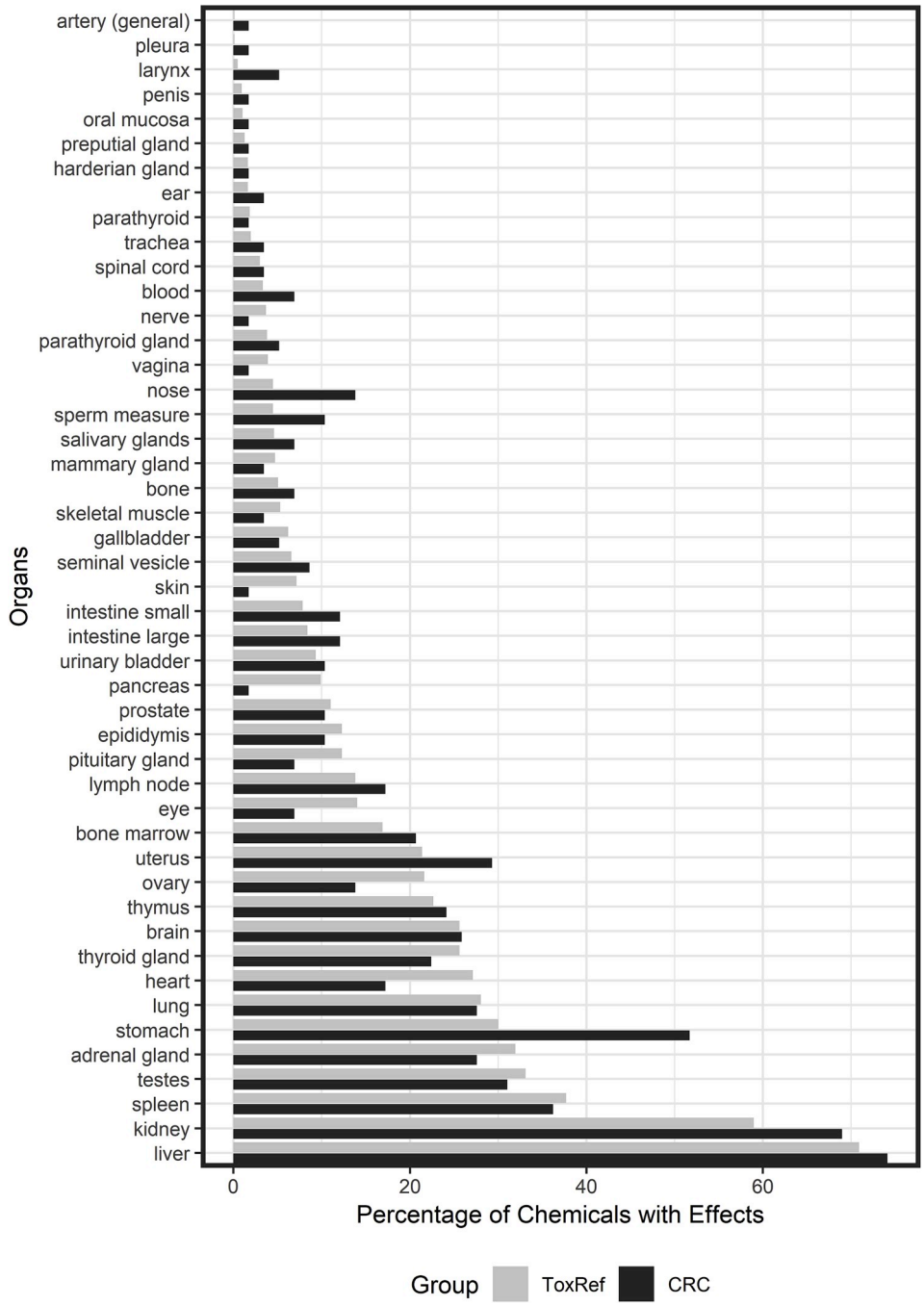

Fig. 2. Target Organs Affected by CRCs. Histogram displaying the percentage of cosmetic (58 total) and ToxRefDBv1 (870 total) chemicals affecting each of the target organs.

any findings, including weight, gross pathology, and/or microscopic pathology changes, for CRCs. All of the study types with repeat-dose systemic toxicity information available in ToxRefDBv1, including sub-acute (SAC), subchronic (SUB), chronic (CHR), developmental (for F0 generation only, DEV), developmental neurotoxicity (for F0 generation only, DNT), and multigenerational reproductive (for F0 generation only, MGR) were used for the study type comparison. In the third analysis, distributions of the LEL values for the CRCs and the remaining ToxRefDBv1 chemical library were compared for the six study types used in the analysis herein.

A non-parametric Wilcoxon-Mann-Whitney test (wilcox.test in the R stats library) was performed to address whether the means of the LEL values for the CRC and ToxRefDB chemical sets were different for each of the 14 target organs and each of the six study types (see Supplemental Files 1 and 2).

2.5. *In vitro* data source

The ToxCast *in vitro* database (invitrodbv2, October 2015) was used as a source for HTS bioactivity data. A key feature of the ToxCast program is public access to the chemical and screening data, available in user interfaces at the CompTox Chemicals Dashboard (Williams et al., 2017) and as downloads (https://www.epa.gov/chemical-research/exploring-toxcast-data-downloadable-data). Invitrodbv2 stores all of the curve-fit information generated using the ToxCast Data Pipeline (tcpl), which has been made publicly available as an R package

(tcpl, CRAN (Filer et al., 2017);). Curve-fitting was performed using tcpl, and the summary information including the 50% activity concentration (AC50) for any assay endpoint is stored in invitrodbv2. For each of the 58 CRCs, the AC50 values and cytotoxicity burst information were retrieved. This distribution of AC50 values, on a $\log_{10}$ scale, was then compared to the *in vitro* concentration anticipated to result in cytotoxicity based on results in a battery of 35 cytotoxicity and cell stress assays, referred to as the "burst" value (Judson et al., 2016). The active hit rate (%) per chemical was calculated to indicate the percent of assay endpoints that were positive out of the total number of screened assay endpoints.

2.6. In vitro to in vivo comparison using reverse dosimetry

To facilitate comparison of AC50 values (in micromolar concentration units) to *in vivo* effect levels from animal studies (in mg/kg/day units), AC50 values were converted from micromolar units to human-based administered equivalent doses (AEDs) using the "calc_mc_oral_equiv" function in the httk R package version 1.8 (Pearce et al., 2017a, 2017b). The httk, or high-throughput toxicokinetics, R package utilizes model assumptions and experimental *in vitro* data for protein binding and hepatic metabolic clearance to develop estimates of the AED required to obtain the observed concentration *in vivo* assuming steady-state kinetics. Within the httk package, the three-compartment model with restrictive clearance and well stirred correction was used. In addition, population variability was stimulated using a Monte Carlo simulation and the 50th percentile was used. Only 17 of the 58 CRCs had publicly-available high-throughput toxicokinetic information to inform estimation of an AED. Of the 870 ToxRefDBv1 non-CRC chemicals, 305 had ToxCast and publicly-available httk data available for comparison.

