Published in final edited form as:
Anal Bioanal Chem. 2020 August ; 412(20): 4931–4939. doi:10.1007/s00216-020-02716-3.

Expanded coverage of non-targeted LC-HRMS using atmospheric pressure chemical ionization: a case study with ENTACT mixtures

Randolph R. Singh[1,2], **Alex Chao**[1], **Katherine A. Phillips**[3], **Xin Rui Xia**[4,5], **Damian Shea**[4,5], **Jon R. Sobus**[3], **Emma L. Schymanski**[2], **Elin M. Ulrich**[3]

[1]Oak Ridge Institute for Science and Education Fellow, United States Environmental Protection Agency, Office of Research & Development, National Exposure Research Laboratory, Research Triangle Park, NC 27711, USA

[2]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 6, avenue du Swing, 4367 Belvaux, Luxembourg

[3]U.S. Environmental Protection Agency, Office of Research & Development, National Exposure Research Laboratory, 109 TW Alexander Dr., Research Triangle Park, NC 27711, USA

[4]Department of Biological Sciences, North Carolina State University, Raleigh, NC 27695, USA

[5]Statera Environmental Inc., 5116 Olde South Road, Raleigh, NC 27606, USA

Abstract

Non-targeted analysis (NTA) is a rapidly evolving analytical technique with numerous opportunities to improve and expand instrumental and data analysis methods. In this work, NTA was performed on eight synthetic mixtures containing 1264 unique chemical substances from the U.S. Environmental Protection Agency's Non-Targeted Analysis Collaborative Trial (ENTACT). These mixtures were analyzed by atmospheric pressure chemical ionization (APCI) and electrospray ionization (ESI) using both positive and negative polarities for a total of four modes. Out of the 1264 ENTACT chemical substances, 1116 were detected in at least one ionization mode, 185 chemicals were detected using all four ionization modes, whereas 148 were not detected. Forty-four chemicals were detected only by APCI, and 181 were detected only by ESI. Molecular descriptors and physicochemical properties were used to assess which ionization type was preferred for a given compound. One ToxPrint substructure (naphthalene group) was found to be enriched in compounds only detected using APCI, and eight ToxPrints (e.g., several alcohol moieties) were enriched in compounds only detected using ESI. Examination of physicochemical

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

Compliance with ethical standards

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

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00216-020-02716-3) contains supplementary material, which is available to authorized users.

Data availability The datasets generated during and/or analyzed during the current study are available in the EPA's ScienceHub and/or the accompanying supporting information.

Disclaimer The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

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

parameters for ENTACT chemicals suggests that those with higher aqueous solubility preferentially ionized by ESI−. While ESI typically detects a larger number of compounds, APCI offers chromatograms with less background, fewer co-elutions, and additional chemical space coverage, suggesting both should be considered for broader coverage in future NTA research.

Keywords

Atmospheric pressure chemical ionization (APCI); Non-targeted analysis(NTA); High-resolution mass spectrometry (HRMS); Electrospray ionization (ESI); EPA's Non-Targeted Analysis Collaborative Trial (ENTACT)

Introduction

Non-targeted analysis (NTA) methods are popular in environmental and exposomic applications, enabling the rapid discovery of previously unknown or understudied compounds [1]. NTA studies often use high-resolution mass spectrometry (HRMS) instrumentation with Orbitrap or quadrupole time-of-flight (QTOF) mass spectrometers, and gas (GC) or liquid chromatography (LC) for compound separation [2]. More volatile compounds are amenable to GC, which typically uses electron ionization (EI, a "hard" ionization technique) at 70 eV and yields reproducible spectra that can be easily matched to spectral libraries (when present). Less-volatile compounds are better suited for LC, which most often uses electrospray ionization (ESI) [2] yielding fewer fragment ions from the parent molecule. The "softer" ionization techniques (e.g., ESI) enable examination of a compound's accurate mass and isotope profile, often allowing characterization at the molecular formula level [3]. Further elucidation of the compound's chemical structure is supported by selective fragmentation such as that produced by tandem mass spectrometry with collision-induced dissociation [2, 4].

Atmospheric pressure chemical ionization (APCI) is another soft ionization technique for mass spectrometric analysis that generates minimal fragmentation and often shows less matrix interference compared with ESI [5, 6]. In addition, APCI provides a wider linear dynamic range and is less prone to form molecular adducts [7–9]. Ion formation in APCI is proposed to occur based on the following simplified mechanisms:

$$\mathrm{M} + e^- \rightarrow \mathrm{M}^{+\bullet} + 2e^- \quad (1)$$

$$\mathrm{S} + e^- \rightarrow \mathrm{S}^{+\bullet} + 2e^- \quad (2)$$

$$\mathrm{S}^{+\bullet} + \mathrm{M} \rightarrow \mathrm{M}^{+\bullet} + \mathrm{S} \quad (3)$$

$$\mathrm{S}^{+\bullet} + \mathrm{S} \rightarrow [\mathrm{S} + \mathrm{H}]^+ + [\mathrm{S} - \mathrm{H}] \quad (4)$$

$$[S + H]^+ + M \rightarrow [M + H]^+ + S \quad (5)$$

where M: analyte molecule S: solvent molecule/reagent gas [8].

