



EPA Public Access

Author manuscript

Environ Sci Technol. Author manuscript; available in PMC 2019 March 06.

About author manuscripts

Submit a manuscript

Published in final edited form as:

Environ Sci Technol. 2018 March 06; 52(5): 3125–3135. doi:10.1021/acs.est.7b04781.

Suspect Screening Analysis of Chemicals in Consumer Products

Katherine A. Phillips^a, Alice Yau^b, Kristin A. Favela^b, Kristin K. Isaacs^a, Andrew McEachran^{c,d}, Christopher Grulke^d, Ann M. Richard^d, Antony J. Williams^d, Jon R. Sobus^a, Russell S. Thomas^d, and John F. Wambaugh^{d,*}

^aNational Exposure Research Laboratory U.S. Environmental Protection Agency, Office of Research and Development, 109 T. W. Alexander Drive, RTP, NC USA 27711

^bSouthwest Research Institute, San Antonio, TX

^cOak Ridge Institute for Science and Education (ORISE), Oak Ridge, TN, USA 37830

^dNational Center for Computational Toxicology U.S. Environmental Protection Agency, Office of Research and Development, 109 T. W. Alexander Drive, RTP, NC USA 27711

Abstract

A two-dimensional gas chromatography-time of flight/mass spectrometry (GC×GC-TOF/MS) suspect screening analysis method was used to rapidly characterize chemicals in 100 consumer products – which included formulations (e.g., shampoos, paints), articles (e.g., upholsteries, shower curtains), and foods (cereals) – and therefore supports broader efforts to prioritize chemicals based on potential human health risks. Analyses yielded 4270 unique chemical signatures across the products, with 1602 signatures tentatively identified using the National Institute of Standards and Technology 2008 spectral database. Chemical standards confirmed the presence of 119 compounds. Of the 1602 tentatively identified chemicals, 1404 were not present in a public database of known consumer product chemicals. Reported data and model predictions of chemical functional use were applied to evaluate the tentative chemical identifications. Estimated chemical concentrations were compared to manufacturer-reported values and other measured data. Chemical presence and concentration data can now be used to improve estimates of chemical exposure, and refine estimates of risk posed to human health and the environment.

TOC Graphic

*Corresponding Author: wambaugh.john@epa.gov, Phone: 1-(919)-541-7641.

Disclaimer

The United States Environmental Protection Agency, through its Office of Research and Development, funded and managed the research described here. However, it may not necessarily reflect official Agency policy, and reference to commercial products or services does not constitute endorsement.



Introduction

Human exposure to chemicals occurs in many environments and along various pathways. For the majority of chemicals, the primary exposure pathway occurs in indoor environments as this is where most time is typically spent¹. Assessing the presence, concentration, and composition of chemicals in consumer product formulations (e.g., household cleaners) and articles (e.g., clothing), and confirming the use of these items in indoor environments, can provide critical information in support of chemical prioritization based on potential health risks²⁻⁴.

Chemicals having “near-field” uses (those found in consumer products used in indoor environments) have been measured at high concentrations in biomonitoring studies relative to chemicals without near-field uses³ and have been identified in indoor dust^{5, 6}. Yet, it has proven difficult to estimate chemical exposures from daily product use due to the large number of unique compounds within certain products and limited data on chemical presence and concentration/weight fractions. Even when ingredients of consumer product formulations are identified on product packaging, the concentrations of these chemicals are rarely known to scientists, public health officials, and consumers. Furthermore, virtually no accessible data exist for proprietary ingredients, mixture-based components (e.g., fragrance formulations), by-products of chemical synthesis, or unintentionally added ingredients^{7, 8}. Even more elusive than product formulations are data on chemicals contained within articles (i.e., furnishings and construction materials). Chemical components of articles do not need to be listed in material safety data sheets (MSDS), as is required for consumer product formulations⁹. In addition, the components within a single product formulation or article can reflect processes and materials of a complex supply chain which can include multiple global suppliers who may or may not communicate all the chemicals involved to the end

manufacturer¹⁰. The use of emerging analytical chemistry analyses such as suspect screening analysis (SSA) and non-targeted analysis (NTA) can aid in determining which chemicals are present at what concentrations for both product formulations and articles.

Recently, SSA and NTA methods have proven useful tools for identifying the presence of chemicals in blood, house dust, drinking water, and other media indicative of human exposure to chemicals^{5, 11–15}. SSA refers to analytical chemistry techniques that compare molecular features observed within samples (reflecting unknown individual compounds, each defined by a neutral accurate mass, retention time, and mass spectrum) to databases of chemical suspects for the purpose of identifying potential matches. NTA is another analytical chemistry technique that seeks to identify compounds in samples without the aid of chemical suspect lists. NTA and SSA methods allow researchers to broadly investigate potentially thousands of chemicals within a sample, as opposed to traditional, targeted analysis methods which are developed for quantifying relatively few chemicals^{14, 16}. Many different implementations of SSA/NTA exist. For example, different chromatographic methods (liquid chromatography vs. gas chromatography) and ionization methods (electrospray, electron, chemical) can be utilized with various mass analyzers (time-of-flight, quadrupole, ion trap). Each implementation of SSA/NTA comes with its own benefits and drawbacks, as well as workflows specific to that implementation^{17, 18}. A review of SSA/NTA methods, instrumentation, applications, and special considerations can be found elsewhere¹⁹.

In an ideal case, SSA/NTA is used to make a tentative identification of a compound in a medium, and then an analytical standard is used to confirm the presence of that compound²⁰. Narrowing the list of chemicals to confirm poses challenges, given the sheer number of candidates that are often tentatively identified via SSA/NTA, and the potential need to synthesize compounds whose standards may not be publically available. Many different methods have been proposed to narrow lists of potential identifications and to prioritize these lists for further confirmation testing^{1, 11, 21–23}. In addition to these methods, taxonomies have been developed for relating the confidence of chemical structure identification²⁰. Recently, researchers at the US Environmental Protection Agency (EPA) have developed metrics⁵ and databases to facilitate more efficient and informed selection of chemicals for confirmation¹⁸.

The use of a chemical in commerce, whether in consumer products, industrial processes, or other applications, dictates the potential exposure pathways by which it may come into contact with human receptors. Categorizations of chemical functional use, product use, and sector of use have been used in the development³ and parameterization^{2, 4} of predictive models of human exposure. Functional use, the role of a chemical in a product, is a marker of the likelihood of the correct identity, the intentionality of an ingredient's inclusion in a product, and an indicator of the concentration of an ingredient in that product^{4, 24}. Knowledge of functional use may be an excellent criterion for assessing tentatively identified compounds in SSA/NTA approaches, as it can provide evidence of the intentionality of an ingredient's inclusion in a product.

The 2017 report by the National Academy of Sciences “Using 21st Century Science to Improve Risk-Related Evaluations” found that uncertainty in chemical source information, such as product composition data, is a “major obstacle to exposure estimation for most chemicals”²⁵. That report recommended that efforts to obtain information on chemical quantities in products and materials “particularly consumer products and materials in the indoor environment, should be expanded substantially”. This research presents a prototype of how SSA workflows can be used to provide these essential chemical exposure data via an analysis of 100 consumer products using two-dimensional gas chromatography time-of-flight/mass spectrometry (GC×GC-TOF/MS) instrumentation, and a novel SSA approach that incorporates chemical functional use information to evaluate tentative chemical identifications. The methods presented in this work are not intended to be a definitive method of SSA/NTA for any conceivable matrix. Rather, they are intended to describe how this and other SSA/NTA workflows can be incorporated into risk-based analysis of chemicals in order to provide risk assessors with a more complete picture of which chemicals are likely found in product formulations and articles used in indoor environments. Although occurrence of a chemical does not indicate exposure, the presence of known toxicants, or the frequent occurrence of previously unstudied chemicals that can be identified in this and other SSA/NTA workflows, will help in chemical risk prioritization.

