RESEARCH PAPER

CrossMark

EPA's non-targeted analysis collaborative trial (ENTACT): genesis, design, and initial findings

Elin M. Ulrich[1] · Jon R. Sobus[1] · Christopher M. Grulke[2] · Ann M. Richard[2] · Seth R. Newton[1] · Mark J. Strynar[1] · Kamel Mansouri[3,4] · Antony J. Williams[2]

Abstract

In August 2015, the US Environmental Protection Agency (EPA) convened a workshop entitled "Advancing non-targeted analyses of xenobiotic chemicals in environmental and biological media." The purpose of the workshop was to bring together the foremost experts in non-targeted analysis (NTA) to discuss the state-of-the-science for generating, interpreting, and exchanging NTA measurement data. During the workshop, participants discussed potential designs for a collaborative project that would use EPA resources, including the ToxCast library of chemical substances, the DSSTox database, and the CompTox Chemicals Dashboard, to evaluate cutting-edge NTA methods. That discussion was the genesis of EPA's Non-Targeted Analysis Collaborative Trial (ENTACT). Nearly 30 laboratories have enrolled in ENTACT and used a variety of chromatography, mass spectrometry, and data processing approaches to characterize ten synthetic chemical mixtures, three standardized media (human serum, house dust, and silicone band) extracts, and thousands of individual substances. Initial results show that nearly all participants have detected and reported more compounds in the mixtures than were intentionally added, with large inter-lab variability in the number of reported compounds. A comparison of gas and liquid chromatography results shows that the majority (45.3%) of correctly identified compounds were detected by only one method and 15.4% of compounds were not identified. Finally, a limited set of true positive identifications indicates substantial differences in observable chemical space when employing disparate separation and ionization techniques as part of NTA workflows. This article describes the genesis of ENTACT, all study methods and materials, and an analysis of results submitted to date.

✉ Elin M. Ulrich
  ulrich.elin@epa.gov

1. National Exposure Research Laboratory, Office of Research & Development, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

2. National Center for Computational Toxicology, Office of Research & Development, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

3. Oak Ridge Institute for Science and Education, Post-Doctoral Participant, National Center for Computational Toxicology, Office of Research & Development, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

4. Present address: Integrated Laboratory Systems, Inc., Contractor to National Toxicology Program, National Institute of Environmental Health Sciences, Morrisville, NC 27560, USA

Introduction

Humans are exposed to thousands of chemicals daily, both intentionally and unintentionally. In order to determine which are mostly likely to pose potential health risks, it is imperative to first identify the specific chemicals to which we are exposed. The field of exposomics aims to characterize important life-stage exposures, and how these exposures impact individual health and well-being [1]. The exposome is inclusive of both external exposures (e.g., environmental, dietary, lifestyle), as well as internal factors (e.g., metabolites, hormones), which can be measured in a range of environmental and biological media and, ultimately, related to health outcomes [2]. It is currently estimated that 10–30% of diseases are caused by genetics, leaving 70–90% likely caused by lifestyle/environmental factors, and gene×environment interactions [3, 4].

Targeted analysis is the gold standard for measuring chemicals within environmental and biological samples.

Such analyses use chemical standards to give a high level of confidence in the accuracy, precision, and reproducibility of the results, but are limited to relatively small (10s to 100s) lists of target chemicals [5–7]. Alternatively, non-targeted analysis (NTA), which we define as analysis that is not targeted including suspect screening analysis (SSA), is a relatively new technique that strives to identify contaminants of emerging concern (CEC) in relevant samples, providing mostly qualitative and some semi-quantitative results [7–9]. NTA methods do not use chemical standards, but instead rely on a typically large (1000s to 100,000s) reference list of chemicals for aiding in chemical identification by drawing on a variety of experimentally or computationally predicted metadata [5, 8]. While targeted analysis techniques are well characterized and routinely undergo rigorous quality control procedures, most non-targeted approaches are in their infancy, with uncertainties regarding quality control and domain of applicability (i.e., determining which methods are better-suited for detecting specific chemical classes) [7]. Studies are therefore needed to define performance benchmarks so that NTA data are better understood and appropriately used in exposomic investigations.

To understand the current state of NTA techniques being applied in exposome studies, the US Environmental Protection Agency (EPA) invited experts in this and related fields (e.g., metabolomics) to a 2015 workshop entitled "Advancing non-targeted analyses of xenobiotic chemicals in environmental and biological media" (https://sites.google.com/site/nontargetedanalysisworkshop/) [9]. The purpose of the workshop was to bring together the foremost NTA experts to discuss state-of-the-science methods for generating, interpreting, and exchanging NTA measurement data. Invited speakers presented on the instrumentation, samples, methods, strategies, software, tools, workflows and databases used in their laboratories. Panel discussions identified short- and long-term needs of the research community, as well as actions that must be taken to address these needs. A group discussion on the final day planted seeds for a research study that would ultimately become *E*PA's *N*on-*T*argeted *A*nalysis *C*ollaborative *T*rial, or ENTACT, which was designed to address these varied needs and actions.

ENTACT is a round-robin or ring-trial style project involving nearly 30 laboratories from academia, government, and the private sector. ENTACT samples consist of ten synthetic mixtures containing 95–365 chemical substances each. The mixtures were created using chemical substances procured for EPA's Toxicity Forecaster (ToxCast) project, a research program within EPA's National Center for Computational Toxicology (NCCT), which employs high-throughput in vitro screening technologies to evaluate thousands of chemicals for potential bioactivity. These mixtures were distributed to participating labs to be analyzed using a variety of NTA methods. In addition to the mixtures, participating labs were provided reference material extracts for three matrices relevant to exposomics: house dust, human serum, and silicone bands. Extracts were prepared from unaltered media, and from media spiked with one of the ten ToxCast mixtures. Finally, individual chemical standard multi-well plates were prepared from the set of ~1200 chemicals used to construct the ten mixtures, as well as from the entire ToxCast library (~4700) [10]. The standards corresponding to the ten mixtures are being provided for follow-up confirmation studies, whereas standards for the entire ToxCast library are being used to generate reference mass spectra for private and public databases across a range of instrumentation. In addition to chemical samples, all participants had access to a current version of the full, publicly available Distributed Structure-Searchable Toxicity (DSSTox) database that underpins the CompTox Chemicals Dashboard website (https://comptox.epa.gov/dashboard; referred to here as the Dashboard), and consisting of over 720 K chemical entries to serve as a reference library, as well as the full ToxCast structure inventory for SSA.

The purpose of this article is to (1) reflect on the discussions that influenced the study design of ENTACT; (2) communicate the methods used in preparing ENTACT samples and data collection/analysis materials; and (3) highlight initial results of ENTACT across a limited data set. The eventual comprehensive analyses of all ENTACT data will allow for careful evaluation of the chemical space covered (and not covered) by each method. Further analyses will allow for determination of the success rates of various NTA approaches and the effects of matrix and mixture complexity on measurement performance. At the completion of ENTACT, collaborating researchers will identify and communicate benchmarks and best-practices for NTA research. This article is the primary point-of-reference for subsequent examinations of ENTACT data, and the foundation for all methods, data, and guidance that emerge from future analyses.

Genesis of ENTACT

The concept of the exposome has been gaining traction since it was first defined by Dr. Christopher Wild in 2005 [11]. Variations on the original definition have been proposed, but it is agreed that the exposome encompasses chemical and non-chemical stressors emanating from external and internal sources. NTA using high-resolution mass spectrometry (HRMS) has emerged as the primary tool for characterizing chemical stressors (both exogenous and endogenous) in environmental and biological samples [5]. To better understand the capabilities and limitations of NTA methods, the EPA convened a 2-day workshop in 2015. Workshop organizers sought to identify NTA research needs from a community of experts, and to organize a collaborative effort that would address critical needs, thus accelerating growth in the field. Over 150 attendees

participated in the workshop, with ten speakers invited to present on emerging NTA techniques, recent NTA applications for characterizing xenobiotics in environmental and biological media, and lessons learned from a recent NTA collaborative trial [12]. Panel and group discussions focused on the state of NTA research as it relates to the generation, interpretation, comparison, and storage of NTA data. Critical needs identified by workshop speakers and participants included:

1) tightly defined ring trials (performance testing with identical sample materials) to evaluate NTA method performance;
2) the availability of custom-made spiked samples for ring trials;
3) exchange of comprehensive suspect lists to enable interoperability; and
4) development of comprehensive spectral libraries to expand screening efforts.