Molecular ions ($M^{+\cdot}$) may form either through direct interaction of the analyte molecule with an electron (Reaction 1) or with a solvent radical ion ($S^{+\cdot}$, Reaction 3), which forms when a solvent molecule encounters an electron (Reaction 2). Because of the dynamic nature of molecules in the gas phase, the solvent radical ions may also interact with another solvent molecule (Reaction 4) giving rise to protonated solvent molecules, which may subsequently react with a neutral analyte molecule to form the protonated molecular ion ($[M + H]^+$, Reaction 5). A more detailed description of these ionization processes may be found in work by Herrera and co-workers [8]. Using this proposed mechanism, the reaction responsible for protonated molecular ion formation is through Reaction 5, whereas the radical ion formation described in Reaction 3 is also known to occur, but to a lesser extent [8]. However, when performing NTA with APCI, assuming that ionization will always occur according to Reaction 5 may lead to incorrect compound identifications. During targeted analysis method development, for selected reaction monitoring, in particular, one can easily determine the identity of the molecular ion and which mechanism it follows from experimental data, making data interpretation straightforward [10, 11]. However, this is not the case when using APCI in a NTA approach for complex mixtures like biological and environmental matrices, where the identities of analyte molecules are not known in advance.

Many applications of APCI coupled with HRMS have focused on the screening of metabolites (e.g., lipids and steroids) that belong to the same class or functional group, which to a limited extent, can be expected to ionize in the same manner as the parent compound [12, 13]. Since the predominant ionization reaction is molecule dependent, it cannot be assumed that all ions formed would be the $[M + H]^+$ species. Formation of $[M]^{+\cdot}$ species complicates data interpretation because its M + 1 ion may contribute to the intensity of the $[M + H]^+$ ion in the cases where both reactions occur. While ample mass resolving power may provide a means by which the two can be discriminated, it requires awareness of the experimenter to take the necessary steps to handle the data appropriately, particularly when using automated data processing. Software settings that assume only $[M + H]^+$ forms when in reality both $[M +H]^+$ and $[M]^{+\cdot}$ occur may lead to incorrect assignment of two different formulas to the same molecule. Correct characterization at the formula level requires that the data obtained have isotopic fidelity, which may be problematic if one reaction is strongly preferred [3].

In this work, we use the wealth of information that can be obtained from the analysis of mixtures of known chemicals to establish which known mechanisms for ion formation still hold when dealing with non-target APCI-HRMS data annotation [14, 15]. The Environmental Protection Agency's Non-Targeted Analysis Collaborative Trial (ENTACT) developed a subset (n = 1269) of the Toxicity Forecaster (ToxCast) chemical library into ten chemical mixtures with varying degrees of complexity [14]. The chemical diversity and sheer number of known compounds in the ENTACT mixtures are uniquely positioned to help establish whether a specific ionization mechanism is preferred within a large domain of

applicability using NTA, and the extent to which each reaction occurs. This will enable more scientifically sound decisions in designing data analysis workflows. Further, this work will help establish the potential advantages of using APCI in NTA as it describes sections of chemical space that are revealed when using APCI vs. ESI. We qualitatively assess the likelihood that individual compounds would be detected using either APCI or ESI using their structural motifs, or ToxPrints. Additionally, we evaluate several readily available predicted physicochemical properties for their importance in determining which chemicals are detected in which ionization mode and polarity.

Materials and methods

Chemicals

EPA's ToxCast chemical library was used to formulate ten chemical mixtures for ENTACT as described in Ulrich et al. [14]. Briefly, the mixtures contained between 95 and 365 chemical substances each and were identified by numeric identifiers ranging from #499 to #508 ($N = 95$ for mixtures #499–502, $N = 185$ for mixtures #503–504, and $N = 365$ for mixtures #505–506). Two mixtures (#507 and #508) were designed to be more challenging (i.e., by including a larger number of isomers, isobars, and low molecular weight compounds) and were not used in the present study. The mixtures were prepared in dimethyl sulfoxide (DMSO) at a concentration of approximately 0.05 mM per chemical. For this analysis, ENTACT mixtures (10 μL) were diluted with 490 μL 50:50 acetonitrile/water. Two 250 μL aliquots were made after vortex mixing for ESI+/− and APCI+/− analyses. Acetonitrile and methanol (MeOH) were HPLC grade purchased from Sigma-Aldrich (St. Louis, MO). Deionized water was prepared by using a PicoPure Water System (Durham, NC).

Instrumentation

Chromatographic separation of the chemical mixtures was carried out using a Waters Acquity ultra-high-performance liquid chromatography (UPLC) system (Milford, MA, USA) equipped with a Hypersil GOLD aQ C18 analytical column (200 mm × 2.1 mm, 1.9-μm particle size, Thermo Fisher Scientific, San Jose, CA). The mobile phase consisted of 0.1% (v/v) formic acid and 4 mM ammonium formate in water (A) and 0.1% (v/v) formic acid and 4 mM ammonium formate in MeOH (B). The analysis started with 20% B for 1 min and ramped linearly to 100% B in 30 min. This composition was held for 10 min, then decreased to 20% B in 0.5 min, followed by a re-equilibration time of 9.5 min (total run time 50 min). The flow rate was 0.3 mL/min and the column temperature was 25 °C. A single replicate using 7.5 μL of each sample and included ENTACT solvent blank was injected using the autosampler module of the UPLC system for approximately 7.5 pmol per compound injected on-column.

