Supporting Information for:

Combining non-targeted analysis with computer-based hazard comparison approaches to support prioritization of unregulated organic contaminants in biosolids

Matthew N. Newmeyer[1], Qinfan Lyu[1], Jon R. Sobus[2], Antony J. Williams[2], Keeve E. Nachman[1,3,4], Carsten Prasse[1,3*]

[1]Department of Environmental Health and Engineering, Johns Hopkins University, Baltimore, MD, 21205, United States

[2]Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, 27709, United States

[3]Risk Sciences and Public Policy Institute, Johns Hopkins University, Baltimore, MD, 21205, United States

[4]Center for a Livable Future, Johns Hopkins University, Baltimore, MD, 21205, United States

*Email: cprasse1@jhu.edu

Number of pages: 23

Number of figures: 8

Number of tables: 1

Table of Contents

Table of Figures

Compounds included in the quality control mixture

A quality control (QC) mixture was prepared to a) evaluate the performance of the QuEChERS extraction, and b) test data processing parameters in Compound Discoverer before processing facility samples. A full list of the chemicals included in the QC mix is provided in Supplemental Table 2.

Our goal was to develop a mixture of chemicals with diverse physicochemical properties that could be detected in both positive and negative ionization modes, and that was representative of chemicals that could be expected to be present in biosolids. Therefore, we compared the distributions of physicochemical properties of the compounds in the QC mix to those from compounds present on the EPA's Chemicals in Biosolids (2022) List (https://comptox.epa.gov/dashboard/chemical-lists/BIOSOLIDS2022) from the CompTox Chemicals Dashboard.[1] Compounds on the Biosolids List without structures and compounds with a molecular weight <100 were filtered out (n = 671 compounds remained). We downloaded physicochemical property values from the Dashboard using the "Batch Search" function[2] and compared distributions for 8 properties: the negative logarithm of the distribution coefficient at pH 5.5 ($logD_{5.5}$) and pH 7.4 ($logD_{7.4}$), the negative logarithm of the octanol-water partition coefficient ($logK_{ow}$), molar volume, the negative logarithm of the apparent acidic acid dissociation constant (*i.e.*, pK_a acidic apparent, pK_{aa}), the negative logarithm of the apparent basic acid dissociation constant (*i.e.,* pK_a basic apparent, pK_{ab}), polarizability, and vapor pressure. Predicted values from ACD/Labs[3] were used for molar volume and polarizability, and OPERA 2.6 predicted values[4] were used for all other properties. Vapor pressure values were $-log_{10}$ transformed for plotting purposes.

Figure S1 compares the distributions of the eight physicochemical properties for all compounds included in the QC mix and compounds in the Biosolids List. The distributions were generally similar, although the compounds in the Biosolids List tended to be more non-polar (higher median $logK_{ow}$). Many of these more non-polar compounds were polyhalogenated biphenyls and diphenyl ethers which are typically detected by gas chromatographic methods.[5,6]

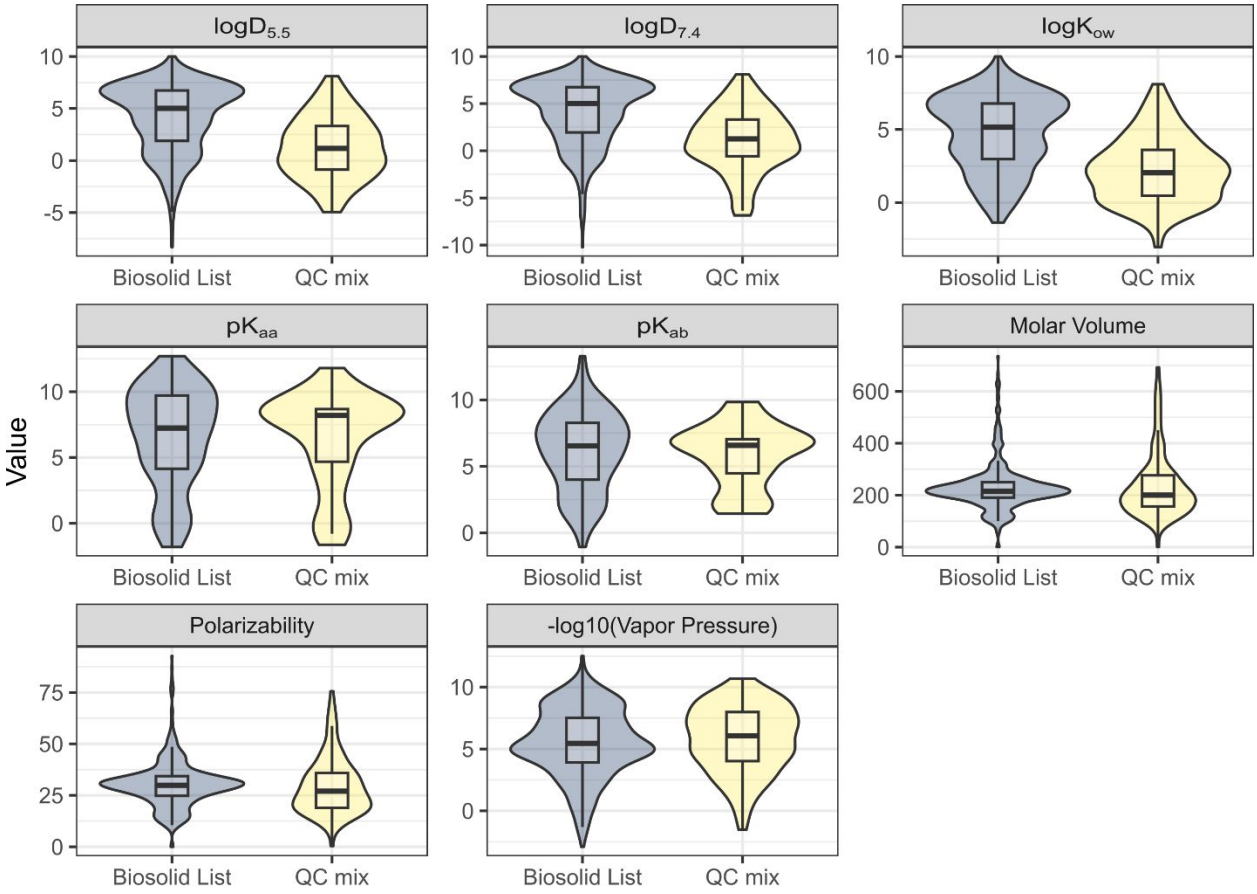


Figure S1. Comparison of the distributions of selected physicochemical properties for compounds present on the Biosolids 2022 List (n = 671) and compounds included in the QC mix (n = 120). The Biosolids List was filtered to exclude compounds without structures and with molecular weights <100. Note that different numbers of compounds were included in each plot depending on data availability for that property.