2.7. Comparison of In vitro and in vivo LELs to exposure and TTC values

The previously estimated exposure (Wambaugh et al., 2014) and thresholds of toxicological concerns (TTC) specifically for cosmetics (Yang et al., 2017), were also compared for the 17 CRCs with AEDs in order to understand if NAM-based approaches could produce conservative point-of-departure estimates for CRCs. Wambaugh et al. estimated the human exposure of 7968 chemicals using five chemical descriptors and inferred ranges of exposure from a subset of chemical as measured by the U.S. National Health and Human Examination Survey (NHANES). The estimated median exposure for the total US population was used in this study. TTC values, as pioneered by Munro et al. (1996), are a threshold for acceptable human exposure based on existing knowledge of point-of-departure values for a set of known chemicals, and then, based on shared chemical descriptors, these TTC estimates are assigned for chemicals that lack toxicity data. Yang et al. expanded upon this work with the inclusion of more cosmetic-related chemicals (COSMOS dataset; http://www.cosmostox.eu), the incorporation of Cramer class (Cramer et al., 1978), and ToxPrint chemotypes (Yang et al., 2015). The federated TTC dataset, as defined by Yang et al. (2017), was used in this study to designate Cramer Class and as a source of TTC values.

2.8. Software

All analysis was performed with open source R version 3.5.1 (R Development Core Team, 2016) and RStudio (Team, 2015). The R scripts are available at: ftp://newftp.epa.gov/comptox/NCCT_Publication_Data/Pham_LyLy/Cosmetics_Survey and supplied here as Supplemental File 2. Supplemental File 3 contains an RData file with all of the ToxRefDBv1 data used in this analysis; this is also available via the FTP address provided.

3. Results

The 58 CRCs were expected to demonstrate a range of bioactivity and *in vivo* effects in this survey, as these chemicals appeared diverse based on manual inspection. To provide additional context for reviewing the survey presented herein, the functional use types for the CRCs are presented using the publicly available, broad use categories developed by ACToR and used in exposure modeling (Wambaugh et al., 2014). The categories include: "antimicrobial", "colorant", "fertilizer", "food additive", "fragrance", "herbicide", "personal care product", "petrochemical", "pesticide inert", "flame retardant", "other", "industrial process no consumer", "consumer no industrial process", "consumer and industrial process", "pesticide active no consumer," and "pesticide active and consumer." The heatmap in Fig. 1 shows diverse usage for the 58 CRCs (13 out of 16 possible categories) and multiple uses for some ingredients. The chemical with the maximum number of associated use categories was 1,2-propylene glycol at seven, while both theophylline and fenarimol only had one. Among the ACToR use categories for the CRCs, the highest number of CRCs were categorized as "pesticide inert" (42), followed by "personal care product" (40), and "consumer and industrial process" (40). The use categories "consumer no industrial process" and "petrochemical" corresponded to only one chemical each. Diversity in bioactivity and/or toxicity profile between CRCs would be anticipated based on the diversity of ACToR use types represented in the CRC set.

Next, the *in vivo* systemic effect data available for the CRCs in ToxRefDBv1 were evaluated and compared to the remaining 928 chemicals in ToxRefDBv1 to understand the relative toxicity of CRCs to other chemicals with data in ToxRefDBv1. The null hypothesis was that the CRCs represented a subset of ToxRefDBv1, and that CRCs would not demonstrate different ranges of LELs by target organ or study type. Indeed, though the CRCs comprise a relatively small number of chemicals, the systemic effects of the CRCs span a similar range of target organs compared to the overall ToxRefDBv1 chemical inventory. The percentage of the CRCs and ToxRefDBv1 chemical library that affected each systemic, organ-level endpoint, including changes in organ weights, gross pathology, and microscopic pathology, were compared (Fig. 2). For 14 of the total 70 organs examined, CRCs and the overall ToxRefDBv1 inventory caused target organ effects at approximately the same rate, i.e., within 10% of each other. The largest difference in the observed organ effects was for stomach where the rate of positive findings for CRCs versus the overall ToxRefDBv1 library was 52% vs. 30%. Chemicals in ToxRefDBv1 affected up to 68 out of 97 organs, whereas the CRC affecting the most organs is 1,2-Dibromo-2,4-dicyanobutane at 46 organs affected. The counts of how many organs are affected for each CRC are available in Supplemental File 1. Count data, with mean and standard deviation of all chemicals and organ type, are also provided in Supplemental File 1 for the chemicals in ToxRefDBv1 and the CRC list, respectively. The results support the inference that CRCs and other chemicals in ToxRefDB appear to affect a range of target organs, and that CRCs do not appear to act as a unique class or set of chemicals in terms of the target organs that were affected.

The top 14 organs in terms of positive incidence for CRCs (from Fig. 2) were adrenal gland, bone marrow, brain, heart, kidney, liver, lung, lymph node, spleen, stomach, testes, thymus, thyroid gland, and uterus. For these 14 organs the distributions of LEL values (log10-mg/kg/day) for the CRCs and the ToxRefDBv1 chemical inventory were compared (Fig. 3). In all cases, the interquartile ranges for the CRCs overlapped with the ToxRefDBv1 interquartile ranges. When considering the LEL range by target organ, the median LEL values for the CRCs tended to be similar to slightly higher than the median LEL values for the remaining ToxRefDBv1 chemicals, with the exception of stomach. A non-parametric Wilcoxon-Mann-Whitney test was performed to examine the possibility of a significant difference between the mean LELs by organ for CRCs and the remaining ToxRefDBv1 chemicals. For 13 of the 14 organs (with stomach being the exception), significant p-

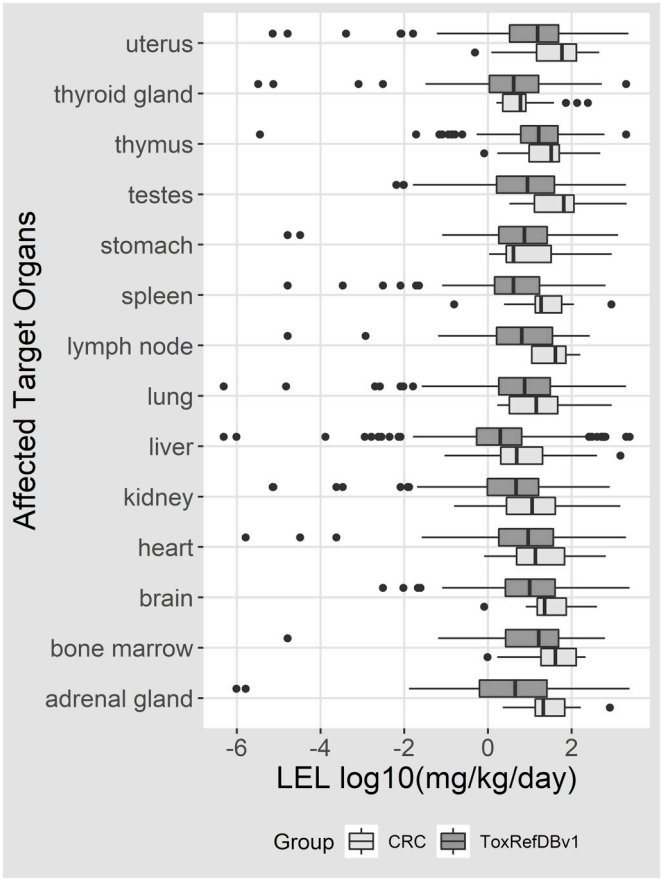

Fig. 3. Distribution of LEL values by Target Organ. Visualization of LEL distributions for each chemical set to illustrate the doses at which specific organs are impacted. The median LEL for CRCs tended to be higher than the median LEL from the remaining chemicals in ToxRefDBv1. This figure has been limited to the 14 target organs with the highest incidence of positive findings associated with CRCs.

values ($p < 0.05$) were obtained, suggesting that the mean LEL values for the CRC chemicals and remaining ToxRefDB chemicals are different. The p-values for the Wilcoxon-Mann-Whitney test were as follows: adrenal gland (< 0.00000001), bone marrow (0.0000004), brain (0.0000002), heart (0.005), kidney (< 0.00000001), liver (< 0.00000001), lung (0.003), lymph node (< 0.00000001), spleen (< 0.00000001), stomach (0.2), testes (< 0.00000001), thymus (0.00001), thyroid gland (0.0005), and uterus (< 0.00000001).