Methods

Sample Extraction and Preparation

The dilution and extraction of chemicals in 100 products across 20 product categories (5 products per category) were carried out using dichloromethane (DCM). Two solvent systems, DCM and 6% diethyl ether in an isomeric hexane solution, were initially evaluated using a subset of the samples (SI Methods). DCM was determined to have lower background noise and the solvent systems were similar in terms of extracted compound classes. Product categories included formulations, articles, and foods. Samples were prepared depending on whether they were solid or liquid/paste: For solid samples (with the exception of vinyl upholsteries and vinyl shower curtains), approximately 1 – 5 g of the material was extracted using a Soxhlet apparatus with 250 mL of dichloromethane for 18 hours. For vinyl upholsteries and vinyl shower curtains, DCM solubilized the vinyl material after 18 hours of extraction. As a result, the same extraction procedure was used on these vinyl materials, however, the solvent was changed to 6% diethyl ether in the hexane isomers due to the low solubility of vinyl material in the diethyl ether and hexanes solution. For liquid or paste samples, a dilute-and-shoot approach was used²⁶. Approximately 1 g of the liquid or paste was extracted with 10 mL of DCM. An aliquot of the DCM extract was removed for GC×GC-TOF/MS analysis. Method blank solutions were prepared by adding the same volume of extraction solvent to the Soxhlet extraction apparatus. The solution was then processed and concentrated in the same manner as the actual extraction samples.

Internal standards were added to the extracts prior to GC×GC-TOF/MS analysis. All internal standards were at 1 µg/g concentration. Ten micrograms each of recovery compounds were added prior to extraction of samples. Chemicals used as internal standards and recovery compounds are provided in SI Methods.

GC×GC-TOF/MS Analysis

GC×GC-TOF/MS analysis was performed using an Agilent 7890 gas chromatograph coupled to a low resolution LECO PEGASUS 4D -TOF (LECO, St. Joseph, Minnesota). Injection volume was 1.0 μ L. The inlet temperature was 260°C and the inlet mode was splitless with a 1 min purge. Separation was achieved using two columns. The primary column (first dimension) was a RXi-1MS (30 m \times 0.25 mm \times 0.25 μ m; Restek, Bellefonte, Pennsylvania) and the second column (second dimension) was a RXi-17SilMS, (1.5 m \times 0.18 mm \times 0.18 μ m; Restek, Bellefonte, PA). The first column was held at 50°C for 3 minutes, ramped to 290°C at a rate of 8°C/min, ramped to 300°C at a rate of 20°C/min and held for 3 minutes. The second column and modulator were offset by 5°C and 20°C, respectively. Helium carrier flow was set to constant flow at 1.0 mL/min. The modulation period was 4 seconds (1.0 s hot, 1.0 s cold with 2 cycles per modulation period) throughout the entire run. The mass spectrometer was operated using electron ionization at 70eV. Spectra were collected from 45–600 m/z with a scan time of 100 spectra/sec.

Spectral data were processed using LECO's Chromatof software (version 4.50.8.0) to deconvolute and integrate peaks, and to match them against reference standards or, for compounds without a reference standard, spectra in the National Institute of Standards and Technology 2008 Mass Spectral Library (NIST 08). The deconvoluted total ion chromatograms were used to generate peak areas. Peaks without a standard were processed using a signal-to-noise cut-off of 150. This cut-off generally yields peaks down to the 10,000 to 20,000 area count level - using a lower cut-off greatly increases the number of low-quality, trace-level mass spectra. A signal-to-noise cut-off of 5 was used for compounds for which a reference standard was assayed. Spectral peaks of compounds in the blanks were excluded from spectral matching unless the area was 10 times greater in the measured sample. Spectral peaks of compounds resulting from solvent and column bleed were also excluded with the exception of some siloxanes which were present at a much greater concentration in the sample extract. Compounds below an area count of 100,000 were generally excluded but categorized and archived for future review, if necessary. Following processing on the LECO Chromatof software system, sample data were exported for use in the analyses presented in this study (Table S1). Chemicals that were deliberately added during sample preparation, such as spike-in chemicals used for approximating concentrations, were eliminated from the reported list of chemicals present in the test objects.

A match factor was calculated based on how closely the peaks of an identification matched the peaks of a NIST 08 spectrum. Match factors could range from 0 to 1000. Preliminary filtering for compounds without an authentic standard was performed using a forward match factor threshold of 650. In this case, scores below 650 were reported as unidentified spectral matches, scores between 650 and 750 were weak indicators of a spectral identification, but possibly correct (particularly for trace level peaks), scores above 750 were considered good indicators of a spectral identification. The spectrum was assigned the name and Chemical Abstract Service Registry Number (CASRN) of the NIST 08 spectra that yielded the highest match factor (resulting in 4270 spectral matches). Spectra with matches greater than 650 were classified as identified spectral matches. Spectral matches then underwent manual

curation to ensure these classifications were correct. In cases where spectra were deemed to be improperly classified, they were reclassified based on expert opinion. Upon manual curation, identified spectral matches were further classified as *tentative chemical class identifications* if the spectrum was indistinguishable from that of potential isomers or if the exact carbon chain length of the identification was unknown. Otherwise, identified spectral matches were classified as *tentative chemical identifications*. For targeted compounds with an authentic standard, preliminary filtering was performed by requiring a reverse match factor threshold of 750. Manual review was used to accept or reject the identification. It should be noted that in most cases the forward and reverse match factor scores are comparable due to the use of deconvolution by the instrument software.

Reported concentrations are estimates, as most compounds were quantified against the internal standard, naphthalene-D₈, response at a single concentration of 1.0 µg/g. This does not take into account the response factor of the compounds. For compounds with a reference standard, the concentration was estimated in two ways, first by using the acenaphthalene-D₁₀ response and second against the reference standard as a single-point calibration. For reference standards, acenaphthalene-D₁₀ was chosen as its retention time was more attuned to the retention times of the reference standards. Concentrations estimated by either the naphthalene-D₈ response, or the acetaphthalene-D₁₀ response (for reference standards), were used for analyses presented in this work and are provided in Table S1. For targeted compounds, relative retention time in both dimensions was used to compensate for minor retention time differences between batches of samples. The relative retention time was calculated using the closest of the six internal standards. While recovery compounds were added to the samples and were detected, the extraction recoveries of these compounds were not calculated as the analytical methodology used here was of a qualitative nature.

Chemical Confirmation with Reference Standards

A set of 200 chemicals from the ToxCast chemical library, chosen for their possible environmental prevalence and toxicity concern, were selected as reference standards for follow-up confirmation (Table S2). Of these 200, 82 chemicals were confirmed to be present in at least one of the test objects. An additional 37 chemicals in the products had reference standards that were previously analyzed by some of the current authors.

Chemical confirmation of reference standards was performed on samples after initial chemical surveillance methods were applied to a sample. Performing these analyses separately can lead to discrepancy between the retention times of the reference standards and the retention times of tentative identifications in the samples due to noise in the chromatographs and mass spectra. Typically, a range of retention times about the retention time of the reference standard (e.g., two modulation periods) is used. However, in some cases, chemicals were classified as “confirmed absent” (i.e., chemical was confirmed not to be in a sample) when it could, in fact, have fallen outside of the retention time window due to signal noise. To account for this possibility while assessing if an identification was negatively confirmed, the retention time of a tentative match was compared to a larger range of retention times centered about the reference standard’s retention time. Details of this procedure are provided in SI Methods. After implementing this method, there were 56

instances (20 unique chemicals across 32 of the products) of a tentative spectral match being confirmed absent. Figure S1 shows the confirmed absent rate of these 20 chemicals.

Results and Discussion

Chemicals identified using Suspect Screening

Using the sample preparation, extraction, and mass spectrometry methods shown in Figure 1, 4270 distinct spectral matches to NIST 08 were found within 100 consumer products (60 formulations, 35 articles, and 5 foods). Based upon the similarity of spectral features to those in NIST 08 and curation of the matched spectra, there were 1602 identified spectral matches. Of these, 119 chemicals were ultimately listed as confirmed chemical identifications using available chemical standards, while the remaining matches are tentative, but have not yet been confirmed. The number of identified spectral matches in a product category ranged from a low of 85 (toothpaste) to a high of 303 chemicals (plastic children's toy). There were also 2668 unidentified spectra with no matches which were excluded from further analysis. The results of the SSA are provided in Table S1.

