Food and Chemical Toxicology 92 (2016) 188–196

Contents lists available at ScienceDirect

Food and Chemical Toxicology

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

Evaluation of food-relevant chemicals in the ToxCast high-throughput screening program

Agnes L. Karmaus [a, b, *], Dayne L. Filer [b], Matthew T. Martin [c], Keith A. Houck [c]

[a] ILSI North America Technical Committee on Food and Chemical Safety Summer Fellow, Washington, DC, USA
[b] Oak Ridge Institue for Science and Education (ORISE) Fellow at the US Environmental Protection Agency, Office of Research and Development, National Center for Computational Toxicology, Research Triangle Park, NC, USA
[c] US Environmental Protection Agency, Office of Research and Development, National Center for Computational Toxicology, Research Triangle Park, NC, USA

ARTICLE INFO

Article history:
Received 16 December 2015
Received in revised form
14 April 2016
Accepted 16 April 2016
Available online 19 April 2016

Keywords:
ToxCast
Food additive
Food contact substance
Pesticide
High-throughput screening

ABSTRACT

Thousands of chemicals are directly added to or come in contact with food, many of which have undergone little to no toxicological evaluation. The landscape of the food-relevant chemical universe was evaluated using cheminformatics, and subsequently the bioactivity of food-relevant chemicals across the publicly available ToxCast highthroughput screening program was assessed. In total, 8659 food-relevant chemicals were compiled including direct food additives, food contact substances, and pesticides. Of these food-relevant chemicals, 4719 had curated structure definition files amenable to defining chemical fingerprints, which were used to cluster chemicals using a selforganizing map approach. Pesticides, and direct food additives clustered apart from one another with food contact substances generally in between, supporting that these categories not only reflect different uses but also distinct chemistries. Subsequently, 1530 food-relevant chemicals were identified in ToxCast comprising 616 direct food additives, 371 food contact substances, and 543 pesticides. Bioactivity across ToxCast was filtered for cytotoxicity to identify selective chemical effects. Initiating analyses from strictly chemical-based methodology or bioactivity/cytotoxicity-driven evaluation presents unbiased approaches for prioritizing chemicals. Although bioactivity *in vitro* is not necessarily predictive of adverse effects *in vivo*, these data provide insight into chemical properties and cellular targets through which foodrelevant chemicals elicit bioactivity.

1. Introduction

An estimated ~10,000 chemicals are directly or indirectly added to food in the United States, serving to enhance and preserve the taste and appearance of foods, prevent spoilage, or act as packaging constituents (Neltner et al., 2013). The addition of such chemicals to human food is allowed by the US Food and Drug Administration (FDA) under the 1958 US Food Additives Amendment. The safe use of new chemicals added to food is determined based on "reasonable certainty in the minds of competent scientists that the substance is not harmful under the intended conditions of use" (21 CFR §170.30) (Rulis and Levitt, 2009). However, for additives commonly used before 1958, safety may have been based on past use/experience rather than scientific data (12 USC §321) (Burdock and Carabin, 2004). Thus, with thousands of food-relevant chemicals approved for use in food, and as many as ~70% of direct additives having minimal to no toxicological guideline study data (Neltner et al., 2013), approaches that offer rapid evaluation to help inform, maintain, and support food safety are needed.

In vitro high-throughput screening (HTS) assays offer a time- and cost-effective platform for the evaluation of large chemical libraries (National Research Council, 2007). Such assays, in combination with computational approaches, provide an opportunity to rapidly gain insight into chemical-elicited effects on biochemical endpoints, cellular processes, and phenotypes as well as support

Abbreviations: CASRN, chemical abstract services registration number; EAFUS, Everything Added to Food in the US; EPA, US Environmental Protection Agency; ER, estrogen receptor; FDA, US Food and Drug Administration; FEMA, Flavor & Extract Manufacturers Association; GRAS, generally recognized as safe; HTS, high-throughput screening; MIE, molecular initiating event; NDGA, nordihydroguaiaretic acid; QSAR, quantitative structure–activity relationship; SDF, structure definition file; SMILES, simplified molecular input line entry system code; SOM, self-organizing map.

* Corresponding author. ILSI North America, 1156 15th Street NW, Suite 200, Washington, DC 20005, USA.
E-mail address: agnes.karmaus@gmail.com (A.L. Karmaus).

http://dx.doi.org/10.1016/j.fct.2016.04.012

the 3 R's (replacement, reduction, and refinement) of animal use in toxicological testing (Russell and Burch, 1958; Ankley et al., 2010b; Thomas et al., 2013). Using HTS data to inform on bioactivity can have great potential impact on both product development and safety testing (i.e., provide a platform for cost-effective high-throughput hazard identification). A significant example is the US Environmental Protection Agency's (EPA) ongoing ToxCast HTS program, which has evaluated a library of over 3000 chemicals in concentration-response format across over 1000 targeted *in vitro* assay endpoints to assess bioactivity *in vitro* (Dix et al., 2007; Kavlock et al., 2012; EPA, 2016b). A subset of 800 chemicals in ToxCast, termed the E1K library, were specifically screened across a subset of endocrine-related endpoints, highlighting that not all chemicals were evaluated in all ToxCast assays. The ToxCast assays cover a broad spectrum of chemical effects including enzyme inhibition, interaction with receptors, induction of cell stress pathways, and overt cytotoxicity (Kavlock et al., 2012).