The analysis of target organ effects utilized LEL values from adult animals from five different study types (CHR, DEV, MGR, SAC, and SUB). To understand relative potency of CRCs in different study types, the distribution of LEL values (in log10-mg/kg/day) for the CRCs and the remaining ToxRefDBv1 chemicals were also compared by study type (Fig. 4). The number of chemicals used to generate each boxplot is printed to the right of each boxplot, as not all chemicals were evaluated in all study types. The total chemical count in this figure can be greater than 58 (CRC) or 870 (ToxRefDBv1) because some chemicals are studied in multiple study types or more than once in a study type. The median LELs for the CRCs were slightly greater than the median LELs for the remaining chemicals in ToxRefDBv1 when considering study type; however, with the exception of SAC, the medians do overlap with the interquartile range of the corresponding study type for all groups, and it is possible that the LEL distribution for SAC studies is skewed due to the limited number of CRCs (only 7) with associated SAC studies. The median LEL values in the three study types with the greatest number of CRCs (SUB, DEV, and CHR) (Fig. 4, number line) are within the interquartile range of the LELs for the rest of the chemicals in the

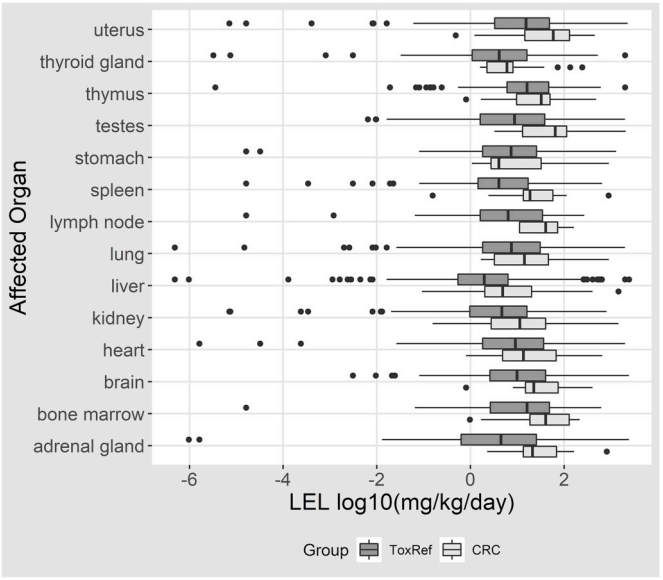

Fig. 4. Distribution of LEL values by Study Type. Overlay of LEL distributions for each chemical set to illustrate relative potency of effects in specific study types. The number on the far right indicates the number of chemicals used to generate the plot. in the boxplots demonstrate the interquartile range and median.

ToxRefDBv1 database. Similar to the analysis of LEL values by organ, a Wilcoxon-Mann-Whitney test was applied to address whether the mean LEL value might differ between the CRC and ToxRefDBv1 sets. Significant p-values were obtained for all study types except DNT; however, there were only 2 DNT studies available in the dataset for the CRCs, and so due to the very small sample it is not appropriate to use any statistical testing in this comparison. The p-values for the Wilcoxon-Mann-Whitney test were as follows: CHR (< 0.00000001), DEV (< 0.00000001), DNT (0.7), MGR (< 0.00000001), SAC (< 0.00000001), and SUB (< 0.00000001).

The CRCs were also evaluated across ToxCast HTS assay endpoints to assess overall bioactivity. Performance in the ToxCast HTS assays demonstrated that the CRCs are generally bioactive *in vitro* (Fig. 5), with hit-rates from less than 1% of the assays screened, e.g. boric acid, to 40% active, e.g. zinc pyrithione, a known cytotoxic agent (Fig. 5B). The AC50s across all ToxCast assays ranged from −5.3 to 2.8 log10-μM (5.01^{-6} to 631 μM), but the majority of the AC50s are between 0 and 2 log10-μM (1–100 μM) as shown in Fig. 5A. Total counts and hit rates are available in Supplementary File 1.

Previous work suggests that AC50 values near or greater than the cytotoxicity "burst" threshold for a given chemical may be attributable to non-specific effects secondary to cytotoxicity (Judson et al., 2016); therefore, the cytotoxic "burst" for each chemical is also indicated. In general, chemicals with cytotoxic activity appeared to demonstrate higher hit rates in the available ToxCast assays. However, the majority of CRCs (37 of 58 chemicals) failed to demonstrate cytotoxicity in the available assays and as such were not amenable to computing a chemical-specific cytotoxic "burst" concentration, rather they were assigned a default of 1000 μM, above the concentration range screened in ToxCast. For the 21 CRCs with computed "burst" cytotoxicity points, cytotoxicity occurred within 1–100 μM, most between 10 and 100 μM. Furthermore, for these 21 chemicals, the "burst" AC50 was less than or within the same concentration range as the majority of AC50s observed. *In vitro* effects observed within the same concentration range of the burst may have been confounded by cytotoxicity, whereas *in vitro* bioactivity at much lower concentrations than the burst may represent selective bioactivity (though other forms of assay and curve-fitting interference are still possible).