Instrumental analysis

The UltiMate 3000 RSLCnano system consisted of an SRD-3400 solvent degasser, NCS3500RS pump module with column oven, and WPS-3000TPL RS autosampler. The system was controlled by Xcalibur v4.1 software. The same chromatography gradient was used for both positive and negative mode analyses: the gradient started at 10% B from 0-1 min, increased to 98% B from 1-16 min, was held at 98% B from 16-22 min, decreased to 10% B from 22-22.1 min, and was held at 10% B from 22.1-30 min; the flow rate was 0.1 mL/min, injection volume was 5 µL, the column temperature was 45 °C, and the autosampler temperature was 10 °C. LC flow was diverted to the mass spectrometer from 2-28 min.

Batches were designed based on previous recommendations.[7] First, two solvent blanks were injected. Then, the pooled QC was injected ten times to “condition” the column and system to improve retention time reproducibility; the extraction blank was injected once after the 4th conditioning injection to measure instrument background. Next, the pooled QC was injected two additional times, followed by two injections of the fortified control, and then the facility samples. Facility samples were each injected once and in a random order. The pooled QC and fortified control were each injected once after 1/3 and 2/3 of the facility samples were analyzed, and each were injected twice after all facility samples were analyzed (total pooled QC and fortified control injections = 6 each). Then, the extraction blank was injected a second time, followed by solvent blanks. Samples were first analyzed in positive mode, then the organic mobile phase was switched, the system equilibrated, and samples analyzed in negative mode. The

second pooled QC and fortified control extracts were used for injections in the negative mode, and the facility sample injection order was re-randomized.

Chemical space evaluation

We conducted a series of experiments testing the effect of different instrumental conditions on the detectable chemical space[8] of our analysis. We made the decision *a priori* to evaluate the positive and negative modes separately with the goal of finding one set of conditions among those tested that produced the most diverse set of detected chemical features for each polarity mode.

Two conditions were evaluated:

1. the organic mobile phase composition (n = 4)
    a. methanol, with and without 0.1% formic acid
    b. acetonitrile, with and without 0.1% formic acid
2. the ionization mode (n = 2)
    a. heated-electrospray ionization (H-ESI)
    b. atmospheric pressure chemical ionization (APCI)

Therefore, there were eight total combinations of conditions evaluated for each polarity mode.

To perform the evaluation, five biosolid samples and one MQW sample (as an extraction blank) were extracted using the protocol detailed in the main text, with slight modifications detailed in Figure S1. After adding the salt pouches and centrifuging, only

4 mL of the acetonitrile supernatant was transferred to a 15-mL centrifuge tube containing dSPE sorbent, but this was done twice for each sample, creating an “A” set and a “B” set. Then, after mixing and centrifuging, 1 mL of the supernatant was removed twice from each tube and transferred to separate vials for the concentration step, resulting in “A1”, “A2”, “B1”, and “B2” sample sets. By doing this we generated four extracts from each original sample that could be analyzed under the different experimental conditions. This way the variability in sample extraction could be reduced compared to extracting four aliquots of the same sample. Next, 200 µL of MQW was added to all samples and then samples were dried to 200 µL, after which 250 µL MQW and 50 µL methanol were added. Finally, tubes were centrifuged, and supernatants were filtered. All extracts were stored at -20 °C until analysis.

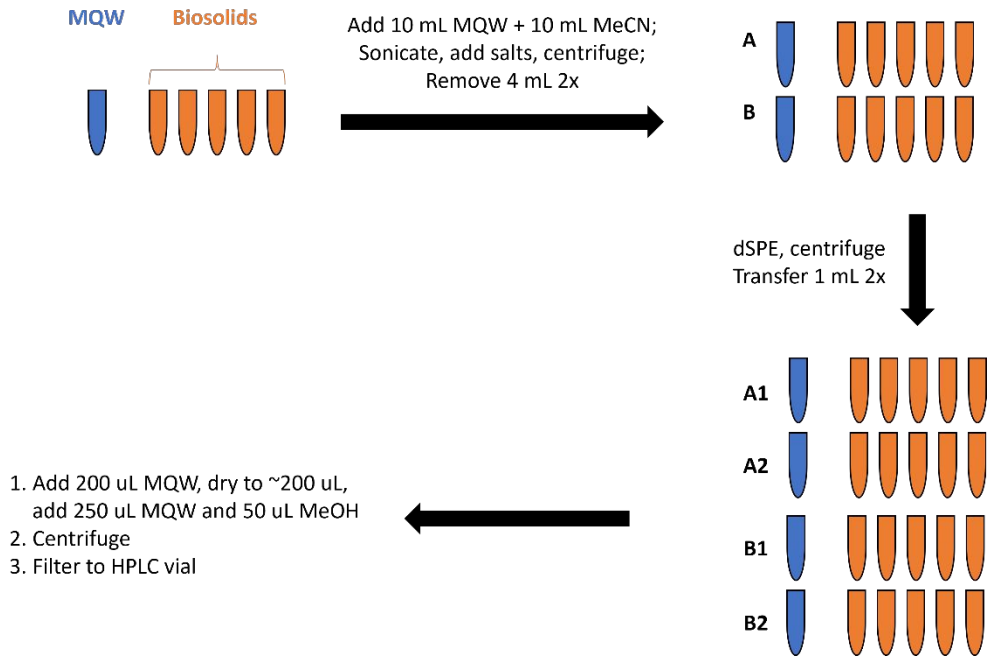


Figure S2. Modified QuEChERS procedure for chemical space evaluation

All experimental conditions were tested across four analytical batches, with one “set”, e.g., set “A1”, analyzed in each batch. To accommodate this, the instrumental method

incorporated polarity switching. We took advantage of the ternary pumping system on our LC instrument to conduct the experiments. Line A pumped the aqueous mobile phase (1 mM ammonium fluoride in MQW; constant across all conditions), Line B pumped the methanolic mobile phase, and Line C pumped the acetonitrile-based mobile phase. A single batch consisted of analyzing all samples with the aqueous/methanol condition followed by flushing the system with the aqueous/acetonitrile condition and then re-analyzing all the samples. Samples were injected in duplicate and in a random order; the order was re-randomized for each experimental condition. This sequence template was repeated for each extraction set, changing the mobile phase or ionization method. Before each new batch the instrument was cleaned and calibrated. All experimental conditions were tested within ten days of sample extraction.