The UPLC system was coupled to a Q-Exactive Plus Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with an IonMax Atmospheric Pressure Ionization (API) source. Full scan mass spectra MS1 were acquired at a mass resolution of 70,000 at 200 m/z. The mass range was 100 to 1500 Da using positive/negative ion switching mode. A heated electrospray interface (HESI−II) was used to acquire the ESI+/− data with optimal

ionization source working parameters (for the given LC flow rate): a sheath gas flow of 32 arbitrary units (au), auxiliary gas flow of 7 au, a spray voltage of 3500 V, a capillary temperature of 310 °C, and a vaporizer temperature of 200 °C. The optimal MS parameters of the Q-Exactive Plus MS were set at a S-lens RF-level of 50, a S-Lens voltage of 21 V, and a skimmer voltage of 15 V. The full scans were applied by targeting an automatic gain control (AGC) of 10^6 and a maximum injection time of 100 ms for ESI+ and 200 ms for ESI −. The same instrument conditions were employed for APCI data acquisition except using a sheath gas flow of 30 au, auxiliary gas flow of 5 au, a spray voltage of 4000 V, a capillary temperature of 320 °C, and a vaporizer temperature of 30 °C. Instrument control and data acquisition were carried out with Xcalibur 4.0 (Thermo Fisher Scientific, USA).

Data analyses

MS-Ready structures for the ENTACT mixture chemicals were generated according to the protocol developed in McEachran et al [16]. which uses Konstanz Information Miner (KNIME) to de-salt, de-solvate, and remove stereochemistry from a chemical structure [16]. Masses for each intentionally spiked compound were calculated for all ionization mechanisms and polarities by adding (ESI/APCI+) or subtracting (ESI/APCI-) the mass of a proton (1.0079 amu) to/from the MS-Ready monoisotopic masses. For APCI+, the loss of an electron (0.0005 amu) was also considered. Xcalibur software was used to extract chromatograms for the calculated masses in the relevant sample run and peaks were manually inspected in an unblinded fashion. Features [a chromatographic peak with retention time, masses (i.e., m/z peaks related to the monoisotopic ion and associated isotopologues, adducts, etc.), and abundances] matching multiple compounds in the ENTACT mass list due to the presence of isomeric or isobaric compounds were dropped due to the inability to unequivocally identify the compound in the context of the present work. Spiked compounds were deemed present after expert review and inspection of the data for exact mass match (< 5 ppm), isotope pattern, peak shape, minimum peak width of 0.1 min, and peak intensity greater than 3× the level in the blank. Peaks were excluded from further analysis if they did not meet all criteria. Because this set of samples was created by intentionally adding known chemical substances and isomers/isobars were excluded, matching accurate mass from MS1 features was deemed sufficient for compound determination in the absence of additional fragmentation information. Correlation of retention times between ionization modes was also considered.

To assess different structural features that can contribute to the ionization of a chemical via one or both ionization methods, odds ratios were calculated for detected chemicals that only ionized by +/− APCI or +/− ESI with specific structural fingerprints. A detailed description is provided in the Supporting Information (SI, Odds ratio calculation section). ToxPrint fingerprints were used to identify canonical substructures in all compounds [17, 18]. These fingerprints consist of 729 chemical substructures, such that each chemical’s fingerprint is a binary array of length 729. A value of 1 is assigned to an element in the array if the chemical has that substructure; otherwise a value of 0 is assigned.

After removing duplicates, 1264 unique compounds comprised the MS test set for ENTACT mixtures #499–#506. Twelve available OPEn structure-activity Relationship App (OPERA)

predicted physicochemical property estimates were obtained using Distributed Structure-Searchable Toxicity (DSSTox) substance identifier (DTXSID) for the spiked substances via the U.S. EPA CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard) [19, 20]. MS-Ready monoisotopic mass, melting point (MP), K_{ow}, and water solubility were selected as properties potentially affecting type of ionization (APCI, ESI) and polarity of ionization (+, -). Based on the different modes of ionization used for analysis, results from NTA experiments were classified into six unique chemical categories. Specifically, each MS-Ready compound was classified into the following groups: (1) observed using multiple methods; (2) observed using APCI+ only; (3) observed using APCI- only; (4) observed using ESI+ only; (5) observed using ESI− only; or (6) not observed using any method. Physicochemical property estimates for MS-ready compounds for groups 2–6 were compared using the Kruskal-Wallis test [21]. Post hoc pairwise comparisons were then performed using Dunn's multiple comparisons test [22]. Statistical testing was performed using GraphPad Prism 7 (San Diego, CA); significant differences were reported when $p < 0.05$.

