Identification, Curation, and Prioritization of Food-Use Chemicals In ToxCast

M Krishan[1], TD Trautman[2], DF Kapraun[3], LA Fix[4], and AL Karmaus[5]

[1]ILSI NA, Washington, DC, USA; [2]Retired, General Mills, Minneapolis, MN, USA; [3]EPA/ORD/NCCT, RTP, NC, USA ; [4]Unilever, Englewood Cliffs, NJ, USA; [1]ILS, RTP, NC, USA

ABSTRACT

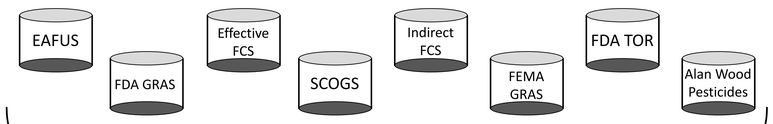
Evaluating the thousands of chemicals that are directly added to or come in contact with food poses a great challenge due to the time, cost, and sheer volume of data necessary to thoroughly conduct comprehensive toxicological testing. This study compiled a list of food-use chemicals in the United States (U.S.) and demonstrates approaches amenable to the evaluation of this large and diverse chemical inventory. 8,659 unique food-use chemicals were compiled from 8 public sources and mined against the ToxCast in vitro high-throughput screening inventory to identify 1,530 food-use chemicals with in vitro assay data. Each of these chemicals was then manually evaluated for current registration status and categorized based on exposure likelihood from food in the U.S. into four categories: direct food additive, indirect additive, pesticide/residue, or non-food. Ultimately, 319 chemicals were categorized as non-food and removed from the list, leaving 556 direct additives, 339 indirect additives, and 406 pesticides/residues. The cytotoxicity elicited by the curated list of food-use chemicals in ToxCast revealed that only 10% of direct additives elicited cytotoxicity, while 24% of indirect additives and 41% of pesticides/residues were cytotoxic. To address the need to prioritize chemical mixtures, we used frequent itemset mining (FIM) to identify which individual chemicals or combination of chemicals appear with the greatest frequency in the U.S. Food and Drug Administration's Effective Food Contact Substances (EFCS) database. The EFCS database comprises 978 registrations, containing 715 unique chemicals; in total, 189 of the registrations contain mixtures of ≥2 chemicals and 110 registrations contain ≥3 chemicals, up to a maximum of one registration containing 9 chemicals. The FIM approach identified acetic acid, hydrogen peroxide, and peracetic acid as the most abundant co-occurring chemicals in EFCS registrations, each occurring in 4.3% (42/978) of registrations. The combination of peracetic acid and hydrogen peroxide occurred most frequently, appearing in 21% (40/189) of registered mixtures. The current inventory and analysis of ToxCast cytotoxicity and EFCS mixture prioritization represent the first evaluation of food-use chemicals on this scale, providing insight into this overlooked but critical chemical inventory.

INTRODUCTION AND SPECIFIC AIMS

A large diversity of chemicals can end up in food. These chemicals come from a variety of sources including direct addition to food, food contact from packaging or handling, and residues from food production. The mechanisms documenting and regulating such food-relevant chemicals are also diverse, with no single resource available where such chemicals are listed in one place. Furthermore, many chemicals lack exposure or toxicological data. This study sought to address these challenges by:

- Compiling a list of relevant food-use chemicals
- Curating the list for exposure likelihood from food in the USA
- Evaluating any available ToxCast high-throughput screening data for the identified food-use chemicals
- Prioritizing food-use chemical mixtures warranting further evaluation by mining food contact notifications

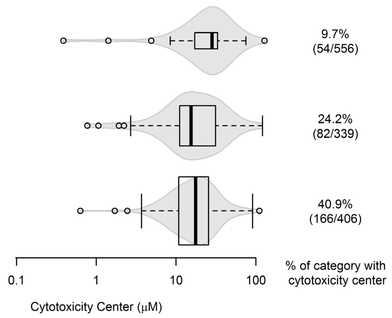
COMPILATION OF A FOOD-USE CHEMICAL INVENTORY