Chemicals that had either a confirmed or tentative chemical identification were compared to lists of chemicals known to be in consumer products or of toxicological interest for chemical prioritization (see Table 1). These lists included chemicals from manufacturer-reported MSDSs and those selected by EPA and other US government agencies for high-throughput screening for *in vitro* biological activity; combined there are more than 10,000 chemicals in these lists. Only 30.5% of the chemical identifications in the 100 products were found on those chemical lists in Table 1. Figure 2 focuses on the identifications in the SSA that were also in the Consumer Product Chemical Profiling database (CPCPdb). This database provides aggregated reported weight fractions obtained from MSDSs for a total of 1797 unique chemicals in general household products sold by a major US retailer²⁷. However, for the 100 products tested in this study, only 170 of the 836 confirmed or tentatively identified chemicals matched chemicals included in CPCPdb. While it is expected that many chemical identifications from the SSA would not be represented in CPCPdb due to the lack of information available on the composition of product articles, one would expect chemical identifications from product formulations studied in the SSA to overlap with the chemicals in CPCPdb. Of the 60 products tested that were also formulations, however, only 130 of the 486 confirmed or tentative identifications were found in CPCPdb. This suggests that there could be many more chemicals in product formulations than are identified in the MSDS (Table 1). This is, of course, not surprising as MSDSs are only required to report chemicals known to pose hazards to human health, not every ingredient. It should be noted that while chemicals that pose hazards to health are the minimum requirement for reporting, the manufacturer is allowed to include any chemical in the product on the MSDS (whether hazardous or not), which is why MSDSs of products are used to assess chemicals in a product.

Figure 2 also summarizes the estimated concentration of identified chemicals (both tentative and confirmed chemical identifications) across all 20 product categories. Cereals had relatively lower estimated chemical concentrations than other product categories. However, the concentrations of chemical identifications in products that were articles (mean: 35.8

µg/g, median: 0.439 µg/g) tended to be less than those of products that were formulations (mean: 286 µg/g, median: 9.26 µg/g). Due to the homogenization process of articles, extraction efficiency (how much of the total amount of a chemical was extracted from the sample) for articles and foods could have been lower than that of formulations thus resulting in lower estimated concentrations for individual compounds.

Figure 3 shows estimated concentrations of each identified spectral match across all 100 products. Of the confirmed chemical identifications, many of the most ubiquitous chemicals identified in the products were those listed in the ToxCast inventory (e.g., hexadecanoic acid and diethyl phthalate), which is to be expected since many of the reference standards were provided by the ToxCast project. The 732 tentative chemical identifications include many that are not covered by existing EPA toxicology testing libraries (ToxCast, Tox21²⁸, and ToxRefDB^{21, 29–31}). A total of 1848 rows are plotted in Figure 3, reflecting that some chemicals were confirmed in some sampled products but only tentatively identified in others (i.e., neither confirmed nor confirmed absent), due to the time difference between sample analysis and chemical standard analysis.

Not surprisingly, many of the same chemicals are found in each product within a product category, regardless of manufacturer. This is shown to some extent by the proximity of products in the same category in Figure 3, where the columns have been sorted by products with the largest number of identified chemicals. This overlap is shown more clearly in Figure S2, where chemicals are shown sorted by product category rather than prevalence in tested products.

Flame retardants and those chemicals identified as potential estrogen receptor (ER) agonists were chosen as two chemical classes of interest, due to their prevalence in consumer products and their potential to cause adverse health effects^{32, 33}. In addition to these chemical classes, ToxCast chemicals and prevalent chemicals (i.e., chemicals that were tentatively identified in 25 or more products) are shown in Figure 3.

The majority of the 1602 identified spectral matches are found in only a single product (Figure S3), however, the presence of 43 prevalent chemicals, which occurred in 25 or more products should be noted (Table S3). The prevalent chemicals may be present for a variety of reasons: 1) they are the result of sample packaging or preparation (e.g., unintentional contamination); or 2) they are extremely common across the products that were sampled. Common occurrence of chemicals in multiple product categories potentially leads to much higher aggregate exposure than expected from any one source, and thus might lead to greater concern making them potential priorities for exposure and toxicity assessment. The most prevalent chemical in the study, hexadecanoic acid (CASRN 57–10-3, palmitic acid, a saturated fatty acid), was found across 74 products and 18 product categories. The ubiquity of hexadecanoic acid is unsurprising as it is one of the most prevalent fatty acids reported in use³⁴; indeed mere skin contact while manufacturing or purchasing some products (such as fabrics) could be enough to leave a film of hexadecanoic acid on a product.

The sample preparation process may have unintentionally introduced some chemicals, but such contamination would likely be apparent as a group of chemicals present across many or

most of the samples. Of the 43 chemicals that occurred in greater than 25 test objects (Table S3), many do not appear to be the result of contamination as they are known to be present in consumer products, although some such as diethyl phthalate (CASRN 84–66-2) cannot be ruled out as contaminants, ironically because they are known to be ubiquitous²⁷.

At least one confirmed or tentative chemical identification was made for one of the 64 potential ER agonists in 14 product categories: air freshener, baby soap, carpet padding, cereal, cotton clothing, fabric upholstery, hand soap, lipstick, plastic children's toy, shampoo, shower curtain, skin lotion, toothpaste, and vinyl upholstery. Bisphenol A (CASRN 80–05-7) was detected in 4 products. Confirmed identifications were found in a shampoo, shower curtain, and vinyl upholstery. A tentative chemical identification could not be ruled out for a plastic children's toys marked "Bisphenol A Free". The plastic children's toy also contained a tentative identification for the fluorinated bisphenol A replacement chemical bisphenol AF (CASRN 1487–61-1). In three of these products, exposure would not necessarily be expected, except, perhaps, due to any mouthing action on the children's toy. However, the other product containing measurable quantities of bisphenol A was a shampoo (indicating likely contamination from packaging or processing), from which significant dermal exposure might be expected.

Finally, Figure 3 shows chemicals that are labeled as flame retardants. Phosphate flame retardants were observed across a range of product categories including: carpet, carpet padding, cereal, cotton clothing, fabric upholstery, hand soap, plastic children's toy, shower curtain, skin lotion, and vinyl upholstery. Initially, the most surprising was the detection of tributyl phosphate (CASRN 126–73-8) in a breakfast cereal; however, this chemical has multiple functions reported in EPA's functional use database (FUse) including use as an anti-foaming agent, colorant, defoamer, and plasticizer. As a result, it is likely that tributyl phosphate was included for reasons other than its flame retardant abilities, however, the fact that it could be an unintentional addition to the product (at some point along the manufacturing process) cannot be ruled out.

It should be noted here that confirmed presence of a chemical in a consumer product is a prerequisite for implicating exposure via consumer product usage, but is by no means a guarantee that significant exposure has occurred. Samples analyzed in the present study were homogenized and extracted with a solvent, which in many cases (such as plastic articles) might extract chemicals that should reasonably be expected to be sealed within the product. If a chemical of concern is found within a product, the emissivity of that chemical from the specific formulation, within a realistic environment, must be measured³⁵ or modeled³⁶. Measurements would most likely require targeted analytical chemistry, and therefore would not be high-throughput. However, it is possible to conceive potential non-targeted experimental workflows that could aim to capture these emissivity values in a more high-throughput fashion. In addition, different solvents could be used in which could more accurately assess chemicals likely to be extracted through normal use of a product.

Assignment of Functional Use to Support SSA Identification of Chemicals

Many of the tentative chemical and chemical class identifications from the SSA were chemicals not found to be associated with databases cataloguing chemicals known to be

associated with consumer products. For these chemicals, some form of classification was desired to rule out false positive identifications. Hence, the potential functional role of the chemicals was used to evaluate the credibility of tentative spectral identifications.

Two different (but related) approaches were used to obtain functional uses for tentatively identified spectra in the SSA: retrieving reported functional uses from FUse and predicting functional uses from quantitative structure-use relationship (QSUR) models³⁷. These methods are further described in SI Methods. As isomers and chemicals of different chain length (but still having the same functional groups) would likely serve similar, if not the same functional uses in products, all tentatively identified spectra were assigned functional uses, where possible. Functional uses of confirmed chemicals were not considered for identification purposes as the chemicals were known to be in the product. Of the 1529 tentatively identified spectra, 538 were found to have at least one reported functional use in FUse, which was considered sufficient information to support a chemical identification in a product. The structural descriptors needed for QSUR predictions were available for 562 of the remaining 991 tentatively identified chemicals, but only 317 had valid QSUR predictions of at least one function.

