Food and Chemical Toxicology 103 (2017) 174–182

Contents lists available at ScienceDirect

Food and Chemical Toxicology

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

Curation of food-relevant chemicals in ToxCast

Agnes L. Karmaus [a, *], Thomas D. Trautman [b], Mansi Krishan [c], Dayne L. Filer [d], Laurel A. Fix [e]

[a] Integrated Laboratory Systems Inc., Research Triangle Park, NC, USA
[b] Retired, Toxicology & Regulatory Affairs, General Mills, Minneapolis, MN, USA
[c] ILSI North America, Washington, DC, USA
[d] University of North Carolina, Chapel Hill, NC, USA
[e] Unilever, Englewood Cliffs, NJ, USA

ARTICLE INFO

Article history:
Received 6 September 2016
Received in revised form
23 February 2017
Accepted 2 March 2017
Available online 7 March 2017

Keywords:
Food additive
Indirect food additive
ToxCast
CPCat
High-throughput screening

ABSTRACT

High-throughput *in vitro* assays and exposure prediction efforts are paving the way for modeling chemical risk; however, the utility of such extensive datasets can be limited or misleading when annotation fails to capture current chemical usage. To address this data gap and provide context for food-use in the United States (US), manual curation of food-relevant chemicals in ToxCast was conducted. Chemicals were categorized into three food-use categories: (1) direct food additives, (2) indirect food additives, or (3) pesticide residues. Manual curation resulted in 30% of chemicals having new annotation as well as the removal of 319 chemicals, most due to cancellation or only foreign usage. These results highlight that manual curation of chemical use information provided significant insight affecting the overall inventory and chemical categorization. In total, 1211 chemicals were confirmed as current day food-use in the US by manual curation; 1154 of these chemicals were also identified as food-related in the globally sourced chemical use information from Chemical/Product Categories database (CPCat). The refined list of food-use chemicals and the sources highlighted for compiling annotated information required to confirm food-use are valuable resources for providing needed context when evaluating large-scale inventories such as ToxCast.

1. Introduction

High-throughput approaches for *in vitro* bioactivity profiling informing on hazard or toxicity pathways together with exposure estimates are rapidly evolving; the future integration of these data support the development of high-throughput risk models (Kavlock et al., 2012; Wambaugh et al., 2013; Edwards et al., 2016; EPA, 2016c, 2016g, 2016h; Teeguarden et al., 2016). The large scale of these approaches spans thousands of chemicals. Exposure pathways and *in silico* exposure predictions provide previously unavailable chemical usage and exposure estimates, and high-throughput *in vitro* biochemical assays evaluate chemical-mediated bioactivity to further propel the field of toxicology into the big data arena.

For example, the US Environmental Protection Agency (EPA) ToxCast high-throughput screening program contains >3000 chemicals evaluated across >1000 targeted assay endpoints in concentration-response, providing the largest inventory of assay coverage across a large library of chemicals. Furthermore, the Tox21 quantitative high-throughput screening program has evaluated the bioactivity of nearly 9000 chemicals with an automated robotic system using 1536-well plates in 15-point concentration-response across roughly 60 *in vitro* assays (EPA, 2016g). The chemical inventories for these *in vitro* screening programs were compiled from a broad range of sources, including industrial and consumer products, food additives, donated pharmaceuticals, and potentially "green" alternative chemicals.

Abbreviations: ACToR, Aggregated Computational Toxicology Resource; CAS, Chemical Abstracts Service; CFR, Code of Federal Regulations; CPCat, Chemical/Product Categories database; CPCPdb, Consumer Product Chemical Profiles database; EPA, Environmental Protection Agency; EAFUS, Everything Added to Food in the United States; FDA, Food and Drug Administration; FEMA, Flavor and Extract Manufacturers Association; GRAS, generally recognized as safe; MRL, maximum residue limit; PAN, Pesticide Action Network; SPIN, Substances in Preparation in Nordic Countries; USDA, US Department of Agriculture.

* Corresponding author. 601 Keystone Park Drive, Suite 800, Research Triangle Park, NC 27709, USA.
E-mail address: agnes.karmaus@gmail.com (A.L. Karmaus).

http://dx.doi.org/10.1016/j.fct.2017.03.006

Meanwhile, parallel efforts to annotate chemical use and exposure scenarios include the Consumer Product Chemical Profiles database (CPCPdb) and the EPA Chemical/Product Categories database (CPCat) (Goldsmith et al., 2014; Dionisio et al., 2015). CPCPdb is a compilation of information from material safety data sheets obtained from a large retailer, wherein 1797 chemicals were mapped to 8921 consumer products and subsequently 353 use categories (Goldsmith et al., 2014). CPCat is the largest project to annotate chemicals based on use categorization, comprising information for 43,596 chemicals associated to 833 controlled vocabulary terms specifically describing usage (Dionisio et al., 2015). More specifically, CPCat is the result of a high-throughput mining effort whereby information obtained from a dozen data sources was merged and use term annotation was harmonized. In combination, high-throughput screening combined with these sources can help to inform on chemical bioactivity derived by *in vitro* methods and identify the use or exposure routes for chemicals of interest, respectively.

In an attempt to focus on US-relevant food-use chemicals evaluated in ToxCast, a recent study compiled food-relevant chemicals by combining data obtained from relevant publicly available databases identifying 1530 putative food-relevant chemicals in ToxCast (Karmaus et al., 2016). The need for curation to refine this list of food-relevant chemicals, or for independent confirmation of food-use categorization, was emphasized as it was found that high-throughput data mining approaches that integrate entire database inventories across multiple sources resulted in the inclusion of chemicals with historical use but no current tolerances or use in the US. More specifically, a number of chemicals included among the list of 1530 food-relevant ToxCast chemicals were shown to no longer be used in foods or had restricted use in the US. One example cited was nordihydroguaiaretic acid (NDGA), formerly used as an antioxidant and preservative in fats and oils, which was initially approved in 1943 but withdrawn in 1968 due to nephropathy in rats (Lü et al., 2010). NDGA, included due to its listing in the US Food and Drug Administration (FDA) Everything Added to Food in the United States (EAFUS) database, had elicited concentration-dependent effects in a number of ToxCast *in vitro* assay endpoints; however, it no longer presents any exposure concern via food in the US and should be annotated as historical food-use and be removed from consideration when evaluating results for food-use chemicals of current relevance in the US.