These four needs were the initial drivers of ENTACT. Discussions around these needs took place on the final day of the workshop. Participants worked together to develop a research "strawman" with details later filled in by EPA staff and management. The following sections highlight workshop discussions that influenced the final study design.

Only a few collaborative trials, to date, have attempted to compare NTA results across labs [12–14]. In most cases, an overarching hurdle has been limited access to highly controlled samples with which to assess method performance (e.g., true positives and false negatives). EPA's ToxCast project manages a physical library of ~ 4700 chemical substances that are used for in vitro bioactivity screening [10]. Workshop discussions focused heavily on the possible use of these resources for ENTACT. Participants ultimately agreed on the use of ToxCast substances for the trial, with some recommending use of a small number of substances (more manageable) and others recommending use of the full set (more comprehensive). It was ultimately decided that a series of synthetic chemical mixtures would be created, with the number of unique ToxCast chemicals varying across mixtures, and a subset of chemicals serving as replicates across the mixture set.

In addition to synthetic mixtures of individual chemical substances, there was interest in ENTACT including "real" samples. There was disagreement as to how samples should be shared. Some wanted to receive raw samples, thus allowing evaluation of extraction and cleanup methods. Others preferred sharing of sample extracts, giving more focus to instrument and data processing methods. It was ultimately decided to include extracts of several standardized media, including two National Institute of Standards and Technology (NIST) standard reference materials (SRM 2585 - *Organic Contaminants in House Dust*; and SRM 1957 – *Organic Contaminants in Non-Fortified Human Serum*) and silicone bands (prepared by Oregon State University). It was further decided to include extracts of media that had been spiked with a mixture of ToxCast substances. Comparisons of fortified vs. unfortified extracts would provide insight into a lab's ability to identify a wide range of unknown compounds at environmentally relevant levels.

Sharing of a comprehensive, well-curated suspect list was agreed to be critical to the success of ENTACT. Discussions took place regarding the optimal size of a suspect list. Some lobbied for a small focused list of compounds, given limitations of certain data processing workflows. Others requested access to a larger list of substances, hoping to identify as many compounds as possible in ENTACT samples. To accommodate all parties, EPA ultimately shared both a manageable list of ToxCast substances (ca. 4700 at the time of the study, see Electronic Supplementary Material (ESM) Table S1), and a larger list of substances registered in EPA's DSSTox Database (ca. 720,000 at the time of distribution). EPA further provided "MS-Ready" structures [15], formulae, and monoisotopic masses for all substances on these lists. The generation and distribution of these data (vide infra) are what allow spiked substances (often salts, multi-component mixtures, etc.) to be correctly identified using various mass spectrometry (MS) platforms.

Perhaps the largest need of the NTA research community, as articulated by the workshop attendees, is quality reference spectra for high-interest compounds. When included in reference libraries, experimental spectra can enable broad and accurate suspect screening. They can further function as a training set when building spectra prediction models (e.g., Competitive Fragmentation Modeling for Metabolite Identification, CFM-ID: http://cfmid.wishartlab.com/). Shortly after the 2015 workshop, EPA decided to make all ToxCast substances available to a group of vendors and software developers to facilitate the development of reference libraries and NTA tools. Furthermore, EPA has the ability to provide a subset of the individual ToxCast substances (i.e., those included in the ten synthetic mixtures) to a limited number of labs participating in ENTACT. These materials enable the generation of reference data (e.g., MS^2 spectra, method specific retention times, collision cross sections), and are intended to facilitate rigorous self-evaluation of NTA results for ENTACT mixtures and spiked samples.

Materials and methods

ENTACT consists of three experimental parts and two phases, each with different primary goals, and addressing different aspects of non-targeted analysis. The three parts of ENTACT are illustrated in Fig. 1. Part 1 makes use of EPA's large ToxCast chemical library, drawing from over 1200

Fig. 1 Three parts of ENTACT

substances to create ten mixtures of varying size (i.e., numbers of chemicals per mixture) and complexity (i.e., numbers and types of chemicals), and including a small set of 5 replicates across all 10 mixtures, and 90 replicates across 3 of the mixtures. The main purpose of part 1 is to evaluate how well NTA methods perform in a best-case scenario (i.e., without interference from a sample matrix), and with a large list of "true positive" chemicals to identify.

Part 2 of ENTACT makes use of standardized sample extracts (unaltered and fortified house dust, human serum, and silicone bands). Each matrix was fortified with a different mixture, though there are overlaps in chemicals across the spiking solutions. The main purpose of part 2 is to determine how a matrix affects a method's ability to detect and identify compounds. Finally, part 3 of ENTACT makes use of individual ToxCast substances on multi-well plates. The main purpose of part 3 is to expand reference spectra available through vendor (e.g., Agilent Personal Compound Database and Library, PCDLs), in-house (laboratory specific), and open libraries (e.g., MassBank, Metlin, mzCloud), and to facilitate informative self-evaluation studies that will enhance NTA workflows.

Phase I of ENTACT includes blinded analyses of the 10 synthetic mixtures from part 1 and the six sample extracts (three unaltered and three fortified) from part 2. Submitted phase I reports from each participating lab include listings of features (equating to individual compounds, defined by a neutral accurate mass, retention time, and mass spectrum), identified at the mass, formula, and/or compound level, that were observed in the unknown samples. Phase II of ENTACT involves revealing which substances were intentionally added into each synthetic mixture (part 1) or sample (part 2). Having knowledge of spiked substances allows participants to retrospectively analyze their data and calculate performance statistics (e.g., true positive and false positive rates). It further provides each lab a means by which to optimize method parameters to best achieve a desired level of performance. The following sections describe methods used to prepare all ENTACT synthetic mixtures, sample extracts, multi-well plates, and data files used for compound identification and reporting.

ToxCast chemicals

Because ENTACT is intended to evaluate the ability of NTA methods to identify "true positives" in the chemical mixtures, the quality control (QC) results from previous liquid chromatographic (LC) MS and gas chromatographic (GC) MS analyses of ToxCast samples were included in the design of the mixtures. In particular, the majority of the approximately 1200 chemicals selected for inclusion in the ten mixtures were selected from a subset of ToxCast chemicals that: 1) could be represented by a single DSSTox chemical structure (i.e., they were not mixtures); 2) did not contain inorganics and organometallics; and 3) QC results confirmed the sample parent mass and indicated purity was ≥ 90%. Chemicals intentionally added to each mixture are listed in ESM Table S2. First, it was agreed that a set of five "control" chemicals known to be detectable by NTA methods (based on a pilot study performed on a mixture of 100 ToxCast chemicals [16]) would be included in every mixture. By adding these controls in every mixture sample, we hoped to obtain information regarding the reproducibility of detection for a chemical within a NTA method.

Part 1 of the Phase I trial consisted of blinded analysis of the 10 mixtures (referred to *infra* as Mixtures 1–10). Constituents for the majority of mixtures were selected with the intent of providing mixtures that were highly "amenable" to evaluation using standard HRMS techniques. For these amenable mixtures (Mixtures 1–8), constituents were selected only from ToxCast samples where the QC results unambiguously provided evidence of parent identity, and exceeded 90%

purity. In addition, the test constituents included in Mixtures 1–8 were selected to avoid identification difficulties caused by similar monoisotopic mass (with a threshold of 5 ppm difference in the monoisotopic mass for the largest component of a substance). These selections were made outside of the 5 control samples added to all mixtures and the "replicate" samples described below (resulting in a small number of isomeric/isobaric constituents in these mixtures). Though some ToxCast substances consisting of stereochemical isomer mixtures were included, they were considered a single constituent in the mixture. The amenable mixtures (1–8) were further designed to cover the breadth of logP and monoisotopic mass values for chemicals from the ToxCast screening library that met QC criteria. In this way, we hoped to provide a general dataset for evaluating the detectability of chemicals by various NTA methods based on chemical space analysis.