Inventory Source	Entries in Inventory	CASRN in Inventory	Putative Use Category	CASRN in Use Category	CASRN in ToxCast
FDA EAFUS	3968	3277	**Direct Food Additives**	**3888**	**616**
FDA SCOGS	378	320			
FDA GRAS Notices	603	349			
FEMA GRAS	2796	2659			
FDA Effective FCS	1205	715	**Food Contact Substances**	**3039**	**371**
FDA Indirect in FCS	3229	2555			
FDA TOR	50	56			
Alan Wood Pesticides	1813	1808	**Pesticides**	**1732**	**543**
TOTAL	14,042	11,733	**TOTAL**	**8,659**	**1530**

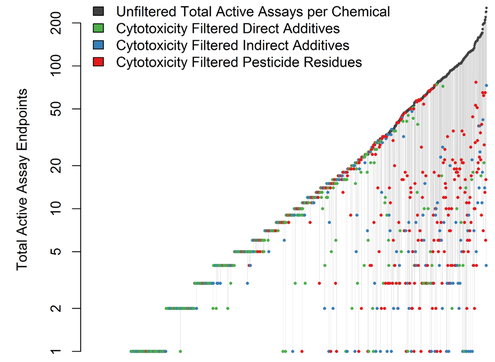
▲ **Figure 1: Compiling food-use chemicals from publicly available resources.** The complete inventories from 8 publicly available resources were mined for chemicals, identifying 8,659 unique chemicals, which were grouped into three putative categories: (1) direct food additives, (2) food contact substances, and (3) pesticides. For chemicals appearing in multiple resources, duplicates were removed and preference was given to classification as a direct food additive > food contact substance > pesticide. Of this food-relevant chemical inventory, 1530 were in ToxCast. Abbreviations: FDA, Food and Drug Administration; EAFUS, Everything Added to Food in the United States; SCOGS, Select Committee on GRAS Substances; GRAS, Generally Recognized As Safe; FCS, Food Contact Substances; TOR, Threshold of Regulation Exemptions.

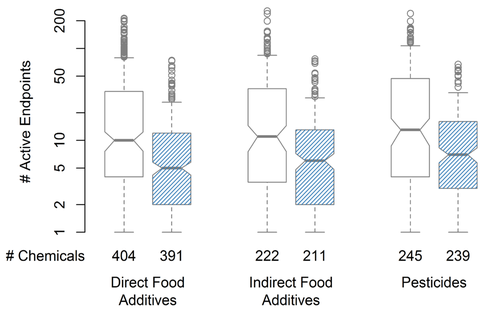
EVALUATION OF FOOD-USE CHEMICALS IN TOXCAST



◀ **Figure 3: Evaluation of cytotoxicity in ToxCast across food-relevant chemical categories**. The curated ToxCast food-use chemical inventory was evaluated for cytotoxicity across 35 assays. While all 1211 food-use chemicals were evaluated in at least 14 out of the 35 cytotoxicity assays, chemicals were required to elicit concentration-dependent cytotoxicity in at least three assays in order for a cytotoxicity center to be determined. This analysis reveals that a far lower proportion of direct food additives elicited cytotoxicity (10%) than either indirect food additives (24%) or pesticides/residues (41%). These proportions are visualized as the width of the boxplots, while the shaded area represents the density of chemicals in the distribution. The box edges represent the 25th and 75th percentiles, the bold center represents the mean, and the whiskers denote the 5th and 95th percentiles. Chemicals that were outliers to the distribution are plotted as individual points.

▶ **Figure 4: Total assay endpoints active per chemical before and after cytotoxicity filtering**. For each of the 1211 food-use chemicals in ToxCast, a gray line indicates the connection between the total number of assay endpoints in which the chemical was active (black point) and the cytotoxicity filtered number assay endpoints (colored points). In several cases where there was no effect of filtering, the colored point is overlapping with the black one. In general, chemicals with >50 assays hit are also cytotoxic and most affected by cytotoxicity filtering. Furthermore, direct additives (green) are active in fewer assay endpoints overall and less affected by the filtering versus pesticides and residues (red) which elicit activity in more assays are most affected by filtering. For the purpose of this plot, the chemicals were listed across the x-axis in order of increasing number of unfiltered active assays endpoints, and only one category was allowed per chemical.