The study presented herein is the first to evaluate strictly food-relevant chemicals across the entire ToxCast HTS program. Initially, a comprehensive inventory of food-relevant chemicals was compiled identifying 8659 food-relevant chemicals that was divided into three lists based not only on use but also chemistry: (1) direct food additives, (2) food contact substances, and (3) pesticides. The compiled food-relevant chemical list was then mined against the entire ToxCast chemical inventory identifying 1530 food-relevant chemicals evaluated in ToxCast. The bioactivity of these 1530 chemicals across all tested assay endpoints was assessed, and we demonstrated that filtering bioactivity using cytotoxicity can help hone in on potential selective chemical-mediated bioactivity to aid in prioritization and characterization of chemical effects. Combined, the results suggest that large HTS programs such as ToxCast are a valuable resource that can help inform on chemical prioritization and can have potential use as support for food safety testing.

2. Materials and methods

2.1. Identification of food-relevant chemicals

The inventory from publicly accessible databases was mined for chemicals, identified by their chemical abstract services registration numbers (CASRNs), to compile a comprehensive list of chemicals having any use associated with food. Accessed databases included the following FDA resources: Everything Added to Food in the US (EAFUS) (FDA, 2016a); Generally Recognized as Safe (GRAS) Notice Inventory (FDA, 2016a); Select Committee on GRAS Substance Database (SCOGS) (FDA, 2016c); List of Indirect Additives Used in Food Contact Substances (FDA, 2015); Inventory of Effective Food Contact Substances (FDA, 2016b); and Threshold of Regulation (TOR) Exemptions (FDA, 2016d). In addition, the Flavor & Extract Manufacturers Association GRAS inventory (FEMA, 2016) and the Aland Wood Pesticide database comprising active ingredients in pesticides which were assumed to be food use for the purpose of this study (Wood, 2015) were also included. Any defined chemical mixtures encountered were separated into the individual components and listed as unique CASRN for the purposes of this study. The compiled list of food-relevant chemicals including all source lists are summarized in Supplementary File S1, and a summary of the inventories is provided in Table 1. The food-relevant chemical list was cross-referenced against the entire publicly available ToxCast program chemical inventory comprising 3784 chemicals (EPA, 2016b). More specifically, the "ToxCast & Tox21 Chemicals Distributed Structure Searchable Toxicity Database (DSSTox files)" dataset (DSSTox_20151019 released October 2015) was downloaded; chemicals evaluated in ToxCast were obtained from "DSSTox_ToxCastRelease_20151019.xlsx".

2.2. Chemical clustering

Manually curated, high-quality, quantitative structure–activity relationship (QSAR)–ready simplified molecular input line entry system codes (SMILES) curated by Mansouri et al. were obtained from DSSTox for 4719 of the 8659 food-relevant chemicals (EPA, 2015; Mansouri et al., 2016). More specifically, the DSSTox Data was downloaded and SMILES were retrieved from the "DSSTox-All_20151019.xlsx" file. DSSTox does not contain SMILES for metals, polymers, and unstable stereoisomers as they were not amenable to the requirements for QSAR-ready structure definition file (SDF) generation, and were omitted from these analyses. Furthermore, it is important to note that while SMILES may exist for more of the food-relevant chemicals, the current study only obtained SMILES from DSSTox for consistency and reliability as DSSTox is a trustworthy manually curated resource. Using the rcdk package in R software (Guha, 2007), the SMILES were used to generate SDFs from which fingerprints were subsequently calculated using the same rcdk package. The generated molecular fingerprints describe a chemical's structure in a series of zeros or ones representing the presence or absence of a substructure descriptor which were defined using two descriptor sets: MACCS comprised of 166 descriptors and PubChem comprised of 881 descriptors. The MACCS descriptors are commonly used to evaluate chemical similarity describing general chemical substructure features, the PubChem fingerprints also describe substructural features summarizing a diversity of structural valence-bond forms. In total, 874 descriptors were associated with at least one food-relevant chemical and hence included for analysis (162 from MACCS and 712 from PubChem). The kohonen package in R (Wehrens and Buydens, 2007) was used to cluster the chemicals based on fingerprints across the 874 descriptors provided to the algorithm to form a self-organizing map (SOM), which groups the most similar chemicals together and displays cluster relationships in map form. The SOM generated from the 4719 chemicals was used to visualize chemical use categories as well as the ToxCast results. Supplementary File S2 provides the CASRN and SMILES for the 4719 chemicals in the SOM as well as which bin each chemical was in after clustering. Supplementary File S3 provides performance metrics from the SOM clustering. All analyses were conducted in R v3.1.3, with all scripts including analysis and each figure's code compiled into the source package "karmaus.fct.2016" attached as Supplementary File S4.

2.3. ToxCast HTS data

ToxCast data were retrieved from the publicly available download files (EPA, 2016b). For reproducibility, a self-contained R package with all pertinent data, analysis scripts, and figure generation scripts is provided as Supplementary File S4. Using the karmaus.fct.2016 R package, all data can be viewed, and all analyses and figures can be reproduced. To create this package, the "MySQL Database" (invitrodb_v2, released in October 2015) and the "R Package" (tcpl_1.0 released in November 2015) were downloaded and used as the foundation for all work (EPA, 2016b). Additionally, all the ToxCast data used for this study are also available for download as Excel files using the "ToxCast & Tox21 Summary Files" download link (for invitrodb_v2 released October 2015), the pertinent files used in the current study to evaluate ToxCast results are "tested_Matrix_151020.csv", "modl_ga_Matrix_151020.csv", "hitc_Matrix151020.csv", and "zscore_Matrix_151020.csv"; ToxCast data can also be viewed using the iCSS ToxCast Dashboard (EPA, 2016a). Chemicals evaluated in the ToxCast program were screened in concentration-response across 1157 assay endpoints,