Results and discussion

Ionization mode summary

The detection results for the eight (#499 to #506) ENTACT mixtures analyzed by APCI and ESI were compared. A complete listing of all ENTACT chemicals found, and the corresponding ionization mode and polarity with which they were observed, can be found in Table S1 in the Electronic Supplementary Material (ESM). The percent of compounds that can be detected in each mixture is shown in ESM Fig. S1 and ranges between 60 and 78% for APCI and 78–88% for ESI and did not appear to be systematically impacted by the complexity of the mixture. A Venn diagram of the breakdown for ionization modes and polarities is shown in Fig. 1. For APCI, 935 (74%) substances were observed, including 44 compounds unique to APCI. For ESI, 1072 (85%) substances were observed, including 181 compounds unique to ESI, about four times that which were unique to APCI. Out of the total 1264 ENTACT substances analyzed, 148 were not detected in any mode/polarity, and 185 were observed in all four tested modes. Non-detection in the context of this work can either mean that (a) the compound is not detectable using a specific ionization mode at environmentally relevant concentrations or (b) the limit of detection is higher than the sample concentration. As the samples in this work were prepared approximately at the higher end of environmentally relevant concentrations with environmental applications in mind, the compounds for which no peak was observed were considered not detected.

As expected, more compounds ionized in positive polarity (blue bar) than in negative polarity (orange bar; ESM Fig. S1). Considering only positive mode results, 22 and 119 unique compounds were detected using APCI and ESI, respectively. Considering only negative mode results, 19 and 37 unique compounds were detected using APCI and ESI, respectively. Finally, considering detections across both polarities, 3 unique compounds were detected using APCI and 25 unique compounds were detected using ESI.

Assessing results in the context of both types of ionization Reactions (3 and 5) was imperative to prevent data misinterpretation. Upon initial investigation, 30 chemicals

appeared to ionize only via Reaction 3 ($[M]^{+\cdot}$ formation); however, further inspection revealed that most of these compounds have a permanent positive charge due to having a quaternary N or P, or trivalent O atom. Once compounds with a permanent positive charge were removed from the list, we identified eight chemicals that ionize through Reaction 3 instead of Reaction 5. These chemicals include: chlorpropham (DTXSID7020764), lactofen (DTXSID7024160), 1, 3-dinitronaphthalene (DTXSID9025164), 3-hydroxy-2-naphthoic acid (DTXSID3026560), aloe-emodin (DTXSID2030695), nitrothal-isopropyl (DTXSID5037579), bismaleimide (DTXSID8044381), and profluralin (DTXSID2044559). Thus, the results demonstrate that more compounds ionized through the conventional Reaction 5 ($[M +H]^+$) mechanism rather than the molecular ion formation (Reaction 3). Overall, less than 1% of the analytes present in the ENTACT mixtures ionized by molecular ion formation ($M^{+\cdot}$). Having such a low percentage of compounds ionizing through $[M]^{+\cdot}$ formation over a broad range of chemicals indicates that simply screening for the $[M + H]^+$ is when developing data analysis workflows is a reasonable assumption, although both should be kept in mind.

Despite multiple ionization reaction possibilities in APCI (see SI In-house mixture section, and references [23–29] therein) the majority of compounds identified were ionized by the acquisition or loss of a proton, $[M + H]^+$ or $[M H]^-$, respectively. While it is tempting to conclude that the mechanism observed in ESI is the same driving force for ion formation in APCI, the predominant process is affected by many factors including but not limited to: choice of mobile phase solvent, additives, spray conditions, and the inherent properties of the molecule under study [30]. While specific changes can be implemented to promote one reaction over the other, it does not guarantee that all analytes in a mixture or sample would undergo the reaction being promoted. One of the conclusions from this work is that the experimenter must keep both Reactions 3 and 5 in mind when interpreting APCI data, in addition to other special cases, because of the unknown extent by which each reaction may contribute to the overall ionization process. While the reactions presented in this work are not exhaustive of all potential ion formation mechanisms, they demonstrate the importance of considering reactions beyond those most common. This also highlights the benefit of MS libraries to ease proper identification for real world samples.

Structural properties determining observation in APCI or ESI

The variety of chemical substructures along with their many combinations are likely to play a role in determining if a compound is observed by APCI and/or ESI. Enriched odds ratios were calculated for substances that were identified via only one of the two ionization methods as a way of determining if there were significant structural differences between compounds identified via only ESI (181 compounds) or via only APCI (44 compounds). It is hoped that by studying these enriched features, it will be possible to determine which ionization method would be appropriate to use when seeking to identify specific types of compounds in future analyses.

Chemotypic enrichment analysis of the chemicals in the ENTACT data set that were observed only by a single ionization type and polarity was performed using computed odds ratios to identify functional groups or structural features that were more likely to be

observed in one ionization mode versus the other (ESM Table S2). Furthermore, the enrichment of substructures with the odds ratios was determined by statistical testing which is further described in the SI (Odds ratio calculation section). Only one ToxPrint was found to be enriched in compounds only detected using APCI (Table 1). Two fused benzene rings (i.e., a naphthalene group) were more than seven times as likely to be found in compounds detected only via APCI as compounds detected only via ESI. There are 37 compounds containing this naphthalene group out of 1263 compounds (DTXSID1047364 does not have a structure available in DSSTox, which is required to determine ToxPrints). Of those 37, 5 were identified only via APCI, another 3 were identified only via ESI, and the remaining 29 were identified using both ESI and APCI. The naphthalene feature was considered enriched in APCI as explained by the equation for odds ratio (Equation SI1) and noting that out of the 44 compounds observed in APCI alone, 5 of them have the naphthalene substructure. On the other hand, only 3 out of 181 compounds observed only via ESI had the naphthalene substructure. This then renders the naphthalene group as being considered enriched in APCI compared with ESI. These results are consistent with studies where naphthalene-like chemicals, such as polyaromatic hydrocarbons, were detected using an APCI source [31, 32].