Screening of functional uses has been applied because confirmation of identified spectral matches is not high-throughput, and reference standards may not always be available. Therefore, reported and predicted functional uses have been applied to chemicals tentatively identified in the SSA to improve confidence in their identification. Eight hundred fifty-five chemicals that had a tentative spectral match were reported or predicted to have a functional use that is common to chemicals known to be in consumer products. There were 674 chemicals that had no reported or predicted functional uses, which indicates that it is less likely that these chemicals would be in consumer products (particularly product formulations, where a great deal more is known about the roles chemicals play than in articles or foods).

Chemicals with the functions of fragrance or ubiquitous (i.e., chemicals typically having 10 or more reported functions) were identified in all product categories (Figure 4). Chemicals with functional uses of masking agent, UV absorber, fragrance, ubiquitous, solvent, emollient, surfactant, and colorant were found in 75% of the product categories examined. Five or more surfactants were identified in such product formulations as baby soap, glass cleaner, hand soap, indoor house paint, shampoo, and shaving cream. Surfactants in these products can be used to remove oils and sediments with water (soaps and cleaners), stabilize foams (shaving cream), or improve spreading of a liquid (paint). More than 5 surfactants were also identified in both cotton clothing and fabric upholstery, likely due to factory washing of fabrics prior to sale. More identifications were reported to be fragrances than any other functional use. It should be noted that fragrances make up roughly 20% of chemicals in FUse, making it the largest functional use category in FUse.

For this study, there were 473 chemicals in product formulations (e.g., glass cleaners, skin lotion) found in FUse compared to 436 chemicals in articles (e.g., carpet, shower curtains) and 132 chemicals in foods (cereals). This is unsurprising as many manufacturers in the household and personal care product sectors have begun reporting functional use of

ingredients in their products, whereas this is not necessarily common practice for articles or foods. The toothpaste category had the largest ratio of tentative spectral matches to chemicals with known functional uses (54 of 73 tentative matches in this category were in FUse). The carpet padding category had the lowest ratio of chemicals with known functional use information (34 of 197 tentative matches in this category were in FUse).

Of the 538 tentative spectral matches in FUse, 307 were reported as fragrances. Fragrances are ubiquitous across product categories and can have multiple roles within a consumer product. Fragrances also do not have to be identified by chemical name on an ingredient list of cosmetic products. Therefore, chemicals that serve as fragrances likely are not represented in databases such as CPCPdb, but are represented in FUse due to data sources that disclose lists of fragrances. Reported and predicted functional uses of tentative spectral matches are provided in Table S1.

SSA Concentration Comparison with Existing Weight Fraction Data

A comparison of chemicals in a product that were either tentatively identified or confirmed that were also present on that product's ingredient list can be found in SI Results. In addition to this comparison, estimated weight fractions (WFs) of chemicals in sampled products were compared (Figure 5) with values from similar products (i.e., products in the same product category) either from manufacturer-reported MSDSs (formulations) or the State of Washington's Product Testing Database (formulations and articles)³⁸. In addition, estimated WFs were compared with label-reported values for active ingredients in the sampled products. For product formulations, 37 CASRN-product combinations (20 unique CASRNs) occurred in both the SSA and MSDS data. CASRNs from the SSA were taken from all spectral matches; chemicals included solvents, emollients, and preservatives. For articles, 9 common chemical-product pairs were found (in cotton clothing, carpet padding, and plastic toys); 8 unique CASRNs included phthalates and flame retardants. Eleven reported active ingredient occurrences were found in the SSA. All were in sunscreens (all 5 samples were represented) and comprised 3 unique chemicals. Log-transformed mean estimated WFs from the SSA were moderately correlated with log-transformed mean MSDS-reported values ($r=0.52$, $p<0.001$) and with Product Testing Database values ($r=0.67$, $p=0.05$). Estimated WFs in product formulations were mostly 1 to 2 orders of magnitude smaller than MSDS values (median ratio=0.1), consistent with a relatively low extraction efficiency. Differences were greater for articles (median ratio of estimated WF to the Product Testing Database values was 0.02), perhaps indicating a lower quantity of extraction. A linear regression of the estimated WF for the active ingredients yielded an $R^2=0.90$ ($p<0.001$); the slope of the regression indicated an estimated effective extraction efficiency from these sunscreen formulations of approximately 19%.

The methods used here are approximate with respect to predicting chemical concentration and identity. Concentration is approximated relative to chemicals spiked in at known concentrations, but this causes the dynamic range to be unknown. Due to low extraction efficiencies achieved in analyzing formulation products, the estimated concentrations from the SSA for these products likely underestimate true values (Figure 5). Thus, if there is the potential for transfer of a given chemical from a product to a human or the environment,

these concentration measurements likely underestimate the exposure potential. Even though underestimates are expected, however, there was a relationship between weight fractions of ingredients reported in MSDSs and the estimated concentration of those same ingredients in products studied in this SSA. Again, the presence of a chemical (even with a high estimated concentration) does not imply that a product user would be exposed to the chemical. For those chemicals shown in Figure 5, exposure seems possible since the chemical has been found on ingredient lists or in other analytical screens of the same or similar products. However, chemicals would need to be examined in the specific formulation that they occur to determine the likelihood of leaching or vaporization into the product's surroundings, which would produce a potential exposure to a user (as is typically done within a chemical exposure model).

Informing Future Research Efforts

A key need for SSA, as identified by this research, is standardized approaches for reporting of large numbers of analytes. Manual collation of thousands of analytes across 100 items is vulnerable to errors due to manual copying of the data between spreadsheets. This was addressed with manual curation, but left room for unidentified errors. While basic chemistry, quantitative structure-property relationships, and matching against other data sources provide increased confidence in the data as reported, improved data workflows (such as automation of the data collection process) for conversion of spectral data into identified chemical structures could be implemented to address these issues. Ongoing research within the EPA is striving to provide harmonized data and workflows for use in chemical identification via newly released tools in the publicly available CompTox Chemistry Dashboard (<https://comptox.epa.gov/dashboard>).

Chemical confirmation with 200 available reference standards was performed after surveillance methods. As a result, noise in the chromatographs and mass spectra was enough, in some cases, to cause the retention time of a chemical in a sample to fall outside an acceptable window of the retention time of the reference standard. Such a situation can lead to a failure to confirm the presence of a chemical that is in fact present in a product. Ranges of retention time centered about the reference standard's retention time were applied to reduce the possibility of incorrectly negating the presence of a compound. Running both chemical confirmation and surveillance methods in parallel would reduce this possibility in future research efforts, but requires *a priori* knowledge of which chemicals are to be confirmed.

Even among SSA/NTA methods, there is chemical class detection specificity. For instance, many chemicals are detectable with only gas or liquid chromatography, with a specific ionization method, or a combination of both; only gas chromatography with electron ionization was used in the analysis here. Supposing a third class of chemicals that are not amenable to either method (metals, chemicals that elute with the solvent, etc.), the chemicals identified here are almost certainly an underestimate of the total number of chemicals in the test objects. Further, samples were extracted with only one of two solvents (DCM or 6% diethyl ether in isomeric hexanes); use of other solvents may either extract other compounds, or extract compounds with greater or lesser efficiency, changing the estimate of

concentration in the sample, and perhaps more appropriately representing chemical extraction in indoor environments. Finally, a low-resolution TOF/MS was used here; by using a high-resolution mass spectrometer, greater confidence could be given to matches between tentative identifications and the NIST 08 mass spectral library.

All test objects were intended to be methodologically challenging (e.g., solid plastics, aerosolized sprays, textile fabrics) rather than statistically representative of either products or product categories. Their selection was arbitrary, and follow-up studies, using replication and larger numbers of samples, are required before these approaches can be said to have appropriately characterized an individual product or category of products. Given the success with these methods reported herein, and the surprising number of chemicals identified across both product and functional use categories, it seems warranted to continue and expand these investigations.

