Supporting Information for:

Expanded coverage of NT-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, RTP, NC, United States

[2] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg

[3]US Environmental Protection Agency, Office of Research & Development, National Exposure Research Laboratory, RTP, NC, United States

[4]Department of Biological Sciences, North Carolina State University, Raleigh, NC, United States

[5]Statera Environmental, Raleigh, NC, United States

§Corresponding Author- 109 TW Alexander Dr., Maildrop D205-05, RTP, NC 27711; ulrich.elin@epa.gov, 919-541-3717

Contents:

In-house mixture

An additional in-house mixture containing 521 compounds including pesticides, pharmaceuticals, and polyaromatic hydrocarbons was also analyzed, though in a more qualitative manner for anecdotal discussion rather than in-depth analysis. A list of the components in this mixture is provided in Table S4. Neutral lipids like methyl decanoate (DTXSID4026842) and methyl dodecanoate (DTXSID5026889) were only detected in APCI+. This is in agreement with previous work showing the utility of APCI in the analysis of neutral lipids and sterol-like compounds.[1-3] Volatile chlorinated and brominated substances, historically analyzed using gas chromatography mass spectrometry (GC-MS), were also detected using APCI-. There are studies reporting their detection using APCI but mostly as employed in a targeted approach.[4,5] Some of the halogenated compounds we were able to detect in APCI- include (but are not limited to): 1,2,4-trichlorobenzene (DTXSID0021965), 2,6-dinitrotoluene (DTXSID5020528), 2,4,5-trichlorophenol (DTXSID4024359), 2,3,4,6-tetrachlorophenol (DTXSID9021716), pentachlorobenzene (DTXSID7024247), dichlorodiphenyldichloroethane (DTXSID4020373), chlorthal-dimethyl (DTXSID0024000), endosulfan I (DTXSID9037539), endosulfan II (DTXSID8037540), and endosulfan sulfate (DTXSID3037541). As expected, most of the above-mentioned compounds ionized by virtue of a proton loss, a result of the electron-withdrawing effect of phenyl-ring halogens. More interestingly, endosulfan I and endosulfan II ionized differently in APCI-, one by gaining a hydride ion while the other by radical anion formation (Fig. S2). The uncommon reaction of hydride gain may be explained by nucleophilic addition that has been reported to occur in negative ion chemical ionization.[6] When a strong nucleophile is present, gas phase molecules are susceptible to nucleophilic attack thus forming the negatively charged species. In the context of the compounds where this mechanism was observed, the

presence of multiple halogens, which are electronegative, leave portions of the electron cloud with a slight positive charge which is then susceptible to nucleophilic attack.

Forty-two polybrominated diphenyl ethers (PBDE) and all 209 polychlorinated biphenyl congeners were also included in the in-house mixtures. Despite positive detections, no chromatographic separation was achieved within these groups due to the very similar physicochemical properties of the different congeners. This challenge was addressed by Zheng et al. in work that demonstrated how APCI can be paired with ion mobility spectrometry and mass spectrometry to discriminate between isobaric flame retardant congeners.[7] PBDE data analysis is not straightforward, as these molecules may lose a bromine and acquire an oxygen atom during the ionization process, as was the case here, demonstrating an additional reaction beyond those included earlier.[8] APCI also allowed for detection of endrin (DTXSID6020561), dieldrin (DTXSID9020453), and heptachlor (DTXSID3020679). These molecules fragment in source, however, losing two to three chlorine atoms, thus making the base peak different from the molecular ion.[9] Taken together, the APCI results for halogenated and neutral lipid compounds underscore its utility for detecting contaminants of emerging concern and legacy pollutants that would typically be measured using gas chromatography, as long as special ionization cases are considered.[10-13]

