Environment International 88 (2016) 269–280

Contents lists available at ScienceDirect

Environment International

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

Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring

Julia E. Rager [a], Mark J. Strynar [b], Shuang Liang [a], Rebecca L. McMahen [a], Ann M. Richard [c], Christopher M. Grulke [d], John F. Wambaugh [c], Kristin K. Isaacs [b], Richard Judson [c], Antony J. Williams [c], Jon R. Sobus [b,*]

[a] Oak Ridge Institute for Science and Education (ORISE) Participant, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States
[b] U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States
[c] U.S. Environmental Protection Agency, Office of Research and Development, National Center for Computational Toxicology, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States
[d] Lockheed Martin, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States

ARTICLE INFO

Article history:
Received 12 August 2015
Received in revised form 3 December 2015
Accepted 9 December 2015
Available online xxxx

Keywords:
Non-targeted
Suspect screening
Exposome
ExpoCast
ToxCast
Dust

ABSTRACT

There is a growing need in the field of exposure science for monitoring methods that rapidly screen environmental media for suspect contaminants. Measurement and analysis platforms, based on high resolution mass spectrometry (HRMS), now exist to meet this need. Here we describe results of a study that links HRMS data with exposure predictions from the U.S. EPA's ExpoCast™ program and *in vitro* bioassay data from the U.S. interagency Tox21 consortium. Vacuum dust samples were collected from 56 households across the U.S. as part of the American Healthy Homes Survey (AHHS). Sample extracts were analyzed using liquid chromatography time-of-flight mass spectrometry (LC–TOF/MS) with electrospray ionization. On average, approximately 2000 molecular features were identified per sample (based on accurate mass) in negative ion mode, and 3000 in positive ion mode. Exact mass, isotope distribution, and isotope spacing were used to match molecular features with a unique listing of chemical formulas extracted from EPA's Distributed Structure-Searchable Toxicity (DSSTox) database. A total of 978 DSSTox formulas were consistent with the dust LC–TOF/molecular feature data (match score ≥ 90); these formulas mapped to 3228 possible chemicals in the database. Correct assignment of a unique chemical to a given formula required additional validation steps. Each suspect chemical was prioritized for follow-up confirmation using abundance and detection frequency results, along with exposure and bioactivity estimates from ExpoCast and Tox21, respectively. Chemicals with elevated exposure and/or toxicity potential were further examined using a mixture of 100 chemical standards. A total of 33 chemicals were confirmed present in the dust samples by formula and retention time match; nearly half of these do not appear to have been associated with house dust in the published literature. Chemical matches found in at least 10 of the 56 dust samples include Piperine, N,N-Diethyl-m-toluamide (DEET), Triclocarban, Diethyl phthalate (DEP), Propylparaben, Methylparaben, Tris(1,3-dichloro-2-propyl)phosphate (TDCPP), and Nicotine. This study demonstrates a novel suspect screening methodology to prioritize chemicals of interest for subsequent targeted analysis. The methods described here rely on strategic integration of available public resources and should be considered in future non-targeted and suspect screening assessments of environmental and biological media.

Abbreviations: ACToR, EPA's Aggregated Computational Toxicology Resource; AHHS, American Healthy Homes Survey; AhR, aryl hydrocarbon receptor; AR, androgen receptor; CASRN, Chemical Abstract Services Registry Number; DI, deionized; DSSTox, EPA's Distributed Structure-Searchable Toxicity database; ERα, estrogen receptor alpha; GC × GC–TOF/MS, two-dimensional gas chromatography coupled with time-of-flight mass spectrometry; HPLC, high performance liquid chromatograph; HPV, high-production volume; HT, high-throughput; HTS, high-throughput screening; LC-Si, liquid-chromatography/silica; LC–TOF/MS, liquid chromatography time-of-flight mass spectrometry; HRMS, high resolution mass spectrometry; MFE, Molecular Feature Extraction; MS, mass spectrometry; MW, molecular weight; NFκB1, nuclear factor of kappa light polypeptide gene enhancer in B cells 1; NHANES, U.S. National Health and Nutrition Examination Survey; PPARγ, peroxisome proliferator-activated receptor gamma; RSD, relative standard deviation; RT, retention time; SPE, solid-phase extraction; ToxPi, Toxicological Priority Index.

* Corresponding author at: U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States.
E-mail address: sobus.jon@epa.gov (J.R. Sobus).

http://dx.doi.org/10.1016/j.envint.2015.12.008

1. Introduction

Over the past ~15 years, an enormous research effort has focused on the application of 'omics-based technologies to better understand genome-wide effects of environmental exposures (Rager and Fry, 2013). Paralleling this effort is the study of the human exposome, conceptualized in 2005 as the compilation of all life-course environmental exposures from the prenatal period onwards (Wild, 2005). Interest in the human exposome has grown rapidly since 2005, leading to more than 100 exposome-related articles in the published literature, and several exposome research centers/programs worldwide. These programs have invested in new tools, technologies, and studies to better characterize the breadth of human exposures, and the linkages between exposure and disease. As primary research drivers, it has been recognized that exposure data are sparse for many existing chemicals (Egeghy et al., 2012), and that knowledge-driven approaches alone are unlikely to meet the demands of this rapidly evolving field of research (Rappaport and Smith, 2010). Exposure scientists have therefore begun to advance exposome research efforts, in part, by expanding environmental monitoring through the application of "non-targeted" and "suspect screening" analyses. Suspect screening involves the detection of analytes in samples using existing chemical inventories and software matching algorithms (based on accurate mass and isotope patterns) (Krauss et al., 2010; Schymanski et al., 2014). Non-targeted screening involves the detection of analytes in samples given no *a priori* information — that is, no list of suspected or targeted chemicals (Krauss et al., 2010; Schymanski et al., 2014; Zedda and Zwiener, 2012). The goals of these complementary efforts are to more fully characterize the chemicals to which humans are frequently exposed, ultimately allowing systematic evaluation of associations between chemical exposures and incidence of human disease (Bell and Edwards, 2015; Patel and Ioannidis, 2014).