The results of the pilot study described in this work are a first step to addressing the need of more certainty in product composition data for exposure estimates. This was achieved by performing SSA of a wide variety of consumer products that can be found in homes, offices, or other indoor environments where humans tend to spend the bulk of their time (therefore, implying that exposure to these chemicals are greater than exposures in outdoor environments). While evaluating the results of this workflow pointed out significant areas of improvement for future research efforts, these methods have pushed knowledge of chemicals in products beyond current publically available data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We acknowledge exceptional efforts from Elaine Cohen-Hubal, Peter Egeghy, David Murphy, Nicole Hairston, Matthew T. Martin, and Stephen Little in developing and awarding the ExpoCast contract between SWRI and the U.S. EPA. The authors would like to thank Drs. Mark Strynar and Kathie Dionisio for their helpful review of the manuscript. Data were analyzed and plotted with Pandas⁴⁴, Matplotlib⁴⁵, and the R statistical computing language⁴⁶ (including ggplot2⁴⁷).

References

1. Little JC; Weschler CJ; Nazaroff WW; Liu Z; Cohen Hubal EA, Rapid methods to estimate potential exposure to semivolatile organic compounds in the indoor environment. *Environ. Sci. Technol* 2012, 46 (20), 11171–11178. [PubMed: 22856628]
2. Isaacs KK; Glen WG; Egeghy P; Goldsmith M-R; Smith L; Vallero D; Brooks R; Grulke CM; Özkaynak H, SHEDS-HT: an integrated probabilistic exposure model for prioritizing exposures to chemicals with near-field and dietary sources. *Environ. Sci. Technol* 2014, 48 (21), 12750–12759. [PubMed: 25222184]
3. Wambaugh JF; Wang A; Dionisio KL; Frame A; Egeghy P; Judson R; Setzer RW, High throughput heuristics for prioritizing human exposure to environmental chemicals. *Environ. Sci. Technol* 2014, 48 (21), 12760–12767. [PubMed: 25343693]
4. Isaacs KK; Goldsmith M-R; Egeghy P; Phillips K; Brooks R; Hong T; Wambaugh JF, Characterization and prediction of chemical functions and weight fractions in consumer products. *Toxicol. Rep* 2016, 3, 723–732. [PubMed: 28959598]

5. Rager JE; Strynar MJ; Liang S; McMahan RL; Richard AM; Grulke CM; Wambaugh JF; Isaacs KK; Judson R; Williams AJ; Sobus JR, Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring. *Environ. Int* 2016, 88, 269–280. [PubMed: 26812473]
6. Mitro SD; Dodson RE; Singla V; Adamkiewicz G; Elmi AF; Tilly MK; Zota AR, Consumer product chemicals in indoor dust: a quantitative meta-analysis of U.S. studies. *Environ. Sci. Tech* 2016, 50 (19), 10661–10672.
7. Seltnerich N, A hard nut to crack: reducing chemical migration in food-contact materials. *Environ. Health Perspect* 2015, 123 (7), A174. [PubMed: 26133041]
8. Xu Y; Liang Y; Urquidi JR; Siegel JA, Semi-volatile organic compounds in heating, ventilation, and air-conditioning filter dust in retail stores. *Indoor Air* 2015, 25 (1), 79–92. [PubMed: 24766478]
9. Hazard Communication, Code of Federal Regulations, Part 1910.1200, Title 29, 2012.
10. Becker M; Edwards S; Massey RI, Toxic chemicals in toys and children's products: limitations of current responses and recommendations for government and industry. *Environ. Sci. Technol* 2010, 44 (21), 7986–7991. [PubMed: 20873709]
11. Schymanski EL; Singer HP; Longrée P; Loos M; Ruff M; Stravs MA; Ripollés Vidal C; Hollender J, Strategies to characterize polar organic contamination in wastewater: exploring the capability of high resolution mass spectrometry. *Environ. Sci. Technol* 2014, 48 (3), 1811–1818. [PubMed: 24417318]
12. Park YH; Lee K; Soltow QA; Strobel FH; Brigham KL; Parker RE; Wilson ME; Sutliff RL; Mansfield KG; Wachtman LM, High-performance metabolic profiling of plasma from seven mammalian species for simultaneous environmental chemical surveillance and bioeffect monitoring. *Toxicol* 2012, 295 (1), 47–55.
13. Vergeynst L; Van Langenhove H; Joos P; Demeestere K, Suspect screening and target quantification of multi-class pharmaceuticals in surface water based on large-volume injection liquid chromatography and time-of-flight mass spectrometry. *Anal. Bioanal. Chem* 2014, 406 (11), 2533–2547. [PubMed: 24633561]
14. Hilton DC; Jones RS; Sjödin A, A method for rapid, non-targeted screening for environmental contaminants in household dust. *J. Chromatogr. A* 2010, 1217 (44), 6851–6856.
15. Schymanski EL; Singer HP; Slobodnik J; Ipolyi IM; Oswald P; Krauss M; Schulze T; Haglund P; Letzel T; Grosse S; Thomaidis NS; Bletsou A; Zwiener C; Ibáñez M; Portolés T; de Boer R; Reid MJ; Onghena M; Kunkel U; Schulz W; Guillon A; Noyon N; Leroy G; Bados P; Bogialli S; Stipani ev D; Rostkowski P; Hollender J, Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. *Anal. Bioanal. Chem* 2015, 407 (21), 6237–6255. [PubMed: 25976391]
16. Soltow QA; Strobel FH; Mansfield KG; Wachtman L; Park Y; Jones DP, High-performance metabolic profiling with dual chromatography-Fourier-transform mass spectrometry (DC-FTMS) for study of the exposome. *Metabolomics* 2013, 9 (1), 132–143.
17. Rathahao-Paris E; Alves S; Junot C; Tabet J-C, High resolution mass spectrometry for structural identification of metabolites in metabolomics. *Metabolomics* 2015, 12 (1), 10.
18. Sobus JR; Wambaugh JF; Isaacs KK; Williams AJ; McEachran AD; Richard AM; Grulke CM; Ulrich EM; Rager JE; Strynar MJ; Newton SR, Integrating Tools for Non-Targeted Analysis Research and Chemical Safety Evaluations at the US EPA. *J. Exposure Sci. Environ. Epidemiol* 2018.
19. Andra SS; Austin C; Patel D; Dolios G; Awawda M; Arora M, Trends in the application of high-resolution mass spectrometry for human biomonitoring: An analytical primer to studying the environmental chemical space of the human exposome. *Environ. Int* 2017, 100 (Supplement C), 32–61. [PubMed: 28062070]
20. Schymanski EL; Jeon J; Gulde R; Fenner K; Ruff M; Singer HP; Hollender J, Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ. Sci. Technol* 2014, 48 (4), 2097–2098.
21. Aalizadeh R; Thomaidis NS; Bletsou AA; Gago-Ferrero P, Quantitative structure–retention relationship models to support nontarget high-resolution mass spectrometric screening of emerging