Odds Ratio calculation

Structural features, or chemical fingerprints provide ways to encode the absence or presence of specific bond, ring, or atom types within a compound in such a way as to be usable by data analytics methodologies such as structure-activity relationship and structure similarity

searching. Each chemical fingerprint is essentially a binary array of an of length *n*, where *n* is the number of bonds, rings, atoms, etc whose presence or absence are enumerated in the fingerprint. If an element in the array is 0, that means the chemical does not have that specific bond, ring or atom type. If an element in the array is 1, that means the chemical does have that specific feature. Specific sets of atoms, bond, and/or rings that make up a fingerprint have been developed for many purposes and are typically designed to cover large libraries of pharmaceutical compounds. However, the ToxPrint fingerprints were designed with non-pharmaceutical chemical libraries, like Tox21, in mind.[14] Therefore, this set of fingerprints should cover sub-structures (i.e., elements in the fingerprint array) much more broadly than a set of fingerprints designed for pharmaceutical compounds.

ToxPrint fingerprints were used to compute odds ratios for each of the 729 elements in a ToxPrint fingerprint array for each compound that was identified via only one ionization type (i.e., each compound has an associated binary array of length 729). Here, an odds ratio is the ratio of odds that a compound will be detected using one of the ionization methods (either ESI or APCI) in the presence of one of the 729 elements in a ToxPrint fingerprint and the odds that that a compound will be detected with the same ionization method in the absence of that substructure. Odds ratio values are calculated for each substructure (i.e., each element in the fingerprint) and each ionization method using the following formula:

$$\text{Odds Ratio} = \frac{a \times d}{b \times c} \qquad \text{(Eqn SI 1)}$$

where *a* is the number of compounds that have a substructure and were detected with an ionization method, *b* is the number of compounds that did not have a substructure but were still detected, *c* is the number of compounds that have a substructure but were not detected, and *d* is the number

of compounds that did not have a substructure and were not detected. Odds ratio values have a lower bound of 0, but no upper bound. An odds ratio > 1 implies that compared to the absence of a substructure, the presence of the substructure raises the odds of detecting a compound with one of the ionization methods. Odds ratios equal to 1 imply there is no relation between detecting a compound based on the presence or absence of a substructure. Odds ratio values < 1 indicate that the absence of a substructure raises the odds of detecting a chemical with the ionization method. While odds ratio values > 1 do imply relationships, more stringent conditions have been placed on odds ratio values examined here to ensure a stronger relationship, or enrichment, between the occurrence of a substructure and the ionization method used.[15] Namely, for a substructure to be considered as enriched it must: 1) have an odds ratio value greater than 3 (i.e., odds must be more than three times higher than a compound with a substructure that was identified using a specific ionization method compared to the odds of identifying that compound in the absence of the substructure), 2) have a Fisher exact score of less than 0.05 (this ensures that the odds ratio is statistically significant for the relatively low sample sizes of compounds containing a substructure and being identified with an ionization method), and 3) more than 3 compounds must have a substructure and have been identified by the ionization method.[16]

Figure S1. *Summary of ENTACT components found using a) atmospheric pressure chemical ionization and b) electrospray ionization in positive and negative mode. Bars show found component counts using the left y-axis, and red diamonds show the percent components found using the right y-axis.*