As shown in Fig. 2, four amenable mixtures were created with 95 constituents each (Mixtures 1–4, 90 test substances +5 control samples), two with 185 constituents (Mixtures 5–6, 180 test substances +5 control samples), and two with 365 constituents (Mixtures 7–8, 360 test substances +5 control samples), in part to examine if increased number of constituents contributed to a lower rate of true positive detection. However, to control for the increased variability in chemical properties naturally associated with a larger chemical set, three of these amenable mixtures (one of each size) were specially conceived to include a consistent subset of chemicals. For these, 90 "replicate" substances from the smallest mixture were embedded into the two larger size mixtures. Hence, Mixture 1 consisted of 95 total substances with 90 replicate compounds and 5 control compounds; Mixture 5 consisted of 185 total substances with 90 new test compounds, 90 replicate compounds from Mixture 1, and the 5 control compounds;
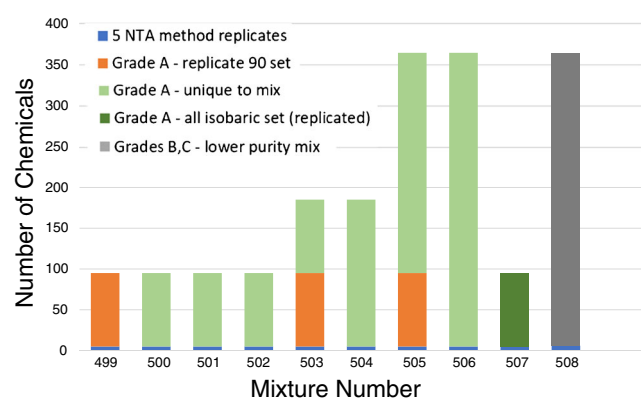
and Mixture 7 consisted of 365 total substances with 270 new test compounds, 90 replicate compounds from Mixture 1, and the 5 control compounds.

It was expected that the amenable mixtures would, for the most part, provide a fair, simple test for examining how chemical properties may affect NTA detection; however, there was concern that such contrived mixtures would prove to be a poor substitute for the complexity of real world samples. Although that concern would be addressed to some degree by Part 2 of the study, a more controlled examination of potential confounding factors was built into the two remaining mixtures. Mixture 9 was crafted specifically to contain pairs or groups of isomeric and isobaric compounds (i.e., yielding mass conflicts) that had also been embedded individually in the amenable mixtures. Mixture 10 was populated with some isomeric and isobaric compounds, but the majority were additional ToxCast substances where QC data did not indicate pristine quality (i.e., < 90% purity). It was thought that these "challenge" mixtures (9–10) might more closely approximate the reality of what would be seen in a real sample where the degradation products, variable concentration and isomeric/isobaric conflicts might obscure the ability to identify some chemicals. The isomeric/isobaric conflicts were specifically intended to test the ability of a NTA method to discern mass-conflicted substances when presented simultaneously in a single mixture vs. presented in different mixtures.

Mixtures for ENTACT were generated by Evotec who maintain EPA's ToxCast chemical library. Each ToxCast stock solution was nominally 20 mM in dimethyl sulfoxide (DMSO); the final concentration of each substance varied depending on the concentration in the ToxCast stock solution, but most were at the nominal concentration. DMSO would not be the first choice of solvent for chromatographers, especially given the room temperature freezing point. DMSO was used because the original purpose for ToxCast was in vitro assays in which DMSO provides both a good range of solubility and biological system amenability. Most compounds included in the mixtures were previously QC tested and confirmed to have concentrations in the correct range indicating good solubility in DMSO [10]. A total of 7.3 mL of each mixture was created by adding 20 μL of each requested ToxCast stock solution into a clear glass vial. Each mixture was then diluted to the final volume in DMSO for a final nominal concentration of 0.05 mM for each mixture constituent; the DMSO volume added varied (0–5400 μL) depending on the number of substances included in the mixture. EPA prepared 30 aliquots of 100 μL of each mixture using an Agilent 7696 Sample Prep Workbench (Santa Clara, CA) and a solvent blank using a different stock of DMSO than was used to prepare the mixtures. While not ideal, this type of blank can help correct for vial, cap, and aliquot process contaminants.

Single chemical well plates were generated in the same manner regardless of whether they contained only ENTACT

Fig. 2 Composition of ten synthetic mixtures. Blue bars represent five control compounds repeated in every mixture. Orange bars indicate 90 replicate compounds used in three mixtures of increasing number of substances added. Light green bars show amenable compounds (defined in text), while the gray bar represents more challenging compounds defined by lower purity or low concentration. Dark green bars indicate isomers and isobaric compounds selected to challenge NTA methods

substances (plate map listed in ESM Table S3) or the complete ToxCast library (plate map listed in ESM Table S4). ENTACT plates contained the mixtures as well as the individual substances, and substances that were repeated across multiple mixtures were repeated identically in the well-plates. Well-plates formatted with 384 cells (Greiner #781280) with seals (Agilent #24210-001) were used. Master plates were created using 20 μL of the 20 mM ToxCast stock solution, which was diluted with 180 μL of DMSO for a final volume of 200 μL and concentration of 2 mM. Daughter plates were created using 10 μL of the master plate solution, which was then diluted with 40 μL of DMSO for a final volume of 50 μL and concentration of 0.4 mM for each chemical sample on the daughter plate. For a compound with a molecular weight of 100 amu, this translates to a concentration of 40 ppm or ng/μL, with heavier compounds having higher concentration by weight. The plates were constructed in this manner to give labs enough volume to easily manipulate the solutions. Furthermore, high initial concentrations allowed for collection of high quality mass spectra even after additional dilution (if needed). Daughter plates for the ENTACT substances were shipped on dry ice to EPA and stored at − 80 °C until delivered to each participating laboratory. Daughter plates for the complete ToxCast library were shipped on dry ice from Evotec directly to participating laboratories.

Reference materials

Extracts of house dust were prepared at EPA laboratories (Research Triangle Park, NC) using *SRM 2585 Organic Contaminants in House Dust* from NIST (Gaithersburg, MD, USA) [17]. Twenty aliquots of 300 mg ± 10 mg dust were weighed in Falcon tubes (Becton Dickinson, Franklin Lakes, NJ). Ten of the tubes were spiked with 10 μL of a nominally 0.05 mM ENTACT mixture 7. Four empty test tubes were included as method blanks to undergo the same procedure as the dust samples. Methanol was added to each sample until the 13 mL mark. Samples were vortexed for approximately 1 min, until the dust at the very bottom tip could be seen moving in solution. Samples were then sonicated for 30 min and vortexed again for approximately 1 min. They were then centrifuged at 10,000 rpm for 10 min. Aliquots of 4 mL were applied to precleaned 3 cm³ liquid chromatography/silica (LC-Si) cartridges (Supelco, Bellefonte, PA, USA) and the eluent was collected. An additional 2 mL of methanol was added for further elution. Samples were evaporated to approximately 0.5 mL under gentle nitrogen except blank samples which were evaporated to 1.0–1.25 mL. Samples of the same type (spiked, unspiked, or blanks) were combined and the volume of each was adjusted to 15 mL with methanol. Samples were stored at − 20 °C. Prior to preparing aliquots to send to trial participants, and after storing in the freezer for 3 days, a precipitate was noticed

so the samples were centrifuged again to remove solids. The supernatant was poured into a new tube and the volume adjusted to 15 mL with methanol (less than 2 mL was needed). The final dust samples and blanks were distributed as 400 μL aliquots. Aliquots were stored at − 20 °C prior to shipment to ENTACT participants.