- contaminants in environmental samples. *J. Chem. Inf. Model* 2016, 56 (7), 1384–1398. [PubMed: 27266383]
22. Sjerps RMA; Vughs D; van Leerdam JA; ter Laak TL; van Wezel AP, Data-driven prioritization of chemicals for various water types using suspect screening LC-HRMS. *Water Res* 2016, 93, 254–264. [PubMed: 26921851]
 23. Steger J; Arnhard K; Haslacher S; Geiger K; Singer K; Schlapp M; Pitterl F; Oberacher H, Successful adaption of a forensic toxicological screening workflow employing nontargeted liquid chromatography-tandem mass spectrometry to water analysis. *Electrophoresis* 2016, 37 (7–8), 1085–1094. [PubMed: 26791338]
 24. Chevillotte G; Ficheux AS; Morisset T; Roudot AC, Exposure method development for risk assessment to cosmetic products using a standard composition. *Food Chem. Toxicol* 2014, 68, 108–116. [PubMed: 24636944]
 25. National Academies of Sciences, Engineering, and Medicine, *Using 21st Century Science to Improve Risk-Related Evaluations*. The National Academies Press: Washington D.C., 2017.
 26. Mol HG; Plaza-Bolaños P; Zomer P; de Rijk TC; Stolker AA; Mulder PP, Toward a generic extraction method for simultaneous determination of pesticides, mycotoxins, plant toxins, and veterinary drugs in feed and food matrixes. *Anal. Chem* 2008, 80 (24), 9450–9459. [PubMed: 19072261]
 27. Goldsmith MR; Grulke CM; Brooks RD; Transue TR; Tan YM; Frame A; Egeghy PP; Edwards R; Chang DT; Tornero-Velez R; Isaacs K; Wang A; Johnson J; Holm K; Reich M; Mitchell J; Vallero DA; Phillips L; Phillips M; Wambaugh JF; Judson RS; Buckley TJ; Dary CC, Development of a consumer product ingredient database for chemical exposure screening and prioritization. *Food Chem. Toxicol* 2014, 65, 269–279. [PubMed: 24374094]
 28. Tice RR; Austin CP; Kavlock RJ; Bucher JR, Improving the human hazard characterization of chemicals: a Tox21 update. *Environ. Health Perspect* 2013, 121 (7), 756–765. [PubMed: 23603828]
 29. Knudsen TB; Martin MT; Kavlock RJ; Judson RS; Dix DJ; Singh AV, Profiling the activity of environmental chemicals in prenatal developmental toxicity studies using the US EPA's ToxRefDB. *Reprod. Toxicol* 2009, 28 (2), 209–219.
 30. Martin MT; Judson RS; Reif DM; Kavlock RJ; Dix DJ, Profiling chemicals based on chronic toxicity results from the US EPA ToxRef Database. *Environ. Health Perspect* 2009, 117 (3), 392. [PubMed: 19337514]
 31. Martin MT; Mendez E; Corum DG; Judson RS; Kavlock RJ; Rotroff DM; Dix DJ, Profiling the reproductive toxicity of chemicals from multigeneration studies in the toxicity reference database. *Toxicol. Sci* 2009, 110 (1), 181–190. [PubMed: 19363143]
 32. Judson RS; Magpantay FM; Chickarmane V; Haskell C; Tania N; Taylor J; Xia M; Huang R; Rotroff DM; Filer DL; Houck KA; Martin MT; Sipes N; Richard AM; Mansouri K; Setzer RW; Knudsen TB; Crofton KM; Thomas RS, Integrated model of chemical perturbations of a biological pathway using 18 in vitro high-throughput screening assays for the estrogen receptor. *Toxicol. Sci* 2015, 148 (1), 137–154. [PubMed: 26272952]
 33. Sjödin A; Patterson DG; Bergman Å, A review on human exposure to brominated flame retardants—particularly polybrominated diphenyl ethers. *Environ. Int* 2003, 29 (6), 829–839. [PubMed: 12850099]
 34. Liu Q-T; Chen R; McCarry BE; Diamond ML; Bahavar B, Characterization of Polar Organic Compounds in the Organic Film on Indoor and Outdoor Glass Windows. *Environ. Sci. Technol* 2003, 37 (11), 2340–2349. [PubMed: 12831015]
 35. Liu C; Liu Z; Little JC; Zhang Y, Convenient, rapid and accurate measurement of SVOC emission characteristics in experimental chambers. *PLoS One* 2013, 8 (8), e72445. [PubMed: 24015246]
 36. Biryol D; Nicolas CI; Wambaugh J; Phillips K; Isaacs K, High-throughput dietary exposure predictions for chemical migrants from food contact substances for use in chemical prioritization. *Environ. Int* 2017, 108, 185–194. [PubMed: 28865378]
 37. Phillips KA; Wambaugh JF; Grulke CM; Dionisio KL; Isaacs KK, High-throughput screening of chemicals as functional substitutes using structure-based classification models. *Green Chem* 2017, 19, 1063–1074.

38. State of Washington Department of Ecology Product Testing Data <https://fortress.wa.gov/ecy/ptdbpublicreporting/Default.aspx> (accessed November 2016),
39. U. S. Environmental Protection Agency Endocrine Disrupter Screening Program <https://www.epa.gov/endocrine-disruption> (accessed November 2016),
40. Krowech G Brominated and chlorinated organic chemical compounds used as flame retardants (BFRs and CFRs) http://www.biomonitoring.ca.gov/sites/default/files/downloads/120408flame_pres.pdf),
41. Centers for Disease Control and Prevention National Health and Nutrition Examination Survey <https://www.cdc.gov/nchs/nhanes/> (accessed November 2016),
42. Obach RS; Lombardo F; Waters NJ, Trend analysis of a database of intravenous pharmacokinetic parameters in humans for 670 drug compounds. *Drug Metab. Dispos* 2008, 36 (7), 1385–1405. [PubMed: 18426954]
43. Dix DJ; Houck KA; Martin MT; Richard AM; Setzer RW; Kavlock RJ, The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci* 2007, 95 (1), 5–12. [PubMed: 16963515]
44. McKinney W In *Data structures for statistical computin in python*, Proceedings of the 9th Python in Science Conference, 2010; van der Walt S; Millman J, Eds. 2010; pp 51 – 56.
45. Hunter JD, *Matplotlib: a 2D graphics environment*. *Comput. Sci. Eng* 2007, 9 (3), 90–95.
46. R Core Team R: A Language and Environment for Statistical Computing Vienna, Austria, 2015.
47. Wickham H, *ggplot2: Elegant Graphics for Data Analysis* Springer-Verlag: New York, 2009.

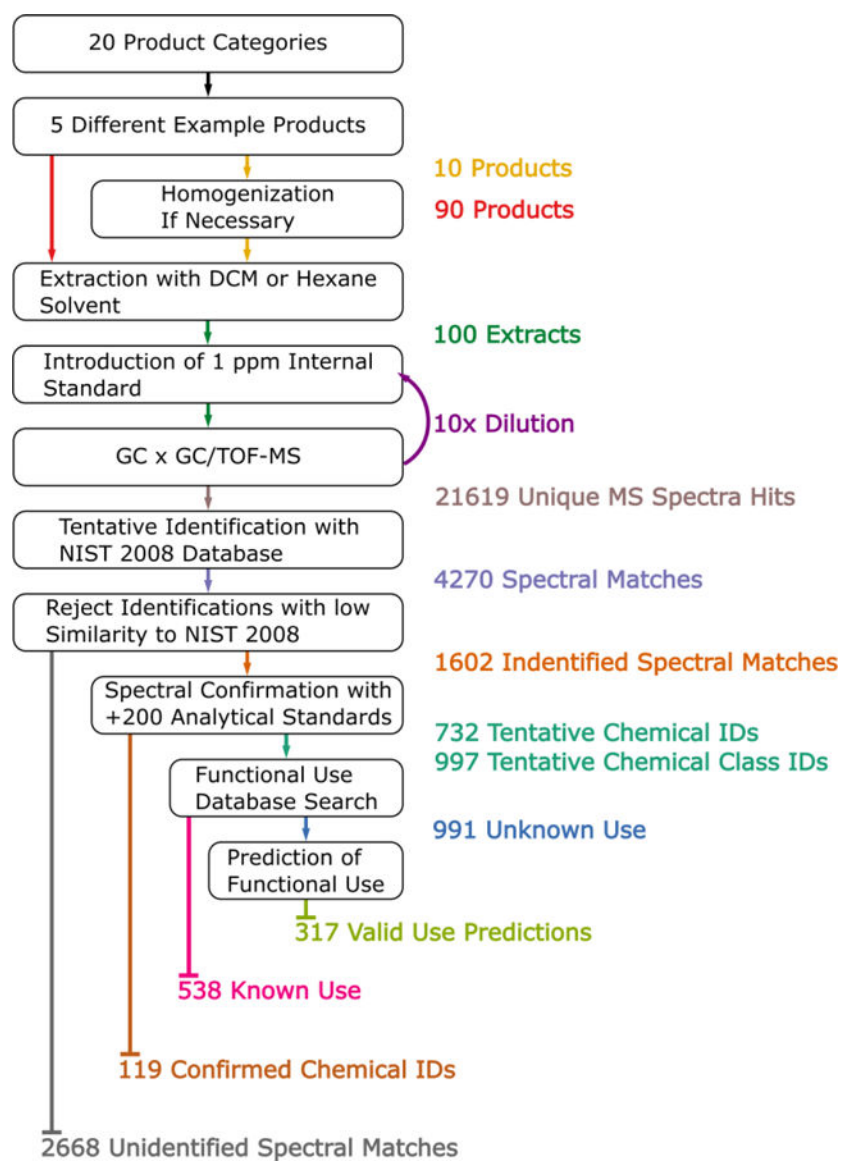


Figure 1. Workflow of suspect screening analysis for 100 consumer products. The color of the text on the right hand side of the figure corresponds to the arrows in the workflow. Samples were homogenized (if necessary) and extracted. Chemical standards were added at 1 $\mu\text{g/g}$ and then GC \times GC/TOF-MS was used to characterize the chemicals in the samples via their mass spectra. The spectra were matched, if possible, using the NIST 08 database, and a subset were confirmed using analytical standards. Reported and predicted functional uses of chemicals were used as additional support of chemical identification.