In response to the stated need for curation to determine current relevance, and in an effort to incorporate context regarding exposure concern from food-use in the US, the purpose of the current study was to demonstrate the value of a comprehensive manual review on the identification of food-relevant chemicals in ToxCast. Though manual review focused on US current-day use, comparison to independently sourced global chemical usage information from CPCat was also conducted. Comparison to CPCat largely confirmed chemical inventory; however, manual review still had a significant effect on refining chemical categorization. Several important considerations for identifying and confirming food-use were identified and addressed herein. These included determining which chemical identifiers are most appropriate for searching across resources and which resources are of value for confirming US or global food-use. The inventory of putatively food-relevant chemicals in ToxCast was sorted into four categories based broadly on exposure likelihood from food-use in the US: (1) direct food additives, (2) indirect food additives, (3) pesticide residues, and (4) non-food. Extensive manually curated comments have been provided to give context on food-use and current use allowance in the US for the entire chemical inventory evaluated. This curation effort provides a more accurate picture of current US exposure to chemicals via foods and refines the ToxCast food-relevant chemical inventory.

2. Materials and methods

2.1. Chemical inventory sources

Two independently sourced inventories were integrated to identify current-day US food-use chemicals in ToxCast. The first was a list of US food-relevant chemicals identified by Karmaus et al. (2016) and the second was a publicly available chemical use annotation database called the Chemical/Product Categories database (CPCat; Dionisio et al., 2015; EPA, 2016c).

The list of food-relevant chemicals identified by Karmaus et al. (2016) was compiled by amalgamating inventories across eight US food-relevant databases, yielding 11,733 chemicals. Upon removing redundancies across the inventory, a list of 8965 unique food-relevant chemicals with current or former use in the US was identified. Subsequently, the published study identified 1530 of these food-relevant chemicals were in ToxCast, and chemicals were grouped into three use categories based strictly on their database of origin: (1) direct additives, (2) food contact substances, and (3) pesticides (Karmaus et al., 2016).

The recently released CPCat database is a large-scale aggregation across a dozen global data sources containing 833 mapped controlled vocabulary use categorization terms, not limited to food, for nearly 43,600 chemicals intended to inform on and classifying chemical use (Dionisio et al., 2015; EPA, 2016c). Though our focus was explicitly on US food-use, CPCat effectively summarizes many possible uses for chemicals worldwide, including food-use. The CPCat database inventory was downloaded (cpcat_v04; EPA, 2016c) and mined for any chemicals with “food” in the Description_Source or Description_CPCat fields. To further refine the inventory of interest and focus in on ToxCast specifically, the list of “food” chemicals in the CPCat was narrowed down to only those present in the ToxCast inventory (EPA, 2016d), ultimately resulting in the independent identification of 1749 ToxCast “food”-relevant chemicals from CPCat. The list of 1749 ToxCast chemicals and their “food”-containing CPCat description terms along with the manual curation categories per chemical are provided in Supplementary File S1.

2.2. Manual curation: confirming food-use and chemical categorization

Chemicals evaluated by manual curation were evaluated to confirm food-use as well as current use allowance in the US. The manual curation of the entire inventory was conducted by one scientist to maximize consistency; rationale for defining food-use and subsequent categorization were clearly articulated and adhered to, as described herein. Non-food classification was assigned to chemicals that do not to present any human exposure through foods in the present-day US. Though some of these chemicals may be widely used in the US for non-food uses, the current study seeks only to incorporate food-relevant usage. Criteria for the “non-food” designation are summarized in Table 1.

Chemicals confirmed as current US food-use were categorized based on criteria that considered likelihood of exposure from food-use, the categories are ranked 1–3 from highest to lowest exposure likelihood from food-use: (1) direct food additives, (2) indirect food additives, and (3) pesticide residues. The first category, direct food additives, was a definitive assignment issued primarily by confirming current status either from direct food or color additive listings in Title 21, Code of Federal Regulations (CFR; 21 CFR, Parts 1–190) or from authoritative sources such as FDA Generally Recognized as Safe (GRAS) listings (FDA, 2016), the Flavor and Extract Manufacturers Association (FEMA) GRAS listings (FEMA, 2016), or Food Contact Substance Notifications (FDA, 2015e). Additionally, chemicals occurring naturally in foods, and therefore directly present (Burdock, 1996; FEMA, 2016; WHO, 2016), were also included in the direct food additives category (e.g., citric acid or linoleic acid). The second category, indirect food additives, was also based on CFR listings and Food Contact Substance Notifications, which were found for most packaging components, lubricants, sanitizers, processing aids, and so forth (21 CFR, Parts 1–190). The pesticide residue category comprises pesticides currently registered for food use in the US as well as chemicals with no identifiable current registration for agricultural use but that had current tolerances for specific foods (i.e., as residues). Pesticides that are not currently registered in the US but may have foreign uses and current import tolerances were also included in the pesticide residue category. Notably, the categorization scheme herein permitted multiple category assignments per chemical where justified (i.e., chemicals were not restricted to annotation into only one category). Detailed notes are provided in the chemical inventory summary table available online (http://ilsina.org/curation-of-food-relevant-chemicals-in-toxcast/; © 2017 ILSI North America, all rights reserved).

The publicly available resources manually searched to identify current food-use status are summarized in Table 3. It is important to note that in order to comprehensively obtain information across resources, any synonyms used as alternate chemical names had to be considered because chemical names are often inconsistent across data sources. For example, the chemical name "trimethyl benzene-1,2,4-tricarboxylate" was used in the ToxCast inventory, whereas FDA files cited the more common synonym "trimethyl trimellitate". Although chemicals names were the primarily used for searching, the Chemical Abstracts Service registration numbers (CASRN) were most often a good resource for harmonizing any textual inconsistencies in chemical name. However, it is important to note that CASRN can also present challenges in cases where chemicals have multiple associated CASRN, in which case aggregation of multiple CASRN was required when searching resources.

3. Results

3.1. Identification of food-use chemicals in ToxCast

To focus on food-use chemicals in ToxCast, the two independently sourced inventories were restricted to only include chemicals that were (1) food-relevant and (2) in ToxCast. The Karmaus et al. (2016) inventory was already sourced from US food-relevant resources and had been narrowed down to 1530 ToxCast chemicals. On the other hand, the CPCat inventory required two steps: first CPCat was subsetted by searching for the word "food" in the Description_Source or Description_CPCat fields of the database, this identified 10,972 "food" chemicals (Fig. 1). This list of 10,972 chemicals was then compared to the ToxCast program inventory, revealing that 1749 of the chemicals associated with food in CPCat were also in the ToxCast program. The total of 1749 food-relevant chemicals in ToxCast derived from CPCat is very close in number to the 1530 identified as US food-relevant in ToxCast (Karmaus et al., 2016), and a direct comparison of these chemical lists identified 1276 chemicals in common to both inventories (Fig. 1).