The same LC conditions described in the main text were used, other than the organic mobile phase composition being tested. Raw data were acquired using the Full MS/dd-MS^2 method described in Supplemental Table 3 with slight modifications. Because the method used polarity switching, the mass resolving power was 60,000 and the top N was 5. H-ESI source settings were the same as described in Supplemental Table 3; for APCI, the sheath gas was 20, aux gas was 0, spray current was 3 µA for both polarity modes, capillary temperature was 300 °C, S-lens RF level was 60, and vaporizer temperature was 300 °C.

Each condition and polarity mode (e.g., methanol-ESI-positive, methanol w/ 0.1% formic acid-ESI-positive, etc.) were processed separately with Compound Discoverer, resulting in sixteen datasets. Our metric for evaluation was the number of unique molecular formulas observed in each data set. We processed the datasets similarly to what is

described in the main text by assigning annotations and corresponding confidence levels. Since we focused on molecular formulas, confidence Level 5 features were excluded from the analysis. After processing, each dataset was exported for analysis using in-house R scripts.

The unique molecular formulas for each condition were compiled and compared. First, the methanolic and acetonitrile-based conditions were separated and compared within-group, and then the conditions with the highest percentage of unique molecular formulas were compared. Figure S3 shows an example for the positive mode. Of all unique molecular formula observed in the methanolic conditions, 72.9% were detected when using methanol + 0.1% formic acid and ESI, and 57.7% of all molecular formulas detected among the acetonitrile-based mobile phase conditions were detected using acetonitrile + 0.1% formic acid and APCI. When those two conditions were directly compared, 85.3% of the molecular formulas were detected with methanol + 0.1% formic acid and ESI. Therefore, this was the condition used for the analysis of positive mode compounds. In negative mode, the methanolic and acetonitrile-based conditions in which the largest percentage of unique molecular formulas were observed were methanol and ESI (70.6%) and acetonitrile and ESI (67.0%), respectively; when directly compared, 84.6% of the unique molecular formulas were detected with acetonitrile and ESI. This condition was chosen for the analysis of negative mode compounds.

Data processing

Pooled QC replicates were included in the processing workflow to correct for sequence effects[7] using the "Apply QC correction" node. This correction was performed for each feature that was detected in the pooled QC. When detected, the variability in the feature's peak area across the pooled QC injections was minimized by fitting a regression model, and that model was used to correct the peak areas for that feature in the "unknown" facility samples. Importantly, this was performed *feature-by-feature*, i.e., no global correction was performed. This correction is performed by the software. The criteria for applying the pooled QC-based correction are in Supplemental Table 4. The processing workflow included a "Fill Gaps" node which attempts to re-detect a missing feature in samples at a reduced threshold if the feature was detected in *any* sample and imputes missing peak areas when the feature isn't re-detected. The workflow also calculated a "Peak Rating" for each feature detected in a sample. The Peak Rating is a weighted score from 0-10 that rates the quality of the chromatographic peak, with 10 the highest quality. Importantly, if a missing feature was not re-detected in a sample with the "Fill Gaps" node (and the peak area was imputed) a Peak Rating is not calculated.

Detected features were filtered to generate a list for further evaluation. First, the dataset was filtered to only include features that were detected in all pooled QC replicates (n = 6) with a software-generated Peak Rating ≥6. Then, the resulting features table was exported and processed in R (v 4.2) to further filter features that were detected in ≥80% of biosolid samples. A feature was considered detected in a sample if it fulfilled two criteria. First, the feature in a facility sample had to have a peak area at least 5x greater than the peak area in each of the extraction blanks (if the feature was detected in the

blank). Second, the feature had to have a software-calculated Peak Rating ≥6. For the samples that were extracted in duplicate, a feature had to fulfill both criteria in both extraction replicates to be considered detected. For the purposes of determining whether a feature was detected in at least 80% of samples, each of the 4 pairs of extraction replicates were considered one sample (so 80% was determined based on an *n* of 16, not 20).

We assigned annotations and corresponding confidence levels to each "detected" feature based on previous recommendations.[9] Level 1 confidence was assigned to annotations where the identity of the compound was confirmed by comparison to a reference standard and the retention time and fragmentation spectra matched (see below). Level 2a confidence was assigned to annotations where the feature's experimental spectrum had a single match to Thermo's mzCloud spectral library with a score ≥70; this is the highest level of confidence possible without analyzing and matching to a reference material. Level 2b confidence was assigned to annotations where the feature's fragmentation spectrum had a match in the MassBank of North America spectral library (MoNA, https://mona.fiehnlab.ucdavis.edu) with a score of ≥70. We decided to differentiate between matches in the different spectral libraries due to our fragmentation settings. We collected fragmentation spectra using stepped normalized collision energies (stepped NCE) at 30, 65, and 100. Searching in the mzCloud database will generate a composite spectrum (using individual spectra collected using similar collision energies) that more closely resembles our experimental spectrum. However, searching using the MoNA database finds matches to single spectra where the collision energy and setting (*i.e.*, absolute vs. normalized) is not always obvious or

available. Therefore, we considered annotations based on matches to the MoNA database to be of slightly lower confidence. Level 3 confidence was assigned to annotations with spectral library matches but a single structure could not be assigned, *e.g.*, when the library matches are different isomers of a compound. When no structural annotation could be assigned for a feature but when a single molecular formula could be determined, the formula served as the feature's annotation and was assigned confidence Level 4. To evaluate if a single formula was determined, all the software-predicted formula were sent to a search using the "Search ChemSpider" node and were marked if there was a match; if only one molecular formula had a match to a compound in the database, then that formula was used as the compound annotation. Finally, when no structural annotation could be assigned, and either no molecular formula or multiple molecular formulas had matches using the "Search ChemSpider" node, the feature's observed m/z served as its annotation and assigned confidence Level 5. Fluorine was omitted from the "Predicted Compositions" node because, in our experience, many fluorine-containing formula candidates are generated per feature which may lead to other plausible candidates being "pushed out" of the top 10 candidates. This means any detected per- or poly-fluorinated alkyl substances (PFAS) would be annotated at confidence Level 5 in this analysis. The remaining features were manually reviewed to remove any remaining low-quality features or suspected in-source fragments.