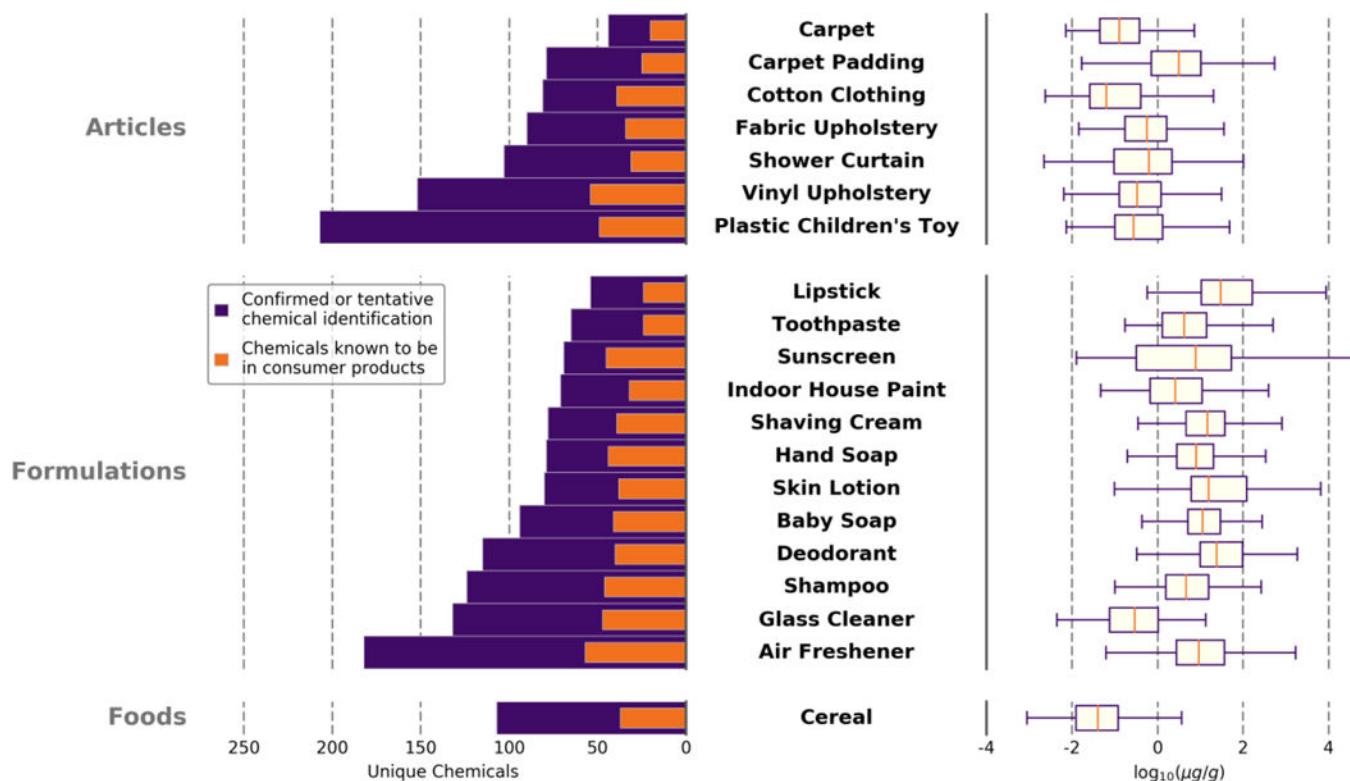


Figure 2. The number of unique chemicals within each category is shown by the bar chart (left). The total number of confirmed or tentative chemical identifications are compared to the number of those chemicals already known to be in consumer products (i.e., CPCPdb). The range of estimated concentrations are shown by the box plot (right) for all confirmed or tentative identifications for each product category. The upper pane represents articles, the middle represents formulations, and the lower pane represents foods.

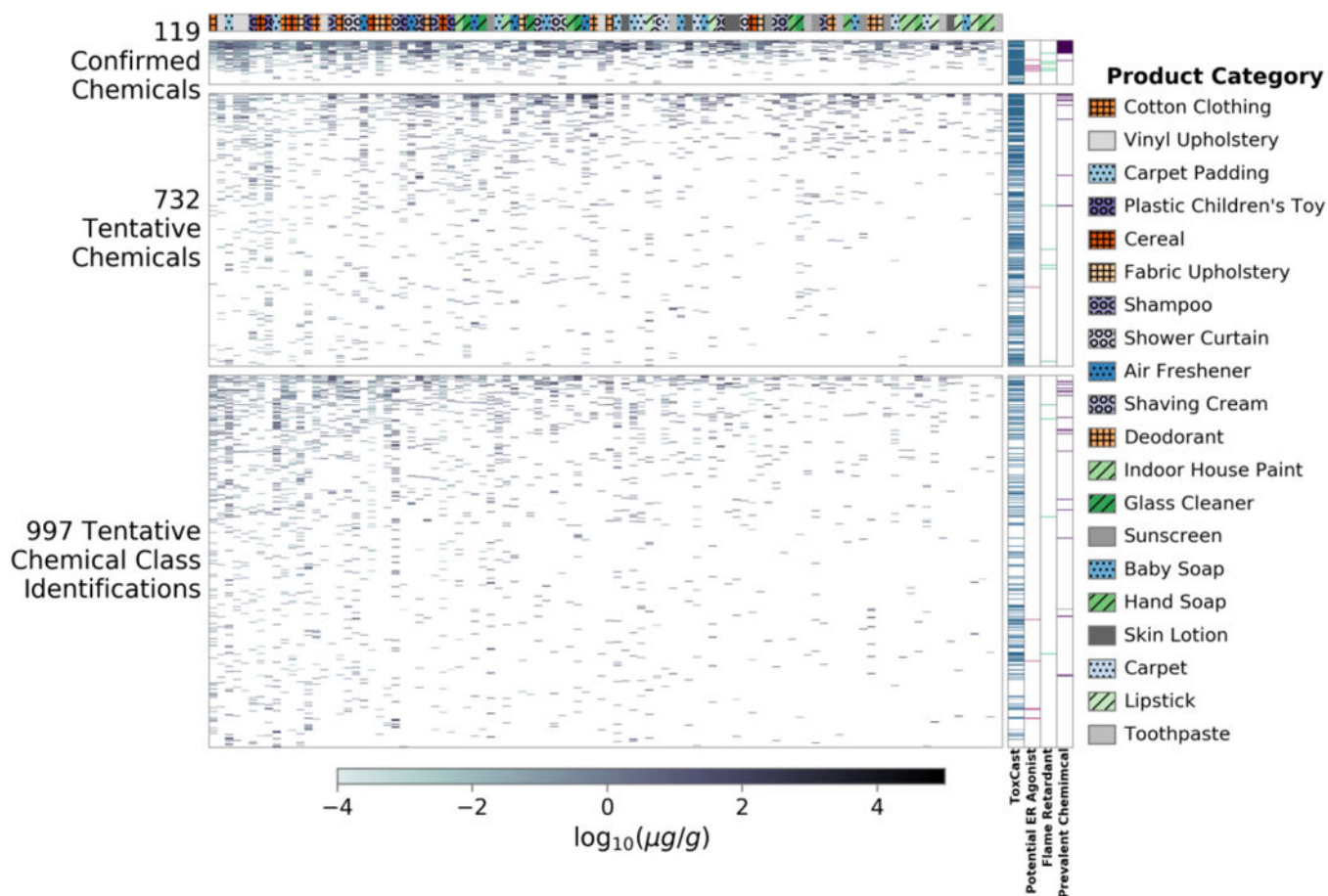
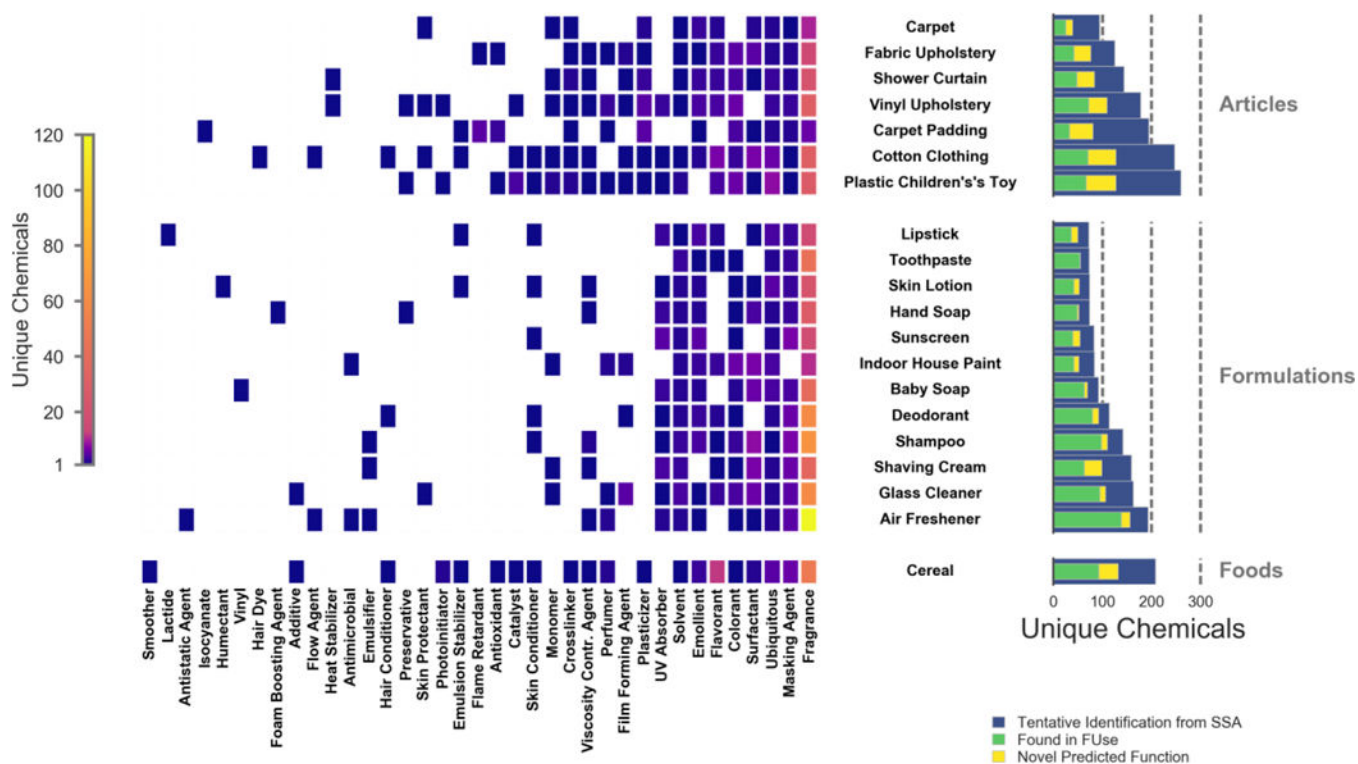