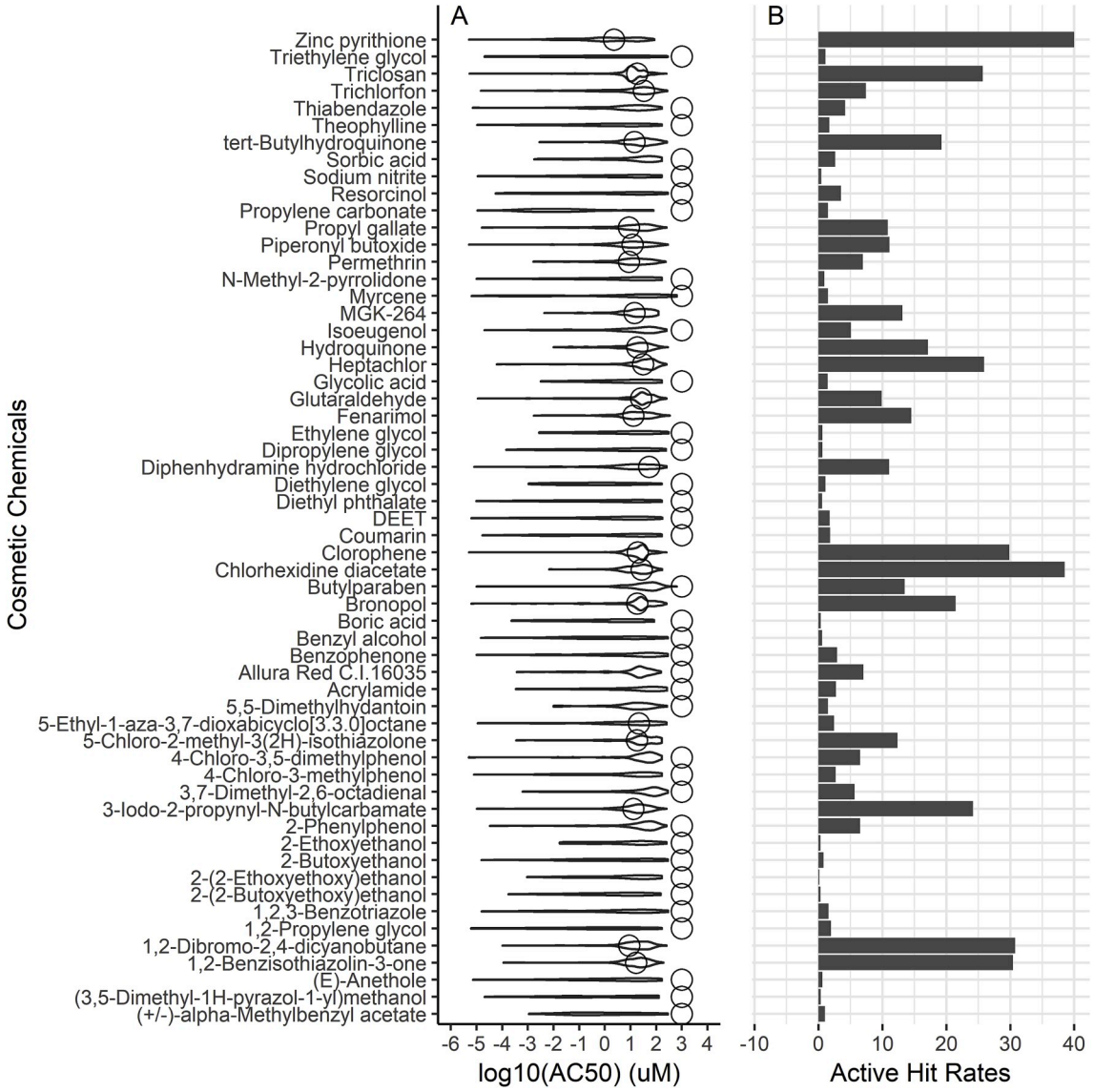

Fig. 5. Distribution of AC50 values for CRCs Across ToxCast *In Vitro* Assay Endpoints. A. Violin plots showing the range and density of AC50s for the 58 CRCs across all available ToxCast assay endpoints; circles indicate the chemical-specific cytotoxicity "burst" AC50 (Judson et al., 2016). **B.** Percent of active ToxCast assay endpoints per chemical.

The concept of reverse dosimetry was applied using high-throughput toxicokinetic (HTTK) information to convert the AC50 values (micromolar concentrations) to a human-based, administered equivalent dose (AED) values in mg/kg/day units to facilitate an *in vitro* to *in vivo* comparision. Only 17 of the 58 CRCs had the necessary, publicly-available toxicokinetic information to perform this calculation. The LELs from the *in vivo* studies were all alometrically scaled to human doses. The ranges of the AEDs from ToxCast and the *in vivo* systemic LELs from ToxRefDBv1 were compared (Fig. 6A). The AEDs and LELs failed to overlap for only four CRC chemicals (triclosan, permethrin, MGK-264, and 2-Phenylphenol). For all 17 CRCs, the minimum AED served as a conservative estimate of doses that could cause *in vivo* effects, as the minimum human-based AED was less than the minimum LEL retrieved from ToxRefDBv1 animal studies. The estimated exposure values from Wambaugh et al. (2014) and the cosmetic-relevant TTC (Yang et al., 2017) are also indicated. For TTC, all CRCs were either in Cramer Class I or III. In general, exposure values are either lower than, or at, the lower end of the computed AED values. The TTC values are all lower than the *in vivo* LEL values but are all within the AED range. All LELs and AED values are available in Supplemental File 1. The dose ranges of the AED and *in vivo* LELs for the remainder of ToxRefDBv1 chemicals associated with ToxCast and httk information (305 of the total 766 ToxRefDBv1 chemicals) were also compared (Fig. 6B). These density plots demonstrate that for most of the chemicals, the AED range (using the minimum and maximum AC50 values) and *in vivo* LEL ranges overlap. Overall, the upper range of computed AEDs tended to overlap with active *in vivo* doses. All LELs and AED values are available in Supplemental File 1.

4. Discussion

With a major shift in toxicity testing for CRCs to utilize NAMs in the EU, the need to evaluate and further develop these methods for screening-level assessments has become increasingly important. However, prior to method development, it is critical to evaluate available existing *in vivo* data that can be used as a reference when considering the performance of NAMs. In this study, we surveyed a list of 58 CRCs, including chemicals used as cosmetic ingredients as well as some chemicals that are trace contaminants, for *in vivo* and *in vitro* activity in public data repositories, ToxRefDB (version 1) and ToxCast