Identities of detected features were confirmed by comparison to commercially available reference standards. First, features in the facility samples were compared to analytes added in the fortified control, which allowed us to confirm the identities of compounds during the initial analysis. After processing data from the initial analysis, we conducted

another analysis to confirm the identities of additional compounds. In this follow-up confirmation analysis, a mixture of analytes was prepared and diluted with MQW and analyzed with a pooled QC extract from the initial analysis; the extract was stored at -20 °C immediately following the end of the initial analysis. The same instrumental methods described above were used for the follow-up confirmation analysis. Identities were considered confirmed if the retention times of the reference material and the unknown feature matched within ±0.2 min and they had matching fragmentation spectra.

Literature search for compounds present in biosolids

Searching was performed with Web of Science using the Web of Science Core Collection Database. To keep the results comparable to the EPA Biennial Reviews, we narrowed our search to only include manuscripts reporting results in biosolids produced in either the United States or Canada. Studies that only evaluated activated sludges or fortified samples with target compounds were excluded (consistent with Biennial Review criteria). The “core” search terms utilized in all searches were **ALL=(biosolid OR sludge) AND CU=(USA OR Canada) AND DT=(Article)**. The synonyms for each searched chemical were downloaded from the Dashboard using the “Batch Search” feature and were included as search terms. For some compounds, the name from the spectral libraries was not on the synonym list, in which case the library name was manually added. If any manuscripts were found, they were reviewed for adherence to the criteria above.

We also carefully reviewed studies using non-targeted and suspect screening approaches to characterize chemical contamination in biosolids. We found that performing the "per compound" search described above initially missed some papers. For example, a search of **ALL=(sludge) AND ALL=(curcumin) AND CU=(USA OR Canada) AND DT=(Article)** did not return the paper from Black et al. in which curcumin was detected in California biosolids and listed in Table 1[10] because tables are not a "searchable field".

Calculating scores from Hazard Comparison Module data

Hazard data were compiled by searching confidence Level 1, 2a, and 2b compounds on the EPA's Cheminformatics Hazard Comparison Module (HCM, https://www.epa.gov/comptox-tools/cheminformatics). Additional details on the data structure underlying the HCM can be found in the manuscript by Vegosen and Martin.[11] We utilized the "trump" score shown on the hazard profile for calculating scores.

The table below shows three hypothetical chemicals with data for five endpoints to illustrate how the average hazard score, average quality score, completeness score, and quality-adjusted hazard score were calculated.

Table S1. Example data for three hypothetical compounds and five endpoints (E) to demonstrate calculation of different scores.

Chemical	E1	E2	E3	E4	E5	Avg. Hazard score	Avg. Quality score	Completeness Score	Quality-adjusted score
A	**VH**	L	I	*M*	*H*	2.50	1.75	0.80	4.38
B	L	**L**	*H*	L	N/A	1.50	2.00	0.80	3.00
C	**H**	L	N/A	N/A	N/A	2.00	2.50	0.40	5.00

Individual hazard scores: VH: Very High; H: High; M: Medium; L: Low; I: Inconclusive; N/A: not available
Data source authority is given by font: **Authoritative**, Screening, *QSAR Model*

Individual hazard scores and data source authorities were converted to numerical values. For hazard scores: VH = 4, H = 3, M = 2, and L = 1; for data source authority: Authoritative = 3; Screening = 2; and QSAR Model = 1. Inconclusive and N/A scores were omitted for calculating average hazard and quality scores.

Below are the equations used to calculate the scores with Chemical A as an example:

$$Avg.\,Hazard\ score = \frac{\sum individual\ hazard\ scores}{number\ of\ endpoints\ with\ data\ available} = \frac{4+1+2+3}{4} = 2.5$$

$$Avg.\,Quality\ score = \frac{\sum individual\ quality\ scores}{number\ of\ endpoints\ with\ data\ available} = \frac{3+2+1+1}{4} = 1.75$$

$$Completeness\ score = \frac{number\ of\ endpoint\ with\ data\ available}{number\ of\ endpoints\ searched} = \frac{4}{5} = 0.8$$

$$Quality-adjusted\ hazard\ score = Avg.\,Hazard\ score * Avg.\,Quality\ Score = 2.50 * 1.75 = 4.38$$

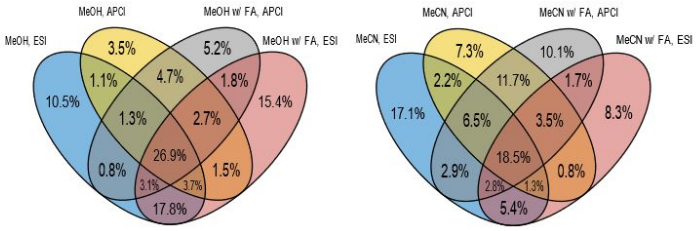
Compounds were assigned to one of 4 groups based on the completeness score:

1. High: completeness score ≥0.75
2. Medium high: 0.50 ≥ completeness score < 0.75
3. Medium low: 0.25 ≥ completeness score < 0.50

4. Low: completeness score <0.25

A

	MeOH	MeOH + 0.1% FA	MeCN	MeCN + 0.1% FA
ESI	65.2%	72.9%	56.7%	42.3%
APCI	45.4%	46.5%	51.8%	57.7%



B

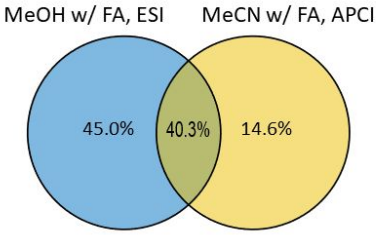


Figure S3. Comparison of unique molecular formulas observed under different instrumental conditions in the positive ionization mode. (A) The methanolic and acetonitrile-based conditions were separated and compared within-group, and the condition with the highest percentage of observed unique molecular formulas were selected (red ovals). Venn diagrams show the distribution of formulas across the conditions and the table shows the cumulative percentage for each condition; therefore, percentages in the table will sum to >100%. (B) The selected methanolic and acetonitrile-based conditions were directly compared. The condition with the higher percentage of observed formulas was selected for the analysis of all facility samples.