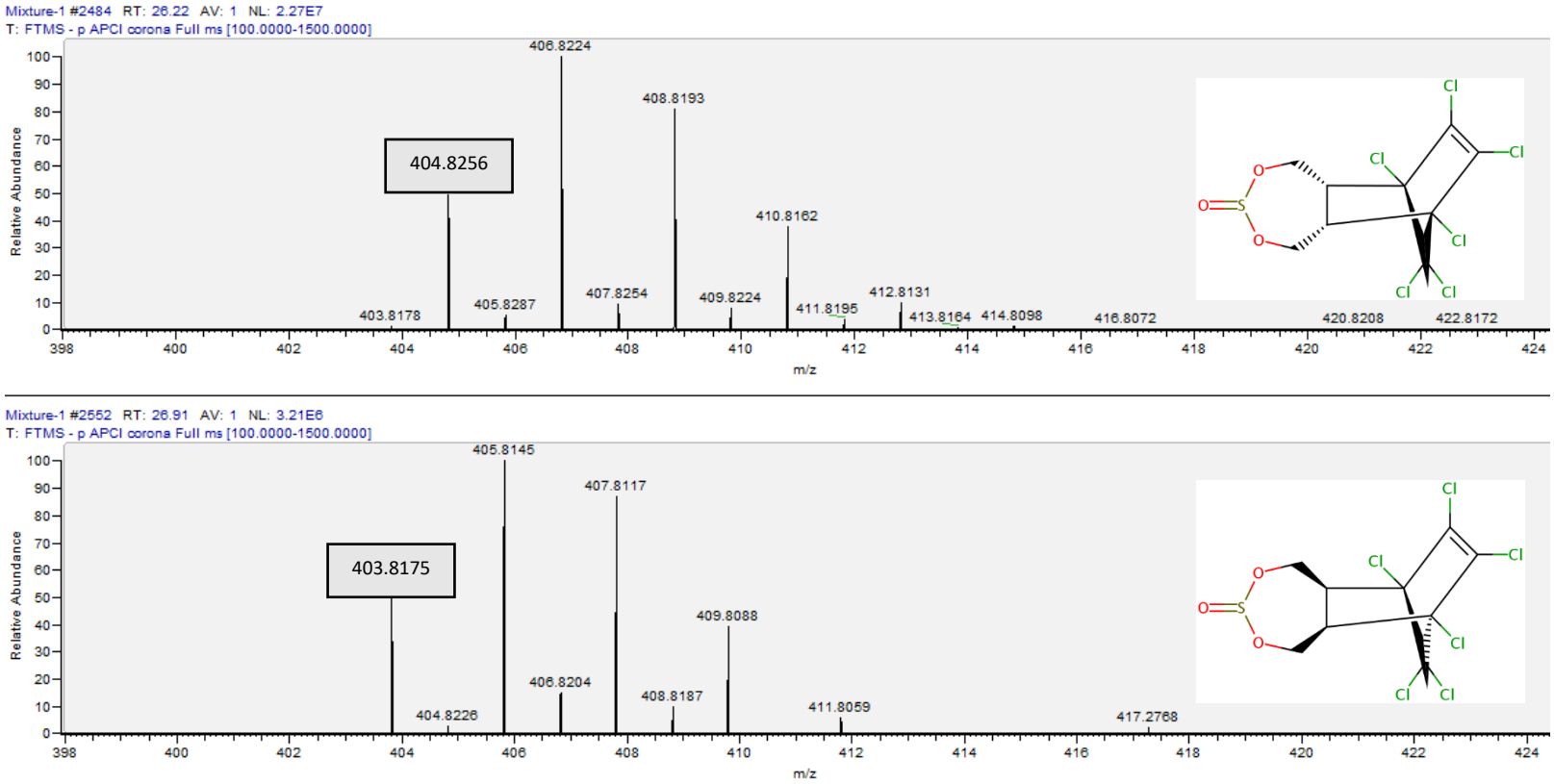

Figure S2*. Mass spectra for A) endosulfan I and B) endosulfan II in APCI-. These isomers have different m/z despite having only stereochemistry differences in the structures.*

References

1. Donot F, Cazals G, Gunata Z, Egron D, Malinge J, Strub C, Fontana A, Schorr-Galindo S (2013) Analysis of neutral lipids from microalgae by HPLC-ELSD and APCI-MS/MS. Journal of Chromatography B 942-943:98-106. doi:https://doi.org/10.1016/j.jchromb.2013.10.016.

2. Rozenberg R, Ruibal-Mendieta NL, Petitjean G, Cani P, Delacroix DL, Delzenne NM, Meurens M, Quetin-Leclercq J, Habib-Jiwan J-L (2003) Phytosterol analysis and characterization in spelt (Triticum aestivum ssp. spelta L.) and wheat (T. aestivum L.) lipids by LC/APCI-MS. Journal of Cereal Science 38 (2):189-197. doi:https://doi.org/10.1016/S0733-5210(03)00022-5.

3. Murphy RC, Leiker TJ, Barkley RM (2011) Glycerolipid and cholesterol ester analyses in biological samples by mass spectrometry. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids 1811 (11):776-783. doi:https://doi.org/10.1016/j.bbalip.2011.06.019.