Serum extracts were prepared at EPA laboratories by reconstituting *NIST SRM 1957 Organic Contaminants in Non-fortified Human Serum* [18] in 10.7 mL deionized water per the instructions. Twenty-six aliquots of 750 μL serum each were added to Falcon tubes. Thirteen samples were spiked with 10 μL of a nominally 0.05 mM ENTACT mixture 1. Three empty tubes were included for method blanks to undergo the same procedure as the serum samples. The samples received 1500 μL of 0.1 M formic acid and were vortexed for approximately 1 min. The samples then received 10 mL cold acetonitrile (kept at −20 °C until used) and were vortexed briefly. They were centrifuged at 10,000 rpm for 10 min and the supernatant was poured into new tubes. Samples of the same type (spiked, unspiked, or blanks) were combined and the volume of each was adjusted to 19.5 mL (except the blank which was adjusted to 15 mL). Prior to preparing the aliquots, and after storing in the freezer for 3 days, a precipitate was noticed so the samples were centrifuged again to remove solids. The supernatant was poured into a new tube and the volumes were adjusted to their original volume before centrifugation. The final serum samples and blanks were distributed as 400 μL aliquots. Aliquots were stored at − 20 °C prior to shipment.

Silicone band extracts were prepared at Oregon State University (Corvalis, OR). Sixteen silicone bands were cleaned by a water rinse and thermal conditioning [19]. Cleaned bands were stored in airtight jars or bags until use. Bands were then deployed as passive air samplers in a semi-rural outdoor environment for 18 days. After the sampling period, bands were sealed and transported in polytetrafluoroethylene (PTFE) bags and stored at − 20 °C until further processing. Before extraction, bands were cleaned by sequential rinses in high purity deionized water and isopropanol then placed individually in extraction jars for dialysis. Eight bands were spiked with 20 μL of a nominally 0.05 mM ENTACT mixture 5 by applying the mixture to the surface of each band. To remove the DMSO solvent from the ENTACT mixture, spiked bands were sealed in a glass jar and heated for 15 min then cooled to room temperature. All bands were spiked with 500 ng each of ten isotopically labeled and three non-labeled standards (DTXSID indicates the substance identifier in the DSSTox database (vide infra)): naphthalene-D8 (DTXSID10894058), acenaphthylene-D8 (DTXSID00109466), acenaphthene-D10 (DTXSID40893473), phenanthrene-D10 (DTXSID60893475), fluoranthene-D10 (DTXSID20893476), chrysene-D12 (DTXSID00893474), benzo[a]pyrene-D12

(DTXSID00894062), benzo[ghi]perylene-D12 (DTXSID40894066), polychlorinated biphenyl (PCB) 100 (DTXSID8073504), PCB 209 (DTXSID4047541), 9-fluorenone-D8 (DTXSID60894068), 2-methyl-1,4-naphthalenequinone-D8 (DTXSID90703033), and tetrachlorometaxylene (DTXSID6075433). Each band was then submerged in 100 mL of ethyl acetate and placed on an orbital shaker for at least 2 h. The extraction solvent was removed, the extraction was repeated, and the two extraction solvents were combined. The volume was then reduced to 1 mL using a large volume closed cell TurboVap (Biotage, Charlotte, NC) and a small volume nitrogen blowdown TurboVap. Samples of the same type (spiked, unspiked, and blanks) were combined and the volume of each was adjusted to approximately 12 mL. Extracts were shipped overnight to the EPA lab and were kept frozen at − 20 or − 80 °C until aliquots were prepared. The final samples were distributed as 400 μL aliquots. Samples were stored at − 20 °C until shipment. Band blank samples from a different stock of ethyl acetate were prepared at EPA laboratories using 400 μL of ethyl acetate dispensed into the same vial type as samples, using the same pipette and tip stock. While not ideal, this type of blank can help correct for vial, cap, and aliquot process contaminants.

Participants

Institutions in five countries (Canada, Czech Republic, Switzerland, UK, and USA), representing eight government (California Dept. of Public Health, California Dept. of Toxic Substances Control, Eawag, EPA, NIST, Pacific Northwest National Laboratory, Research Centre for Toxic Compounds in the Environment, US Geological Survey); five industry (AB Sciex, Agilent, Leco, Thermo, Waters); and 15 academic (Colorado School of Mines, Cornell Univ., Duke Univ., Emory Univ., Florida International Univ., Icahn School of Medicine at Mt. Sinai, North Carolina State Univ., San Diego State Univ., Scripps Research Institute, Univ. of Alberta, Univ. of Birmingham, Univ. of California at Davis, Univ. of Florida, Univ. of Washington, WI State Laboratory of Hygiene) laboratories are participating in the ENTACT trial.

Shipping

In order to participate in ENTACT, groups were required to either be under contract with EPA or have a material transfer agreement (MTA) with the Agency. A copy of the MTA template is available in ESM Section 1 and describes what each organization received and was responsible for. Once these arrangements were in place, the EPA shipped the samples on dry ice for overnight delivery (domestic) or express delivery (international) using one of two commercial carriers. In cases where shipments were delayed and samples arrived warm, duplicate samples were shipped to minimize sample differences due to increased storage temperatures. Each shipment was confirmed by the recipient using chain of custody forms. Some groups chose to receive and analyze only a subset of ENTACT samples (i.e., ten synthetic mixtures), given their specific research interests and expertise.

Experiments

After samples were received, each participating laboratory was instructed to follow the Standard Operating Procedure (SOP) provided by EPA (ESM Section 2). Samples were stored at ≤ − 20 °C until analysis. Participating laboratories were instructed to analyze study samples (including blanks) in accordance with their existing SOPs and/or methods. Dilution, concentration, and/or solvent exchange was performed at the discretion of each lab, with the expectation of being documented and reported. Participants were provided data (ESM Table S5) and method (ESM Table S6) templates to be used to return results to the EPA, alongside raw data files. File transfer occurred by several methods in practice, but the recommended method was to upload files to an EPA file transfer protocol (FTP) site into individual folders for each participating laboratory. The requested timeline for Phase I was within 180 days from receipt of the samples, and within 270 days for Phase II. This timeline proved overly ambitious, as the samples and analyses are complex, and resource (personnel, instrumentation) shortages were common.

Information returned

Data and method templates (ESM Tables S5 and S6) were designed to standardize the information returned to the EPA for the study. Participants were instructed to complete the templates to the best of their ability, realizing that some information would not be applicable to their analyses and could be left blank, and that some additional information might be provided by their workflow and could be added. Input on the design of the templates was provided by several participants and informed by previous NTA ring trials, covering both GC and LC methods.

EPA requested supplemental results files (e.g., ".d" files from Agilent systems, ".raw" files from Thermo systems, and generic ".mzML" files) from all participants. These files were requested to allow future systematic review using unified data processing techniques. Additionally, these files can provide information on chemicals detected but not intentionally added to either the mixtures or exposure extracts, possibly arising due to impurities, reactions, and laboratory process contamination.

All data returned to the EPA is stored on an FTP site regardless of the mode of transfer. Additional copies of the data may be stored on EPA computers and servers for a variety of purposes. All data template information will be stored in a

relational database, which is currently under development. It is intended that the information will be de-identified and made public for deeper investigations of the data.

EPA's DSSTox chemical database and CompTox chemicals dashboard advanced tools

EPA's DSSTox chemical database represents one of the largest, publicly available sources of manual and auto-curated chemical structure information available to environmental toxicology and exposure researchers, currently exceeding 760,000 registered chemical substances, of which almost 95% are associated with a unique chemical structure representation (mol file). Distinguishing features of this database, making it particularly suited to serve as a reference database for NTA investigations, include: enforcement of a 1:1:1 CAS-name-structure requirement for substance registrations; expert-manual curation effort focused on high interest environmental chemicals (particularly those with limited information or significant uncertainty in the public domain); cheminformatics infrastructure enforcing strict structure-based representations and rules; and focus on chemical content of particular relevance to environmental exposure and toxicology, and of interest to EPA researchers and programs. DSSTox chemical substance records are assigned a unique identifier (DTXSID) and, where possible, associated with a uniquely defined chemical structure (and identifier, DTXCID). DTXSID provides ENTACT participants with an unequivocal way to report the structures found within ENTACT samples. Generic substances and their associated 2D structures are most often registered in association with lists, and at the level of specificity indicated from the source identifiers (typically a CAS and/or name from a publication, document, list, or collaborator). Structures, in turn, can include details pertaining to stereochemistry (E; Z; mixture of E,Z; or relative or absolute stereochemistry at chiral centers), salt or hydrate form, and stoichiometric mixtures. The 1:1:1 CAS-name-structure requirement for a DSSTox substance record, in association with this level of structural specificity, in turn, enables clearer linkages to be made from chemical structure to source data records associated with exposure and toxicological outcomes than is typical of other large chemical databases [20]. Also, pertinent to the ENTACT project, all ToxCast library chemicals are subject to expert-manual curation review at both the supplier-sample (i.e., confirming chemical identity details from supplier documentation) and generic chemical levels (to establish accurate CAS-name-structure assignment) prior to registration and mapping [10].