Manual curation (described below) to categorize food-use chemicals based on current-day US food-use were evaluated for the 1276 chemicals identified by both approaches identifying 122 non-food chemicals. This finding highlights the need for manual curation, as two independent large-scale database agglomeration resources had unanimously identified these 122 chemicals as food-use despite not being food-relevant when current-day US food-use context was explicitly required. After eliminating the 122 non-food chemicals from the 1276 chemicals identified by both approaches, the remaining 1154 food-use chemicals comprised 550 categorized as direct food additives, 319 categorized as indirect additives, and 375 categorized as pesticide residues (Fig. 1; note that 90 of the chemicals were annotated with two categories).

Of the chemicals not identified by both methods (Fig. 1), the 254 identified only by Karmaus et al. (2016) were integrated since these chemicals were obtained from US-relevant food resources warranting further evaluation. Manual curation confirmed 57 food-use chemicals in current-day US, and when added to the 1154 commonly identified food-use chemicals, a sum of 1211 food-use chemicals was confirmed. The 473 CPCat-identified "food" chemicals in the ToxCast inventory, which were not identified by Karmaus et al. (2016), were largely annotated as food from non-US sources; namely, 233 of these chemicals were from the Substances in Preparation in Nordic Countries (SPIN) database. The categories associated with these 473 chemicals largely broke down into 155 chemicals with the terms "food" or source description of "food & drug product", 93 with the term "food_contact", 180 with the term "food_additive", and 135 associated with the terms "pesticide", "manufacturing", "food_service", or "food_production". Although future studies may benefit from the evaluation of these 473 chemicals for relevance to exposure in the US from food, for the purposes of this study, these chemicals were not incorporated because the majority were obtained from non-US sources.

3.2. Manual curation to identify current-day US food-use chemicals

Manual curation was used to identify food-use chemicals based on up-to-date use as well as exposure concern from food-use in the United States. Chemicals that were not confirmed as food-use in current-day US were deemed non-food for the purpose of this study. It is important to note that these chemicals may have foreign food-use, or may be widely used in the US for non-food purposes. The chemical inventory containing all ToxCast chemicals reviewed by manual curation herein, complete with categorization and notes on food-use and current allowance in the US can be found online at http://ilsina.org/curation-of-food-relevant-chemicals-in-toxcast (© 2017 ILSI North America, all rights reserved).

Table 1
Criteria for "non-food" designation.

Criteria	Number of chemicals
Pesticides cancelled in the United States, no import tolerances	87
Pesticides for foreign-use only, no import tolerances[a]	98
Pesticides for non-agricultural or veterinary use	25
Pesticides for crack and crevice use only	3
Pesticides for animal feed use, no food tolerance	3
Various non-food: drugs (19), cosmetics (6), fragrance (6), industrial chemicals (35), medical devices (2)	68
Banned chemicals	11
Miscellaneous non-food–related use or lacking information	24

[a] May have had import tolerances at one time.