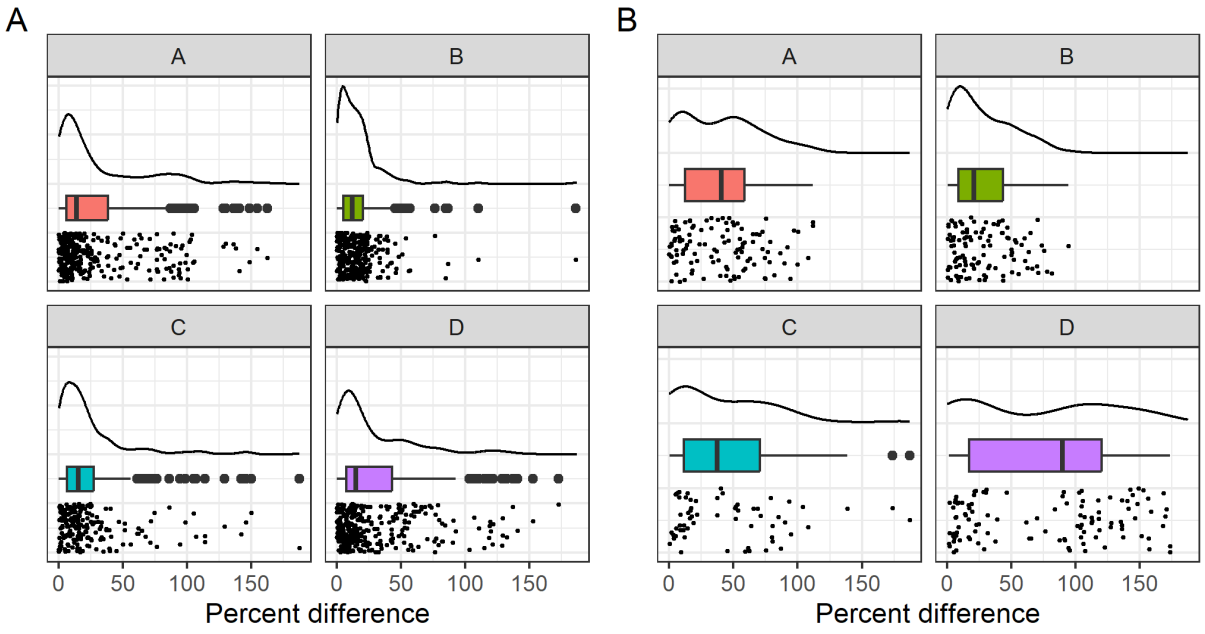


Figure S4. Distribution of peak area percent differences for features detected in extraction duplicates in positive mode (panel A) and negative mode (panel B). Individual graphs labeled A-D are for each pair of duplicate samples. Each graph shows a density plot (top plot), a boxplot (middle plot), and the individual data points (bottom plot).