EPA's CompTox Chemicals Dashboard was developed within EPA's NCCT as a vehicle to surface and add functionality and content linkages to the DSSTox structure database, and as an integrative platform for other NCCT and external content databases [20]. In particular, the Dashboard has become the central hub for displaying ToxCast bioassay data

results, as well as providing tools to explore experimental and predicted physicochemical data, in vivo, in vitro and in silico toxicology data, exposure data and coupled with literature search capabilities across the more than 760,000 chemical substances. Information contained in the Dashboard can be used to support structure identification for ENTACT and general HRMS NTA data and workflows. The Dashboard was initially released in April 2016 and has become a critical component of the informatics systems supporting MS efforts in the agency [16, 21–24]. Specific searches that have been delivered to support MS are available via the Advanced Search page (https://comptox.epa.gov/dashboard/dsstoxdb/advanced_search) and, for batch searching (https://comptox.epa.gov/dashboard/dsstoxdb/batch_search). These include supporting molecular formula searching, mass searching and generation of matching formulae on the Dashboard based on an input mass. These capabilities are summarized via an online video [25].

The level of structure-specificity of DSSTox substance records creates some challenges for the NTA community, which employs a variety of spectroscopic methods to detect desalted, parent ions with monoisotopic mass. In addition to their experimental and predicted properties, the Dashboard provides "MS-ready forms" of the chemical structures that are most relevant to NTA and SSA. ENTACT participants had access to DSSTox MS-Ready structures by downloading DSSTOX_MS_Ready_Chemical_Structures.zip from ftp://newftp.epa.gov/COMPTOX/Sustainable_Chemistry_Data/Chemistry_Dashboard. MS-Ready structures are a more generalized, processed version of the original DSSTox structures that are produced by separating mixtures into individual components, desalting, removing of solvents of hydration, removing stereobonds, and neutralizing structures. These resulting MS-Ready structures retain their mappings to the original DTXCID structures in the database so that a single MS-Ready form of a chemical can map to many DTXSID substances (mapped 1:1 to DTXCID). For example, for Atrazine the MS-ready form maps to 16 unique substances (https://comptox.epa.gov/dashboard/dsstoxdb/ms_ready_mixture?cid=112&gsid=20112&name=Atrazine). A complete description of how MS-ready structure forms are used to facilitate the identification of chemicals based on database searching is reported in McEachran et al. [15]. The MS-Ready structures are generated using a free and open source automated workflow similar to the quantitative structure-activity relationship (QSAR)-ready workflow used for OPEn quantitative structure-activity Relationship Application (OPERA) predictions and is available on Github (https://github.com/kmansouri/MS-ready). MS-Ready files are updated as the DSSTox content grows and versioned releases are available at https://figshare.com/articles/DSSTox_MS_Ready_Mapping_File_11_14_2016/5588575."

Results

As of this writing, 16 groups have submitted data for phase I (blinded analysis) of parts 1 (ToxCast mixtures) and, to a lesser degree, 2 (media extracts) of ENTACT. Seven groups have submitted data for phase 2, and one group has submitted data for part 3 (spectra collection). A summary of the instrumental methods used are presented in Table 1. Two laboratories used GC methods, with both one- and two- dimensional chromatography used, and both low and high-resolution MS. Thirteen laboratories used LC methods and the vast majority used a C18 column (including the two called T3). Seven of the LC labs chose methanol and five chose acetonitrile as their organic phase, with

Table 1 Instrumental methods used for ENTACT to date

Lab #	Separation	Mobile phase	MS type	MS/MS
Gas chromatography				
1	Agilent GC × GC, Restek Rxi-5 ms (30 m × 0.25 mm × 0.25 μm) + Restek Rxi-17Sil MS (0.6 m × 0.25 mm × 0.25 μm)	Helium	Leco HRT+ ToF in EI and CI for confirmation	NA
2	Agilent GC, Agilent J&W VF-5MS (30 m × 0.25 mm × 0.25um)	Helium	Agilent 7010 Triple Quad in EI+	NA
3	Agilent GC, Agilent HP-5 ms Ultra Inert (30 m × 0.25 mm × 0.25 μm)	Helium	Agilent ToF in EI	NA
Liquid chromatography-time of flight MS				
4	Agilent LC, Agilent Zorbax Eclipse Plus C8 (50 × 2.1 mm, 1.8 μm)	Water, methanol, ammonium formate (AF)	Agilent 6530 QToF in ESI ±	Data dependent 10, 20, 40 collision
5	Agilent LC, ZORBAX Eclipse Plus C18 (100 × 2.1 mm, 1.8 μm)	Water, methanol, ammonium acetate, acetic acid	Agilent 6530 QToF in ESI ±	Data dependent 10, 20, 40 collision
6	Agilent LC, Agilent Zorbax RRHD Eclipse Plus C18 (150 × 2.1 mm, 1.8 μm)	Water, acetonitrile, formic acid (FA)	Agilent 6550 QToF in ESI ±	Data dependent 10, 20, 40 collision
7	Agilent LC, Agilent Zorbax Eclipse Plus C18 (100 × 2.1 mm, 1.8 μm)	Water, methanol, ammonium acetate	Agilent 6550 QToF in ESI ±	Data dependent 10, 20, 40 collision
8a	Drift tube ion mobility spectrometry	NA	Agilent 6560 QTOF in nESI ± & APPI ±	NA
9	Dionex LC, Waters XSelect HSS T3 (150 × 3 mm, 3.5 μm)	Water, acetonitrile, FA	Bruker maXis II TOF in ESI +, Bruker maXis II UHR QTOF in ESI +	Data dependent 35 collision energy
10	Waters LC, ACQUITY UPLC BEH C18 (50 × 2.1 mm,1.7um)	Water, methanol, AF	Waters Xevo G2-XS QToF in ESI ±	Data independent, Low = 4, High = ramp 10–45
Liquid chromatography-orbitrap MS				
11a	Drift tube ion mobility spectrometry	NA	Thermo Orbitrap Elite in nESI ±	NA
11b	Thermo LC, Thermo Accucore C30 (150 × 2.1 mm, 2.6 μm)	Water, acetonitrile, isopropanol, AF, FA	Thermo Q Exactive in HESI ±	Data dependent 30, 60 collision
12	Dionex LC, MAC-MOD Analytical ACE Excel C18-PFP (125 × 3 mm, 2 μm)	Water, acetonitrile, FA or ammonium hydroxide	Thermo Q Exactive in ESI ±, APCI ±	Data dependent 15, 30, 45 collision
13	Dionex LC, Waters XBridge BEH C18 (50 × 2.1 mm, 3.5 μm)	Water, methanol, FA	Thermo Q Exactive in HESI ±	Data dependent 50 collision, and varied by m/z
14	Thermo LC, Thermo Hypersil GOLD aQ C18 Polar Endcapped (100 × 2.1 mm, 1.9 μm)	Water, acetonitrile, FA	Thermo Q Exactive in HESI ±	Data independent, stepped 30
15	Dionex LC, Waters Atlantis T3 (150 × 3 mm, 3 μm)	Water, methanol, isopropanol, FA	Thermo Q Exactive Plus in ESI ±	Data Independent, varies 15–120; Dependent 20, 50, 90
16	Waters LC, Thermo Hypersil Gold aQ C18 Polar Endcapped (200 × 2.1 mm, 1.9 μm)	Water, methanol, FA, AF	Thermo Q Exactive Plus in HESI ±, APCI ±	NA
Liquid chromatography-other MS				
8b	Direct infusion	NA	Thermo Velos Pro +21T FT-ICR in ESI ±	NA

the most common modifiers being formic acid and ammonium formate.