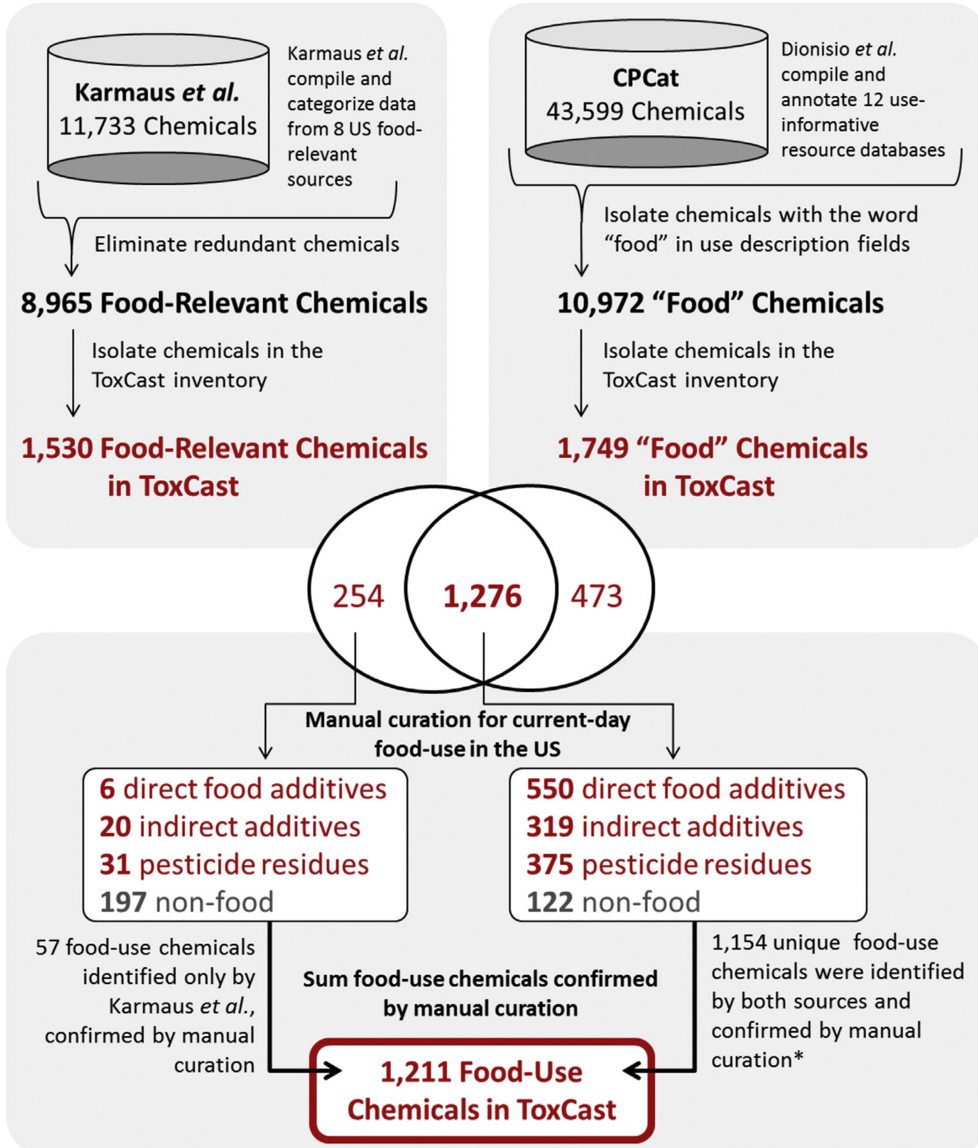

Fig. 1. Workflow for the identification of US-relevant ToxCast food-use chemicals. Food-use chemicals in ToxCast identified by Karmaus et al. (2016) and CPCat were identified, respectively, and compared. Based on manual curation, the common inventory of 1276 food-use chemicals contained 122 chemicals deemed non-food in current-day US. The inventory unique to Karmaus et al. (2016) was also manually curated as this set of chemicals arose from US food-related resources, but the majority was pesticides found to not be current-day food-use. Nonetheless, there were 57 food-use chemicals identified solely by Karmaus et al. and 1154 chemicals from the consensus between both sources were confirmed as food-use by manual curation (calculated by removing the 122 non-food from 1276; note that of the 1154, there were 90 chemicals assigned to two use categories), yielding a total of 1211 confirmed current-day US food-use chemicals in ToxCast.

In total, 319 of the manually curated chemicals were designated as "non-food"; these included cancelled or foreign-use pesticides, drugs, components of cosmetics, and industrial chemicals (Table 1). The majority of the non-food category consists of 219 chemicals formerly identified as pesticides by Karmaus et al. (2016), removed due to no current tolerance or non-food crop use only. Examples of "non-food" chemicals include *para*-aminobenzoic acid (PABA), a sunscreen ingredient; diethyl-*meta*-toluamide (DEET), an insect repellant; warfarin, an anticoagulant drug; phenothiazine, an antipsychotic drug; and several antineoplastic drugs. Exposure to some of these non-food chemicals can still be significant, but they were removed from the current inventory since exposure does not result from food consumption. There were 11 chemicals that are now banned in the US and 24 chemicals judged to be non-food due to lack of information on food-use. Examples include 1,2,3-trichloropropane, a solvent and intermediate in the manufacture of pesticides (no food tolerances or other food-relevant regulations were identified); chlorophene, a biocide and preservative used in cosmetics; and diethyl sulfate, an ethylating agent that is deemed a likely carcinogen.

Definitive assignments were made for 1321 of the 1530 chemicals evaluated, meaning current references were found for either use in or prohibition from foods in the US. The remaining 209 chemicals were categorized through inference and expert judgment due to a lack of use information, outdated information, or both. Of these, the majority (160 chemicals) were categorized as non-food, whereas 49 conservatively relied on postings in FDA food registries or GRAS listings for category assignment. An example of the latter is sucrose octaacetate, which is the bitter component of nail-biting formulations and an alcohol denaturant. However, it can be found on FEMA's GRAS List 3, issued in 1965, because of two reports at that time for use in beverages at up to 20 ppm (FEMA, 2016). Current use in food for sucrose octaacetate seems unlikely, but sucrose octaacetate is inferred to still be a direct food additive for the purposes of this study. Similar examples include flavors in apparent use according to the European Commission Fl@voring Substances database, such as dimethyl malonate or methyl decanoate, for which no evidence confirming use in the US was found. Conversely, an example of a chemical that was deemed non-food is oxycarboxin, a fungicide used in the US only on greenhouse ornamentals. Oxycarboxin did not undergo a dietary assessment by the EPA in the 2004 RED due to its rapid breakdown and limited exposure, even though it is a metabolite of carboxin, which is used on food crops. We inferred from this that oxycarboxin should be re-categorized from pesticide to non-food. Finally, the fungicide epoxiconazole represents an example of a chemical with no current use in the US yet it was kept in the food-use inventory because it does have residue tolerances set for coffee and bananas. For this example, exposure is limited but possible via imported foods, placing it in the pesticide residue category under the current categorization scheme.

3.3. Categorization of confirmed food-use chemicals

Manual curation confirmed current US food-use for 1211 ToxCast chemicals, which were categorized based on use and exposure likelihood into three categories: (1) direct additives, (2) indirect additives, and (3) pesticides and/or residues (Table 2). There were 556 chemicals classified as direct food additives based on likely exposure via current food-use in the United States. This result varied from the previous categorization of direct food additive chemicals when categorization was merely based on the resource database (Karmaus et al., 2016). To demonstrate the significant impact of manual curation for categorization the former assignments were compared to the current categorization results (Fig. 2). For example, 616 chemicals were formerly reported as direct additives (Karmaus et al., 2016), manual curation resulted in 54 of these chemicals now being identified as indirect food additives, 39 being classified as pesticides or residues, and 28 chemicals were deemed non-food (Fig. 2A). Meanwhile, the former food contact substances category of 371 chemicals (Karmaus et al., 2016) contributed 281 to the current indirect food additives category, 6 were newly classified as direct food additives, 46 were re-classified as pesticide residues, and 72 were deemed non-food (Fig. 2B). Finally, the former 543 chemicals in the pesticides category (Karmaus et al., 2016) resulted in the most non-food chemical designations, with 219 considered non-food use; meanwhile, 321 remain pesticide residues and 4 are now classified as indirect additives (Fig. 2C).

An important consideration for categorization using the scheme described herein is the allowance of multiple category assignments per chemical, when necessary. In total, 90 chemicals in the confirmed 1211 food-use ToxCast chemical inventory have multiple category designations: 20 are both direct and indirect food additives, 32 are both direct food additives and pesticide residues, and 38 are both indirect food additives and pesticide residues. Notably, 70 of the 406 chemicals categorized as pesticide residues herein were found to have other uses warranting classification in more than one category. This includes chemicals registered by the US EPA as inert ingredients of various pesticides that also had regulated uses as flavors, emulsifiers, or plasticizers. Accordingly, these chemicals were also listed in the appropriate direct or indirect food additive category in addition to being classified as pesticide residues. The US EPA usually exempts inert ingredients of pesticides from the requirement of a food tolerance because the significance or likelihood of exposure via food is low (40 CFR180.900) (EPA, 2016a). Such details on usage and categorization are all summarized in the detailed notes provided in the chemical inventory summary table (http://ilsina.org/curation-of-food-relevant-chemicals-in-toxcast/; © 2017 ILSI North America, all rights reserved). Specific examples include trimethylamine and *o*-cresol, which are GRAS flavors that can be added directly to foods but are also listed in FDA's indirect additive regulations. Similarly, propylene glycol is a polyester resin packaging precursor but is also listed by FDA as a direct GRAS additive. In such cases, multiple uses are food-relevant and important to consider, thus the chemicals were listed as both direct food additives as well as indirect additives.