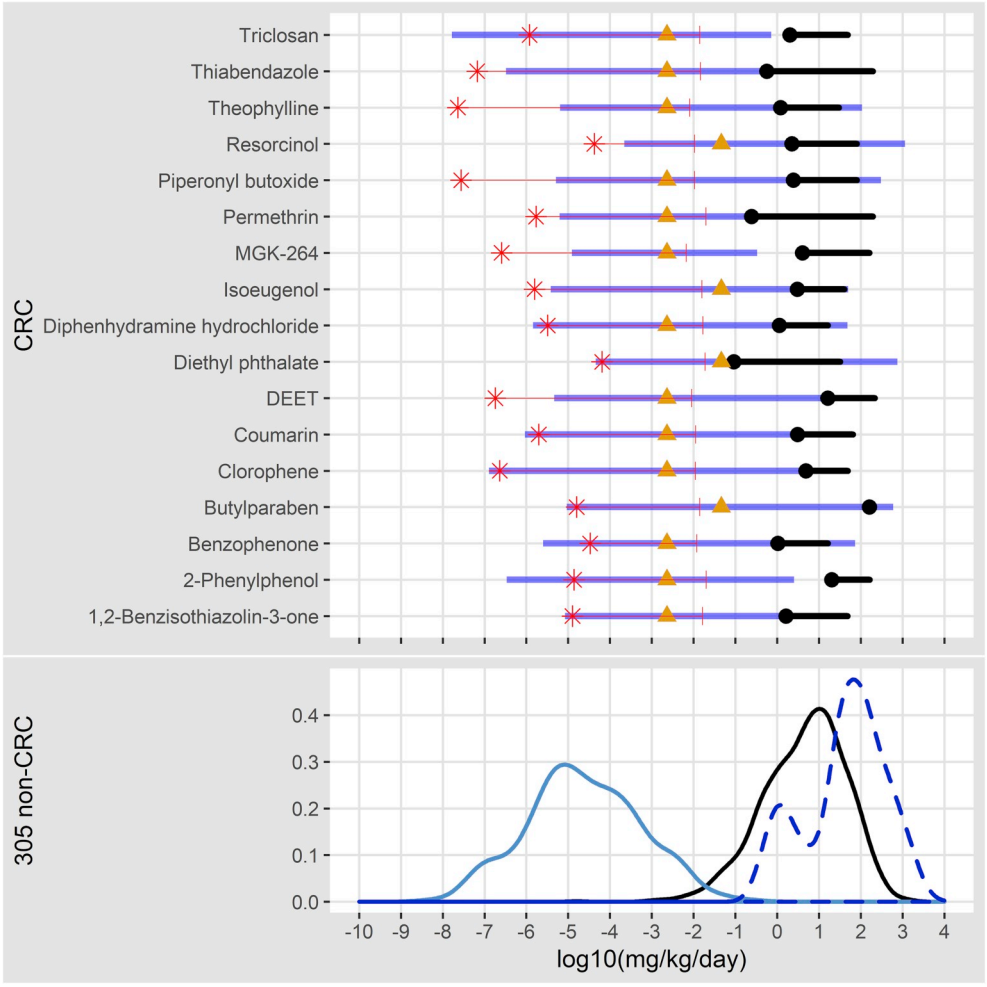

Fig. 6. Human-based administered equivalent doses (AEDs) and LELs. A. The AED 50th quantile was calculated for the minimum and maximum AC50 from all active ToxCast assay endpoints for the 17 CRCs (blue) and plotted along with the *in vivo* LELs at the study level (black) for each chemical. The filled black dot shows the LEL at the chemical level, no spread indicates that the LEL was either consistent across all studies or that only one study was available for the chemical. Red stars indicate estimated total median human exposure, with the 95% upper confidence interval (CI) indicated by a red line (Wambaugh et al., 2014) Gold triangles show cosmetics-specific TTC values (Yang et al., 2017). **B.** Density plots for data from the 305 non-CRC chemicals in ToxRefDB, ToxCast, and with available HTTK information; these density plots show the *in vivo* LELs (black), AEDs for the minimum AC50 (light blue) and maximum AC50 (dashed dark blue) across all active ToxCast assay endpoints. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

(invitrodbv2), respectively. The many use categories for the CRCs demonstrated that these chemicals are functionally diverse, and as such, a range of *in vitro* and *in vivo* activities were anticipated. This study demonstrated a novel application of publicly available data to first characterize and compare organ-level toxicities, and the doses at which these occurred, for CRCs and other chemicals. In surveying ToxRefDB and ToxCast, our goal was to address several hypotheses: (1) the types of *in vivo* effects from CRCs were similar to those observed across the remaining ToxRefDB chemicals, i.e. exposures to cosmetics ingredients did not correspond to increased rates of particular positive findings; (2) the doses at which *in vivo* hazards were observed for CRCs were generally greater than or equal to the median dose at which *in vivo* effects were observed for the remaining ToxRefDB chemicals; (3) different study types did not demonstrate major differences between the LEL values of CRCs versus the ToxRefDBv1 chemical library; and, (4) a NAM-based approach using *in vitro* bioactivity or cosmetics TTC values can be used to inform conservative point-of-departure estimate for CRCs. The comparison presented herein for AEDs based on *in vitro* bioactivity, exposure estimates, TTC values, and *in vivo* effects is an important demonstration to bolster understanding of how a NAM-based approach might compare to traditional approaches for screening-level chemical assessments.

Overall, we found that the 58 CRCs evaluated in this study appeared to be a representative subset of the bioactivity and hazard observed for the other chemicals with data in ToxCast and ToxRefDB. The target organs affected by CRCs and other chemicals in ToxRefDB were similar, with no specific set of organs demonstrating increased incidence of findings after CRC exposure. The range of LELs for *in vivo* systemic effects following CRC exposure tended to be similar or perhaps slightly higher than the LELs for the other chemicals in ToxRefDB. With respect to *in vitro* bioactivity, two-thirds of the CRCs were not cytotoxic and demonstrated low rates of positives *in vitro,* whereas about one-third of the CRCs demonstrated higher *in vitro* hit rates and some cytotoxicity, typically with the bioactivity and cytotoxicity overlapping in concentration. It is possible that *in vitro* bioactivity that occurs at concentrations lower than cytotoxicity would be considered "selective," or related to a specific biological target, whereas activity near or around cytotoxicity may be due to cell stress, cell death, and nonspecific responses related to these processes (Judson et al., 2016; Thomas et al., 2013). Though bioactivity data as aggregated in this work cannot predict a specific adverse outcome, the concentration range in which both selective and/or nonselective bioactivity occur appears to be a useful starting point for estimation of a range of AEDs that would be a threshold for *in vivo* effects. Herein, we found that either cosmetic-relevant TTC values or AEDs from *in vitro* bioactivity assays would conservatively estimate points-of-departure for CRCs when compared to LELs from legacy *in vivo* studies.

Though the analysis of LELs by target organ and study type suggests that in general CRCs appear to be slightly less potent than the remaining chemicals from the ToxRefDBv1 chemical inventory, due to the limited number of identified CRCs, it is also possible that any of the small differences between the distributions of the LEL values for the CRCs and remaining chemicals in ToxRefDB are due to random chance in subsampling. The Wilcoxon-Mann-Whitney testing should not be overinterpreted for this reason. With respect to the ToxCast data, a lack of burst activity could indicate that the majority of CRCs evaluated

were not overtly cytotoxic, or perhaps not very active in the assay platforms tested. However, it is important to note that this conclusion is limited by the caveat that the assay endpoints in ToxCast currently cover only a portion of biological space and may not have the diversity of assay endpoints necessary to reflect all possible bioactivity *in vivo*. Regardless of the specific bioactivities observed, the range of bioactive concentrations appears useful in defining a point-of-departure for screening level assessments. Indeed, the range of AED values derived from AC50 values overlap with the LEL values from *in vivo* studies; this overlap was typically seen at the higher end of the computed AED values for all 17 CRCs evaluated. This suggests that the minimum AED could be a conservative surrogate for point-of-departure. In general, the median exposure estimates for the 17 CRCs are either lower or at the lower end of the calculated AED estimates, showing that most bioactivity of the chemicals happened at higher levels than anticipated exposure.

It is worth noting that in previous work that compared ExpoCast to AED estimates, these values overlapped for very few chemicals (Wetmore, 2015), whereas here there is some overlap between the ExpoCast credible interval for all 17 CRCs for which a comparison could be made. The convergence of the ExpoCast (Wambaugh et al., 2014) estimates and the AED estimates in this work suggest that the bioactivity:exposure ratio may be small for CRCs. This may be due to CRCs presenting some near-field exposures, which increases the exposure prediction from the ExpoCast model employed here (Wambaugh et al., 2014); more refined exposure models for CRCs may be needed to pursue a NAM-based approach. Further, many of the *in vitro* bioactivity assays included provide estimates of activity rather than adversity, making the *in vitro* bioactivity dataset a conservative input for AED calculation. The AED range also provided more conservative estimates of hazard when compared to TTC values derived from the work of Yang et al. (Yang et al., 2017) that refined TTC values for cosmetic chemicals specifically. Most of the TTC values were within the 95% credible interval of the estimated exposure and greater than 1 log10 unit from *in vivo* estimations. Overall, these findings suggest that CRCs demonstrate *in vivo* outcomes similar to other chemicals in ToxRefDB, though perhaps at slightly higher doses; that AEDs and/or TTC may be useful in developing estimates of point-of-departure; and, that refinements to exposure prediction models and *in vitro* to *in vivo* extrapolation may enhance the feasibility of using AEDs as surrogates for point-of-departure.