Eight ToxPrint substructures were enriched for compounds detected only via ESI (Table 1). Of these substructures, four ToxPrints indicate that compounds with alcohol groups (namely primary, secondary, or generic alcohol groups) are 3 to 8 times more likely to be identified with ESI than other compounds, depending on which type of alcohol group is present. In addition to alcohol groups, three alkane chain ToxPrints had infinite odds of being identified via only with ESI: chain:alkaneCyclic_pentyl_C5, chain:alkaneLinear_hexyl_C6, and chain:alkaneLinear_octyl_C8. Additionally, compounds with the chain:alkaneCyclic_ethyl_C2(connect_noZ) were 6 times more likely to be detected only via ESI. While ~ 80% of the compounds with these four ToxPrints were identified using both ESI and APCI (199/244), only one of the 244 compounds was detected only with APCI. This could indicate that detection of compounds with these ToxPrint features is heavily favored by ESI. As with compounds containing a fused naphthalene structure in APCI, a large number of compounds identified using both methods do not necessarily discount the enrichment of the feature in only one mode or the other. Substances containing fused naphthalene rings (APCI) or alkane chains or rings (ESI) are not the only ToxPrints in these compounds. It is likely that another part of the substance’s structure contributes to identification in the other ionization mode, rather than the enriched feature contributing to detection in both modes. It should be noted, however, that while application of feature enrichment highlights ToxPrints that are more abundant in compounds identified in one method compared with another, it does not mean that an enriched feature is the cause of a compound being identified via one method over another.

Physicochemical properties determining ionization

Cognizant of the chemical diversity of the ENTACT mixtures, physicochemical properties were obtained from the U.S. EPA CompTox Chemicals Dashboard. Table S1 (see ESM) provides a statistical summary and Fig. 2 shows box and whisker plots of the results for MS-ready monoisotopic mass, MP, water solubility, and K_{ow} values across the two different

ionization modes (APCI/ESI), two different polarities (+/−)and non-detected substances. There was no overall effect of measurement group on monoisotopic mass. Median monoisotopic mass values across the five groups spanned 228 Da (for *ESI− only*) to 270 Da (for *APCI- only*), and individual chemicals ranged from 85.0 to 642 Da. An overall significant group effect was observed for MP ($p = 0.009$), with significant pairwise differences observed between *ESI+* only and *ESI−only*, and between *ESI− only* and not observed. The highest MP values were observed for *ESI− only* compounds (median = 170 °C), and the lowest values for *ESI+ only* (median = 122 °C) and *not observed* (median = 121 °C). MP is a reflection of the strength of a molecule's intermolecular forces of attraction in its pure state and thus it is counterintuitive why there is some correlation with the ionization mode. Having said that, the intermolecular forces of attraction that govern MP might shed more light on the ionization process and should be further investigated. Acid disassociation constants (pK_a) are known to impact whether a molecule is ionized by APCI or ESI [30]; however, pK_a was not one of the OPERA predicted properties that was available from the U.S. EPA CompTox Chemicals Dashboard at the time of this analysis.

Water solubility was observed to significantly vary across groups ($p < 0.0001$), with pairwise differences between *ESI+ only* and *ESI− only*, and between *ESI− only* and not observed. Here, the lowest values were clearly observed for *ESI+ only* (median = 0.789 mmol/L) and the highest values for *ESI−only* (median = 34.07 mmol/L). Results for K_{ow} were very similar to those for MP, and the inverse of those for water solubility. Specifically, we observed an overall significant group effect ($p = 0.0006$), as well as pairwise differences between *ESI+ only* and *ESI− only*, and between *ESI− only* and *not observed.* The highest and lowest median values were observed for *ESI+ only* ($K_{ow} = 2.69$) and *ESI− only* ($K_{ow}= 0.73$), respectively. No pairwise difference for any physicochemical properties were observed for the *APCI+ only* or *APCI- only* groups. This result likely stems, in part, from the lower number of observations in these groups ($n \leq 22$) relative those in *ESI+ only* ($n =118$)and *not observed* ($n = 148$) groups. Despite having a low number of observations specific to APCI, the data suggests that for the subsets of ENTACT chemicals examined, there was no significant difference between the physicochemical properties of the molecules for which APCI or ESI are amenable. The lack of significant difference between the ionization-relevant physicochemical properties for the subsets of ENTACT chemicals examined, and the large number of compounds detected by both ionization types suggests comparable performance of ESI and APCI for NTA for the ENTACT mixtures examined. As other physicochemical properties (i.e., pK_a [33]) become available through the U.S. EPA CompTox Chemicals Dashboard, similar analyses can be performed to identify additional relevant factors for determining a chemical's favored ionization mode.