3.4. Evaluation of overall bioactivity for food-use chemicals in ToxCast assays

Ultimately, the current project refined the categorization of food-use chemicals in the ToxCast program, with the intent of optimizing chemicals warranting evaluation if US food-use is of interest. Thus, we evaluated the overall bioactivity of these chemicals across the ToxCast assays.

Based on the hypothesis that non-specific bioactivity can be observed around testing concentrations near and greater than the concentration at the cytotoxicity center (Judson et al., 2016), it has been suggested that filtering *in vitro* activity by incorporating information about cytotoxicity may help identify assays in which a chemical is eliciting selective activity. As such, any assay in which a chemical's AC_{50} is above the determined cytotoxicity center (per chemical) would not be considered chemical-selective activity and disregarded. This approach ultimately helps to identify the more specific, chemical-mediated effects (Judson et al., 2016; Karmaus et al., 2016). To visualize the effect of such filtering, the total number of assays across all of ToxCast with concentration-dependent effects per chemical before and after cytotoxicity filtering was evaluated (Fig. 3). The inventory of 1211 food-use chemicals ranged from chemicals having elicited concentration-dependent effects in 0–255 assay endpoints across ToxCast. Chemicals that elicited

Table 2
Refined food-use categories for manual curation.

Category	Number of chemicals[a]	Criteria
Direct food additives	556	Chemicals directly added to food for functional purposes
Indirect food additives	339	Chemicals that may migrate to food from packaging, processing, or cleaning chemicals, etc.
Pesticide residues	406	Distinct from other indirect additives, these chemicals have no firm conclusion regarding the degree of exposure compared to other indirect additives (but registered pesticide uses generally are designed to minimize crop residues and finished foods often show no residues)

[a] The number of unique chemicals is 1211. However, 90 chemicals were designated to two categories: 20 are both direct and indirect food additives, 32 are both direct food additives and pesticide residues, and 38 are both indirect food additives and pesticide residues.

Table 3
Resources used for determining the current status of chemicals with regard to food-use in the United States.

	Resource	Reference
1	EPA ACToR database	EPA (2016b)
2	EPA REDs	EPA (2016f)
3	Encyclopedia of Food and Color Additives	Burdock (1996)
4	Legal Information Institute website	Cornell University (2016)
5	Fl@voring Substances database	European Commission (2016)
6	FEMA GRAS determination listings	FEMA (2016)
7	21 Code of Federal Regulations	
	21 CFR181 Prior Sanctioned Food Ingredients	FDA (2015a)
	21 CFR182 Substances GRAS in Foods	FDA (2015b)
	21 CFR184 Substances Affirmed as GRAS in Food	FDA (2015c)
	21 CFR186 Substances Affirmed as GRAS for Use in Food Packaging	FDA (2015d)
8	40 CFR180.900: Protection of Environment: Exemptions From the Requirement of a Tolerance	EPA (2016a)
9	FDA 2005 Glossary of Pesticide Chemicals	FDA (2005)
10	FDA EAFUS database	FDA (2013)
11	PAN Pesticides Database	PAN (2016)
12	EPA DSSTox database	EPA (2016d)
13	EPA Pesticide InertFinder	EPA (2016e)
14	FDA Inventory of Effective FCS Notifications	FDA (2015e)
15	FDA List of Indirect Additives Used in FCS	FDA (2015f)
16	Good Scents Company Ingredient Database	Good Scents Company (2016)
17	ChemBioFinder.com	PerkinElmer (2016)
18	Global MRL Database	USDA (2016)

Abbreviations: ACToR, Aggregated Computational Toxicology Resource; CFR, Code of Federal Regulations; DSSTox, Distributed Structure-Searchable Toxicity; EAFUS, Everything Added to Food in the United States; EC, European Commission; EPA, US Environmental Protection Agency; FCS, food contact substance; FDA, US Food and Drug Administration; FEMA, Flavor and Extract Manufacturers Association; GRAS, generally recognized as safe; MRL, maximum residue limit; PAN, Pesticide Action Network; RED, Reregistration Eligibility Decision; USDA, US Department of Agriculture.

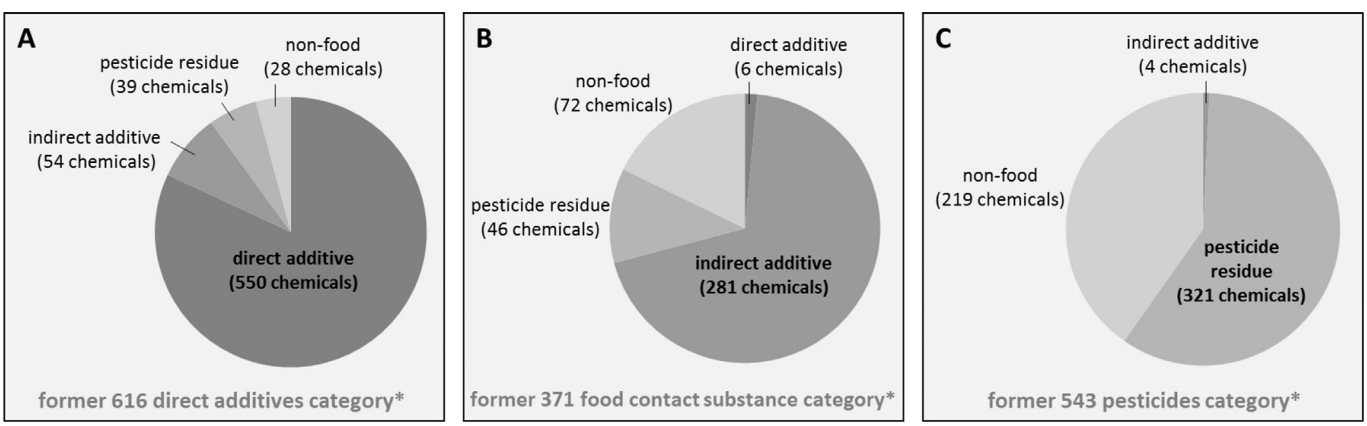

Fig. 2. Effects of manual curation on categorization. Manual curation of the 1530 chemicals formerly identified as food-relevant in ToxCast (Karmaus et al., 2016) resulted in 20%–40% of chemicals per category being re-categorized. (A) Breakdown of the former 616 direct additives demonstrating 550 of 616 remain direct additives with the current classification criteria, whereas 54 were re-assigned as indirect additives, 39 are now considered pesticide residues, and 28 were deemed non-food. (B) Similarly, the distribution of the 371 chemicals from the former food contact substances category. (C) Distribution of the formerly 543 chemicals designated as pesticides. *The former categories summarized in this figure refer to classification from Karmaus et al. (2016), which grouped chemicals into categories based solely on the database from which the chemical was sourced.