Non-targeted and suspect screening methods can be implemented using numerous analytical platforms, across a broad range of chemicals, to examine a variety of media. For example, methods based on gas chromatography–mass spectrometry (GC–MS) and/or liquid chromatography–mass spectrometry (LC–MS) have recently been used to screen for emerging contaminants in wastewater treatment plant effluent (Schymanski et al., 2014), lake sediment cores (Chiaia-Hernandez et al., 2014), food (Díaz et al., 2012), marine mammalian tissues (Shaul et al., 2015) and other biological specimens (Díaz et al., 2012; Sana et al., 2008), and in various sample extracts for effect-directed analysis (Simon et al., 2015). Chemical groups observed in these studies include biocides, disinfectants, flame retardants, food additives, mycotoxins, pharmaceuticals, pesticides, and surfactants, among others (Chiaia-Hernandez et al., 2014; Díaz et al., 2012; Schymanski et al., 2014; Shaul et al., 2015; Simon et al., 2015). Research consortia are now being developed to integrate these data across time, space, media, and analytical platforms. An example of such an effort is the NORMAN network, a consortium of scientists from over 50 laboratories and authorities across Europe and North America. This group facilitates the integration of information on emerging environmental substances and contributes to the harmonization and validation of monitoring methods and tools (NORMAN, 2015). Efforts of this scale will ultimately be necessary if exposome-level analyses are to become ingrained in environmental health research and implemented in public health policy.

Household dust has been the focus of many "targeted" research studies in recent years (Butte and Heinzow, 2002; Stapleton et al., 2009; Wu et al., 2007). In these studies, individual chemicals are selected for examination based on existing information or a specific research hypothesis, and are generally analyzed quantitatively using external and internal standards. Dust is an important environmental medium, with respect to human exposure, because it acts as a repository for various compounds that originate indoors, as well as for those that are transported into the home from the outdoor environment (Butte and Heinzow, 2002). Compounds that are present in household dust include biologically derived materials (e.g., animal dander, fungal spores, pollen, insect parts, skin fragments), building materials (e.g., flame retardants, textile fibers), particulate matter from indoor aerosols and soils brought in by foot traffic, and other volatile and semivolatile organic compounds, among others (Butte and Heinzow, 2002; Stapleton et al., 2009). Exposure to household dust can occur through several routes. Specifically, chemicals in dust may enter the body via inhalation of resuspended particles, dermal absorption, and non-dietary ingestion. Of particular concern, dust ingestion rates for infants and toddlers are estimated to be twice as high as those for adults because of their high rates of hand-to-mouth contact and floor contact from crawling (Butte and Heinzow, 2002). The comprehensive characterization of compounds in dust is therefore of high interest to better understand impacts of dust exposure on human health.

To date, non-targeted and suspect screening of chemicals in dust has been carried out by a limited number of studies. In 2010, Hilton et al. tested a method to screen for certain compound classes, specifically polycyclic aromatic hydrocarbons, phthalates, halogen containing compounds, and nitro compounds, using two-dimensional gas chromatography coupled with time-of-flight mass spectrometry (GC × GC–TOF/MS) (Hilton et al., 2010). This proposed method was tested using a National Institute of Standards and Technology (NIST) dust reference material certified to contain specified amounts of compounds belonging to the classes investigated. The study identified 370 chromatographic peaks of interest, 273 of which showed spectra indicative of the classes of compounds investigated (Hilton et al., 2010). Given that reference material was the focus of this analysis, additional research is needed to better characterize chemical constituents in diverse samples of house dust. Specifically, research is needed to help identify emerging contaminants in dust that have not been characterized in existing reference materials, or analyzed using targeted methods.

In light of these needs, the goal of this study was to develop and apply a novel suspect screening method using samples of house dust collected throughout the U.S. A high resolution mass spectrometry (HRMS) platform was used to generate MS data which were first matched to a suite of chemical formulas. Predicted formulas were then mapped to possible chemical structures using an existing U.S. EPA chemical database that provides highly curated structures for environmental chemical inventories of regulatory and toxicological interest. Prioritization algorithms, considering measurement data (i.e., detection frequencies and abundances), high-throughput (HT) predictions of chemical exposure, and HT measures of bioactivity, were then used to select individual chemicals for follow-up confirmatory analysis. These methods lay a foundation for characterizing and prioritizing measurement data from non-targeted and suspect screening studies, and are applicable to a variety of environmental media, and perhaps biological media (e.g., human blood).

2. Materials and methods

2.1. Chemicals for dust sample analysis

Methanol (B&J Brand High Purity Solvent) was purchased from Honeywell Burdick & Jackson (Muskegon, MI, USA) and ammonium acetate from Sigma Aldrich (St. Louis, MO, USA). Ultrapure deionized (DI) water was generated in-house from a Barnsted Easypure UV/UF (Dubuque, IA, USA) coupled with activated charcoal and ion exchange resin canisters.

2.2. Sample collection

Dust samples were collected as part of the American Healthy Homes Survey (AHHS), conducted by the U.S. Department of Housing and Urban Development between June 2005 and March 2006 (HUD, 2011). The survey was designed to assess a nationally-representative sample of permanently occupied, non-institutional homes throughout