◀ **Figure 5: Effect of cytotoxicity filtering across food-relevant chemical categories**. The curated ToxCast food-use chemical inventory was evaluated for the effect of cytotoxicity filtering. Briefly, the distribution for total number of active assay endpoints per chemical in each category was summarized using a boxplot. White boxplots show the total number of assay endpoints per chemical before cytotoxicity filtering, whereas the blue shaded boxplots summarize cytotoxicity filtered sums. Filtering impacts all categories, with the means reflecting a trend of a larger number of assay endpoints active after filtering for pesticides than for either indirect additives or direct additives. The boxplots represent the 25th-75th percentiles, with the mean represented by the bold line in the center of the boxplot; the whiskers mark the 5th and 95th percentiles. Chemicals outside this distribution are represented by hollow circles.

CURATION OF THE FOOD-USE CHEMICAL INVENTORY IN TOXCAST

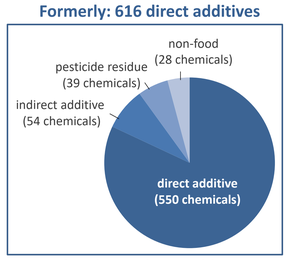
▶ **Figure 2: Manual curation of the food-use chemicals in ToxCast.** Focusing on the 1530 food-relevant chemicals in ToxCast identified from the compilation of publicly available databases (summarized in Figure 1), manual curation was conducted to assess current-day food-use in the United States. Subsequently, chemicals were re-categorized based on exposure likelihood and current use information. Across the former putative categories, 20%-40% of chemicals were re-categorized, 90 chemicals were assigned to two categories based on usage, and a total of 319 were removed based on no apparent application in foods an no expected exposure from food in current-day United States. Ultimately, 1211 ToxCast chemicals remain "food-relevant": 556 are direct additives added to food for functional purposes and expected to have the highest food exposure likelihood; 339 are indirect food additives that may migrate to food from packaging, processing, or handling; and 406 are pesticides or residues generally presenting the lowest exposure likelihood from food and having no firm conclusion regarding degree of exposure with finished foods often showing no registered residue concern.

1530 food-relevant chemicals in ToxCast identified by database mining

↓

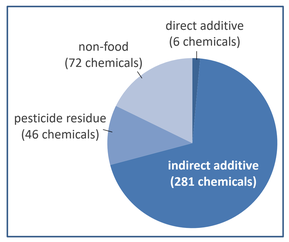
1,211 chemicals confirmed as having current-day food-use in the US, 319 deemed non-food after manual curation

Formerly: 616 direct additives



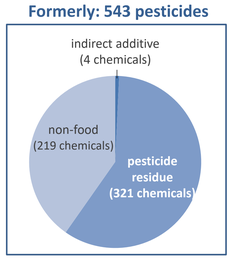
Refined: 556 direct additives

Formerly: 371 food contact substances



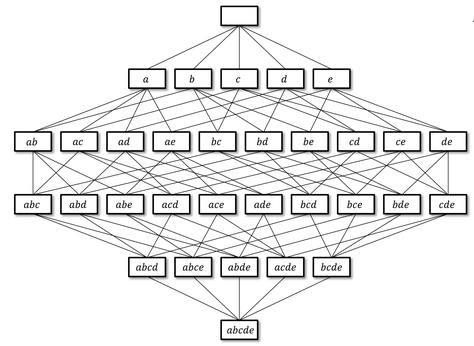
Refined: 339 indirect additives

Formerly: 543 pesticides



Refined: 406 pesticides and residues

FREQUENT ITEMSET MINING ANALYSIS



◀ **Figure 6: Hasse diagram.** This type of diagram depicts all possible combinations that can be formed from a given set of chemicals and illustrates the Apriori property of frequent itemsets. In the Hasse diagram at left, we consider a mixture comprised of chemicals *a*, *b*, *c*, *d*, and *e*. In the first tier, all possible single chemical subsets of the mixture are listed, in the second tier all possible binary (two chemical) subsets are shown, etc. Most FIM algorithms operate on the Apriori property: no superset of an infrequent itemset can be frequent. So, for example, if it were determined that $\{d, e\}$ is infrequent, there would be no need to calculate the frequencies of $\{a, d, e\}$, $\{b, d, e\}$, $\{c, d, e\}$, $\{a, b, d, e\}$, $\{a, c, d, e\}$, $\{b, c, d, e\}$, or $\{a, b, c, d, e\}$, as all of these are necessarily infrequent (as shown in the hierarchy in the Hasse diagram). Diagram adapted from Borgelt 2012 (*WIREs Data Mining Knowl Discov*. 2: 437-456).