Eight labs used time-of-flight (TOF) style instruments, six used Orbitrap style instruments, one used a triple quadrupole, and one used an ultra-high resolution Fourier transform-ion cyclotron resonance (FT-ICR) style instrument. Two laboratories included dissimilar methods that have been designated by "a" and "b" subtypes. Two laboratories also included ion mobility as an orthogonal technique. Both laboratories performing GC analyses used electron ionization (EI) with one adding chemical ionization (CI), while all laboratories performing LC used electrospray ionization (ESI), with three adding atmospheric pressure chemical ionization (APCI); nearly all labs used both positive and negative ionization modes. Nearly all labs included MS/MS experiments, with most using data dependent acquisition (DDA).

As depicted in Table 2, most laboratories report more features than the actual number of spiked substances (red). There are several possible reasons for this, including double counting of reported features (e.g., across ionization modes, multiple isomers) and additional compounds present in the samples that were not intentionally added. The additional compounds could originate from impurities in the neat standards used to prepare the mixtures, reaction or breakdown products from chemicals added to the mixtures, or laboratory contaminants added during the handling of the samples. During further analysis of the results, we will carefully consider the number of laboratories that reported the same feature for chemicals that were not intentionally added to provide some evidence for additional true positive results.

In most cases, a given lab would consistently report a number of chemicals either under (< 75% of added), near (± 25% of added), or over (> 125% of added) the number added to the sample. Results from five labs entirely fell in one category, while seven labs produced results in two of the categories. There were, however, four labs whose results fell into each category, possibly indicating different levels of review across the sample set. Whereas it is tempting to use these values to determine "better" methods, it is not appropriate to do so since

Table 2 Preliminary results for the number of features (defined here as rows in the data reporting template, characterized by retention time, mass spectra, and abundances) reported during phase I of ENTACT for each mixture and fortified matrix. The actual number of chemical substances intentionally added are listed under the mixture numbers. Color coding and text format reflects the percent reported compared to the number of spiked substances: blue, *italics* < 75% (under-report), green, **bold** > 75% and < 125% (near actual), red, normal > 125% (over-report). NR = not reported

| | ToxCast Mixtures | | | | | | | | | | Fortified Matrices | | |
	1	2	3	4	5	6	7	8	9	10	Dust	Serum	Band
Act.	95	95	95	95	185	185	365	365	95	365	365	95	185
Lab 1	128	148	166	187	292	269	**318**	470	177	**410**	NR	NR	NR
2	142	154	**102**	129	250	242	**401**	**399**	**105**	452	NR	NR	NR
3	48	40	48	59	110	101	97	130	37	**109**	NR	NR	NR
4	301	130	375	341	408	404	719	687	198	**327**	NR	NR	NR
5	*65*	*66*	**74**	**72**	*105*	*118*	*193*	*215*	*54*	*162*	NR	NR	NR
6	587	552	596	554	798	846	1327	1274	509	1176	NR	NR	NR
7	**93**	**114**	**116**	**106**	**182**	**201**	**360**	**374**	*73*	**330**	*236*	**92**	*124*
8	337	372	303	365	321	363	466	505	510	463	*259*	222	313
9	135	130	125	154	**188**	**195**	**284**	**295**	**100**	*153*	270	*54*	*101*
10	*70*	*57*	*64*	*66*	*105*	*115*	*176*	*125*	*35*	*159*	NR	NR	NR
11a	595	486	571	630	746	669	899	910	588	792	1009	614	NR
11b	*66*	170	*51*	*41*	272	*116*	*214*	*101*	163	**404**	861	145	557
12	*51*	*37*	*35*	*39*	*74*	*59*	*124*	*109*	*42*	*105*	*124*	*52*	*76*
13	137	*65*	*45*	**74**	*68*	234	**413**	**408**	120	**317**	**389**	178	*88*
14	215	249	212	249	**207**	275	*245*	*254*	140	*253*	NR	NR	NR
15	1298	1258	1304	1209	1651	1641	2520	2538	1202	2193	NR	NR	NR
16	153	217	221	199	254	321	523	651	496	**396**	NR	NR	NR

the number of correct features has not been considered for this preliminary analysis, and reporting qualifications (i.e., Schymanski et al. confidence level [26], detection limits, etc.) could be highly variable between laboratories. It is also of note that nine of 15 labs did not analyze or report values for the spiked matrices; this is likely due to time constraints, and/ or laboratory research interests. Overall, the information returned is highly variable, likely due to the differences in instrumentation, approaches, and workflows across labs. We consider the variation in methodology to be a strength of ENTACT, but it will also make thorough evaluation of the data complex and challenging.

In order to compare the chemical space covered by the different types of methods, we examined results from two labs that have completed both phases I and II, and published a thorough self-evaluation [24, 27] as a case study to provide initial findings. Results from these labs cover GC and LC methods (ESI+ and ESI− were considered separately). It is important to note that these analyses will be repeated after results from all laboratories are submitted. There were a total of 1269 unique substances added to one or more of the ToxCast mixtures; of those, 1074 substances were detected and correctly identified by at least one of the methods. As shown by Fig. 3, 195 substances (15.4%) were not identified by any method; however, likely explanations for missing some compounds include the fact that the GC method relied heavily on matches to two databases [27], and there are known issues with LC detection such as in-source fragmentation and lack of ionizable groups [24]. For both separation techniques, it is quite possible that other methods and laboratories may be able to detect and identify more of these 195 not detected compounds. The majority of correctly identified substances were detected by only one of the three methods (575 compounds, 45.3%). A slightly smaller number of substances were identified with two methods (462 compounds, 36.4%). Finally, the smallest number of substances were identified by all three methods (37 compounds 2.9%). Including replicate detections across chromatography and ionization methods, 809 substances were detected by GC, and 801 by LC (539 by ESI+ and 262 by ESI−).
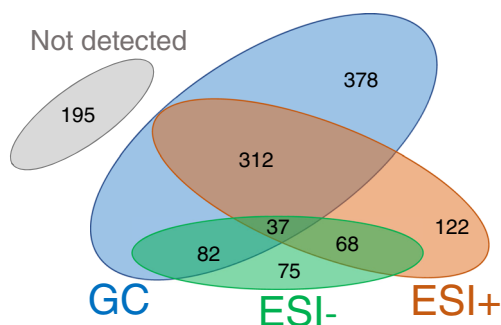

Fig. 3 Methods that correctly identified a subset of 1074 substances (out of 1269 total spiked substances) in the ToxCast mixtures

OPERA is a free and open-source suite of models for predicting physicochemical properties and environmental fate endpoints of relevance to regulatory procedures [28]. OPERA models were applied to the "QSAR-ready" form of the more than 700 K DSSTox structures (in the majority of cases, identical to the "MS-ready" desalted form) and the predictions are available on the Dashboard. OPERA predictions for the ToxCast chemicals were downloaded from the Dashboard to investigate some aspects of the chemical space covered by two laboratories participating in ENTACT. We recognize that these two laboratories are not necessarily representative of all GC and LC methods, and further analysis using results from all ENTACT participants will provide a much richer data set and definitive conclusions. However, this preliminary analysis provides some insight into the chemical space covered by typical NTA methods.