The evaluation of a small subset of CRCs presented herein provides an interesting example for characterization of screening methodologies for cosmetic ingredients, trace contaminants, and other consumer product chemicals. One of the keys to transitioning from traditional animal studies to NAMs is the ability to do an *in vitro* to *in vivo* comparison (Wetmore, 2015). Only 17 of the 58 CRCs, and 305 of the 870 non-cosmetic chemicals, had enough publicly-available toxicokinetic information to do this comparison, highlighting that additional HTTK information would facilitate a larger set of CRCs for comparison. An additional limitation of the *in vitro* data is the lack of total biological coverage; ToxCast assays in invitrodb v2 relate to a relatively small portion of the human transcriptome, and as such, may not include all of the biological targets of interest for screening of CRCs. In the future, high-throughput transcriptomic profiles (Subramanian et al., 2017) and cellular phenotypic profiling (Kremb and Voolstra, 2017) would provide improved coverage of biological space and could be used in conjunction with bioactivity assay data, including ToxCast data, to estimate AEDs.

A further limitation is the composition of the *in vivo* database, ToxRefDBv1; the database contains largely oral repeat dose toxicity studies, which may not be as realistically informative for CRCs as would dermal repeat dose toxicity studies. However, the current lack of dermal data in ToxRefDBv1, and the fact that the TTC values were based on oral repeat dose toxicity studies (Yang et al., 2017), makes the current study an appropriate comparison based on the state of the science. Other efforts are being made to estimate TTC values for CRCs that are delivered dermally (Kroes et al., 2007; Williams et al., 2016).

The TTC values used were derived from analysis of available chemical structural information and toxicology studies and designed to be protective and represent an acceptable level of exposure; therefore, they are expected to be below the *in vivo* LELs. All CRCs were either a ToxTree (v2.6.0) Cramer Class I or III, which correspond to a TTC value of −1.37 log10-mg/kg/day or −2.64 log10-mg/kg/day in Fig. 6, respectively. The Cramer Class I CRCs corresponded to a TTC value that did not overlap with the estimated exposure values. In contrast, all Cramer Class III chemicals are within the 95th percentile of the credible interval for the median exposure estimate, suggesting that human exposure to those chemicals might be higher than the estimated TTC; typically, this might indicate that these chemicals would be relevant for further review. However, for those Cramer Class III CRCs that demonstrated overlap between the TTC and the exposure estimate interval, there were no *in vivo* effects observed at the same dose as the TTC value, though some *in vitro* assays were active. Therefore, these chemicals might be relevant for further refinements to the NAM data inputs, including the bioactivity, HTTK, exposure, or TTC estimates. In some cases, *in vitro* bioactivity data may be available for mixtures, substances with undefined structure, or substances that fall outside of the domain of applicability for TTC; further, it is possible *in vitro* bioactivity data may provide an opportunity for experimental refinement of predicted doses with bioactivity. If certain bioactivity types indicative of a specific hazard were of greater interest, these data could inform a more specific threshold dose. Conversely, TTC provides a general and rapid, *in silico* alternative for prediction of a threshold dose. Overall, for the CRCs in this study, the TTC and AED values demonstrate a protective estimate of systemic effects when compared to LEL values, suggesting that both TTC and AED values may have important roles in screening level chemical safety evaluation.

Additional limitations in this study, and possible needs in future work, are evaluation of product formulations or mixtures and consideration of how cosmetic-relevant products are actually used by consumers. In the current chemical safety paradigm, chemicals are often evaluated individually, but finished cosmetic products are typically packaged as mixtures in liquid, solid, or aerosolized phases. These phases and mixtures may affect the safety profile of the chemicals contained in the product due to potential chemical interactions or by affecting absorption rate (Kienzler et al., 2016). The sheer number of possible combinations of mixtures makes adoption of NAMs an important step to help identify and prioritize chemicals that require examination in screening models of greater biological complexity. Furthermore, the high-throughput nature of NAMs might make them more conducive to experiments with mixtures; screening individual chemicals, and relevant mixtures of these chemicals, is simply insurmountable using resource-intensive traditional methodologies, and would be best facilitated with a NAM approach. Recent work suggests that NAM-based methods to address knowledge gaps regarding co-exposures observed in the human population (Kapraun et al., 2017) could be combined with NAMs for bioactivity screening (Bopp et al., 2019) in order to promote rapid mixtures evaluation. Other future work for NAM-based approaches that would further benefit the evaluation of CRC safety includes updates to exposure predictions and *in vitro* to *in vivo* extrapolation to account for dermal absorption rates. Cosmetic products are often applied topically, and as such, dermal absorption would greatly limit the amounts of CRCs that would be bioavailable, and so typical use patterns could be further considered to better quantitatively inform estimates of exposure based on dermal application, thereby building on ongoing exposure work (Ring, 2019; Dionisio, 2018; Isaacs, 2014).

As REACH implementation continues, other collections of *in vitro* and *in vivo* data may become available for retrospective analyses like the one presented herein. Currently, there are more validated *in vivo* data than *in vitro* data for cosmetic chemicals (Vinardell, 2015), but that is expected to change as more *in vitro* methods are developed and more

chemicals are evaluated. The workflow for a screening level assessment demonstrated here with a set of CRCs used publicly available data to facilitate reproducibility and transparency, but the evaluation of new CRCs using such an approach could make use of non-public data or new sources of *in vitro* bioactivity or *in silico* predictions of bioactivity data. Establishing confidence in NAMs will likely require some comparisons with existing *in vivo* outcomes or more directly to human health effects. Thus, the question of how *in vitro* and *in vivo* data for CRCs compare, and how to establish confidence NAMs, will continue to be a critical question in safety assessment. Great progress has already been made in developing NAMs for skin sensitization, in particular the local lymph node assay (LLNA) (Casati et al., 2018; Dumont et al., 2016; Hoffmann et al., 2018; Kleinstreuer et al., 2018) as a replacement for *in vivo* studies. Comparisons like the one herein and others (Karmaus et al., 2016; Shah and Greene, 2014) intended to evaluate the utility of HTS assays for different uses of chemicals can further support the use of NAMs by highlighting the current capabilities and scientific gaps that prevent full animal replacement. Based on currently available data evaluation herein, the *in vitro* HTS bioactivity data, combined with HTTK information, or TTC, can yield protective points-of-departure for CRCs in screening-level assessment.

Conflicts of interest

This work was in part supported financially by L'Oreal, a company involved in the development and sale of cosmetics. Gladys Ouedraogo and Sophie Loisel-Joubert are employed by Loreal. Ly Ly Pham and Lisa Truong received support from Loreal through a Cooperative Research and Development Agreement (CRADA #507-A-11) between L'Oreal and the US EPA National Center for Computational Toxicology.