activity in very few assays typically did not result in cytotoxicity; as such, no cytotoxicity center was determined and the results were not affected by filtering. Conversely, those chemicals having concentration-dependent effects in >100 assays typically also elicited cytotoxic responses and the number of active endpoints was greatly affected by applying cytotoxicity filtering, such that approximately 20 or less assay endpoints remained after applying the filtering. This approach can help narrow down the assays that reflect chemical-mediated specific and potent effects. Interestingly, there is a visually discernable trend showing that direct additives were overall active in fewer assays compared to the pesticide residues which typically had a larger number of active assay endpoints. This trend is seen by noting that the large majority of green points for direct additives are on the far left of the plot versus red points for pesticide residues appearing to be generally enriched toward the right or to have high active assay counts (Fig. 3).

4. Discussion

The need for curation and contextualization in order to interpret and analyze large datasets was highlighted in the current study, demonstrating the importance of properly defining the context in which data shall be interpreted. For example, the current study sought to confirm a list of chemicals that were (1) food-use, (2) currently used or have tolerance in the United States, and (3) in the ToxCast inventory. These criteria incorporate current registration for food-use and potential exposure likelihood in the US. By incorporating exposure considerations into the classification of chemicals, the resulting updated chemical list, in combination with the ToxCast *in vitro* assay data, is more helpful in identifying and prioritizing food-use chemicals that may warrant further toxicological evaluation. However, these requirements were difficult to address given the historical or globally-sourced inventories integrated in publicly available large-scale chemical usage information resources. Ultimately, downloading complete database inventories can result in large chemical lists and an over-estimation of chemicals relevant to the context of interest, reiterating the need for clearly defined rationale to refine chemicals needing manual curation.

The identification and categorization of food-use chemicals in ToxCast as conducted by Karmaus et al. (2016) and by pulling food-relevant descriptors from CPCat both resulted in the inclusion of chemicals that were not relevant to current-day US food-use. This was demonstrated by the 122 chemicals that were manually identified as not current-day US food-use despite being among the overlapping inventory of 1276 chemicals identified by both methods. Reasons for such over-estimation include the fact that integrating complete inventories of databases can result in content beyond the desired context. For example, the inclusion of globally-sourced information or historical data in large repositories, these can be very useful in some cases and very misleading in other instances. As such, retrieved data may be dated or imply usage that is not relevant to exposure through foods, because many databases also serve as repositories of historical information and may not be routinely updated. While we have focused on current-day US-relevance, the data provided herein can be used as a starting point to identify global food-use chemicals in ToxCast. More specifically, by considering all CPCat-identified "food"-annotated ToxCast chemicals (ie. including the 473 chemicals that were omitted herein from Fig. 1) and adding chemicals deemed non-food by our manual curation due to only foreign use (identified in the detailed notes provided in manual curation at http://ilsina.org/curation-of-food-relevant-chemicals-in-toxcast; © 2017 ILSI North America, all rights reserved, one can begin identifying chemicals that have global food-use and not necessarily current US food-use.

Manual review revealed significant added value by providing context to the data retrieved from large inventories. Each of the three categories formerly defined by Karmaus et al., 2016, had 20%–40% of the chemicals re-categorized. Namely, the former pesticides category had the largest number of chemicals that were removed in the current categorization, with 219 chemicals becoming non-food. This shuffling of chemicals highlights how critical manual curation can be, particularly when specific criteria (e.g., exposure from food and current use in the US) are incorporated. For example, a search for a pesticide can bring up tolerances in 40 CFR180 that imply use (EPA, 2016a), but the document may be decades old with the latest CFR not listing the chemical at all. Furthermore, 40 CFR may not contain all pesticides that are used internationally but may have import tolerances; in such cases, only EPA resources (EPA, 2016a, 2016b, 2016e, 2016f) can confirm whether the pesticide's registration has been cancelled. In other cases, lack of documentation can pose challenges in identifying uses for chemicals, as was seen with 2-chloroacetophenone, which was pulled into the list of 1530 food-relevant ToxCast chemicals because it was in the "FDA GRAS" database. However, 2-chloroacetophenone is the active component in tear gas and Mace; no searches in the 21 CFR or other databases (Table 3) showed any food use for this chemical. This may be a cross-