Figure 4 compares MS-Ready molecular weight and OPERA predictions for eight physicochemical properties for compounds only detected by GC, ESI+, ESI−, and non-detected compounds, and all chemicals in the ToxCast and ENTACT programs. As expected, the distribution of values is largest for the ToxCast list, as it contains the greatest number of substances, whereas all other boxes are subsets of those substances. The distributions of each bar are different, with ENTACT and LC-only distributions typically shifting in the same direction, and GC-only shifting in the opposite direction. Because the ENTACT substances were selected with emphasis on LC amenable compounds based on QC data previously collected, the similar distribution shifts are not surprising. There are statistical differences for each physicochemical parameter between GC-only, ESI+ only, ESI− only, and non-detected substances (Kruskal-Wallace non-parametric test, $p < 0.0001$), indicating differences in the chemical spaces covered by each analysis method. While the distributions for ESI+ and ESI− compounds are similar, ESI− tends to have a slightly lower distribution than ESI+ (excepting melting point and water solubility). In some cases, the property is higher for GC-only compounds compared to ESI+ or − compounds (Henry's Law, vapor pressure, and water solubility), whereas in the remaining cases, the property is lower for GC-only compounds (MS-Ready molecular weight, log Koa, Koc, Kow, boiling- and melting- point). We hypothesize that the counterintuitive finding that water solubility is higher for GC-only compounds may be a result of using reverse phase liquid chromatography where the very polar (and highly water soluble) compounds elute in or near the void volume, making peak picking and identifications difficult.

Interestingly, the compounds that could not be detected always fall within the range of the GC- or LC-only compounds, thus suggesting some other factors influencing the inability to detect the compounds, perhaps related to ionization in the mass spectrometer. Additionally, there is chemical space covered by the ToxCast library that may not be
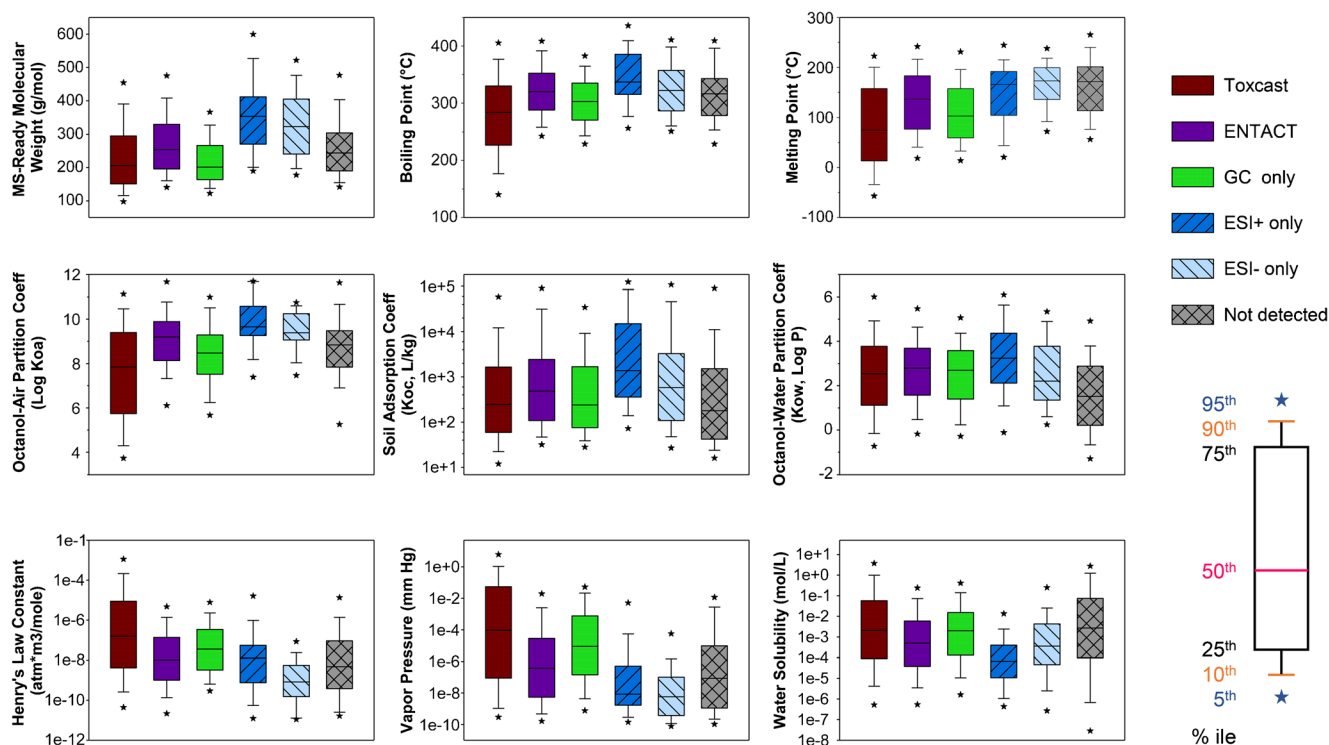
Fig. 4 Box and whisker plots of MS-Ready molecular weight and eight OPERA predicted physicochemical properties segregated by detection method from two laboratories. Represented in red/brick- all ToxCast substances ($n = 4375$); in purple/no pattern- all ENTACT substances ($n = 1254$); in green/confetti- substances only detected by GC ($n = 377$); in dark blue/up-slant- substances only detected by LC ESI+ ($n = 122$); light blue/down-slant- substances only detected by LC ESI− ($n = 75$) in gray/diagonal hatch- substances that were not detected by GC or LC methods ($n = 187$). Percentiles for the boxes and whiskers are shown in the lower right. Molecular weights and OPERA predictions are not available for every chemical substance, which may cause small differences in n values

amenable to GC and LC methods, thus offering the opportunity for further improvements to NTA methodologies.

Discussion

During the EPA's 2015 NTA workshop, we heard details of a previous collaborative non-targeted screening trial, organized by the NORMAN association, which enrolled 18 institutes to analyze a Danube River water extract [12]. Our collaborative trial incorporates many of the NORMAN Network trial's recommendations. On the experimental side of the study, mixtures were developed containing known lists of chemicals from EPA's large ToxCast chemical library. The identity of chemicals in each mixture is made known to participants after initial analysis reports are received, so that retrospective data analysis is built into the trial on both the laboratory and study levels. Another portion of the trial focuses on environmental and biological matrices, to be analyzed both unaltered and after fortification. The samples used were standardized, in two cases using SRMs from NIST, and in the last case using silicone band materials that were deployed together. These samples were delivered as extracts, to remove the variability of extraction methods. Decisions about the mixture and extract preparations lead to more controlled experiments; however, as in the NORMAN trial, we have not been prescriptive about the methods or workflows used by participants. We did not develop separate samples for GC and LC based analyses because of concern that the extracts would be different from the outset, making comparison across labs and methods more difficult.

On the data resource side of the study, we also made decisions to specifically address NORMAN Network trial recommendations. EPA provided the DSSTox database containing ~ 720,000 chemicals at the start of the study to fulfill the NORMAN request for "a more comprehensive suspect list." Both the complete DSSTox database (ftp://newftp.epa.gov/COMPTOX/Sustainable_Chemistry_Data/Chemistry_Dashboard) and a subset of only ToxCast chemicals (ESM Table S1) was shared with all ENTACT participants for use as a suspect screening list. The Dashboard provides data and access to a wide variety of information, including some aspects that specifically support SSA and NTA research (e.g., relevant search functions based on mass and formula). Through some of the relationships developed via ENTACT, many suspect lists have since been added to the DSSTox database and the

Dashboard (https://comptox.epa.gov/dashboard/chemical_lists). Multiple opportunities now exist to develop mass spectral libraries based on the ENTACT project. Instrument vendors and database developers were provided EPA's full library of ToxCast substances as (~ 4700) single samples to allow for the collection of reference mass spectra to serve customers and the public. Additionally, participants who have analyzed the ToxCast mixtures can request (and up to 10 will receive) the corresponding substances as single samples in a well-plate format. These can be used to create in-house spectral libraries, to generate calibration curves for semi-quantitative method development, and to support numerous other research endeavors. Finally, the EPA will create an accessible repository for all the ENTACT data to be used as the basis for retrospective analyses.

Conclusion and next steps

ENTACT analyses are still ongoing at both the data collection and data interpretation levels. Phase I results have been submitted by over half of the participants, but only a handful have completed the self-evaluation for phase II. Because the scope of analyzing individual substances for part 3 is significantly larger than other aspects of the study, only one of those results have been received to date. Once all results for a particular part/phase have been received, a relational database will be constructed so that querying the data and probing specific hypotheses across the entire set of laboratories becomes possible. It is our intention to make that database available to participants so that the wealth of information can be mined to further improve exposomics research and NTA methods and tools. We anticipate that publications from both individual labs evaluating their own results and group publications will be forthcoming. A second workshop (held Aug. 2018) covered both the results of ENTACT to date as well as next steps in the project.