Disclaimer

The United States Environmental Protection Agency (U.S. EPA) through its Office of Research and Development has subjected this article to Agency administrative review and approved it for publication. Mention of trade names or commercial products does not constitute endorsement for use. The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the US EPA.

Acknowledgments

This project was supported in part by an appointment to the Research Participation Program at the National Center for Computational Toxicology, U.S. Environmental Protection Agency, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and EPA. We would also like to acknowledge insightful comments from Keith Houck, Reeder Sams, and Russell Thomas (US EPA), as well as Agnes Karmaus (Integrated Laboratory Systems, Inc).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fct.2019.110718.

Transparency document

Transparency document related to this article can be found online at https://doi.org/10.1016/j.fct.2019.110718.

References

Adler, S., Basketter, D., Creton, S., Pelkonen, O., van Benthem, J., Zuang, V., Andersen, K.E., Angers-Loustau, A., Aptula, A., Bal-Price, A., et al., 2011. Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010. Arch. Toxicol. 85 (5), 367–485.

Attene-Ramos, M.S., Miller, N., Huang, R., Michael, S., Itkin, M., Kavlock, R.J., Austin, C.P., Shinn, P., Simeonov, A., Tice, R.R., et al., 2013. The Tox21 robotic platform for the assessment of environmental chemicals–from vision to reality. Drug Discov. Today 18 (15–16), 716–723.

Bopp, S.K., Kienzler, A., Richarz, A.N., van der Linden, S.C., Paini, A., Parissis, N., Worth, A.P., 2019. Regulatory assessment and risk management of chemical mixtures: challenges and ways forward. Crit. Rev. Toxicol. 1–16. https://doi.org/10.1080/10408444.2019.1579169.

Boyer, I.J., Bergfeld, W.F., Heldreth, B., Fiume, M.M., Gill, L.J., 2017. The cosmetic ingredient review program—expert safety assessments of cosmetic ingredients in an open forum. Int. J. Toxicol. 36 (5_Suppl. 2), 5S–13S.

Casati, S., Aschberger, K., Barroso, J., Casey, W., Delgado, I., Kim, T.S., Kleinstreuer, N., Kojima, H., Lee, J.K., Lowit, A., et al., 2018. Standardisation of defined approaches for skin sensitisation testing to support regulatory use and international adoption: position of the International Cooperation on Alternative Test Methods. Arch. Toxicol. 92 (2), 611–617.

Council, N.R., 2007. Toxicity Testing in the 21st Century: A Vision and a Strategy. The National Academies Press, Washington, DC.

Cramer, G.M., Ford, R.A., Hall, R.L., 1978. Estimation of toxic hazard–a decision tree approach. Food Cosmet. Toxicol. 16 (3), 255–276.

Daneshian, M., Kamp, H., Hengstler, J., Leist, M., van de Water, B., 2016. Highlight report: launch of a large integrated European in vitro toxicology project: EU-ToxRisk. Arch. Toxicol. 90 (5), 1021–1024.

Desprez, B., Dent, M., Keller, D., Klaric, M., Ouedraogo, G., Cubberley, R., Duplan, H., Eilstein, J., Ellison, C., Gregoire, S., et al., 2018. A strategy for systemic toxicity assessment based on non-animal approaches: the Cosmetics Europe Long Range Science Strategy programme. Toxicol. In Vitro : an international journal published in association with BIBRA 50, 137–146.

Dionisio, K.L., et al., 2018. The Chemical and Products Database, a resource for exposure-relevant data on chemicals in consumer products. Nature Scientific Data. https://doi.org/10.1038/sdata.2018.125.

Dumont, C., Barroso, J., Matys, I., Worth, A., Casati, S., 2016. Analysis of the Local Lymph Node Assay (LLNA) variability for assessing the prediction of skin sensitisation potential and potency of chemicals with non-animal approaches. Toxicol. In Vitro : an international journal published in association with BIBRA 34, 220–228.

ECHA, 2016. In: New Approach Methodologies in Regulatory Science, Proceedings of a Scientific Workshop, April 19-20, 2016, Helsinki, Finland. Available at: https://echa.europa.eu/documents/10162/22816069/scientific_ws_proceedings_en.pdf/a2087434-0407-4705-9057-95d9c2c2cc57, Accessed date: 3 September 2018.

EU Regulation (EC) No, 2009. In: Union, O.J.E. (Ed.), 1223/2009 of the European Parliament and of the Council of 30 November 2009 on Cosmetic Products. Europen Union, pp. 59–209.

Filer, D.L., Kothiya, P., Setzer, R.W., Judson, R.S., Martin, M.T., 2017. tcpl: the ToxCast pipeline for high-throughput screening data. Bioinformatics 33 (4), 618–620.

Hartung, T., Blaauboer, B.J., Bosgra, S., Carney, E., Coenen, J., Conolly, R.B., Corsini, E., Green, S., Faustman, E.M., Gaspari, A., et al., 2011. An expert consortium review of the EC-commissioned report "alternative (Non-Animal) methods for cosmetics testing: current status and future prospects - 2010. ALTEX 28 (3), 183–209.

Hoffmann, S., Kleinstreuer, N., Alepee, N., Allen, D., Api, A.M., Ashikaga, T., Clouet, E., Cluzel, M., Desprez, B., Gellatly, N., et al., 2018. Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database. Crit. Rev. Toxicol. 48 (5), 344–358.

Judson, R., Richard, A., Dix, D., Houck, K., Elloumi, F., Martin, M., Cathey, T., Transue, T.R., Spencer, R., Wolf, M., 2008. ACToR - aggregated computational toxicology resource. Toxicol. Appl. Pharmacol. 233, 7–13.

Isaacs, K.K., 2014. SHEDS-HT: an integrated probabilistic exposure model for prioritizing exposures to chemicals with near-field and dietary sources. Environmental Science and Technology. https://doi.org/10.1021/es502513w.

Judson, R., Houck, K., Martin, M., Richard, A.M., Knudsen, T.B., Shah, I., Little, S., Wambaugh, J., Woodrow Setzer, R., Kothya, P., et al., 2016. Analysis of the effects of cell stress and cytotoxicity on in vitro assay activity across a diverse chemical and assay space. Toxicol. Sci. 152, 323–339.

Kapraun, D.F., Wambaugh, J.F., Ring, C.L., Tornero-Velez, R., Setzer, R.W., 2017. A method for identifying prevalent chemical combinations in the U.S. population. Environ. Health Perspect. 125 (8), 087017.

Karmaus, A.L., Filer, D.L., Martin, M.T., Houck, K.A., 2016. Evaluation of food-relevant chemicals in the ToxCast high-throughput screening program. Food Chem. Toxicol. 92, 188–196.

Kavlock, R., Dix, D., 2010. Computational toxicology as implemented by the U.S. EPA: providing high throughput decision support tools for screening and assessing chemical exposure, hazard and risk. J. Toxicol. Environ. Health B Crit. Rev. 13 (2–4), 197–217.

Kienzler, A., Bopp, S.K., van der Linden, S., Berggren, E., Worth, A., 2016. Regulatory assessment of chemical mixtures: requirements, current approaches and future perspectives. Regul. Toxicol. Pharmacol. 80, 321–334.

Kleinstreuer, N.C., Hoffmann, S., Alepee, N., Allen, D., Ashikaga, T., Casey, W., Clouet, E., Cluzel, M., Desprez, B., Gellatly, N., et al., 2018. Non-animal methods to predict skin sensitization (II): an assessment of defined approaches (*). Crit. Rev. Toxicol. 48 (5), 359–374.