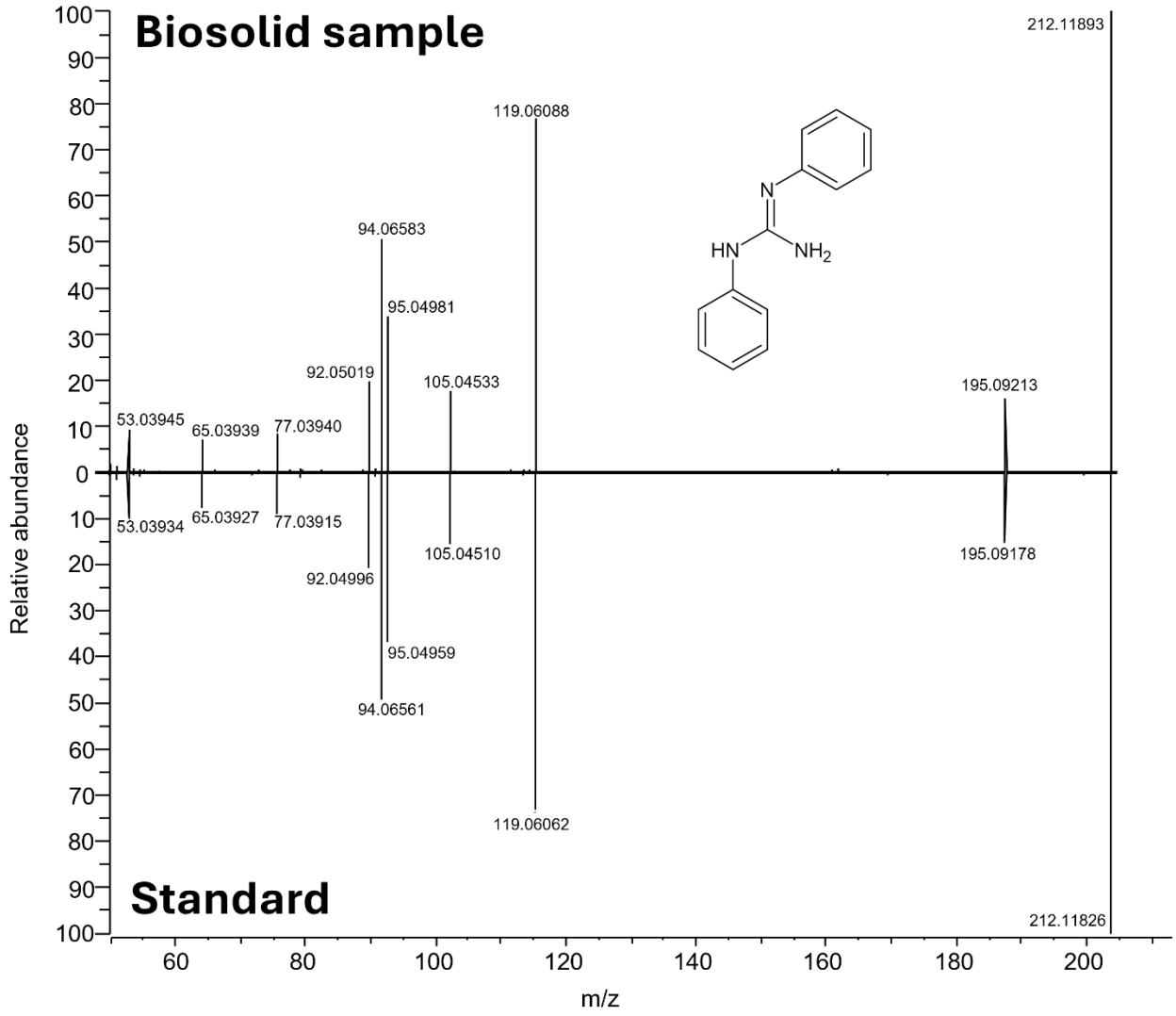


Figure S5. Comparison of fragmentation (i.e., MS^2) spectra for 1,3-Diphenylguanidine between a biosolid sample (top panel) and a reference standard (bottom panel) used for confidence Level 1 structural identification.

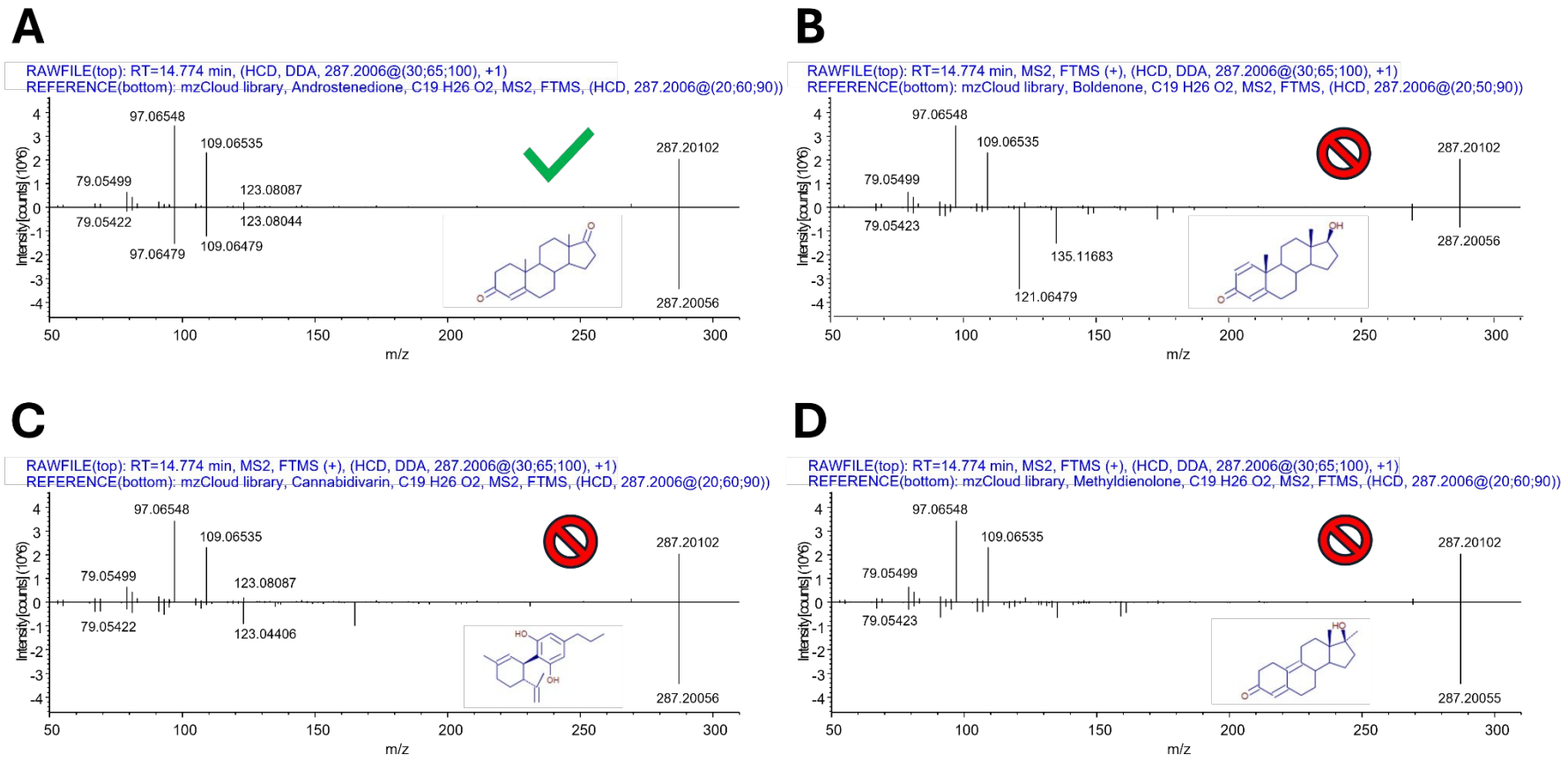


Figure S6. Comparisons of fragmentation (i.e., MS^2) spectra for a feature detected in biosolid samples (top spectrum in all panels) and 4 hits from mzCloud (bottom spectrum in all panels). Manual review of these results reveals the top matching spectrum (panel A) is the only reasonable match, resulting in a confidence Level 2a annotation of Androstenedione for this feature. Notable discrepancies between the spectrum of Boldenone (B), Cannabidivarin (C), and Methyldienoione (D) with the unknown feature's spectrum preclude their consideration as possible structural annotations.