Benefits of APCI for non-targeted analysis studies

This work has provided a unique opportunity to characterize APCI ionization qualitatively with a diverse set of 1264 known chemicals. In this way, it was possible to explore the chemical space revealed by using APCI ionization and the mechanisms involved in ionization. Initial analyses of chemical substructures as well as physicochemical properties have revealed potential factors contributing to the likelihood of ionization of a compound by either ESI or APCI, in either negative or positive mode. The incorporation of additional

properties such as pK_a into the analyses may lead to further under-standing and prediction of preferred ionization mode and polarity for a given compound. While ESI has the very strong benefit of detecting a larger number of compounds, APCI offers additional chemical space coverage, particularly for negative polarity. An increase of 44 (4%) more chemicals were detected by including APCI, compared with using ESI alone. Using APCI in addition to ESI when performing NTA comes with experimental analysis costs, as well as ionization and data analysis complexities. While dual APCI/ESI sources are available for some instruments, these tend to be more expensive than one ionization source. While APCI may not be essential to all NTA because the hypotheses and chemical space of interest may not warrant additional analyses and effort, the current work shows the advantages of inclusion, and gives examples of substructures and physicochemical parameters to consider in future research. Future work could include an investigation into the quantitative similarities and differences of APCI and ESI to provide a deeper comparison of the two techniques.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge support from the Oak Ridge Institute for Science and Education and Luxembourg National Research Fund for project A18/BM/12341006. Special thanks to Chris Grulke and Antony Williams for their support with ENTACT and related data, Daniel Todd for assistance with the LC/MS analysis, and James McCord for scientific discussions.

References

1. 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 U.S. EPA. J Expo Sci Environ Epidemiol. 2018;28:411–26. 10.1038/s41370-017-0012-y. [PubMed: 29288256]
2. 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. 10.1016/j.envint.2016.11.026. [PubMed: 28062070]
3. Kind T, Fiehn O. Seven Golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. Bmc Bioinformatics. 2007;8:105 10.1186/1471-2105-8-105. [PubMed: 17389044]
4. Halket JM, Waterman D, Przyborowska AM, Patel RKP, Fraser PD, Bramley PM. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. J Exp Bot. 2004;56(410):219–43. 10.1093/jxb/eri069. [PubMed: 15618298]
5. Souverain S, Rudaz S, Veuthey J-L. Matrix effect in LC-ESI−MS and LC-APCI-MS with off-line and on-line extraction procedures. J Chromatogr A. 2004;1058(1):61–6. 10.1016/j.chroma.2004.08.118. [PubMed: 15595652]
6. Ismaiel OA, Halquist MS, Elmamly MY, Shalaby A, Karnes HT. Monitoring phospholipids for assessment of ion enhancement and ion suppression in ESI and APCI LC/MS/MS for chlorpheniramine in human plasma and the importance of multiple source matrix effect evaluations. J Chromatogr B. 2008;875(2):333–43.
7. Cai S-S, Syage JA. Comparison of atmospheric pressure photoionization, atmospheric pressure chemical ionization, and electrospray ionization mass spectrometry for analysis of lipids. Anal Chem. 2006;78(4):1191–9. [PubMed: 16478111]

8. Herrera LC, Grossert JS, Ramaley L. Quantitative aspects of and ionization mechanisms in positive-ion atmospheric pressure chemical ionization mass spectrometry. J Am Soc Mass Spectrom. 2008;19(12):1926–41. [PubMed: 18845448]
9. Fernández M, Pico Y, Mañes J. Comparison of gas and liquid chromatography coupled to mass spectrometry for the residue analysis of pesticides in organges. Chromatographia. 2001;54(5):302–8. 10.1007/bf02492674.
10. Zhou SN, Reiner EJ, Marvin C, Helm P, Riddell N, Dorman F, et al. Development of liquid chromatography atmospheric pressure chemical ionization tandem mass spectrometry for analysis of halogenated flame retardants in wastewater. Anal Bioanal Chem. 2010;396(3):1311–20. 10.1007/s00216-009-3279-6. [PubMed: 19957077]
11. Gardinali PR, Zhao X. Trace determination of caffeine in surface water samples by liquid chromatography–atmospheric pressure chemical ionization–mass spectrometry (LC–APCI–MS). Environ Int. 2002;28(6):521–8. [PubMed: 12503918]
12. Donot F, Cazals G, Gunata Z, Egron D, Malinge J, Strub C, et al. Analysis of neutral lipids from microalgae by HPLC-ELSD and APCI-MS/MS. J Chromatogr B. 2013;942–943:98–106. 10.1016/j.jchromb.2013.10.016.
13. Rozenberg R, Ruibal-Mendieta NL, Petitjean G, Cani P, Delacroix DL, Delzenne NM, et al. Phytosterol analysis and characterization in spelt (Triticum aestivum ssp. spelta L.) and wheat (T. aestivum L.) lipids by LC/APCI-MS. J Cereal Sci. 2003;38(2):189–97. 10.1016/S0733-5210(03)00022-5.
14. Ulrich EM, Sobus JR, Grulke CM, Richard AM, Newton SR, Strynar MJ, et al. EPA's non-targeted analysis collaborative trial (ENTACT): genesis, design, and initial findings. Anal Bioanal Chem. 2019;411(4):853–66. [PubMed: 30519961]
15. Sobus JR, Grossman JN, Chao A, Singh R, Williams AJ, Grulke CM, et al. Using prepared mixtures of ToxCast chemicals to evaluate non-targeted analysis (NTA) method performance. Anal Bioanal Chem. 2019;411(4):835–51. [PubMed: 30612177]
16. 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 Cheminform. 2018;10(1):45. [PubMed: 30167882]
17. Yang C, Tarkhov A, Marusczyk J, Bienfait B, Gasteiger J, Kleinoeder T, et al. New publicly available chemical query language, CSRML, to support chemotype representations for application to data mining and modeling. J Chem Inf Model. 2015;55(3): 510–28. [PubMed: 25647539]
18. Molecular Networks GmbH, Altamira LLC ChemoTyper -Chemotype Your Molecular Datasets. 2019 https://www.mn-am.com/products/chemotyper Accessed 9-29-19.
19. Mansouri K, Grulke CM, Judson RS, Williams AJ. OPERA models for predicting physicochemical properties and environmental fate endpoints. J Cheminformatics. 2018;10(1):10.
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 Cheminformatics. 2017;9(1):61 10.1186/s13321-017-0247-6.
21. McKight PE, Najab J. Kruskal-Wallis test. In: The corsini encyclopedia of psychology; 2010 p. 1–1.
22. Dinno A. Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. Stata J. 2015;15(1):292–300.
23. Chusaksri S, Sutthivaiyakit S, Sutthivaiyakit P. Confirmatory determination of organochlorine pesticides in surface waters using LC/APCI/tandem mass spectrometry?. Anal Bioanal Chem. 2006;384(5):1236–45. 10.1007/s00216-005-0248-6. [PubMed: 16432722]
24. Dougherty RC. Negative chemical ionization mass spectrometry. Anal Chem. 1981;53(4):625–36. 10.1021/ac00227a003.
25. Haraguchi K, Kato Y, Atobe K, Okada S, Endo T, Matsubara F, et al. Negative APCI-LC/MS/MS method for determination of natural persistent halogenated products in marine biota. Anal Chem. 2008;80(24):9748–55. 10.1021/ac801824f. [PubMed: 19012416]
26. Jing R, Fusi S, Chan A, Capozzi S, Kjellerup BV. Distribution of polychlorinated biphenyls in effluent from a large municipal wastewater treatment plant: potential for bioremediation? J Environ Sci. 2019;78:42–52. 10.1016/j.jes.2018.06.007.