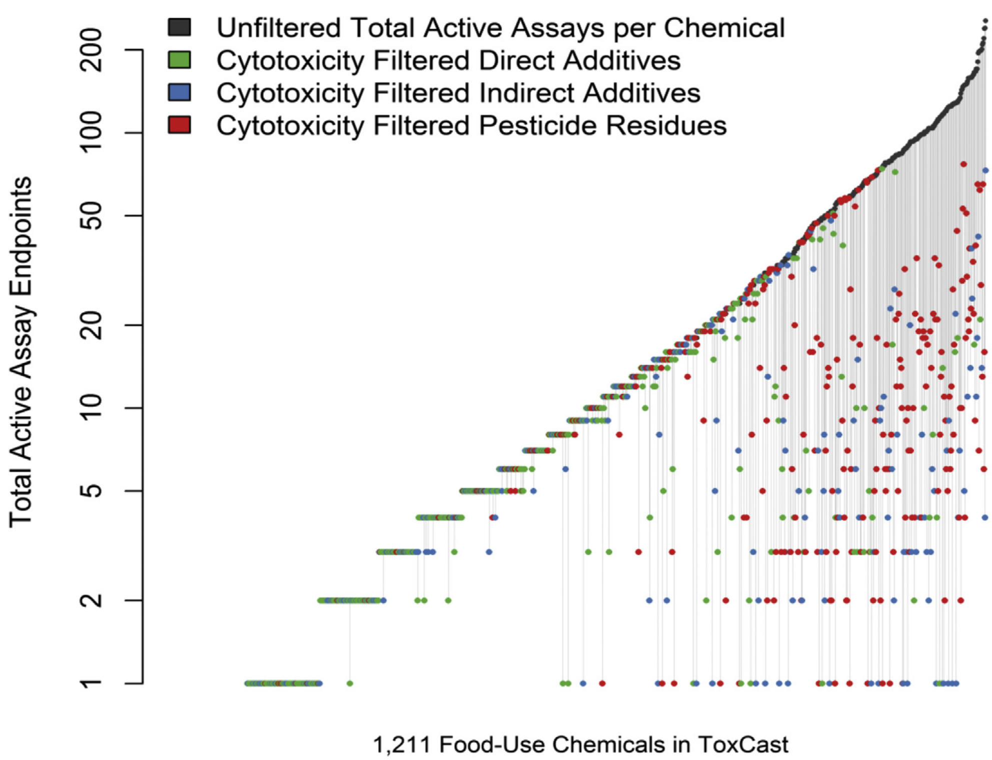

Fig. 3. Total assays per chemical before and after cytotoxicity filtering. For each of the 1211 food-use ToxCast chemicals (*x*-axis), a gray line indicates the connection between the total number of assays in which the chemical was active (black) and the cytotoxicity-filtered number of assays in which the chemical was active (colored circles). In several cases, there was no effect of cytotoxicity filtering, which resulted in the colored circle overlapping with the black one. Evaluation of cytotoxicity across 35 cytotoxicity assays in ToxCast was used to calculate cytotoxicity centers as previously described (Judson et al., 2016; Karmaus et al., 2016) for cytotoxicity filtering purposes. This analysis reveals that chemicals with >50 assays hit are generally also cytotoxic and most affected by cytotoxicity filtering. Furthermore, direct additives (green) are generally active in fewer assays and are not affected by cytotoxicity filtering, whereas pesticide residues (red) elicit activity in more assays and are most affected by cytotoxicity filtering. After incorporating the cytotoxicity filtering approach, the more potent and specific concentration-dependent activity was narrowed down to ~20 assays or less for nearly all the food-relevant chemicals. *Note*: For the illustrative purposes of this plot, only one category was allowed per chemical for coloring such that preference was given to direct additive > indirect additive > pesticide residue in cases where chemicals were assigned more than one category. The R script for analysis and generation of this figure are provided in Supplementary File S2.

listing error due to the common name similarity with the spice, mace. Some other chemicals included in the former inventory of ToxCast "food-relevant" list resulted from GRAS listings. It is critical to highlight that GRAS designations are specific to conditions of intended use and require manual evaluation. For example, sulfamic acid was specifically registered as being intended for indirect use only with no direct addition to foods. Thus, a categorical change for sulfamic acid from direct food additive to indirect additive was made and serves as a reminder that a "GRAS" designation does not automatically assume direct food addition.

The refined food-relevant ToxCast list presented herein is the first comprehensive manually curated inventory encompassing direct additives, indirect additives, and pesticides to incorporate exposure likelihood specifically from food-use in the US. While chemicals may be of high exposure likelihood through non-food routes or be widely-used overall; for the purpose of this inventory and categorization schema, only food-related use and exposure likelihood in the US were integrated. Evaluation of these chemicals across all assay endpoints revealed that over half of the 1211 food-use chemicals in ToxCast had concentration-dependent effects in $\geq$5 assay endpoints, even after refining for potent/specific chemical effects by using the cytotoxicity filtering approach. Yet, even with cytotoxicity filtering, it is important to emphasize that *in vitro* activity is not necessarily indicative of adverse effects *in vivo*. Ultimately, the manual curation approach presented herein was important for identifying chemicals of relevance with current food-use in the US and for prioritizing/categorizing those chemicals, all of which was not possible by simply mining publicly available large-scale chemical use databases.

5. Conclusions

The current study demonstrated that in order to reliably integrate and interpret chemical usage information in large-scale studies, manual curation and thorough review of chemical inventories obtained from database mining is needed. In this case, we have focused on food-relevant chemicals, which encompass a wide diversity of functions from direct additives to indirect food contact and pesticide uses. Although Karmaus et al. (2016) were inclusive and compiled inventories from 8 different US food-relevant chemical data sources, and despite CPCat integrating chemical usage annotation from globally-sourced inventories, manual curation reveals that by comprehensively incorporating up-to-date information the categorization of these chemicals was significantly refined. Most notable considerations included evaluating intended use guidelines for GRAS inventories and integrating updated registration status for older inventory entries. With the ToxCast data being publicly available and serving as a resource of screening data for thousands of chemicals, it is critical to have a thorough foundation/context for identifying chemicals of concern. The final inventory of 1211 food-relevant chemicals in ToxCast is a valuable reference that can help prioritize food-use chemicals of interest and can provide needed context on the use of chemicals included in this high-throughput screening program.

Acknowledgments

The authors thank Christina West for the critical review of this manuscript. T.D.T. received funds from the ILSI North America Technical Committee on Food and Chemical Safety for his work on this manuscript. ILSI North America is a public, nonprofit foundation that provides a forum to advance the understanding of scientific issues related to the nutritional quality and safety of the food supply by sponsoring research programs, educational seminars and workshops, and publications. ILSI North America receives support primarily from its industry membership. The views expressed in this study are those of the authors.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.fct.2017.03.006.

Transparency document

Transparency document related to this article can be found online at http://dx.doi.org/10.1016/j.fct.2017.03.006.

References

Burdock, G., 1996. Encyclopedia of Food & Color Additives. CRC Press, Boca Raton, FL.

Cornell University, 2016. Cornell University Legal Information Institute (accessed 30 March 2016). https://www.law.cornell.edu.

Dionisio, K.L., Frame, A.M., Goldsmith, M.-R., Wambaugh, J.F., Liddell, A., Cathey, T., Smith, D., Vail, J., Ernstoff, A.S., Fantke, P., 2015. Exploring consumer exposure pathways and patterns of use for chemicals in the environment. Toxicol. Rep. 2, 228–237.

Edwards, S.W., Tan, Y.-M., Villeneuve, D.L., Meek, M., McQueen, C.A., 2016. Adverse outcome pathways—organizing toxicological information to improve decision making. J. Pharmacol. Exp. Ther. 356, 170–181.

Environmental Protection Agency (EPA), 2016a. 40CFR180.900 Protection of Environment: Exemptions from the Requirement of a Tolerance (accessed 30 March 2016). http://www.ecfr.gov/cgi-bin/text-idx?tpl=/ecfrbrowse/Title40/40cfr180_main_02.tpl.