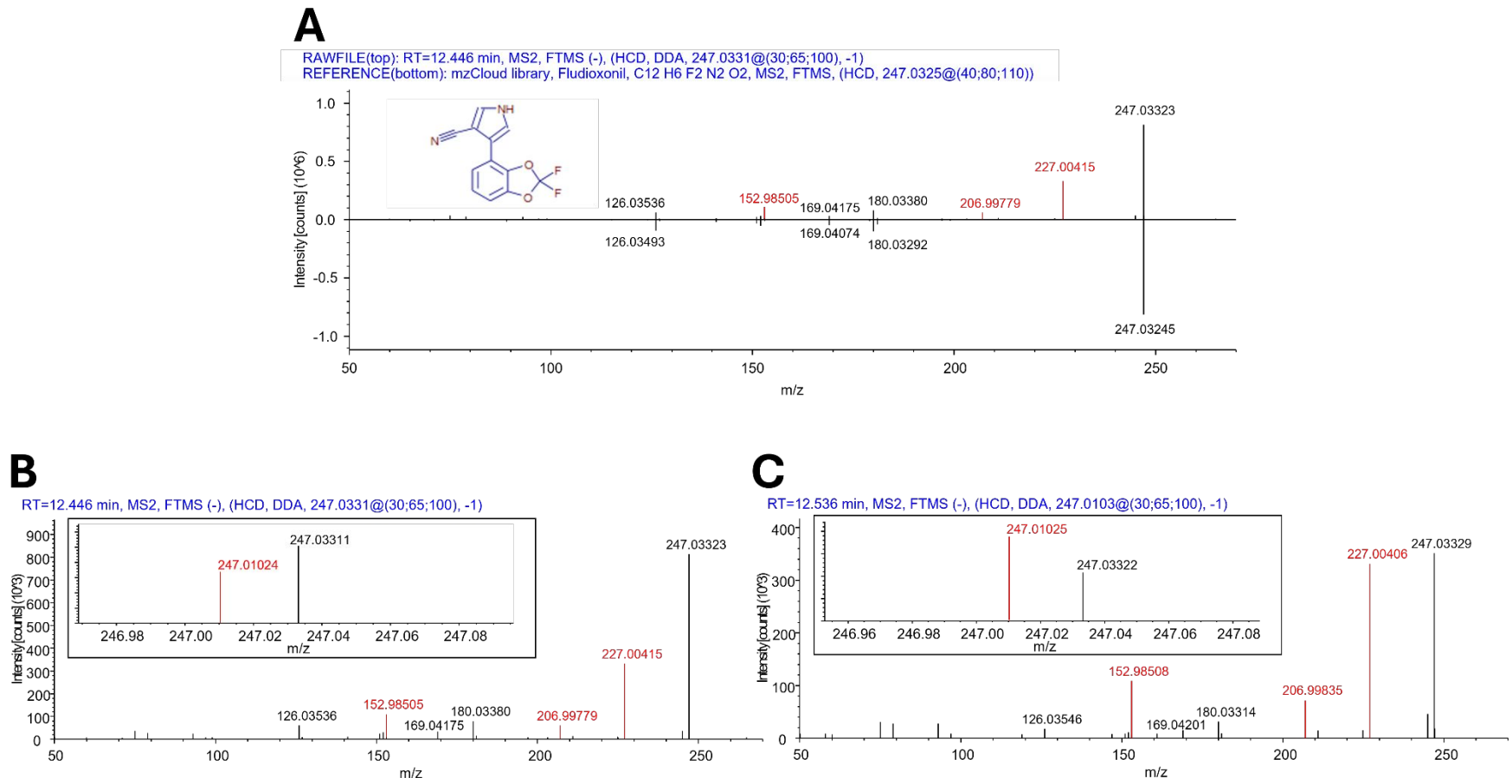


Figure S7. Comparison of fragmentation (i.e., MS^2) spectra between a feature detected in biosolid samples (top spectrum) and an mzCloud hit for Fludioxonil, resulting in a confidence Level 2a annotation (panel A). Fragments in red were identified as belonging to a co-eluting "interfering" ion at m/z 247.0102 (ion highlighted in red in panels B and C inserts showing MS^1 spectra). When the relative intensity of the interfering ion increases (compare insert in panel C to panel B) the relative intensities of three fragments (m/z 227.0041, 206.9978, and 152.9850) also increase. Therefore, these fragments were ignored when comparing the fragmentation spectra between the unknown feature and the mzCloud hit.