4. Kato Y, Okada S, Atobe K, Endo T, Matsubara F, Oguma T, Haraguchi K (2009) Simultaneous Determination by APCI-LC/MS/MS of Hydroxylated and Methoxylated Polybrominated Diphenyl Ethers Found in Marine Biota. Analytical Chemistry 81 (14):5942-5948. doi:10.1021/ac9009885.

5. Lupton SJ, McGarrigle BP, Olson JR, Wood TD, Aga DS (2010) Analysis of hydroxylated polybrominated diphenyl ether metabolites by liquid chromatography/atmospheric pressure chemical ionization tandem mass spectrometry. Rapid Communications in Mass Spectrometry 24 (15):2227-2235.

6. Dougherty RC (1981) Negative chemical ionization mass spectrometry. Analytical Chemistry 53 (4):625-636. doi:10.1021/ac00227a003.

7. Zheng X, Dupuis KT, Aly NA, Zhou Y, Smith FB, Tang K, Smith RD, Baker ES (2018) Utilizing ion mobility spectrometry and mass spectrometry for the analysis of polycyclic aromatic hydrocarbons, polychlorinated biphenyls, polybrominated diphenyl ethers and their metabolites. Analytica chimica acta 1037:265-273.

8. Haraguchi K, Kato Y, Atobe K, Okada S, Endo T, Matsubara F, Oguma T (2008) Negative APCI-LC/MS/MS Method for Determination of Natural Persistent Halogenated Products in Marine Biota. Analytical Chemistry 80 (24):9748-9755. doi:10.1021/ac801824f.

9. Chusaksri S, Sutthivaiyakit S, Sutthivaiyakit P (2006) Confirmatory determination of organochlorine pesticides in surface waters using LC/APCI/tandem mass spectrometry⋄. Analytical and Bioanalytical Chemistry 384 (5):1236-1245. doi:10.1007/s00216-005-0248-6.

10. Jing R, Fusi S, Chan A, Capozzi S, Kjellerup BV (2019) Distribution of polychlorinated biphenyls in effluent from a large municipal wastewater treatment plant: Potential for bioremediation? Journal of Environmental Sciences 78:42-52. doi:https://doi.org/10.1016/j.jes.2018.06.007.

11. Lee S, Song G-J, Kannan K, Moon H-B (2014) Occurrence of PBDEs and other alternative brominated flame retardants in sludge from wastewater treatment plants in Korea. Science of the Total Environment 470:1422-1429.

12. Sánchez-Avila J, Bonet J, Velasco G, Lacorte S (2009) Determination and occurrence of phthalates, alkylphenols, bisphenol A, PBDEs, PCBs and PAHs in an industrial sewage grid

discharging to a Municipal Wastewater Treatment Plant. Science of the Total Environment 407 (13):4157-4167.

13. Wang Y, Zhang Q, Lv J, Li A, Liu H, Li G, Jiang G (2007) Polybrominated diphenyl ethers and organochlorine pesticides in sewage sludge of wastewater treatment plants in China. Chemosphere 68 (9):1683-1691.

14. Yang C, Tarkhov A, Marusczyk Jr, Bienfait B, Gasteiger J, Kleinoeder T, Magdziarz T, Sacher O, Schwab CH, Schwoebel J (2015) New publicly available chemical query language, CSRML, to support chemotype representations for application to data mining and modeling. Journal of chemical information and modeling 55 (3):510-528.

15. Strickland JD, Martin MT, Richard AM, Houck KA, Shafer TJ (2018) Screening the ToxCast phase II libraries for alterations in network function using cortical neurons grown on multi-well microelectrode array (mwMEA) plates. Archives of Toxicology 92 (1):487-500. doi:10.1007/s00204-017-2035-5.

16. Mehta CR, Patel NR (1983) A Network Algorithm for Performing Fisher's Exact Test in r × c Contingency Tables. Journal of the American Statistical Association 78 (382):427-434. doi:10.2307/2288652.