The Food Contact Substance (FCS) Notice inventory contains 978 FCS notices, which are comprised of 715 unique chemicals. In total, 189 of the 978 notices contain more than one chemical, more specifically 110 contain three or more chemicals up to a maximum of 9 chemicals in one registered FCS notice. FIM was applied to the FCS Notice inventory to identify the individual chemical and the combination of chemicals that were most frequent in the inventory. Table 1 summarizes the most frequent mixtures, and Table 2 summarizes the individual chemicals from the FCS notice inventory that are in ToxCast.

Table 1: Mixtures with the highest frequency in the Food Contact Substance (FCS) Notices inventory

Chemicals in FCS notice	Frequency
Peracetic acid, Hydrogen peroxide	40/189
Acetic acid, Peracetic acid	39/189
Acetic acid, Peracetic acid, Hydrogen peroxide	38/189
Etidronic acid, Peracetic acid	37/189
Etidronic acid, Peracetic acid, Acetic acid	35/189
Etidronic acid, Peracetic acid, Hydrogen peroxide	35/189
Etidronic acid, Peracetic acid, Acetic acid, Hydrogen peroxide	34/189
Sulfuric acid, Peracetic acid, Acetic acid, Hydrogen peroxide	20/189
Sulfuric acid, Etidronic acid, Peracetic acid, Acetic acid, Hydrogen peroxide	17/189

Table 2: Individual chemicals in ToxCast with the highest frequency in the FCS Notices inventory

Chemicals in FCS notices and ToxCast	Frequency	Usage information
Acetic acid	39/978	Affirmed GRAS for food use
5-Chloro-2-methyl-3(2H)-isothiazolone	7/978	Antimicrobial for aqueous systems
2,6-Pyridinedicarboxylic acid	5/978	Biocide; food contact surface sanitation
2,5-Dimethyl-2,5-di-(tert-butylperoxy)hexane	4/978	Organic peroxide for crosslinking elastomers, silicone rubbers and thermoplastics
Glycidyl methacrylate	4/978	Production of copolymers (adhesives and resins)
Glycerol	2/978	Affirmed GRAS for food use
2-Ethylhexyl acrylate	2/978	Production of copolymers
Isophorone diisocyanate	2/978	Production of polymers (high-temp and resins)

**Note: an additional 20 chemicals with a frequency of 1/978 were also in ToxCast

SUMMARY

This work represents the first comprehensive evaluation of food-relevant chemicals in the ToxCast inventory. To accomplish this task the entire contents of publicly available databases were compiled, identifying 8659 unique chemicals. This inventory was mined against the chemicals evaluated as part of the ToxCast high-throughput screening program, revealing that 1530 food-relevant chemicals were evaluated in ToxCast. Manual curation was conducted to review the list of 1530 to ensure current-day use in the US and that up-to-date information was used for chemical categorization.

- In total, 1211 chemicals confirmed as current food-use in the US were represented in ToxCast.
  - These 1211 chemicals can be divided into 556 direct food additives, 339 indirect food additives, and 406 pesticides/residues based on usage and exposure likelihood in the US.
- Evaluation of results across 35 cytotoxicity assays in ToxCast reveals that only 10% of direct food additives elicit any cytotoxicity, whereas 24% of indirect food additives and 41% of pesticides/residues elicit cytotoxicity.
  - The mean cytotoxicity across use categories suggests direct food additives are less potently cytotoxic than indirect food additives or pesticides.
- By applying a cytotoxicity filtering approach wherein only assay endpoints with an AC50 below the cytotoxicity center are considered, the average number of active endpoints per chemical was reduced to an average of ~5 assay endpoints/chemical. This effect is most pronounced with pesticides/residues, which had the greatest number of active endpoints but also elicited the most potent cytotoxicity.
- Frequent itemset mining analysis was used to identify the most frequent mixtures of chemicals in the Food Contact Substance Notices inventory. Acetic acid was the most highly represented ToxCast chemical in the FCS Notices inventory.

This poster does not reflect EPA policy.

This work was partially supported by the ILSI North America Technical Committee on Food and Chemical Safety.

Agnes Karmaus
akarmaus@ils-inc.com
919-281-1110 x254