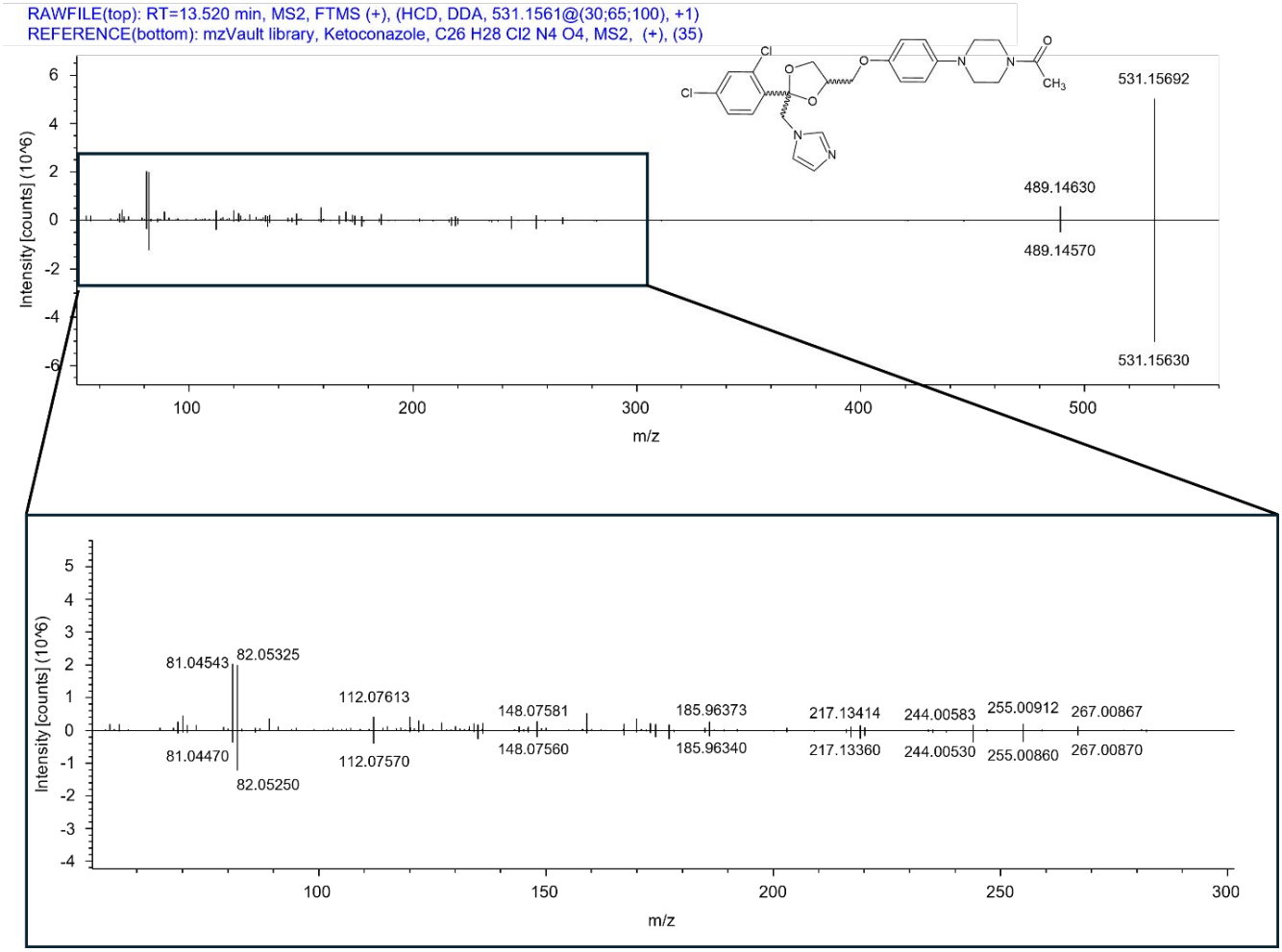


Figure S8. Comparison of fragmentation (i.e., MS^2) spectra between a feature detected in biosolid samples (top spectrum) and a Mass Bank of North America hit for ketoconazole, resulting in a confidence Level 2b annotation.

References

(1) Williams, A. J.; Grulke, C. M.; Edwards, J.; McEachran, A. D.; Mansouri, K.; Baker, N. C.; Patlewicz, G.; Shah, I.; Wambaugh, J. F.; Judson, R. S.; Richard, A. M. The CompTox Chemistry Dashboard: A Community Data Resource for Environmental Chemistry. *Journal of Cheminformatics* **2017**, *9* (1), 61. https://doi.org/10.1186/s13321-017-0247-6.

(2) Lowe, C. N.; Williams, A. J. Enabling High-Throughput Searches for Multiple Chemical Data Using the U.S.-EPA CompTox Chemicals Dashboard. *J. Chem. Inf. Model.* **2021**, *61* (2), 565–570. https://doi.org/10.1021/acs.jcim.0c01273.

(3) *Predict Molecular Properties* | *Percepta Software*. ACD/Labs. https://www.acdlabs.com/products/percepta-platform/ (accessed 2024-03-07).

(4) Mansouri, K.; Grulke, C. M.; Judson, R. S.; Williams, A. J. OPERA Models for Predicting Physicochemical Properties and Environmental Fate Endpoints. *Journal of Cheminformatics* **2018**, *10* (1), 10. https://doi.org/10.1186/s13321-018-0263-1.

(5) Martín-Pozo, L.; de Alarcón-Gómez, B.; Rodríguez-Gómez, R.; García-Córcoles, M. T.; Çipa, M.; Zafra-Gómez, A. Analytical Methods for the Determination of Emerging Contaminants in Sewage Sludge Samples. A Review. *Talanta* **2019**, *192*, 508–533. https://doi.org/10.1016/j.talanta.2018.09.056.

(6) Zuloaga, O.; Navarro, P.; Bizkarguenaga, E.; Iparraguirre, A.; Vallejo, A.; Olivares, M.; Prieto, A. Overview of Extraction, Clean-up and Detection Techniques for the Determination of Organic Pollutants in Sewage Sludge: A Review. *Analytica Chimica Acta* **2012**, *736*, 7–29. https://doi.org/10.1016/j.aca.2012.05.016.

(7) Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, I. D.; Lewis, M. R.; Dunn, W. B. Guidelines and Considerations for the Use of System Suitability and Quality Control Samples in Mass Spectrometry Assays Applied in Untargeted Clinical Metabolomic Studies. *Metabolomics* **2018**, *14* (6), 72. https://doi.org/10.1007/s11306-018-1367-3.

(8) Black, G.; Lowe, C.; Anumol, T.; Bade, J.; Favela, K.; Feng, Y.-L.; Knolhoff, A.; Mceachran, A.; Nuñez, J.; Fisher, C.; Peter, K.; Quinete, N. S.; Sobus, J.; Sussman, E.; Watson, W.; Wickramasekara, S.; Williams, A.; Young, T. Exploring Chemical Space in Non-Targeted Analysis: A Proposed ChemSpace Tool. *Anal Bioanal Chem* **2023**, *415*, 35–44. https://doi.org/10.1007/s00216-022-04434-4.

(9) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, *48* (4), 2097–2098. https://doi.org/10.1021/es5002105.

(10) Black, G. P.; Anumol, T.; Young, T. M. Analyzing a Broader Spectrum of Endocrine Active Organic Contaminants in Sewage Sludge with High Resolution LC-QTOF-MS Suspect Screening and QSAR Toxicity Prediction. *Environ. Sci.: Processes Impacts* **2019**, *21* (7), 1099–1114. https://doi.org/10.1039/C9EM00144A.

(11) Vegosen, L.; Martin, T. M. An Automated Framework for Compiling and Integrating Chemical Hazard Data. *Clean Techn Environ Policy* **2020**, *22* (2), 441–458. https://doi.org/10.1007/s10098-019-01795-w.