Kremb, S., Voolstra, C.R., 2017. High-resolution phenotypic profiling of natural products-induced effects on the single-cell level. Sci. Rep. 7, 44472.

Kroes, R., Renwick, A.G., Feron, V., Galli, C.L., Gibney, M., Greim, H., Guy, R.H., Lhuguenot, J.C., van de Sandt, J.J.M., 2007. Application of the threshold of toxicological concern (TTC) to the safety evaluation of cosmetic ingredients. Food Chem. Toxicol. 45 (12), 2533–2562.

Martin, M.T., Judson, R.S., Reif, D.M., Kavlock, R.J., Dix, D.J., 2009a. Profiling chemicals based on chronic toxicity results from the U.S. EPA ToxRef Database. Environ. Health Perspect. 117 (3), 392–399.
Martin, M.T., Judson, R.S., Reif, D.M., Kavlock, R.J., Dix, D.J., 2009b. Profiling chemicals based on chronic toxicity results from the U.S. EPA ToxRef database. Environ. Health Perspect. 117, 392–399.
Martin, M.T., Mendez, E., Corum, D.G., Judson, R.S., Kavlock, R.J., Rotroff, D.M., Dix, D.J., 2009c. Profiling the reproductive toxicity of chemicals from multigeneration studies in the toxicity reference database. Toxicol. Sci. : an official journal of the Society of Toxicology 110 (1), 181–190.
Munro, I.C., Ford, R.A., Kennepohl, E., Sprenger, J.G., 1996. Correlation of structural class with no-observed-effect levels: a proposal for establishing a threshold of concern. Food Chem. Toxicol. : an international journal published for the British Industrial Biological Research Association 34 (9), 829–867.
Nair, A.B., Jacob, S., 2016. A simple practice guide for dose conversion between animals and human. J. Basic Clin. Pharm. 7, 27–31.
National Academies of Sciences, E., and Medicine, 2017. Using 21st Century Science to Improve Risk-Related Evaluations. The National Academies Press, Washington, DC.
Pearce, R.G., Setzer, R.W., Davis, J.L., Wambaugh, J.F., 2017a. Evaluation and calibration of high-throughput predictions of chemical distribution to tissues. J. Pharmacokinet. Pharmacodyn. 44 (6), 549–565.
Pearce, R.G., Setzer, R.W., Strope, C.L., Sipes, N.S., Wambaugh, J.F., 2017b. 4. Httk: R Package for High-Throughput Toxicokinetics. 2017, vol. 79. pp. 26 (high-throughput; ToxCast; httk; toxicokinetics; pharmacokinetics).
R Development Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna, austria.
SEURAT-1. 2016. Towards the Replacement of in vivo Rpeeated Dose Systemic Toxicity Testing. URL: http://www.seurat-1.eu/.
Richard, A.M., et al., 2016. ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. Chemical Research in Toxicology. https://doi.org/10.1021/acs.chemrestox.6b00135.
Ring, C.L., et al., 2019. Consensus Modeling of Median Chemical Intake for the U.S. Population Based on Predictions of Exposure Pathways. Environmental Science and Technology. https://doi.org/10.1021/acs.est.8b04056.
Shah, F., Greene, N., 2014. Analysis of Pfizer compounds in EPA's ToxCast chemicals-assay space. Chem. Res. Toxicol. 27 (1), 86–98.
Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al., 2017. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell 171 (6), 1437–1452 e17.
Team, R., 2015. RStudio. Integrated Development Enviroment for R.
Thomas, R.S., Philbert, M.A., Auerbach, S.S., Wetmore, B.A., Devito, M.J., Cote, I., Rowlands, J.C., Whelan, M.P., Hays, S.M., Andersen, M.E., et al., 2013. Incorporating new technologies into toxicity testing and risk assessment: moving from 21st century vision to a data-driven framework. Toxicol. Sci. : an official journal of the Society of Toxicology 136 (1), 4–18.
Thomas, R.S., Paules, R.S., Simeonov, A., Fitzpatrick, S.C., Crofton, K.M., Casey, W.M., Mendrick, D.L., 2018. The US Federal Tox21 Program: a strategic and operational plan for continued leadership. ALTEX 35 (2), 163–168.
Tice, R.R., Austin, C.P., Kavlock, R.J., Bucher, J.R., 2013. Improving the human hazard characterization of chemicals: a Tox21 update. Environ. Health Perspect. 121 (7), 756–765.
USEPA, 2014. ToxRefDB version 1. Animal Toxicity Studies: Effects and Endpoints. USEPA.
Vinardell, M.P., 2015. The use of non-animal alternatives in the safety evaluations of cosmetics ingredients by the Scientific Committee on Consumer Safety (SCCS). Regul. Toxicol. Pharmacol. : RTP (Regul. Toxicol. Pharmacol.) 71 (2), 198–204.
Wambaugh, J.F., Wang, A., Dionisio, K.L., Frame, A., Egeghy, P., Judson, R., Setzer, R.W., 2014. High throughput heuristics for prioritizing human exposure to environmental chemicals. Environ. Sci. Technol. 48 (21), 12760–12767.
Wambaugh, J.F., Hughes, M.F., Ring, C.L., MacMillan, D.K., Ford, J., Fennell, T.R., Black, S.R., Snyder, R.W., Sipes, N.S., Wetmore, B.A., et al., 2018. Evaluating in vitro-in vivo extrapolation of toxicokinetics. Toxicol. Sci. 163 (1), 152–169.
Wetmore, B.A., 2015. Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment. Toxicology 332, 94–101.
Williams, F.M., Rothe, H., Barrett, G., Chiodini, A., Whyte, J., Cronin, M.T.D., Monteiro-Riviere, N.A., Plautz, J., Roper, C., Westerhout, J., et al., 2016. Assessing the safety of cosmetic chemicals: consideration of a flux decision tree to predict dermally delivered systemic dose for comparison with oral TTC (Threshold of Toxicological Concern). Regul. Toxicol. Pharmacol. 76, 174–186.
Williams, A.J., Grulke, C.M., Edwards, J., McEachran, A.D., Mansouri, K., Baker, N.C., Patlewicz, G., Shah, I., Wambaugh, J.F., Judson, R.S., et al., 2017. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. J. Cheminf. 9 (1), 61.
Yang, C., Tarkhov, A., Marusczyk, J., Bienfait, B., Gasteiger, J., Kleinoeder, T., Magdziarz, T., Sacher, O., Schwab, C.H., Schwoebel, J., et al., 2015. New publicly available chemical query language, CSRML, to support chemotype representations for application to data mining and modeling. J. Chem. Inf. Model. 55, 510–528.
Yang, C., Barlow, S.M., Muldoon Jacobs, K.L., Vitcheva, V., Boobis, A.R., Felter, S.P., Arvidson, K.B., Keller, D., Cronin, M.T.D., Enoch, S., et al., 2017. Thresholds of Toxicological Concern for cosmetics-related substances: new database, thresholds, and enrichment of chemical space. Food Chem. Toxicol. Int. J. Publ. British Ind. Biol. Res. Assoc. 109 (Pt 1), 170–193.