To the best of our knowledge, ENTACT is the first study to make use of synthetic mixtures and multiple reference media extracts to evaluate the successes (and failures) of NTA methods. We believe that the combination of sample types moves the NTA and exposomics fields forward in an unparalleled manner, tackling multiple needs and challenges at once. Methods used a broad range of approaches that cover the current state of the science for NTA well, with many overlaps providing the possibilities of interesting comparisons between and among separation techniques, MS instrumentation, and MS/MS parameters. ENTACT provides a benchmark for current methods and reveals areas for improvement, development, and further research.

Compliance with ethical standards

Human serum was used in this study. Consent was obtained at the time of original collection. However, its use in the current project is considered exempt human subjects research under 40 CRF 26.101(b)(4) because the source is "publicly available" (purchased from vendor) and the subjects cannot be identified (samples were both pooled and coded). In accordance with federal regulations, IRB approval for this secondary use was not required by either EPA or NIST.

Conflict of interest The authors declare that they have no conflict of interest.

Disclaimer The views expressed in this manuscript are those of the authors and do not necessarily represent the views or policies of the US Environmental Protection Agency. Any mention of trade names, products, or services does not imply an endorsement by the US Government or the US Environmental Protection Agency. The EPA does not endorse any commercial products, services, or enterprises.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutionalaffiliations.

References

1. Wild CP. The exposome: from concept to utility. Int J Epidemiol. 2012;41(1):24–32. https://doi.org/10.1093/ije/dyr236.
2. Wild CP, Scalbert A, Herceg Z. Measuring the exposome: a powerful basis for evaluating environmental exposures and cancer risk. Environ Mol Mutagen. 2013;54(7):480–99. https://doi.org/10.1002/em.21777.
3. Rappaport SM. Implications of the exposome for exposure science. J Expo Sci Environ Epidemiol. 2010;21:5. https://doi.org/10.1038/jes.2010.50.
4. Rappaport SM, Smith MT. Environment and disease risks. Science. 2010;330(6003):460–1. https://doi.org/10.1126/science.1192603.
5. 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;100:32–61. https://doi.org/10.1016/j.envint.2016.11.026.

6. Geer Wallace MA, Pleil JD, Mentese S, Oliver KD, Whitaker DA, Fent KW. Calibration and performance of synchronous SIM/scan mode for simultaneous targeted and discovery (non-targeted) analysis of exhaled breath samples from firefighters. J Chromatogr A. 2017;1516:114–24. https://doi.org/10.1016/j.chroma.2017.07.082.

7. Naz S, Vallejo M, García A, Barbas C. Method validation strategies involved in non-targeted metabolomics. J Chromatogr A. 2014;1353:99–105. https://doi.org/10.1016/j.chroma.2014.04.071.

8. Hollender J, Schymanski EL, Singer HP, Ferguson PL. Nontarget screening with high resolution mass spectrometry in the environment: ready to go? Environ Sci Technol. 2017;51(20):11505–12. https://doi.org/10.1021/acs.est.7b02184.

9. Sobus JR, Wambaugh JF, Isaacs KK, Williams AJ, McEachran AD, Richard AM, et al. Integrating tools for non-targeted analysis research and chemical safety evaluations at the US EPA. J Expo Sci Environ Epidemiol. 2017. https://doi.org/10.1038/s41370-017-0012-y.

10. Richard AM, Judson RS, Houck KA, Grulke CM, Volarath P, Thillainadarajah I, et al. ToxCast chemical landscape: paving the road to 21st century toxicology. Chem Res Toxicol. 2016;29(8):1225–51. https://doi.org/10.1021/acs.chemrestox.6b00135.

11. Wild CP. Complementing the genome with an "Exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomark Prev. 2005;14(8):1847–50. https://doi.org/10.1158/1055-9965.epi-05-0456.

12. Schymanski EL, Singer HP, Slobodnik J, Ipolyi IM, Oswald P, Krauss M, et al. Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. Anal Bioanal Chem. 2015;407(21):6237–55. https://doi.org/10.1007/s00216-015-8681-7.

13. Schymanski E, Neumann S. CASMI: and the winner is. Metabolites. 2013;3(2):412. https://doi.org/10.3390/metabo3020412.

14. Schymanski EL, Ruttkies C, Krauss M, Brouard C, Kind T, Dührkop K, et al. Critical assessment of small molecule identification 2016: automated methods. J Cheminf. 2017;9(1):22. https://doi.org/10.1186/s13321-017-0207-1.

15. McEachran AD, Mansouri K, Grulke C, Schymanski EL, Ruttkies C, Williams AJ "MS-ready" structures for non-targeted high resolution mass spectrometry screening studies. J Cheminf. 2018;10(1):45. https://doi.org/10.1186/s13321-018-0299-2.

16. Rager JE, Strynar MJ, Liang S, McMahen RL, Richard AM, Grulke CM, et al. Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring. Environ Int. 2016;88:269–80. https://doi.org/10.1016/j.envint.2015.12.008.

17. Gonzalez CA, Choquette SJ. Certificate of analysis: standard reference material 2585 organic contaminants in house dust. Gaithersburg: National Institute of Standards and Technology (NIST); 2018.

18. Gonzalez CA, Choquette SJ. Certificate of analysis: Standard Reference Material 1957 Organic Contaminants in Non-Fortified Human Serum (Freeze-Dried). Gaithersburg: National Institute of Standards and Technology (NIST); 2016.

19. Anderson KA, Points GL III, Donald CE, Dixon HM, Scott RP, Wilson G, et al. Preparation and performance features of wristband samplers and considerations for chemical exposure assessment. J Expo Sci Environ Epidemiol. 2017;27:551. https://doi.org/10.1038/jes.2017.9 https://www.nature.com/articles/jes20179#supplementary-information.

20. Williams AJ, Grulke CM, Edwards J, McEachran AD, Mansouri K, Baker NC, et al. The CompTox chemistry dashboard: a community data resource for environmental chemistry. J Cheminf. 2017;9(1):61. https://doi.org/10.1186/s13321-017-0247-6.

21. Newton SR, McMahen RL, Sobus JR, Mansouri K, Williams AJ, McEachran AD, et al. Suspect screening and non-targeted analysis of drinking water using point-of-use filters. Environ Pollut. 2018;234:297–306. https://doi.org/10.1016/j.envpol.2017.11.033.

22. Phillips KA, Yau A, Favela KA, Isaacs KK, McEachran A, Grulke C, et al. Suspect screening analysis of chemicals in consumer products. Environ Sci Technol. 2018;52(5):3125–35. https://doi.org/10.1021/acs.est.7b04781.

23. McEachran AD, Sobus JR, Williams AJ. Identifying known unknowns using the US EPA's CompTox chemistry dashboard. Anal Bioanal Chem. 2017;409(7):1729–35. https://doi.org/10.1007/s00216-016-0139-z.

24. Sobus JR, Grossman JN, Chao A, Singh R, Williams AJ, Grulke CM, Richard AM, Newton SR, McEachran AD, Ulrich EM (2018) Using prepared mixtures of ToxCast chemicals to evaluate non-targeted analysis (NTA) method performance. Anal Bioanal Chem. https://doi.org/10.1007/s00216-018-1526-4.

25. U.S. Environmental Protection Agency (2018) EPA's CompTox chemistry dashboard advanced search how-to. Youtube.com.

26. Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ Sci Technol. 2014;48(4):2097–8. https://doi.org/10.1021/es5002105.

27. Fell L, Richards T, Binkley J. Comprehensive, non-target characterisation of blinded environmental exposome standards using GCxGC and high resolution time-of-flight mass spectrometry. Chromatography Today. 2018;11(1):20–5.

28. Mansouri K, Grulke CM, Judson RS, Williams AJ. OPERA models for predicting physicochemical properties and environmental fate endpoints. J Cheminf. 2018;10(1):10. https://doi.org/10.1186/s13321-018-0263-1.