27. Lee S, Song G-J, Kannan K, Moon H-B. Occurrence of PBDEs and other alternative brominated flame retardants in sludge from wastewater treatment plants in Korea. Sci Total Environ. 2014;470: 1422–9. [PubMed: 23993837]

28. Sánchez-Avila J, Bonet J, Velasco G, Lacorte S. Determination and occurrence of phthalates, alkylphenols, bisphenol A, PBDEs, PCBs and PAHs in an industrial sewage grid discharging to a municipal wastewater treatment plant. Sci Total Environ. 2009;407(13):4157–67. [PubMed: 19362327]

29. Wang Y, Zhang Q, Lv J, Li A, Liu H, Li G, et al. Polybrominated diphenyl ethers and organochlorine pesticides in sewage sludge of wastewater treatment plants in China. Chemosphere. 2007;68(9): 1683–91. [PubMed: 17509654]

30. Thurman E, Ferrer I, Barcelo D. Choosing between atmospheric pressure chemical ionization and electrospray ionization interfaces for the HPLC/MS analysis of pesticides. Anal Chem. 2001;73(22): 5441–9. [PubMed: 11816571]

31. Marvin CH, Smith RW, Bryant DW, McCarry BE. Analysis of high-molecular-mass polycyclic aromatic hydrocarbons in environmental samples using liquid chromatography–atmospheric pressure chemical ionization mass spectrometry. J Chromatogr A. 1999;863(1):13–24. 10.1016/S0021-9673(99)00955-3. [PubMed: 10591460]

32. Titato GM, Lanças FM. Optimization and validation of HPLC-UV-DAD and HPLC-APCI-MS methodologies for the determination of selected PAHs in water samples. J Chromatogr Sci. 2006;44(1):35–40. 10.1093/chromsci/44.1.35. [PubMed: 16599410]

33. Mansouri K, Cariello NF, Korotcov A, Tkachenko V, Grulke CM, Sprankle CS, et al. Open-source QSAR models for pKa prediction using multiple machine learning approaches. J Cheminformatics. 2019;11(1):60 10.1186/s13321-019-0384-1.

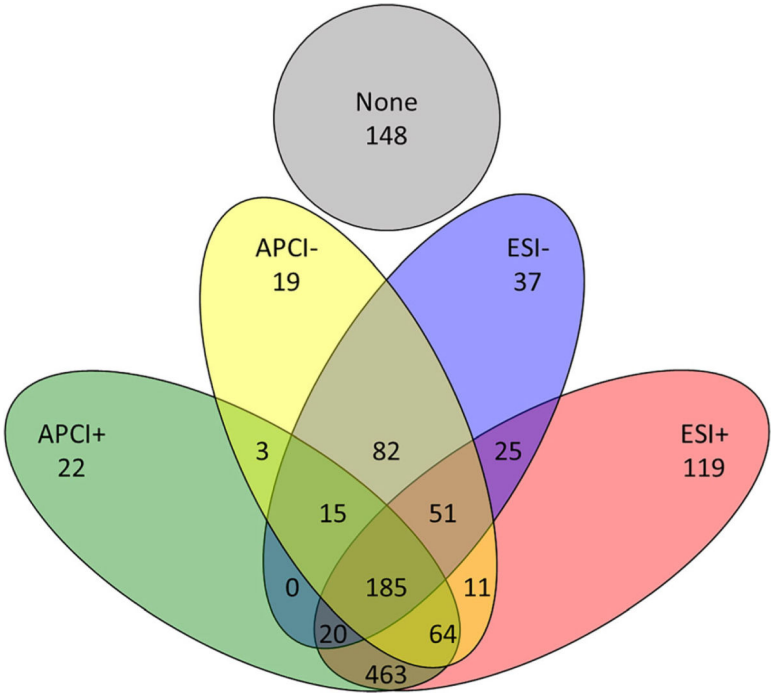


Fig. 1.