Figure 3.

Estimated concentration of chemicals in the tested products. Each one of the product samples is a column of the heat map, while there are 1848 rows corresponding to all identified spectral matches from the SSA. Some chemicals could be confirmed (by analytical standards) in some samples, but only tentatively identified (by spectral match) in others. The bars to the right of the heat maps show if a spectral match appeared in the ToxCast library, was a potential ER agonist, was a known flame retardant, or was prevalent (i.e., identified in 25 or more products) in the products examined in this SSA. The bar to the top of the heat map shows the product category in which a spectral match was identified. Spectral matches in each heat map are sorted such that those occurring in the most products are at the top. Products are sorted such that products with the most spectral matches are on the left. The Product Category legend is ordered by occurrence of a category from left to right. White spaces indicate that the spectral match was not identified in a particular product. Markings (line or circle) of product categories have been added only to aid in distinguishing product types.



Reported Chemical Function

Figure 4. Previously reported (FUse) functional use information for unconfirmed chemicals (tentative chemical or tentative chemical class identifications). Left shows the number of these identifications that were reported to have a given function in each of the product categories. Right shows the number of tentative chemical and chemical class identifications (blue) in each product category, the number of identifications in that category that had a reported functional use (green), as well as the number of identifications with a novel predicted functional use (yellow). The upper pane represents articles, the middle represents formulations, and the lower pane represents foods.

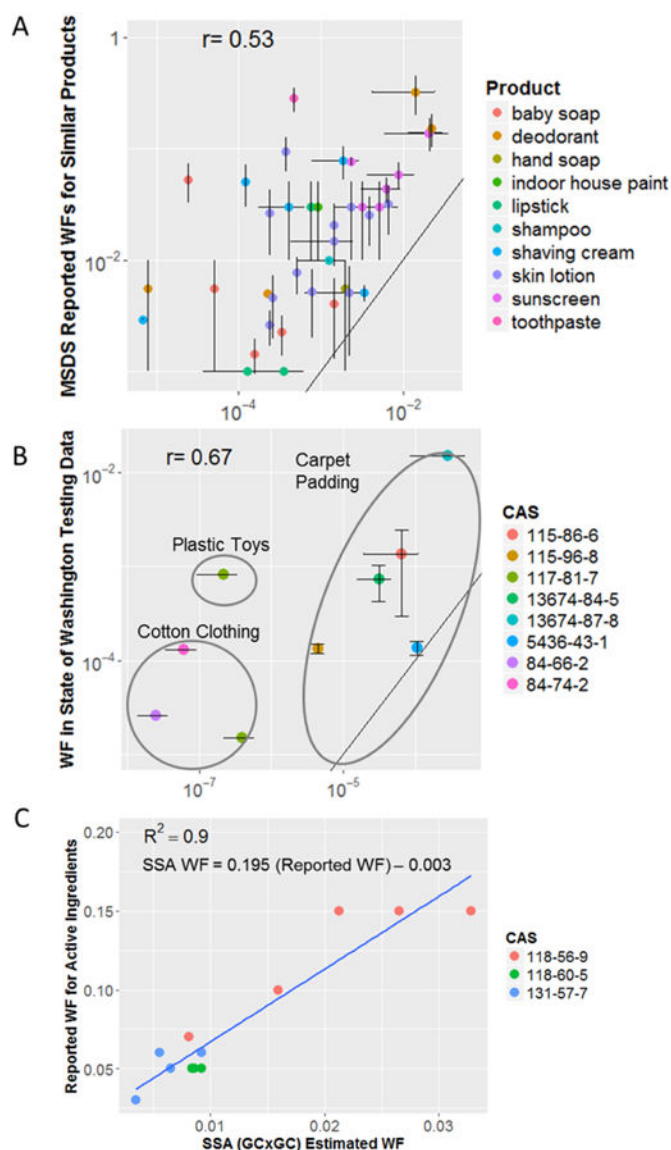


Figure 5. Comparison of SSA estimated product weight fractions (WF) to other sources of weight fraction information. Black lines give the “perfect predictor” identity line. A) Mean estimated WF in products (bars indicate SD across samples) versus mean midrange WF reported on MSDSs for similar products (bars indicate means of reported bounds), and B) Mean estimated WF in products (bars indicate SD across samples) versus mean values reported for similar products in the State of Washington Product Testing Database. C) Estimated WF versus label-reported values for active ingredients in sunscreens (blue line is fitted model; note that the linear regression equation has been algebraically rearrange to be consistent with the regression coefficient provided in the Results and Discussion).

Table 1.

Number of chemicals identified in this study that were also found in data sources that list chemicals related to consumer product ingredients and high-throughput toxicological screening for chemical prioritization.

Chemical List	Number of Chemicals in Source List	Number of Identified Spectral Matches found in Source List
CPCPdb ²⁷	1797	195
EDSP ³⁹	177	19
ToxCast ER Agonist ³²	64	9
Flame Retardants ^{33,40}	67	9
NHANES ⁴¹	452	37
Pharmaceuticals ⁴²	670	1
Tox21 ²⁸	8948	495
ToxCast ⁴³	4745	420
ToxRefDB	1172	105