Environmental Protection Agency (EPA), 2016b. ACToR® Database (Aggregated Computational Toxicology Resource) (accessed 30 March 2016). https://actor.epa.gov.

Environmental Protection Agency (EPA), 2016c. CPCat: Chemical and Product Categories (accessed 30 March 2016). https://actor.epa.gov/cpcat/faces/home.xhtml.

Environmental Protection Agency (EPA), 2016d. Distributed Structure-searchable Toxicity (DSSTox) Database (accessed 30 March 2016). https://www.epa.gov/chemical-research/distributed-structure-searchable-toxicity-dsstox-database.

Environmental Protection Agency (EPA), 2016e. Pesticide Inert Finder Database (accessed 30 March 2016). http://www2.epa.gov/pesticide-registration/inert-ingredients-overview-and-guidance.

Environmental Protection Agency (EPA), 2016f. Reregistration Eligibility Decision Documents (REDs) (accessed 30 March 2016). https://www.epa.gov/ingredients-used-pesticide-products/legacy-chemical-information-decisions-and-documents.

Environmental Protection Agency (EPA), 2016g. Toxicity ForeCaster (ToxCast™) Data (accessed 30 March 2016). https://www.epa.gov/chemical-research/toxicity-forecasting.

Environmental Protection Agency (EPA), 2016h. Toxicology Testing in the 21st Century (Tox21) (accessed 30 March 2016). https://www.epa.gov/chemical-research/toxicology-testing-21st-century-tox21.

European Commission, 2016. Fl@voring Substances Database (accessed 30 March 2016). http://ec.europa.eu/food/food/chemicalsafety/flavouring/database/index.cfm.

Flavor and Extract Manufacturers Association (FEMA), 2016. Flavor and Extract Manufacturers Association GRAS Determination Listings (accessed 30 March 2016). http://www.femaflavor.org/gras.

Food and Drug Administration (FDA), 2005. Glossary of Pesticide Chemicals (accessed 30 March 2016). http://www.fda.gov/Food/FoodborneIllnessContaminants/Pesticides/ucm113891.htm.

Food and Drug Administration (FDA), 2013. Everything Added to Food in the United States (EAFUS) (accessed 30 March 2016). http://www.accessdata.fda.gov/scripts/fcn/fcnNavigation.cfm?rpt=eafusListing.

Food and Drug Administration (FDA), 2015a. 21CFR181 Prior Sanctioned Food Ingredients (accessed 30 March 2016). https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm.

Food and Drug Administration (FDA), 2015b. 21CFR182 Substances GRAS in Foods (accessed 30 March 2016). https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm.

Food and Drug Administration (FDA), 2015c. 21CFR184 Substances Affirmed as GRAS in Food (accessed 30 March 2016). https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm.

Food and Drug Administration (FDA), 2015d. 21CFR186 Substances Affirmed as GRAS for Use in Food Packaging (accessed 30 March 2016). https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm.

Food and Drug Administration (FDA), 2015e. Inventory of Effective Food Contact Substances (FCS) Notifications (accessed 30 March 2016). http://www.fda.gov/Food/IngredientsPackagingLabeling/PackagingFCS/Notifications/ucm116567.htm.

Food and Drug Administration (FDA), 2015f. List of Indirect Additives Used in Food Contact Substances (accessed 30 March 2016). http://www.fda.gov/Food/IngredientsPackagingLabeling/PackagingFCS/IndirectAdditives/ucm115333.htm.

Food and Drug Administration (FDA), 2016. GRAS Notice Inventory (accessed 30 March 2016). http://www.accessdata.fda.gov/scripts/fdcc/?set=GRASNotices.

Goldsmith, M.-R., Grulke, C., Brooks, R., Transue, T., Tan, Y., Frame, A., Egeghy, P., Edwards, R., Chang, D., Tornero-Velez, R., 2014. Development of a consumer product ingredient database for chemical exposure screening and prioritization. Food Chem. Toxicol. 65, 269–279.

Good Scents Company, 2016. The Good Scents Company Ingredient Database (accessed 30 March 2016). http://www.thegoodscentscompany.com/.

Judson, R., Houck, K., Martin, M., Richard, A.M., Knudsen, T.B., Shah, I., Little, S., Wambaugh, J., Woodrow Setzer, R., Kothya, P., Phuong, J., Filer, D., Smith, D., Reif, D., Rotroff, D., Kleinstreuer, N., Sipes, N., Xia, M., Huang, R., Crofton, K., Thomas, R.S., 2016. Editor's highlight: 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.

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., Chandler, K., Houck, K., Hunter, S., Judson, R., Kleinstreuer, N., Knudsen, T., Martin, M., Padilla, S., Reif, D., Richard, A., Rotroff, D., Sipes, N., Dix, D., 2012. Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management. Chem. Res. Toxicol. 25, 1287–1302.

Lü, J.-M., Nurko, J., Weakley, S.M., Jiang, J., Kougias, P., Lin, P.H., Yao, Q., Chen, C., 2010. Molecular mechanisms and clinical applications of nordihydroguaiaretic acid (NDGA) and its derivatives: an update. Med. Sci. Monit. 16. RA93–R100.

PerkinElmer, 2016. ChemBioFinder.com (accessed 30 March 2016). http://www.cambridgesoft.com/databases/login/?serviceid=128.

Pesticide Action Network (PAN), 2016. PAN Pesticides Database (accessed 30 March 2016). http://www.pesticideinfo.org/Search_Chemicals.jsp.

Teeguarden, J.G., Tan, Y.-M., Edwards, S.W., Leonard, J.A., Anderson, K.A., Corley, R.A., Kile, M.L., Simonich, S.M., Stone, D., Tanguay, R.L., 2016. Completing the link between exposure science and toxicology for improved environmental health decision making: the aggregate exposure pathway framework. Environ. Sci. Technol. 50, 4579–4586.

US Department of Agriculture (USDA), 2016. Global MRL Database (accessed 30 March 2016). https://www.globalmrl.com/db#query.

Wambaugh, J.F., Setzer, R.W., Reif, D.M., Gangwal, S., Mitchell-Blackwood, J., Arnot, J.A., Joliet, O., Frame, A., Rabinowitz, J., Knudsen, T.B., 2013. High-throughput models for exposure-based chemical prioritization in the ExpoCast project. Environ. Sci. Technol. 47, 8479–8488.

World Health Organization (WHO), 2016. Concise International Chemical Assessment Documents (accessed 30 March 2016). http://www.who.int/ipcs/publications/cicad.