Venn diagram showing observed mixture compound coverage separated by ionization modes and polarities (n = 1264)

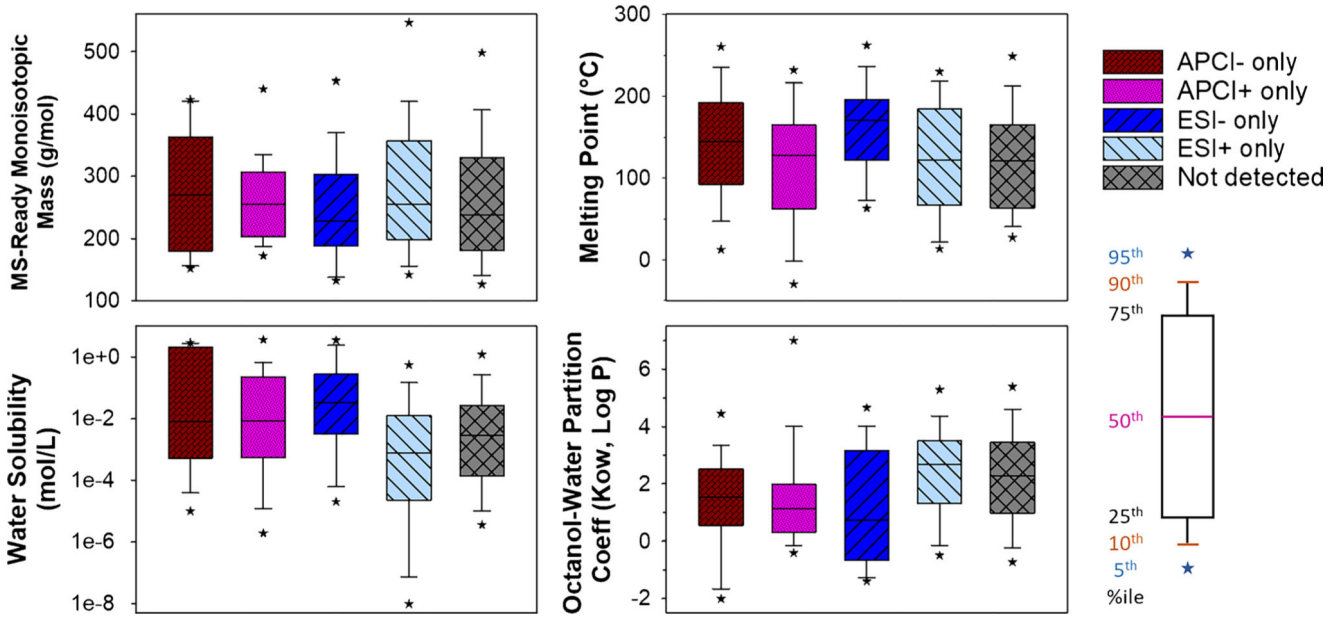
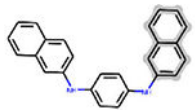
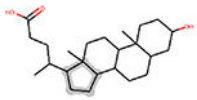
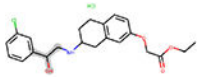
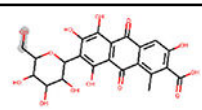
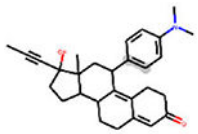


Fig. 2.

Box and whisker plots for physicochemical properties of ENTACT chemicals detected in one of four ionization modes or not detected

Table 1

Enriched ToxPrints for compounds identified with only APCI or with only ESI. The enriched ToxPrint substructure is highlighted in gray for an example compound

Ionization Method	ToxPrint Substructure	Odds ratio	True Positives	Example Compound	Structure with ToxPrint
APCI	ring:fused_[6_6]_naphthalene	7.607	5	DTXSID3020918	
ESI	chain:alkaneCyclic_pentyl_C5	∞^a	15	DTXSID6020779	
ESI	chain:alkaneLinear_hexyl_C6	∞	20	DTXSID0041197	
ESI	chain:alkaneLinear_octyl_C8	∞	17	DTXSID0042454	
ESI	bond:COH_alcohol_sec-alkyl	8.543	30	DTXSID1047360	
ESI	bond:COH_alcohol_pri-alkyl	6.891	25	DTXSID9022817	
ESI	chain:alkaneCyclic_ethyl_C2_(connect_noZ)	6.891	25	DTXSID5023322	

Ionization Method	ToxPrint Substructure	Odds ratio	True Positives	Example Compound	Structure with ToxPrint
ESI	bond:COH_alcohol_generic	3.256	69	DTXSID4047252	
ESI	bond:COH_alcohol_aliphatic_generic	3.116	43	DTXSID3044625	

*a)*There were no compounds found only via APCI containing these features which, computationally, led to infinite odds.